

**Caracterización molecular de profagos en genomas de
Streptococcus spp**

Mónica Marcela Ramírez Ocampo

Asesor:

Juan Pablo Isaza Agudelo
Microbiólogo y bioanalista

Trabajo de grado para obtener el título de:
Microbióloga y bioanalista

Universidad de Antioquia
Escuela de Microbiología
Medellín

2020.

Caracterización molecular de profagos en genomas de *Streptococcus* spp.

Mónica Marcela Ramírez Ocampo¹, Juan Pablo Isaza Agudelo²

1. Escuela de Microbiología, Universidad de Antioquia. Antioquia, Colombia, 2. Biología de sistemas Universidad Pontificia Bolivariana.

Palabras claves: *Streptococcus* spp, profagos, factores de virulencia, factores de resistencia a antibiótico, CRISPR-cas, estudio *in-silico*

Resumen: El objetivo del estudio consistió en evaluar la diversidad genética *in silico* de bacteriófagos atemperados en genomas de *Streptococcus* spp, determinar la presencia de factores de virulencia y genes de resistencia a antibióticos dados por estos y finalmente evaluar la asociación entre el sistema CRISPR-cas con la presencia de bacteriófagos atemperados. Para esto se obtuvieron 402 genomas a partir del NCBI; la identificación de los profagos se hizo mediante el programa PHASTER y fueron categorizados basados en su estructura; la identificación de los factores de virulencia y resistencia a antibióticos se hizo una estrategia de búsqueda por homología en bases de datos, dos correspondientes para factores de virulencia y dos a genes asociados a resistencia a antibióticos; finalmente se analizó la presencia del sistema CRISPR-cas mediante Crisprdisco. Teniendo en cuenta la cantidad de genomas y el amplio número de especies de *Streptococcus* spp, se encontró gran variedad de profagos siendo la categoría incompleta más predominante. Se describió una correlación positiva débil en relación con los factores de virulencia y resistencia a antibióticos bacterianos en comparación de los profagos, lo cual hace pensar que estos pueden estar influenciando otros procesos celulares diferentes. Según la clasificación taxonómica de los profagos completos se demostró que todos pertenecen al orden *Caudoviridae* y de estos el 99,7% a la familia *Siphoviridae*. Del sistema CRISPR cas se sugiere que aquellos genomas con la presencia de los sistemas I, II y III conjuntamente tienen una menor probabilidad de adquirir DNA extracelular de tipo profagico entre su genoma.

Molecular characterization of prophages in genomes of *Streptococcus* spp.

Keywords: *Streptococcus* spp, prophages, virulence factors, antibiotic resistance factors, CRISPR-cas, study *in-silico*

Abstract: The objective of this study was to evaluate the genetic diversity *in silico* of tempered bacteriophages in genomes of *Streptococcus* spp, to determine the presence of virulence factors and antibiotic resistance genes given by these, and finally to evaluate the association between the presence of CRISPR-cas system and the presence of tempered bacteriophages in genomes of *Streptococcus* spp. In this study, 402 species were obtained from NCBI. The identification of the prophages was done through the PHASTER program and they were categorized according to its structure. In the identification of the virulence factors and antibiotics resistance a homology search was implemented in four database, two corresponding for virulence factors and the other two genes associated with antibiotics resistance; finally, the presence of the CRISPR-cas system was analyzed by Crisprdisco. Considering the great number of genomes and a wide number of *Streptococcus* spp species, a great prophages variety was found, being the incomplete category most predominant. A weak positive correlation is described in relation with the virulence factors and bacterial antibiotic resistance compared to the prophages, which suggest that these may influence other different cellular processes. Based on taxonomic classification of the complete

prophages, it demonstrated that all the complete prophages belong to the order *caudoviridae*, and from these, 99.7% belonged to the *Siphoviridae* family. Respect to the CRISPR-cas system, results suggest that those genomes with the presence of systems I, II and III together have a lower probability of acquire DNA extracellular prophage type among their genome.

Introducción.

Los Bacteriófagos son considerados parásitos intracelulares obligados con la capacidad de infectar un amplio espectro de organismos, los cuales liberan su material genético al interior de la célula¹ donde pueden realizar un ciclo lítico o lisogenico. En el ciclo lisogenico, debe ocurrir una fijación del virus e inyección de su material genético, el cual recombina y se integra al cromosoma bacteriano. En este caso el material genético recibe el nombre de profago, y se copia continuamente junto con el material genético de la célula anfitriona². Durante el ciclo lítico, la replicación viral, el ensamblaje de las partículas virales y la subsecuente expresión de proteínas que lisan la bacteria permite liberar los nuevos virus³.

Los bacteriófagos en su mayoría poseen una estructura binaria, una cabeza constituida por el material genético del fago contenido en una cápside y una cola de estructura helicoidal el cual cumple la función de unirse a los receptores de la célula hospedadora. Según la clasificación por el comité internacional de taxonomía de virus (ICTV) se clasifican en un orden *Caudovirales* y diez familias *Myoviridae*; *Siphoviridae*, *Podoviridae*, *Tectiviridae*, *Tortricoviridae*, *Plasmaviridae*, *Microviridae*, *Inoviridae*, *Cystoviridae* y *Leviviridae*⁴ ya sea por su morfología, o por el tipo de contenido de ácidos nucleicos que posea; bien sea DNA o RNA, de doble cadena o de cadena sencilla y la disposición que tenga este contenido genético en la célula hospedadora.

Los profagos como elementos móviles de DNA, responsables de transferencia lateral de genes, están estrechamente relacionados con la evolución de bacterias patógenas, ya que son una fuente importante de factores de virulencia. Han sido asociados a la secreción de toxinas, mejora

en la adherencia, evasión de la respuesta inmune, formación de biopelículas y resistencia a antimicrobianos, entre otros^{2,5}. Sin embargo, el papel de los profagos no se limita solo a la patogénesis ya que en bacterias no patógenas se ha detectado adaptaciones en su genoma mediadas por estos, logrando aumentar aptitudes de la célula huésped; ejemplos de esto se dan en resistencia sérica y mejor supervivencia en macrófagos aumentando el éxito reproductivo de estas bacterias⁶. En ambos casos, ya sean bacterias patógenas o no patógenas, los profagos aportan en gran medida la variabilidad genética encontrada entre diferentes cepas o especies bacterianas, lo que se relaciona a la aparición de nuevos clones virulentos entre cepas no patógenas promoviendo de esta forma la hipótesis sobre la co-evolución entre bacterias y fagos.

El sistema de repeticiones palindrómicas cortas agrupadas y regularmente espaciadas por sus siglas en inglés CRISPR junto con la endonucleasas cas⁷, se han considerado como un sistema de defensa adaptativo de bacterias y archaeas el cual se ha estudiado por su capacidad generada sobre la célula huésped para reconocer y posteriormente destruir el material genético invasor, lo que representa una relación entre la funcionalidad del sistema y la diversidad genética bacteriana. Estos avances informáticos han permitido identificar, caracterizar y clasificar los diferentes sistemas considerando la existencia de variabilidad entre los grupos taxonómicos⁸.

El género *Streptococcus* es un grupo numeroso y heterogéneo de bacterias Gram positivas que pueden ser patógenas como *S. pyogenes*, comensales como *S. oralis* o saprofitas como *S. thermophilus*⁹. En la clasificación taxonómica del genero se han empleado diferentes técnicas tanto moleculares como bioquímicas por lo cual

no se ha logrado un consenso en algunas especies. Por tanto, se crearon los grupos de especies “Pyogenic”, “Mitis”, “Anginosus”, “Bovis”, “Mutans” y “Salivarius” para agrupar especies con características similares. Dicha agrupación es soportada por análisis filogenómicos⁹.

Como género bacteriano es uno de los principales agentes causales de infecciones, siendo una de las razones de muerte de mayor impacto a nivel mundial; a esto se le suma un constante aumento de resistencias al tratamiento por antibióticos, conllevando a estrategias farmacológicas de mayor espectro favoreciendo los fenómenos de resistencia¹⁰. Por otro lado, diferentes factores de virulencia han sido encontrados en el genoma de 135 aislados que representan la mayoría de las especies del género. Se ha determinado también, que algunos factores de virulencia involucrados en adhesión como *pavA* (fibronectin binding proteins), *srtA* (sortase A) y *slrA* (streptococcal lipoprotein rotamase), y otros factores como proteasas *htrA/degP* (Serine protease) y *tig/ropA* (trigger factor) han evolucionado en concordancia con el género⁹.

La presencia de fagos en los cromosomas de aislados clínicos de *Streptococcus* ha sido documentada, y se ha podido observar una predominancia de fagos de la familia *Siphoviridae*. De dicha familia hay representantes como SA1 y SA2 en *S. Agalactiae*⁶ SM1 en *S. Mitis*¹¹, HB-3¹² y MM1¹³ en *S. Pneumoniae* y 315.5 y 315.6 entre otros en *S. Pyogenes*¹⁴. Entre los fagos identificados en *Streptococcus* spp. se ha evidenciado gran diversidad genética incluso entre fagos detectados en la misma especie bacteriana. Por parte de *S. Pyogenes*, los genes involucrados en la replicación de DNA y la lisogenia son poco conservados, mientras que en *S. thermophilus* es uno de los módulos más conservados y la mayor variación genética se encuentra sobre los genes involucrados en la morfogénesis⁶.

El propósito del presente estudio fue evaluar la diversidad genética *in silico* de bacteriófagos

atemperados en genomas de *Streptococcus* spp, identificar y clasificar los bacteriófagos encontrados, determinar la presencia de factores de virulencia y genes de resistencia a antibióticos dados por estos y finalmente evaluar la asociación entre la presencia del sistema CRISPR-cas y la presencia de bacteriófagos atemperados en genomas de *Streptococcus* spp.

Metodología.

Obtención de genomas

402 genomas pertenecientes a 40 especies del género *Streptococcus* spp fueron descargados desde la base de datos de genomas de “National Center for Biotechnology Information” NCBI¹⁵, como criterio de inclusión se utilizó el nivel del ensamblaje, solo fueron incluidos aquellos reportados como genomas completos.

Identificación de profagos

La identificación de profagos se hizo mediante el programa

“Phage Search Tool Enhanced Release”

PHASTER. Los profagos identificados fueron categorizados como completos, cuestionables o incompletos, teniendo en cuenta los criterios de Arndt y colaboradores¹⁶. Adicionalmente se realizó una clasificación taxonómica de los profagos detectados, la cual se basó en la identificación de especies con mayor cantidad de hits entre las proteínas asociadas a virus.

Para agrupar los genomas de los profagos detectados se empleó el programa CD-HIT versión 4.6.6.3.¹⁷ basados en un 95% de identidad y un 95% cobertura de la longitud de la región profagica.

Identificación de genes involucrados en virulencia y en resistencia a antibióticos.

Para evaluar los factores de virulencia asociados a los profagos y a los proteomas bacterianos, se utilizó una estrategia de búsqueda por homología. Para esto se descargó la base de datos de factores de virulencia “Virulence Factors Database” VFDB¹⁸, y “virulence factors”

VICTORS¹⁹ y se comparó con las proteínas identificadas en los profagos por medio del programa Fasta-36.3.8g y la aplicación ggsearch36. Las proteínas que presentaron un alineamiento con un valor de E menor a 1e-6 fueron clasificadas como factores de virulencia. Por otro lado, para identificar las proteínas involucradas en resistencia a antibióticos, se descargó la base de datos “Comprehensive Antibiotic Resistance Database” CARD²⁰ y “National Database of Antibiotic Resistant Organisms” NDARO²¹ y se comparó con las proteínas identificadas en los profagos por medio del programa Fasta-36.3.8g y la aplicación ggsearch36. Las proteínas que presentaron un alineamiento con un valor de E menor a 1e-6 fueron asociadas a resistencia de antibióticos.

Identificación del sistema CRISPR-cas.

Se analizó la presencia del sistema CRISPR-cas mediante Crisprdisco versión Python 2.7⁸ identificando matrices de repetición espaciadora CRISPR y genes cas. Para la asignación de sistemas funcionales se tuvo como criterio la presencia de nucleasas efectoras entre los genomas analizados.

Pruebas estadísticas

Se realizó un análisis descriptivo con valores cuantitativos mínimo, máximo y promedio. Para la validación estadística de la información se empleó el software SPSS statistics versión 25.0 en el cual se realizaron las comparaciones de medianas con de pruebas Spearman y Kruskal – Wallis según correspondiera, para ambas se tomó un nivel de significancia de P menor o igual de 0,05.

Resultados.

Se realizó la descarga de 402 genomas pertenecientes a 40 especies del género *Streptococcus* spp., entre las especies evaluadas se incluyeron un rango de uno a ciento doce genomas por especie. En relación al porcentaje Guanina + Citosina (GC) de los genomas evaluados, se encuentran genomas con un GC

mínimo de 34% *S. parauberis* cepa SPOF3K y un GC máximo de 44% *S. sobrinus* cepa NIDR_6715-7, en promedio el GC es de 39%. En cuanto al tamaño de los genomas, se encuentran en un rango de 1 702 054 en *S. pyogenes* cepa NGAS327 a 2 719 245 pb en *S. suis* cepa HN136, con un promedio de 2 016 790 pb. Respecto a las proteínas anotadas en los genomas, se realizó un recuento global de 757 550 distribuidas entre todos los genomas, entre los cuales el proteomas más pequeño es de 1548 y el más grande de 2371, con un número promedio de proteínas de 1884. (tabla 1).

Posteriormente los datos genómicos fueron procesados con el programa PHASTER en búsqueda de secuencias de profagos, encontrándose un total de 959 con un promedio de 2,4 profagos por genoma evaluado, con un mínimo y un máximo variable entre especies (Figura 1). Sin embargo, de los 402 genomas analizados, en *S. equi* (GCF_000026605); *Streptococcus* sp. I-P16 (GCF_000479315); *S. pneumoniae* G54 (GCF_000019825); *S. pyogenes* (GCF_900475155), no se identificaron elementos profagicos. No se encontró una correlación estadísticamente significativa entre la longitud del genoma bacteriano y la cantidad de elementos profagicos (Spearman, P = 0,191). Por otro lado, al correlacionar el %GC de los genomas bacterianos con la cantidad de profagos, tampoco se encontraron diferencias estadísticamente significativas (Spearman, P = 0,582).

En la distribución de los elementos profagicos a nivel de los grupos se estableció una diferencia estadísticamente significativa entre la cantidad de profagos y los diferentes grupos de *Streptococcus* spp. (Kruskal – Wallis, P = 0,029) (figura 2.a). La cantidad de profagos encontrados entre cada grupo de *Streptococcus* fue diferente, se describe un mayor contenido profagicos en el grupo Pyogenes con 494 profagos, que a la vez tiene una mayor representación por genomas. En contraste, un menor contenido profagico fue

Grupo	Especie	Genomas representativos	Rango %GC	Rango longitud del genoma	Rango proteínas
Pyogenic	<i>Streptococcus pyogenes</i>	112	38,2 - 38,7	1702054 - 1950469	1566 - 1874
	<i>Streptococcus dysgalactiae</i>	9	39,4 - 39,8	2079953 - 2255201	1885 - 2075
	<i>Streptococcus equi subsp. zooepidemicus</i>	9	41,5 - 41,9	2024171 - 2172187	1721 - 1928
	<i>Streptococcus parauberis</i>	2	34,2 - 35,5	2143887 - 2152278	1821 - 2008
	<i>Streptococcus iniae</i>	4	35,1 - 36,8	2086959 - 2116570	1834 - 1886
	<i>Streptococcus uberis</i>	4	36,5 - 36,7	1852352 - 2024265	1710 - 1877
	<i>Streptococcus porcinus</i>	1	36,8	2033046	1894
Bovis	<i>Streptococcus agalactiae</i>	55	35,4 - 36,6	1841946 - 2275139	1637 - 2149
	<i>Streptococcus infantarius</i>	2	36,3 - 37,7	392607 - 1818293	1611 - 1987
	<i>Streptococcus pasteurianus</i>	2	37,4 - 37,4	210007 - 2178089	1869 - 2015
	<i>Streptococcus lutetiensis</i>	3	37,68 - 37,74	1793521 - 1975547	1562 - 1753
	<i>Streptococcus equinus</i>	1	37,4	2311542	2038
	<i>Streptococcus macedonicus</i>	1	36,3	2142762	1798
Mutans	<i>Streptococcus gallolyticus</i>	6	37,5 - 37,7	2052496 - 2493581	1745 - 2333
	<i>Streptococcus sobrinus</i>	6	43,5 - 43,7	2144746 - 2198648	1857 - 1928
Salivarius	<i>Streptococcus mutans</i>	8	36,8 - 37,1	1984441 - 2088369	1839 - 1959
	<i>Streptococcus thermophilus</i>	24	37,1 - 39,2	1731838 - 2102271	1548 - 1852
No agrupados	<i>Streptococcus salivarius</i>	9	37,9 - 40,3	2100988 - 2326348	1825 - 1984
	<i>Streptococcus halotolerans</i>	1	39,2	2182100	1868
	<i>Streptococcus himalayensis</i>	1	41,3	2275471	1987
	<i>Streptococcus suis</i>	39	35,8 - 41,5	1980887 - 2719245	1826 - 2371
	<i>Streptococcus marmotae</i>	1	34,98175479	2309282	2056
	<i>Streptococcus merionis</i>	1	41,75254013	2384128	2115
	<i>Streptococcus pantholopis</i>	1	41,97172001	2241302	1991
	<i>Streptococcus acidominimus</i>	1	39,9	2330811	1980
	<i>Streptococcus ferus</i>	1	42,8	1872314	1773
	<i>Streptococcus sp.</i>	7	35,6 - 42,2	1874337 - 2435494	1739 - 2109
Anginosus	<i>Streptococcus anginosus</i>	5	38,2 - 38,9	1911706 - 2233640	1711 - 1989
	<i>Streptococcus intermedius</i>	6	37,6 - 37,7	1914006 - 2006877	1761 - 1844
	<i>Streptococcus constellatus</i>	3	38,1 - 38,2	1935414 - 1991156	1778 - 1829
Mitis	<i>Streptococcus gordonii</i>	6	40,4 - 40,7	2121309 - 2283306	1970 - 2128
	<i>Streptococcus cristatus</i>	3	42,1 - 42,6	2000351 - 2176084	1879 - 2026
	<i>Streptococcus sanguinis</i>	4	42,9 - 43,4	2296927 - 2449527	2167 - 2320
	<i>Streptococcus australis</i>	1	41,9	2161464	1954
	<i>Streptococcus parasanguinis</i>	2	41,6 - 41,7	2153652 - 2171609	1958 - 2008
	<i>Streptococcus oralis</i>	4	40,9 - 41,1	1889005 - 2053209	1786 - 1910
	<i>Streptococcus mitis</i>	3	39,9 - 40,2	1906344 - 2167922	1787 - 2002
	<i>Streptococcus pseudopneumoniae</i>	1	39,1	2195458	2115
Hyovaginalis	<i>Streptococcus pneumoniae</i>	51	39,5 - 39,9	1983415 - 2287774	1814 - 2198
	<i>Streptococcus pluranimalium</i>	2	38,6 - 38,7	2065522 - 2246024	1930 - 2039

Tabla 1. Descripción general genomas evaluados.

detectado en el grupo Hyovaginalis con 4 profagos (Figura 2.b).

Los profagos se clasificaron estructuralmente de acuerdo a PHASTER en completos, incompletos y cuestionables. De este modo, se encontraron 36% (344) de profagos completos, 11% (107) cuestionables y 53% (508) incompletos. Del total de secuencias genómicas evaluadas, aproximadamente en el 44% de estas presentaron profagos completos, de estos hubo un promedio de 2 profagos por genoma con un máximo de 6; el 23% de los genomas presentaron profagos cuestionables, con un promedio de 1,2 profagos por genoma y un máximo de 4; por último, el

70% de los genomas presentaron profagos incompletos, con un promedio de 1,8 profagos por genoma y un máximo de 6.

En los profagos completos el porcentaje de GC va desde 34% hasta 43% con un promedio de 38% y una longitud en pares de bases desde 11 599 pb hasta 105 200 pb con un promedio de 48 455 pb. En los profagos cuestionables el %GC va desde 32% hasta 45% con un promedio de 39% y una longitud desde 6544 pb hasta 68 280 pb con un promedio de 38 156 pb. Por último, en los profagos incompletos el %GC va desde 31% hasta 46% con un promedio de 39% y una

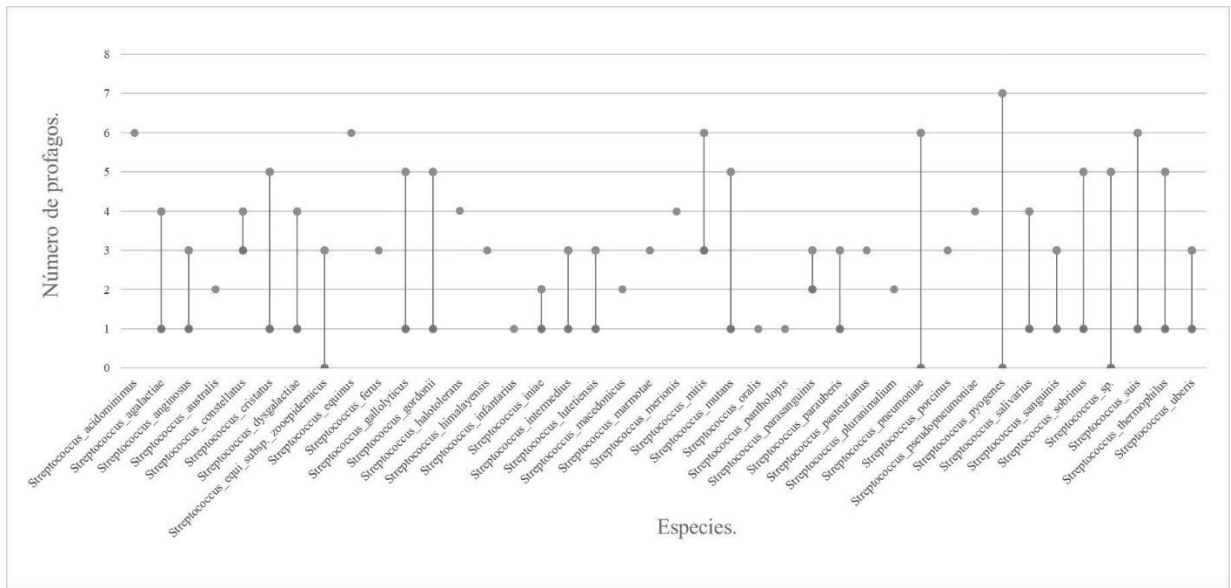


Figura 1. Representación gráfica del rango de la cantidad de profagos encontrados en cada especie del género de *Streptococcus* spp.

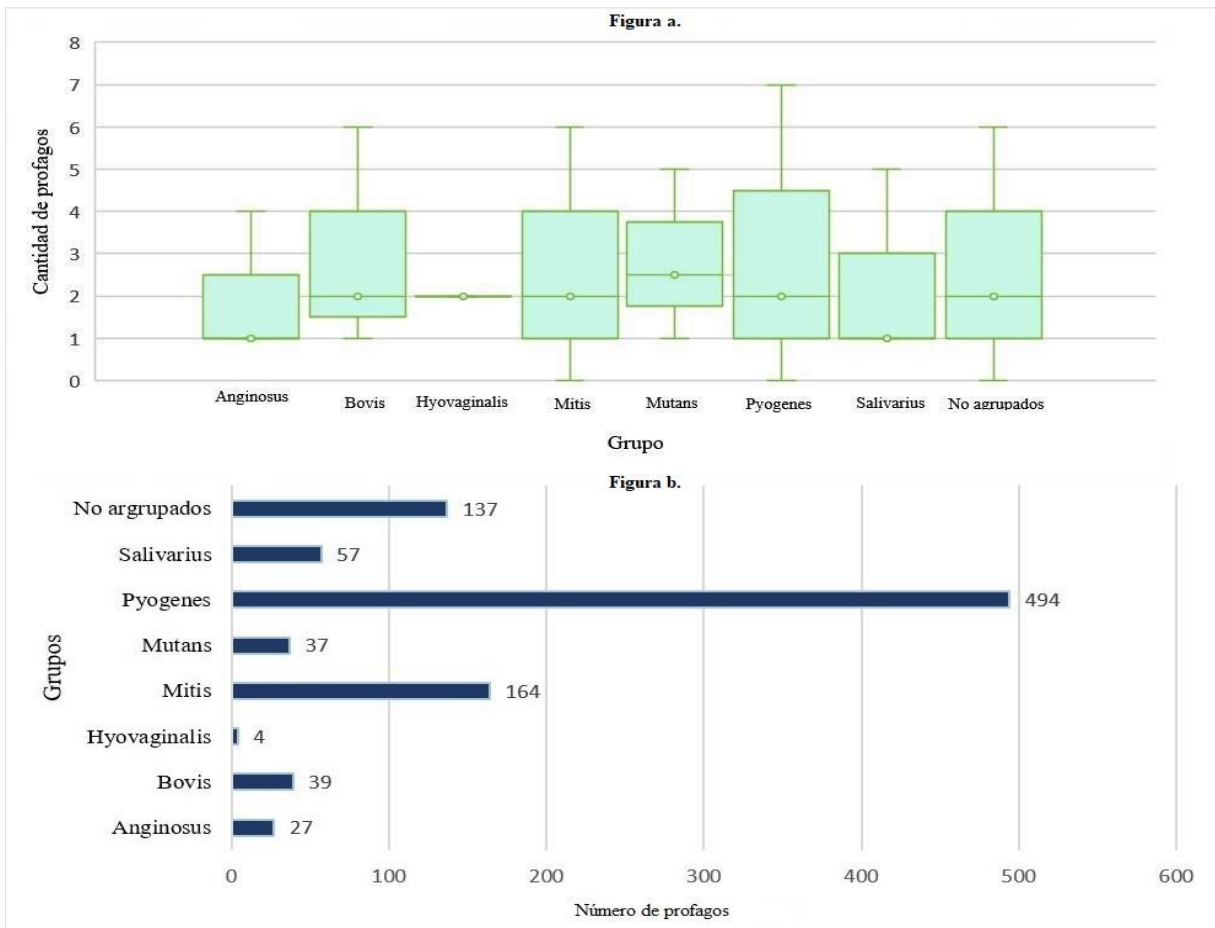


Figura 2. a. Gráfico de cajas y bigotes para representar la distribución de profagos por cada grupo de *Streptococcus* spp. y b. Diagrama de barras para representar el total de profagos por cada grupo de *Streptococcus* spp.

longitud desde 3526 pb hasta 60 537 pb con un promedio de 15 954 pb (figura 3).

En relación a los profagos completos encontrados, se realizó una clasificación taxonómica con el fin de identificar las familias de virus a las cuales pertenecían. Se observó un predominio de la familia *Siphoviridae*, dentro de la cual se encuentran *Streptococcus phage 315.4* con 64 profagos detectados, *Streptococcus phage 315.3* con 61 profagos, y *Streptococcus phage 315.2* con 42 profagos. Se encontró también un representante de la familia *Myoviridae*, *Streptococcus phage EJ-1* (figura 4).

La clusterización de los elementos profagicos encontrados en los genomas de *Streptococcus* spp se realizó con el propósito conocer que tan conservado eran dichos profagos entre los genomas. Para la clusterización se estableció un porcentaje de identidad del 95%, sin embargo, los 344 profagos completos fueron agrupados en 253 clústers, lo que da cuenta de una alta diversidad.

Identificación de genes involucrados en virulencia y resistencia a antibióticos

Entre todos los genomas bacterianos se identificaron 23 770 proteínas clasificadas como factores de virulencia, el menor número de proteínas fue 29 y en el mayor 85, con un promedio de 59 proteínas asociadas a virulencia por cepa. En el caso de proteínas asociadas a resistencia a antibióticos se encontraron 1722 entre todos los proteomas, con una distribución que va desde 0 a 14 proteínas, con un promedio de 4 proteínas por cepa. Entre todos los profagos detectados se identificaron 1179 proteínas asociadas a factores de virulencia, el rango de proteínas codificadas en regiones profagicas por genoma bacteriano va desde 0 a 16, con un promedio de 3 proteínas por cepa. En relación a las proteínas asociadas a los factores de resistencia a antibióticos suman un total de 29 proteínas, con un rango entre 0 y 3. Se encontró una correlación positiva estadísticamente significativa, aunque débil, entre la cantidad de

proteínas codificadas en regiones profagicas asociadas a resistencia a antibióticos y la cantidad de proteínas totales bacterianas asociadas a factores de resistencia a antibiótico (Spearman, $P = 0,012$, coeficiente de correlación = 0,126). Con relación a los factores de virulencia, se encontró una correlación positiva débil pero estadísticamente significativa entre la cantidad de factores de virulencia codificados en las regiones profagicas y el total de factores de virulencia en el genoma bacteriano (Spearman, $P < 0,01$, coeficiente de correlación = 0,406).

Con el propósito de conocer que grupo de especies porta una mayor cantidad de factores de virulencia o de genes de resistencia a antibióticos, codificados en regiones profagicas o en el genoma en general, por cada grupo de especies se normalizó el recuento de proteínas. Para esto la sumatoria de las proteínas por grupo fue dividido por la cantidad de genomas en el grupo, de esta forma se identificaron los grupos de especies con mayor cantidad de proteínas asociadas a factores de virulencia “pyogenes” y de resistencia a antibióticos “No agrupados”. En la distribución de proteínas codificadas en el genoma bacteriano en general, y que estuvieran relacionadas con virulencia y resistencia a antibióticos, a nivel de los grupos se estableció una diferencia estadísticamente significativa entre la cantidad de proteínas y los diferentes grupos de *Streptococcus* spp. (Kruskall –Wallis, $P < 0,01$) (figura 5a). De manera similar, pero al evaluar ambas categorías de proteínas, pero codificadas en regiones profagicas, se encontró una mayor cantidad de proteínas asociadas a factores de virulencia en el grupo “pyogenes” (figura 5.a), mientras que la cantidad de proteínas asociadas a resistencia a antibióticos fue mayor en el grupo “Hyovaginalis” (figura 5.b). En la distribución de proteínas profagicas relacionadas con virulencia y resistencia a antibióticos a nivel de los grupos se estableció una diferencia estadísticamente significativa entre la cantidad de proteínas y los diferentes grupos de *Streptococcus* spp. (Kruskall –Wallis, $P < 0,01$) (figura 5.b).

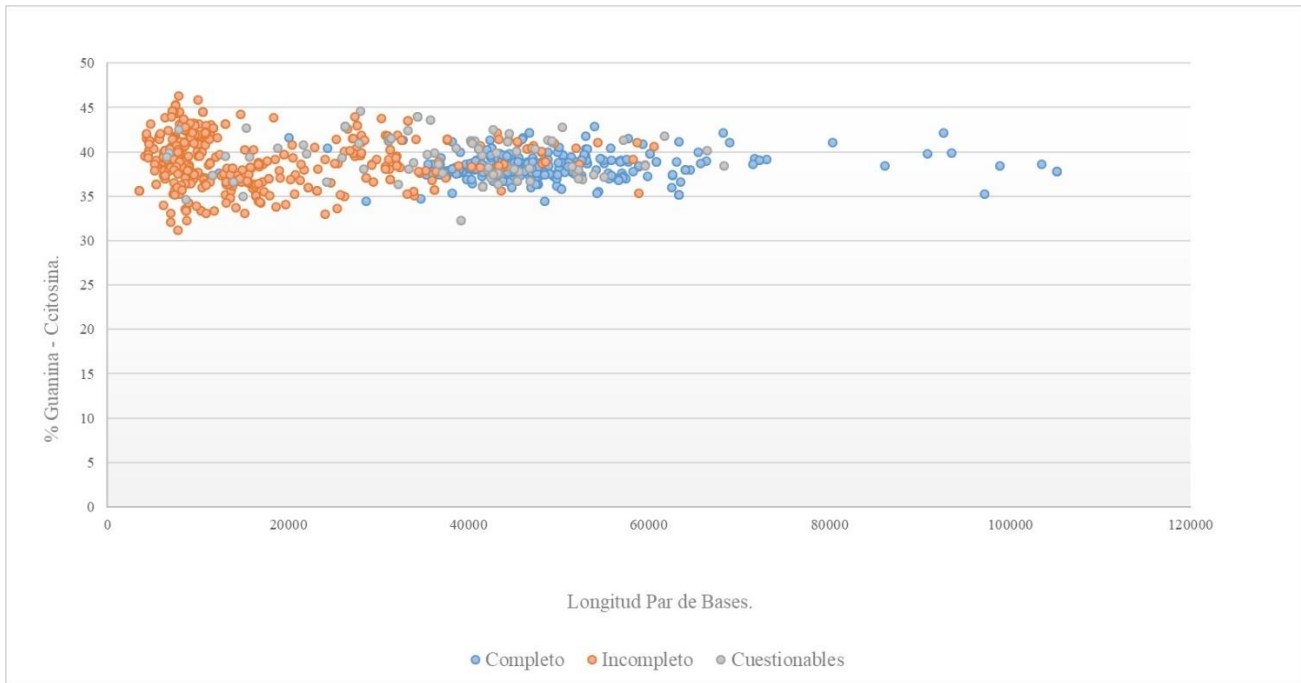


Figura 3. Gráfico de dispersión del %GC y la longitud de profagos entre las diferentes categorías estructurales.

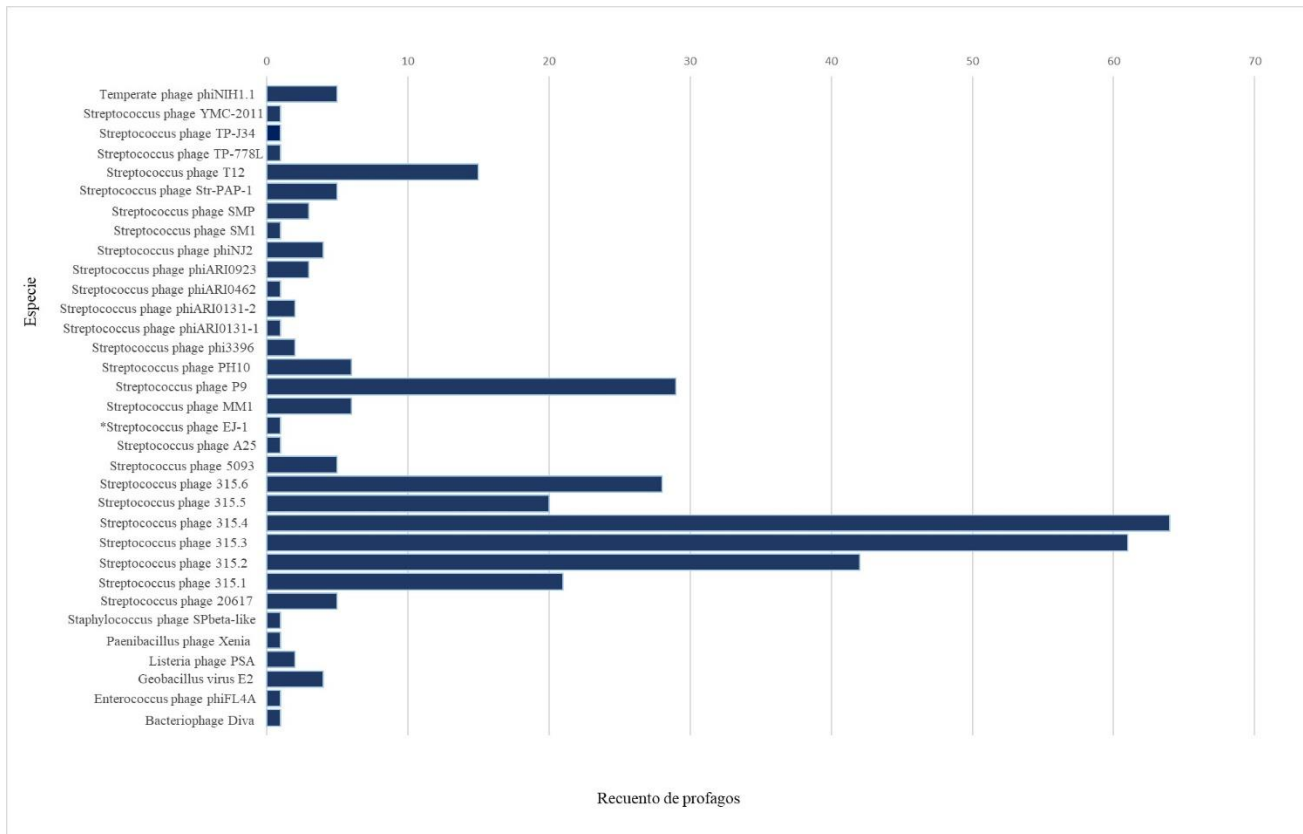


Figura 4. Gráfico de barras para representar la clasificación taxonómica de los profagos de la categoría completos encontrados en el género de *Streptococcus* spp.

Identificación de sistemas CRISPR-cas

Para el análisis de los sistemas CRISPR-cas, se asumió como un sistema funcional, aquellos en los que detectará la presencia de las nucleasas efectoras. Entre los genomas estudiados se cuenta con 165 cepas con ausencia de nucleasas efectoras y 237 cepas donde se detectó presencia de nucleasas efectoras. El sistema CRISPR tiene una serie de clasificaciones de acuerdo a las nucleasas que presenta, entre los sistemas detectados el tipo II fue el dominante, con un porcentaje de 46% (110), seguido está el sistema tipo I y tipo II conjunto con el 35% (84), el sistema tipo I con el 8% (19), el sistema tipo II y tipo III conjunto con el 5% (12), el sistema tipo I, II y III conjunto con el 3% (6), el sistema tipo I y III conjunto con el 1,3% (3), y finalmente el sistema tipo III con el 1,3% (3).

Al evaluar el papel del sistema CRISPR-cas en la protección de la bacteria contra la incorporación al genoma de bacteriófagos, se demostró diferencias estadísticamente significativas entre la cantidad de profagos de acuerdo con el tipo de sistema CRISPR-cas identificado (Kruskall – Wallis, $P = 0,019$) (figura 6). Estas diferencias demuestran que en aquellas cepas donde se encuentra presencia de las nucleasas efectoras Tipo I, II y III conjuntamente, se detecta una menor cantidad de elementos profagicos. Por el contrario, en las cepas en las cuales se encuentra un predominio de nucleasas efectoras de tipo II, se observó una mayor cantidad de profagos.

Discusión.

El estudio de elementos profagicos a nivel del genoma del genero de *Streptococcus* spp ha sido uno de las propuestas de investigación planteados en diferentes momentos de la historia^{22,23}. En estudios realizados como el de Brueggeman y colaboradores (2017), en 482 aislados neumocócicos se logran identificar en el 100% de los genomas bacterianos secuencias asociadas a profagos². En contraste el estudio realizado por García y colaboradores (2009), en una colección

de 108 aislados clínicos de población neumocócica, más de la mitad de los aislados contenían secuencias profagicas de tipo lisogenico²⁴. En la publicación realizada por Javan y colaboradores (2019), un estudio en el que se incluyen 1306 genomas del género *Streptococcus* spp se identificaron 415 profagos de longitud completa y 348 profagos denominados satélites, la prevalencia de profagos detectados varía entre las especies oscilan valores entre 0,4% y 9,5%. en este estudio se realizó una búsqueda exhaustiva de profagos en el género, con un predominio de *Streptococcus pneumoniae* el cual representa el 37% de los genomas evaluados, cabe aclarar que para la investigación se contaba con cepario abundante en representantes de la especie *Streptococcus pneumoniae*²⁵. Por otro lado, nuestro estudio identifica 959 profagos basado en 402 genomas depositados en la base de datos del NCBI, la prevalencia de profagos detectados varía entre especies oscilan valores entre 0,10% y 34%. *Streptococcus pyogenes* presenta un mayor predominio de representantes genómicos con un 29% del total de genomas evaluados. En comparación con el estudio Javan y colaboradores (2019) contamos con especies que no se relacionan a sus análisis entre los cuales se encuentran *Streptococcus porcinus*, *Streptococcus lutetiensis*, *Streptococcus halotolerans*, *Streptococcus himalayensis*, *Streptococcus suis*, *Streptococcus marmotae*, *Streptococcus merionis*, *Streptococcus pantholopis*, *Streptococcus ferus*, *Streptococcus australis* y *Streptococcus pluranimalium*, así mismos ellos estudian especies que para nuestro análisis no fueron incluidas como lo son *Streptococcus urinalis*, *Streptococcus canis*, *Streptococcus azizii*, *Streptococcus pseudopocinus*, *Streptococcus vestibularis*. Es importante resaltar además que en el estudio de Javan y colaboradores (2019) se encuentra una similitud entre los profagos detectados en las especies de *Streptococcus* analizadas y logran asociarlos en cinco grupos principales, en comparación, para el actual estudio realizado la

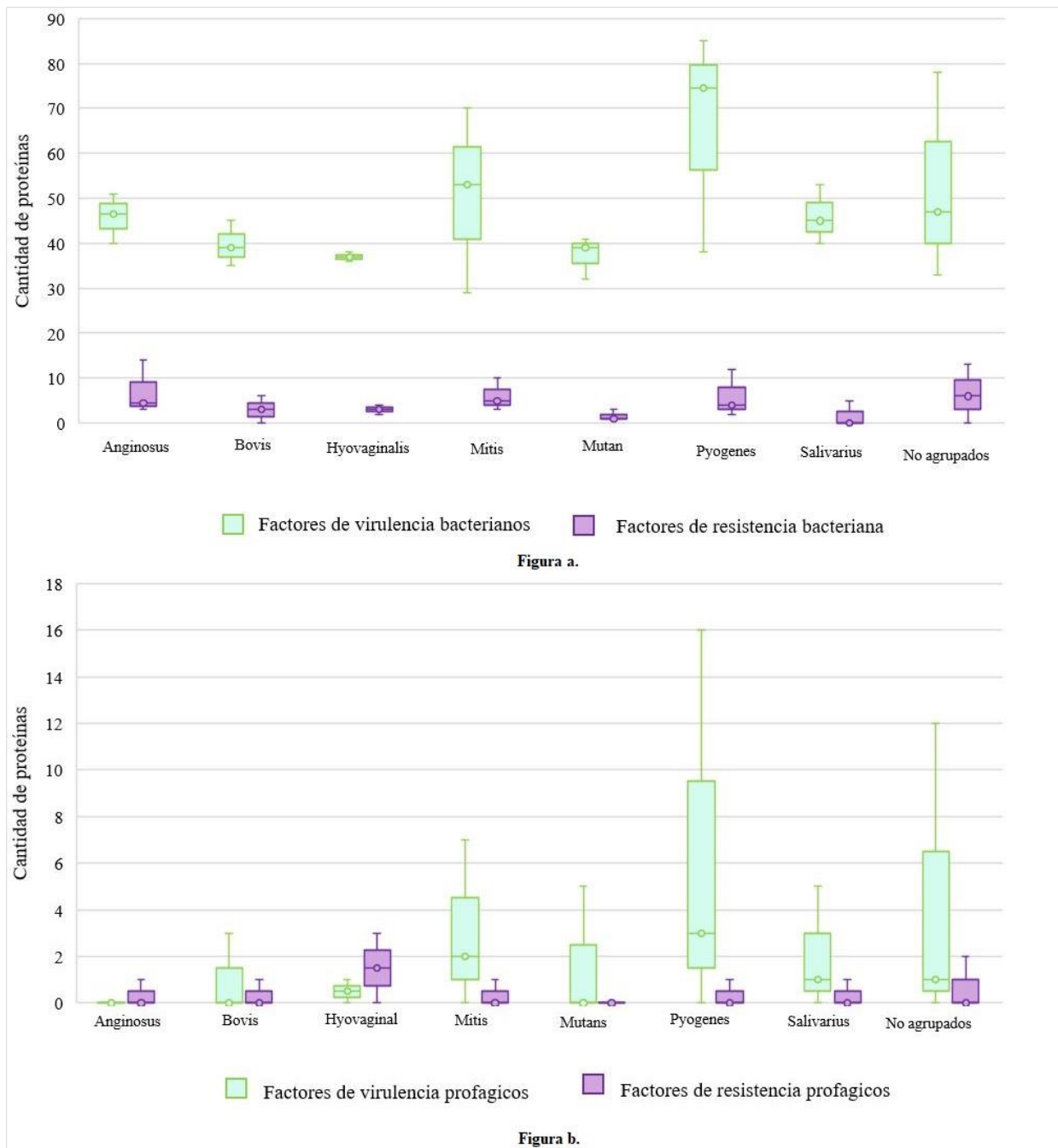


Figura 5. Gráfico de cajas y bigotes para representar la distribución de **a.** factores de virulencia y factores de resistencia a antibióticos entre proteínas genómicas y **b.** factores de virulencia y factores de resistencia a antibióticos codificadas en regiones profágicas de los diferentes grupos del género *Streptococcus* sp.

diversidad encontrada a nivel de los profagos no hizo posible agrupar los mismos. Las herramientas de búsqueda en la detección de profagos a nivel del genoma bacteriano ha sido variable entre los estudios. Los primeros análisis estuvieron basados en la secuenciación e

inspección manual de genes²⁶ o en el cálculo de las abundancias relativas de dinucleótidos el cual tiene en cuenta la heterogeneidad local dentro del genoma bacteriano, lo que ha permitido identificar regiones profágicas dentro del genoma bacteriano²⁷. Recientemente los análisis

in silico emplean como estrategia la comparación de bases de datos con referentes profágicos conocidos, este proceso se relaciona directamente con el estado de actualización de las bases de datos donde se somete la información, lo que puede traer como desventaja pasar por alto información de nuevos profagos por descubrir y caracterizar. En este tipo de búsqueda la composición en nucleótidos y tamaño del total de genomas en estudio, interviene en el total de contenido profágico encontrado^{16,28,29}. En el estudio de Rezaei y colaboradores (2019), la identificación se realizó mediante una estrategia computacional semiautomatizada de búsqueda utilizando PhageMiner, que clasifica los profagos en tres grupos: longitud completa, profagos satélites y regiones desconocidas relacionadas con fagos. La identificación de profagos en el presente estudio se empleó una estrategia de identificación y anotación de secuencias profágicas mediante el programa de búsqueda PHASTER, que realiza comparaciones y alineaciones de secuencia frente a una base de datos y clasifica el resultado profágico en tres grupos: Completos, cuestionables e incompletos. Ambas estrategias de búsqueda se correlacionan en la forma en la que clasifican los profagos detectados, siendo categorías homologas entre sí. Los profagos de longitud completa comparten características estructurales con la categoría denominada por PHASTER completo, asimismo las regiones relacionados con fagos con los profagos cuestionables y los profagos satélites con los profagos incompletos.

Entre la principal limitación de este estudio se encuentra que pese a ser detectadas secuencias profágicas entre los genomas bacterianos, no se tiene la certeza de que estas regiones identificadas sean transcripcionalmente activas, por lo cual para la confirmación de esto es necesario la implementación de estrategias *in vivo* que compruebe la expresión de los mismos.

Diferentes metodologías aplicadas recientemente en los enfoques quimiotaquonómicos como la hibridación genómica y técnicas de

secuenciación de DNA, proporcionan fundamentos sobre las relaciones naturales entre el género *Streptococcus* spp. Esta información ha servido de guía en la taxonomía y nomenclatura del género, formando parte de los lineamientos para la clasificación de los diferentes grupos de especies. La distribución y la evolución entre los grupos tiene un patrón establecido lo cual influye en su diversidad e importancia clínica demostrada en las diferencias encontradas a nivel de los grupos, unos asociados a ser patógenos oportunistas en contraste de otros que son considerados saprofitos o comensales⁹.

En relación al contenido profágico entre los grupos de especies del género de *Streptococcus* spp se ha establecido diferencias, sin embargo, hasta la fecha no existen publicadas investigaciones que a nivel de grupo establezcan este tipo de comparaciones. La presencia de mayor contenido profágico hacia ciertas especies tiene explicaciones plausibles que pueden hacer referencia a la disposición de cada especie bacteriana de ser infectadas según el hábitat y la capacidad de adquisición DNA extracelular de forma lateral que pueda influir en la supervivencia bacteriana³⁰. Al evaluar las diferencias estadísticas entre los datos profágicos encontrados, se destaca una mediana aumentada entre el grupo Mutans conformado por *Streptococcus sobrinus* y *Streptococcus mutans*. Este grupo está representado por 14 genomas con una mediana de profagos de 2,5. En relación a la especie de *Streptococcus mutans* con 8 genomas asociados identificamos un promedio de 2,5 profagos por genoma de los cuales todos presentan características estructurales de ser incompletos. Esta información encontrada, nos permitió comparar los resultados con un estudio que evaluó la presencia de profagos en 24 genomas de *Streptococcus mutans*, y que encontró un promedio de profagos por genoma de 1,4. Al realizar el análisis de las cepas relacionados a los profagos se demostró que algunas de estas podían tener hasta 8 profagos diferentes, de lo anterior llama la atención que en su mayoría comparte las características

estructurales similares a nuestro estudio de ser profagos incompletos³¹. Esto demuestra que nuestro estudio logro detectar de una manera más sensible elementos profagicos asociados a la especie de las cuales comparte características estructurales con estudios previamente reportados^{31,32,33}.

La búsqueda automatizada del contenido profagico en la metodología del presente estudio parte de la cantidad mínima de genes para la formación del profago y la densidad de proteínas

encontrada dentro de este. Una vez detectado el contenido profagico se evalúa su estructura para saber si está completo, cuestionable o incompleto para esto se evalúa si la región contiene genes necesarios para proteínas estructura del fago, regulación de DNA, inserción y lisis. Según la presencia o ausencia de dichos genes será la puntuación de integridad asignada²⁹. El material genético de los profagos replicados por las bacterias durante su división resulta ser desde un beneficio mutuo³⁴, hasta una medida de control

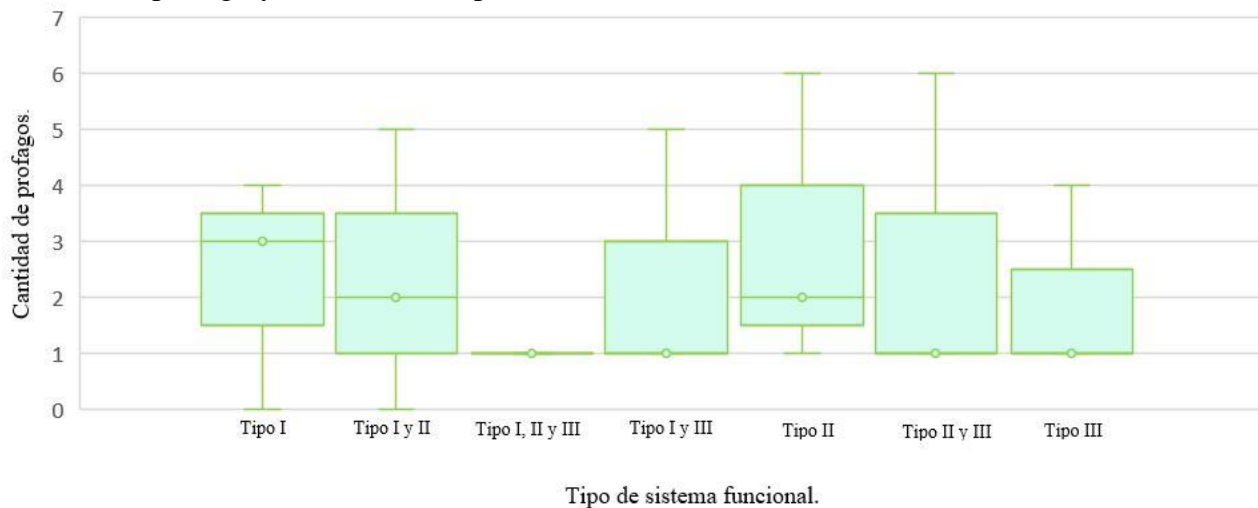


Figura 6. Gráfico de cajas y bigotes para representar la distribución de la cantidad de profagos basado en el tipo sistema CRISPR-cas presente en las cepas de *Streptococcus* spp.

bacteriano^{35,36}. Los profagos puede contribuir a la transferencia de genes e intervenir en diferentes procesos celulares específicos lo cual cuenta como un papel clave en la diversificación bacteriana³⁷. La presencia de profagos denominados como “Completos” determina secuencias proteicas virales codificantes para la invasión, ensamblaje y liberación del posterior contenido viral; En aquellos profagos “Cuestionables” sugiere la presencia de material genético incapaz de producir partículas virales completas lo cual conduce a una alteración del ciclo de vida del fago³⁸; el caso de los profagos “Incompletos” en algunos estudios denominados parcialmente degradados pueden cumplir funciones adaptativas, se ha demostrado que tras la infección pueden proteger de una

sobreinfección, eliminar paquetes de DNA e incluso infectar otras células³⁹.

Las características encontradas en nuestro estudio, como lo son la composición nucleotidica (%GC) y la longitud (Pb) de los profagos identificadas en las diferentes especies de *Streptococcus* spp, han sido previamente reportadas en estudios como en el de Fu y colaboradores (2017)³¹, incluso en otros géneros bacterianos como en lo mencionado por Pratama y colaboradores (2018)⁴⁰, Crispim y colaboradores (2018)¹ y Bobay y colaboradores (2014)³⁹. Respecto a la clasificación estructural planteada para los profagos, en aquellos de longitud completos se describen en promedio de 48 455 pb y un porcentaje Guanina Citosina del 38%, valores más bajos en comparación con lo representado en los genomas hospederos lo que

refleja una integración reciente en la escala de tiempo evolutiva, sin embargo, aún no hay una estimación confiable en la escala de tiempo de la integración de profagos y el equilibrio entre estos y su genoma hospedero⁴⁰. En cuanto al tipo de profagos con mayor predominio detectado entre el género de *Streptococcus* spp, corresponde con los profagos de tipo incompleto con un total del 53% del total del contenido profagico detectado, esta categoría fue característica de presentar un contenido de % de GC menor en comparación de las categorías completas y cuestionables; En relación a la longitud se consideró ser aproximadamente un tercio de la longitud de los profagos de tipo completo. Entre los profagos incluidos en la categoría incompletos se consideran genes alterados y parcialmente degradados en consecuencia de ser un genoma aparentemente pequeño y tener ausencia en genes esenciales para la codificación de proteínas virales. Esta información se relaciona con el estudio realizado por Javan y colaboradores (2019), donde se analizan un gran número de secuencias genómicas del género *Streptococcus* spp y se destaca la prevalencia de este tipo de profagos entre los genomas evaluados, adicionalmente, el estudio señala divergencia entre los profagos de longitud completa y los profagos satélites por lo cual proponen que son dos grupos diferentes y que los profagos satélites que en nuestro estudio denominados incompletos, son considerados una familia aparte de elementos móviles²⁵.

La clasificación taxonómica de las familias de virus asociados al contenido profagico detectado se realizó partiendo de los profagos pertenecientes a la categoría completos, esto porque son el tipo de profagos que tiene mayor cantidad de proteínas, lo que da una mayor certeza que lo identificado corresponde al taxón viral asociado. Referente a la familia de virus predominante entre los profagos detectados, se determinó que todos los profagos completos presentaban componentes estructurales relacionados con el orden *Caudovirales*, fagos con cabeza icosaédrica, DNA bicatenario y colas

largas que interactúan con las proteínas receptoras de la superficie bacteriana. De estos el 99,7% pertenecieron a la familia *Siphoviridae* fagos de colas no contráctiles, esta familia de virus ha sido identificada en múltiples estudios realizados en especies del género *Streptococcus* spp^{32,41,43}, y el 0,030% pertenecieron a la familia *Myoviridae* fagos de colas contráctiles, estos se han relacionada con otro tipo de bacterias hospederas como lo son *Burkholderia*, *Pseudomonas*, *Erwinia*⁴⁰, *Desulfovibrio*¹.

Los factores de virulencia y resistencia a antibióticos bacterianos están directamente relacionados con las capacidades funcionales de la bacteria y la presión selectiva del medio para el desarrollo de sus capacidades. En nuestros resultados los fagos no mostraron una fuerte asociación con la adquisición con mecanismos asociados la virulencia y resistencia a antibiótico. Los profagos pueden estar favoreciendo otros procesos celulares diferentes a los relacionados con la patogenicidad y resistencia bacteriana, por mencionar algunos de los procesos involucrados en base a bibliografía, se ha reportado su intervención de procesos fotosintéticos mediada por bacteriófagos marinos que infectan cianobacterias, los cuales intervienen de forma crucial la fotoinhibición de la célula infectada, lo que permite la continuidad de la fotosíntesis y proporcionar la energía necesaria para la bacteria y al virus para su replicación⁴³, en la modulación del genoma de bacterias no patógenas mediante la resistencia a temperaturas ambientales hostiles⁴⁴ como también se han considerado útiles en la modulación guiada para la implementación de tratamientos terapéuticos⁶. Asimismo, los plasmidos, transposones e islas genómicas representan una gran porción de genomas bacterianos que alteran significativamente la aptitud bacteriana y de los cuales puedan estar aportando de forma más significativa en los procesos de virulencia y resistencia⁴⁵.

Nuestro análisis del sistema de CRISPR-cas estableció la presencia de nucleasas efectoras en

el 59% de los genomas evaluados y entre los cuales se descubrió una variabilidad y diferencias estadísticamente significativamente entre la cantidad de profagos según el tipo de sistema presente. Se sugiere que aquellos genomas que contengan los sistemas I, II y III conjuntamente tienen una menor probabilidad de adquirir DNA extracelular de tipo profagico entre su genoma, esta información se explica debido a que los profagos con espaciadores que coinciden con estas secuencias son más susceptibles a ser detectados y controlados⁴⁶. La detección del sistema CRISPR de forma individual mostro una coexistencia con diferentes tipos de sistemas, Nazir y colaboradores (2012) proponen que dicha coexistencia puede ser consecuencia de un reordenamiento genético a nivel de los genomas bacterianos, pérdida de funcionalidad del sistema CRISPR a causa de su eliminación, un ensamblaje incompleto o deficiente del borrado del genoma⁴⁷, es importante también tener en cuenta que la metodología de búsqueda puede afectar el tipo de sistema CRIPR ya que hay algunos sistemas mayormente estudiados⁴⁸. Este tipo de sistema de inmunidad adaptativa bacteriana ha sido ampliamente estudiado en diferentes géneros bacterianos^{1,8,49,50}, sin embargo, no son los únicos sistemas asociados a este tipo de funciones. A la fecha se desarrollan estudios alternos que tiene como objetivo la identificación de nuevas estrategias de control de material exógeno. Ejemplo de este es el sistema de defensa DISARM sistema de isla de defensa asociado con restricción-modificación, quien altera el genoma en relación con secuencias específicas para evitar la adsorción de los fagos y al tiempo degrada el DNA extraño⁵¹, también el caso del suicidio celular programado o la latencia inducida tras la infección⁵² y además el sistema de infección abortiva o sistema de exclusión del fago el cual causa la muerte celular concomitantemente. En estudios comparativos se ha descrito que algunos genes o nucleasas están ubicados en sitios estratégicos del genoma enriqueciendo otros genes conocidos con funciones de defensa⁵³. Posterior al conocimiento

de este tipo de estrategias de inmunidad se llega a considerar que un genoma bacteriano puede estar codificando simultáneamente múltiples sistemas de defensa de varias clases como puede ser el caso del sistema CRISPR en coexistencia de otros sistemas⁵⁴.

Conflictos de interés.

Los autores no declaran tener ningún tipo de conflicto de interés.

Bibliografía.

1. Crispim, J., Dias, R., Vidigal, P.M., de Sousa, M.P., da Silva, C.C., Santana, M. y de Paula, S.O. (2018). Screening and characterization of prophages in *Desulfovibrio* genomes. *Scientific reports*, 8, 9273.
2. Brueggemann, A. B., Harrold, C. L., Javan, R. R., Van Tonder, A. J., McDonnell, A. J. y Edwards, B. A. (2017). Pneumococcal prophages are diverse, but not without structure or history. *Scientific reports*, 7(1), 1-13.
3. Brüßow, H. y Hendrix, R.W. (2002). Phage genomics: small is beautiful. *Cell*, 108(1), 13-16.
4. Krupovic, M., Prangishvili, D., Hendrix, R. W. y Bamford, D. H. (2011). Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol. Mol. Biol. Rev.*, 75(4), 610-635.
5. Fortier, L.C. y Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, 4(5), 354-365.
6. Canchaya, C., Proux, C., Fournous, G., Bruttin, A., y Brüßow, H. (2003). Prophage genomics. *Microbiol. Mol. Biol. Rev.*, 67(3), 473.

7. Lammoglia-Cobo, M. F., Lozano-Reyes, R., García-Sandoval, C. D., Avilez-Bahena, C. M., Trejo-Revelles, V., Muñoz-Soto, R. B. y López-Camacho, C. (2016). La revolución en ingeniería genética: sistema CRISPR/Cas. *Investigación en Discapacidad*, 5(2), 116-128.
8. Crawley, A. B., Henriksen, J. R. y Barrangou, R. (2018). CRISPRdisco: an automated pipeline for the discovery and analysis of CRISPR-Cas systems. *The CRISPR journal*, 1(2), 171-181.
9. Gao, X. Y., Zhi, X. Y., Li, H. W., Klenk, H. P. y Li, W. J. (2014). Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. *PloS one*, 9(6).
10. Hernández, G. C. y Guadarrama, N. V. (2007). Importancia clínica de *Streptococcus pneumoniae* tolerante a la vancomicina. *Revista de Enfermedades Infecciosas en Pediatría*, 20(79), 51-53.
11. Siboo, I. R., Bensing, B. A. y Sullam, P. M. (2003). Genomic organization and molecular characterization of SM1, a temperate bacteriophage of *Streptococcus mitis*. *Journal of bacteriology*, 185(23), 6968-6975.
12. Romero, A., López, R., Lurz, R. y García, P. (1990). Temperate bacteriophages of *Streptococcus pneumoniae* that contain protein covalently linked to the 5'ends of their DNA. *Journal of virology*, 64(10), 5149-5155.
13. Gindreau, E., López, R. y García, P. (2000). MM1, a temperate bacteriophage of the type 23F Spanish/USA multiresistant epidemic clone of *Streptococcus pneumoniae*: structural analysis of the site-specific integration system. *Journal of virology*, 74(17), 7803-7813.
14. Obregón, V., García, J. L., García, E., López, R. y García, P. (2003). Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. *Journal of bacteriology*, 185(7), 2362-2368.
15. Coordinators, N. R. (2017). Database resources of the national center for biotechnology information. *Nucleic acids research*, 45(Database issue), D12.
16. Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. y Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, 44(W1), W16-W21.
17. Fu, L., Niu, B., Zhu, Z., Wu, S. y Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150-3152.
18. Chen, L., Zheng, D., Liu, B., Yang, J. y Jin, Q. (2015). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic acids research*, 44(D1), D694-D697.
19. Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y, Yang B, Zhang S, Fa Z, Zhao B, Xiang Z, Li Y, Zhao Z, Olszewski MA, Chen L, and He Y. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acid Research*. 2019 Jan 8; 47(D1):D693-D700.
20. Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., McArthur, A.G. (2016). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic acids research*, 45(D1), D566-D573.
21. Sayers, E. W., Beck, J., Brister, J. R., Bolton, E. E., Canese, K., Comeau, D. C., ... & Kitts, P. A. (2020). Database resources of the national center for

- biotechnology information. *Nucleic acids research*, 48(D1), D9
22. Van der Mee-Marquet, N., Diene, S. M., Barbera, L., Courtier-Martinez, L., Lafont, L., Ouachée, A., ... Francois, P. (2018). Analysis of the prophages carried by human infecting isolates provides new insight into the evolution of group B *Streptococcus* species. *Clinical Microbiology and Infection*, 24(5), 514-521.
 23. Brown, L., Kim, J. H. y Cho, K. H. (2016). Presence of a prophage determines temperature-dependent capsule production in *Streptococcus pyogenes*. *Genes*, 7(10), 74.
 24. Romero, P., García, E. y Mitchell, T. J. (2009). Development of a prophage typing system and analysis of prophage carriage in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.*, 75(6), 1642-1649.
 25. Javan, R. R., Ramos-Sevillano, E., Akter, A., Brown, J. y Brueggemann, A. B. (2019). Prophages and satellite prophages are widespread in *Streptococcus* and may play a role in pneumococcal pathogenesis. *Nature communications*, 10(1), 4852.
 26. Burguener, G. F. (2017). *Desarrollo de herramientas bioinformáticas para la anotación de genomas* (Doctoral dissertation, Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales).
 27. Srividhya, K. V., Alaguraj, V., Poornima, G., Kumar, D., Singh, G. P., Raghavenderan, L. ... y Krishnaswamy, S. (2007). Identification of prophages in bacterial genomes by dinucleotide relative abundance difference. *PLoS One*, 2(11), e1193.
 28. Lima-Mendez, G., Van Helden, J., Toussaint, A. y Leplae, R. (2008). Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics*, 24(6), 863-865.
 29. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. y Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic acids research*, 39(suppl_2), W347-W352.
 30. Donati, C., Hiller, N. L., Tettelin, H., Muzzi, A., Croucher, N. J., Angiuoli, S. V., ... Massignani, V. (2010). Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome biology*, 11(10), R107.
 31. Fu, T., Fan, X., Long, Q., Deng, W., Song, J. y Huang, E. (2017). Comparative analysis of prophages in *Streptococcus mutans* genomes. *PeerJ*, 5, e4057.
 32. van der Ploeg, J. R. (2008). Characterization of *Streptococcus gordonii* prophage PH15: complete genome sequence and functional analysis of phage-encoded integrase and endolysin. *Microbiology*, 154(Pt 10), 2970-2978.
 33. Holt, R. G., Abiko, Y., Saito, S., Smorawinska, M., Hansen, J. B. y Curtiss, R. (1982). *Streptococcus mutans* genes that code for extracellular proteins in *Escherichia coli* K-12. *Infection and immunity*, 38(1), 147-156.
 34. Nanda, A. M, Thormann, K. y Frunzke, J. (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol*, 197(3), 410-419.
 35. Krylov, V. N. (2014). Bacteriophages of *Pseudomonas aeruginosa*: long-term prospects for use in phage therapy. *Adv Virus Res*, 88, 227-278.
 36. Górski, A., Międzybrodzki, R., Łobocka, M., Głowacka-Rutkowska, A., Bednarek, A., Borysowski, J. ... Esquemas J. (2018).

- Phage therapy: what have we learned? *Viruses*, 10(6), 288.
37. Roszniowski, B., McClean, S. y Drulis-Kawa, Z. (2018). Burkholderia cenocepacia Prophages—Prevalence, Chromosome Location and Major Genes Involved. *Viruses*, 10(6), 297.
 38. Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far?. *Molecular microbiology*, 49(2), 277-300.
 39. Bobay, L. M., Touchon, M. y Rocha, E. P. (2014). Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences*, 111(33), 12127-12132.
 40. Pratama, A. A., Chaib De Mares, M. y Van Elsas, J. D. (2018). Evolutionary history of bacteriophages in the genus Paraburkholderia. *Frontiers in microbiology*, 9, 835.
 41. Domelier, A. S., Van Der Mee-Marquet, N., Sizaret, P. Y., Héry-Arnaud, G., Lartigue, M. F., Mereghetti, L. y Quentin, R. (2009). Molecular characterization and lytic activities of Streptococcus agalactiae bacteriophages and determination of lysogenic-strain features. *Journal of bacteriology*, 191(15), 4776-4785.
 42. Ouennane, S., Leprohon, P. y Moineau, S. (2015). Diverse virulent pneumophages infect Streptococcus mitis. *PloS one*, 10(2), e0118807.
 43. Mann, N. H., Cook, A., Millard, A., Bailey, S. y Clokie, M. (2003). Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature*, 424(6950), 741.
 44. Bramucci, M. G., Keggins, K. M. y Lovett, P. S. (1977). Bacteriophage PMB12 conversion of the sporulation defect in RNA polymerase mutants of *Bacillus subtilis*. *Journal of virology*, 24(1), 194-200.
 45. Juhas, M. (2015). Horizontal gene transfer in human pathogens. *Crit Rev Microbiol*, 41(1), 101-108.
 46. Touchon, M., Bernheim, A. y Rocha, E. P. (2016). Genetic and life history traits associated with the distribution of prophages in bacteria. *The ISME journal*, 10(11), 2744-2754.
 47. Nazir, R., Zhang, M., de Boer, W. y van Elsas, J. D. (2012). The capacity to comigrate with *Lyophyllum* sp. strain Karsten through different soils is spread among several phylogenetic groups within the genus Burkholderia. *Soil Biology and Biochemistry*, 50, 221-233.
 48. Makarova, K. S. y Koonin, E. V. (2015). Annotation and classification of CRISPR-Cas systems. *Methods Mol Biol*, 1311, 47-75.
 49. Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., ... Nakagawa, I. (2011). CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. *PloS one*, 6(5), e19543.
 50. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., ... Koonin E. V. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13(11), 722-736.
 51. Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G. y Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature microbiology*, 3(1), 90-98.
 52. Labrie, S. J., Samson, J. E. y Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, 8(5), 317-327.

53. Makarova, K. S., Wolf, Y. I., van der Oost, J. y Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology direct*, 4(1), 29.
54. Makarova, K. S., Wolf, Y. I., Snir, S. y Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of bacteriology*, 193(21), 6039-6056.