



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS  
Centro de Investigación en Dinámica Celular  
Área Terminal de Bioquímica y Biología Molecular

# **METAGENÓMICA VIRAL DE AMBIENTES EXTREMOS E IDENTIFICACIÓN DE SUS POSIBLES HOSPEROS MEDIANTE CRISPR CAS**

## **TESIS**

Que presenta:

**MARÍA GUADALUPE AYALA RODRÍGUEZ**

Para obtener el grado de  
LICENCIADO EN CIENCIAS CON ÁREA TERMINAL EN  
BIOQUÍMICA Y BIOLOGÍA MOLECULAR

Director de tesis:  
Dr. Sonia Dávila Ramos

Sinodales:

Dr. Agustín Pérez Reyes UAEM (Presidente).

Dr. Ramón Alberto Batista García UAEM (Secretario).

Dra. Sonia Dávila Ramos UAEM (Vocal).

Dr. Hugo Gildardo Castelán Sánchez CONACYT  
(Suplente).

Dr. Luis Fernando Lozano Aguirre Beltrán CCG-UNAM  
(Suplente).

## **Agradecimientos**

Dra. Sonia Dávila Ramos gracias por su orientación y atención durante este proceso, además gracias permitirme desarrollar este proyecto, por confiar en mis habilidades y por brindarme las oportunidades para crecer.

Dr. Hugo Castelán Sánchez gracias por su asesoría, paciencia, dedicación, disposición, apoyo bioinformático y por compartir su conocimiento conmigo. Sin su ayuda, este logro no habría sido posible. Muchas gracias.

Dr. Agustín Pérez Reyes por sus consejos, apoyo personal, por las sugerencias recibidas.

Dr. Luis Fernando Lozano Aguirre Beltrán por las observaciones y orientación brindada.

Dr. Ramón Alberto Batista García por sus clases fueron fundamentales para desarrollar mi interés en el tema.

A Frida por iluminar cada momento, acompañarme en cada uno de mis logros, llantos y alegrías, a Luz por estar dispuesta escucharme, por tu amistad. A Jess por creer en mí. A Luis por apoyarme y la confianza otorgada.

A Tae, por recordarme que el amor empieza por quererse a uno mismo. Saranghae.

A mis papás que siempre estuvieron ahí para entenderme, a mi mamá por su paciencia, comprensión y cariño, a mi papá por apoyarme incondicionalmente. A mi hermano por acompañarme.

## **Resumen**

El estudio de los ambientes extremos ha cobrado gran interés debido a la presencia de organismos conocidos como extremófilos, los cuales habitan en ambientes con condiciones físicas, químicas y geológicas extremas.

Dentro de estos ambientes extremos se encuentran los ambientes hipersalinos los cuales presentan una concentración de sales superior a la salinidad del mar. Los ambientes salinos son similares en todo el mundo, difieren en nutrientes y condiciones climáticas; los organismos tienen adaptaciones que les permiten crecer de manera óptima en condiciones extremas físicas, químicas y geológicas (temperatura, presión, radiación, alta salinidad, metales pesados, pH, oxígeno, gases) estas adaptaciones se relacionan con las funciones metabólicas, derivadas de modificaciones enzimáticas y en la membrana plasmática.

Para comprender mejor la biodiversidad en ambientes hipersalinos, la metagenómica se ha convertido en la técnica de elección. Esto se debe a que no requiere el aislamiento y cultivo previo de los organismos, lo que resulta más eficiente en comparación con otras técnicas. Sin embargo, debido a las condiciones fisicoquímicas únicas de estos ecosistemas, puede ser difícil crear las condiciones ideales para el crecimiento de los organismos y para el establecimiento de los hospederos de los virus que forman parte de la comunidad.

El metagenoma se comprende como el total de material genómico de una muestra ambiental, con esta información es posible identificar y conocer la

estructura de las comunidades microbianas, así como los genes que las componen, resultando ser la estrategia ideal para el conocimiento de comunidades de microorganismos y virales en estos ambientes en particular.

En este trabajo se realizó una búsqueda de posibles hospederos procariontes de virus utilizando distintas estrategias bioinformáticas para identificar las secuencias crisper y la relación hospedero-virus que existen en estos ambientes

## **Lista de siglas, símbolos y abreviaturas.**

ARN: Ácido ribonucleico

ARNt: Ácido ribonucleico de transferencia

ARNr: Ácido ribonucleico ribosomal

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

DNA: Ácido desoxirribonucleico

UV: Radiación ultravioleta

pH: Potencial de hidrógeno

M: Molaridad

NaCl: Cloruro de sodio

NGS: Next Generation Sequencing

PHASTER : Phage Search Tool Enhanced Release

SRA: Sequence Read Archive

## **Lista de figuras**

Figura 1 Ambientes extremos alrededor del mundo.

Figura 2 Mecanismos de adaptación extremófilos de los extremófilos en respuesta al estrés ambiental.

Figura 3 Distribución global de lagos hipersalinos.

Figura 4 Flujo de trabajo de metagenómica.

Figura 5 Ejemplos de familias de fagos agrupados en función de su material genético.

Figura 6 Ubicación geográfica de las muestras metagenómicas de ambientes hipersalinos.

Figura 7 Asignación taxonómica de orden jerárquico.

Figura 8 Análisis de diversidad.

Figura 9 Análisis de clúster.

Figura 10 Abundancia relativa de ambientes hipersalinos. Especie más abundantes Halovirus.

Figura 11 Análisis de similitud de genes.

## **Lista de Tablas**

Tabla 1 Clasificación de halófilos.

Tabla 2 Acceso SRA de las secuencias metagenómicas de NCBI.

Tabla 3 Tabla de Completeness

Tabla 5 Identificación de hospederos por distintas herramientas bioinformáticas

## Índice general

	Páginas
Capítulo 1. Introducción	1
1.2 Ambientes extremos	1
1.3 Microorganismos extremófilos	2
1.3.1 Adaptaciones celulares y moleculares	4
1.4 Ambientes hipersalinos	7
1.4.1 Microorganismos halófilos	8
1.5 Estudio de muestras ambientales	9
1.6 Metagenómica	9
1.6.1 Análisis metagenómico	11
1.6.2 Metagenómica viral	12
1.7 Virus en ambientes extremos	13
1.8 Búsqueda de hospederos virales	13
1.9 CRISPR- Cas	14
Capítulo 2. Fundamentación teórica	16
2.1 Antecedentes	16
2.2 Justificación	17
2.3 Hipótesis	18
2.4 Objetivo general y objetivos particulares	18
Capítulo 3. Metodología	
3.1 Obtención de muestras	19
3.2 Análisis de control de calidad	20
3.3 Recorte de calidad de las secuencias y remoción de adaptadores	21



3.4	Ensamble de metagenomas	21
3.5	Identificación de genomas virales	21
3.6	Identificación de profagos	22
3.7	Evaluación de genomas	22
3.8	Predicción de genes	22
3.9	Evaluación comparativa de genes	23
3.10	Asignación taxonómica	23
3.11	Identificación de hospederos	24
3.11.1	Identificación de operones CRISPR CASy relación de hospederos	24
3.12	Clasificación taxonómica y comparación de genomas	24
3.13	Abundancia relativa	25
Capítulo 4.	Resultados y discusión de resultados	25
4.1	Clasificación taxonómica y evaluación de la diversidad.	25
4.1.2	Análisis de diversidad	28
4.1.3	Análisis de agrupamiento	29
4.2	Comparación de datos metagenómicos.	30
4.3	Identificación fagos y similitud de genes	32
4.4	Identificación de posibles hospederos	34
4.5	Identificación de hospederos mediante estrategias bioinformáticas	35
Capítulo 5.	Conclusiones	37
Referencias		38

## **Capítulo 1. INTRODUCCIÓN**

Los ambientes hipersalinos son aquellos en los que la concentración de sales en el agua es significativamente más alta que la del agua del mar. Estos ambientes pueden encontrarse en todo el mundo, incluyendo regiones desérticas, lagunas costeras, salinas y pantanos salinos.

La concentración de sales es mayor al promedio que tiene el agua de mar la cual es de aproximadamente del 3.5 %, donde los iones más abundantes son el cloro y sodio, aunque también se encuentran en concentraciones significativas otros iones como el magnesio, calcio y potasio.

Para conocer la ecología microbiana se han empleado técnicas como la metagenómica mediante la secuenciación masiva, esto ha permitido analizar la composición de los organismos presentes sin necesidad de cultivarlos por lo cual se da la detección de microorganismos que no son identificados mediante las técnicas tradicionales y la posibilidad de analizar una gran cantidad de muestras de forma simultánea.

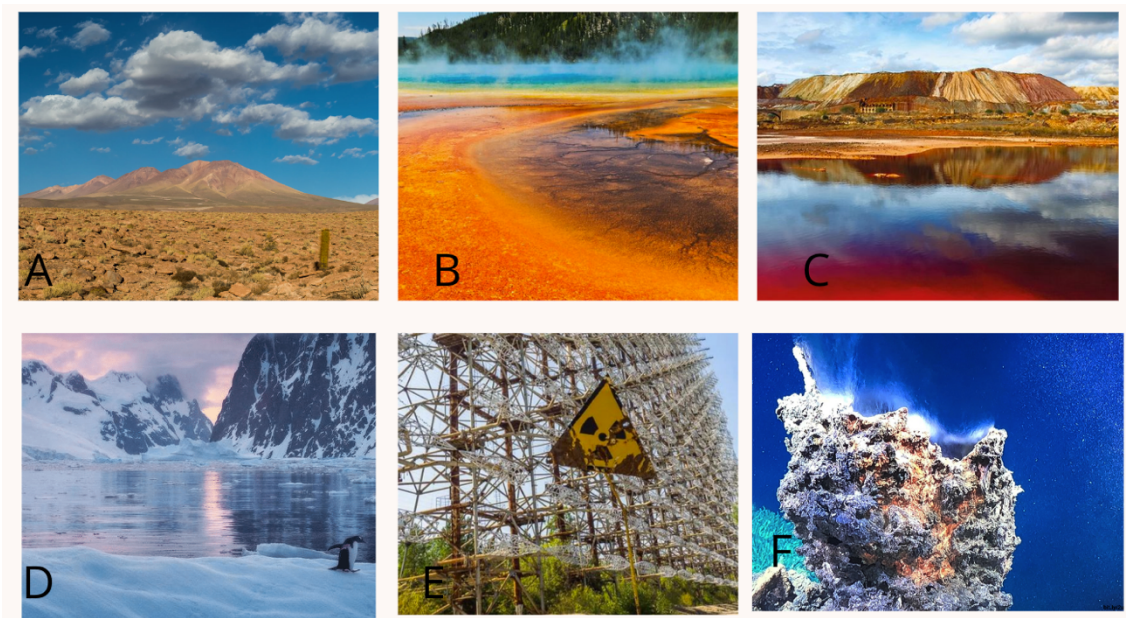
Por lo anterior el presente trabajo está enfocado al estudio de ambientes hipersalinos empleando la metagenómica y herramientas bioinformáticas para la identificación de virus y a su vez poder identificar a los posibles hospederos.

### **1.2 Ambientes extremos**

La vida en la Tierra se encuentra en donde están presentes las condiciones favorables que permiten el desarrollo y crecimiento de los organismos, así los

parámetros fisicoquímicos delimitados por rangos específicos, definen los distintos ambientes donde puede existir la vida.

Nuestro planeta incluye distintos ambientes relacionados con valores extremos que pueden incluir tensiones físicas, como la temperatura o radiación; geoquímicas, como la desecación o salinidad y/o biológicas, como la disponibilidad limitada de nutrientes (Rothschild & Mancinelli, 2001), que definidos desde un punto de vista antropocéntrico son aquellos ambientes donde el hombre no podría vivir (Figura 1).



**Figura 1.** Ambientes extremos alrededor del mundo A) Desierto de atacama Chile, B) Yellowstone, Estados Unidos C) Río Tinto España, D)Antártida E) Chernobyl, Ucrania F) Ventilales hidrotermales.

### 1.3 Microorganismos extremófilos

Los microorganismos extremófilos viven en ambientes donde a primera vista parece no existir la vida. Sin embargo, en las últimas décadas se han

descubiertos organismos capaces de vivir bajo condiciones extremas de temperatura, pH, radiación UV, y salinidad, por poner algunos ejemplos. Conocer a estos organismos que desafían los límites de la vida, es importante ya que permite entender algunos procesos evolutivos que se dan en las condiciones ambientales que se asemejan a la Tierra primitiva; o conocer los procesos biológicos que les permiten establecerse de manera óptima en estos ambientes, lo cual podría ser interesante para algunos desarrollos biotecnológicos.

Los extremófilos incluyen miembros de los tres dominios de la vida. La mayoría de los extremófilos son microorganismos, como las bacterias, arqueas y eucariotas como protistas y organismos multicelulares (p. ej., algas, hongos y protozoos) (Rampelotto, 2013).

Los organismos extremófilos son clasificados dependiendo del ambiente en donde se dan las condiciones óptimas para su crecimiento en:

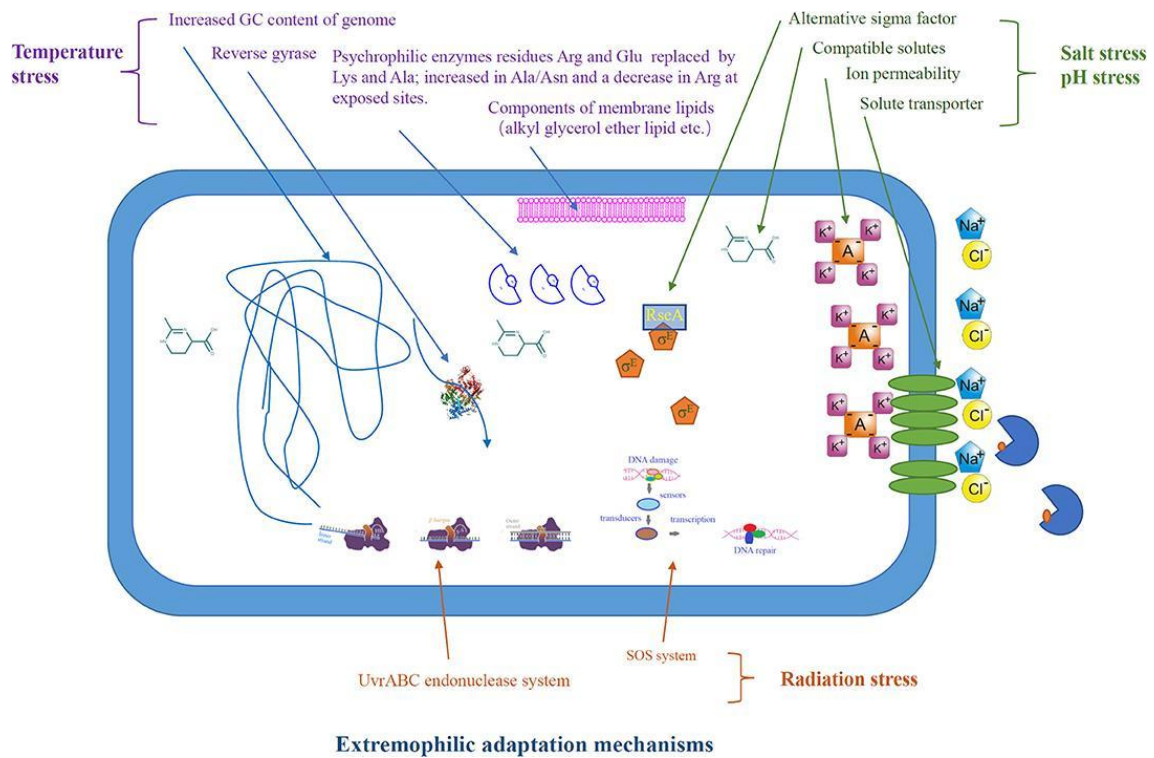
- Psicrófilos: Se encuentran en ambientes con temperatura  $< 15\text{ }^{\circ}\text{C}$ . Los psicrófilos se encuentran en ambientes de baja temperatura, como océanos polares, lagos y permafros (Williams et al., 2014).
- Termófilos: Los termófilos se encuentran en fuentes termales volcánicas y respiraderos hidrotermales oceánicos con un rango de temperatura  $60\text{ }^{\circ}\text{C}$  termófilos y  $80\text{ }^{\circ}\text{C}$  hipertermófilos.
- Acidófilos y alcalófilos: Se desarrollan en  $\text{pH} < 3$  para acidófilos y  $\text{pH} > 9$  alcalófilos. Los acidófilos se encuentran en aguas termales volcánicas y

drenajes ácidos de minas. Los alcalófilos se encuentran en lagos ricos en minerales y fuentes alcalinas hidrotermales (Banciu & Muntyan, 2015).

- Xerófilos: Los xerófilos se encuentran en ambientes con escasez de agua como son desiertos cálidos y fríos, a menudo también tienen la propiedad de ser radio-resistentes y tener una actividad de agua menor a 0.8 aw.

### 1.3.1 Adaptaciones moleculares y celulares

Para sobrevivir a estos ambientes extremos, los organismos extremófilos han desarrollado varias estrategias para lidiar con él.



**Figura. 2** Mecanismos de adaptación de los extremófilos en respuesta al estrés ambiental (Daochen Zhu et al., 2020).

Los psicrófilos viven en condiciones de frío extremo, estos han desarrollado mecanismos celulares de adaptabilidad al frío, como: la regulación de proteínas de choque, pequeñas proteínas de unión a ARN y sustancias poliméricas extracelulares para proteger a las células contra la ruptura mecánica de la membrana celular causada por la baja temperatura. Además, el genoma de los psicrófilos contiene regiones ricas en G+C superiores a las de otros organismos, también codifican para ARNt, factores de elongación y ARN polimerasas, por otro lado, se ha establecido una relación entre la presencia de plásmidos y elementos genéticos transponibles/móviles con la biosíntesis de ácidos grasos insaturados. Un ejemplo es la bacteria *Polaromonas vacuolata* aislada en la Antártida, la cual tiene una temperatura óptima de crecimiento de 4 °C (Derui Zhu et al., 2020).

Los microorganismos termófilos son capaces de restaurar su estructura y función proteica mediante la producción de chaperonas o termosomas, que les permiten resistir la destrucción de proteínas por las altas temperaturas (Annamalai et al., 2016). Para resistir el despliegue de proteínas causado por las altas temperaturas, las bacterias termófilas han desarrollado adaptaciones como es el caso de las bacterias termófilas presentan una mayor cantidad de puentes salinos y/o puentes disulfuro adicionales, lo que contribuye a la formación de una estructura más estable (Derui Zhu et al., 2020). Esta

adaptación asegura la estabilidad térmica manteniendo el transporte de nutrientes y sus actividades básicas. *Pyrococcus furiosus* es un archae que crece óptimamente a una temperatura de 100 °C y puede sobrevivir a temperaturas de 120 °C, sintetiza enzimas que la vuelven estable, estas a su vez han sido utilizadas en aplicaciones para catalizar reacciones en algunos bioprocesos que tienen que llevarse a cabo a altas temperaturas.

Con el fin de sobrevivir, los acidófilos, mantienen un pH neutro utilizando una bomba de flujo de protones y membranas deflectoras de protones (Siliakus et al., 2017), mientras que los halófilos utilizan anti portadores de  $Na^+/H^+$  lo cual permite acumular protones intracelulares y debido a eso mantiene un pH neutro. *Thiobacillus ferrooxidans* tiene la capacidad de metabolizar hierro y azufre en ácido sulfúrico lo que ayuda a equilibrar el pH.

Los xerófilos usan adaptaciones similares a los halófilos de acumulación de osmolitos compatibles como son la glicina, betaína, glicerol. Además, los xerófilos a menudo son radiorresistentes, ya que al parecer comparten los mecanismos para lidiar con la desecación, la resistencia a la radiación ionizante y estrés oxidativo (Hallsworth, 2018).

Un ejemplo de este tipo de extremófilo es la bacteria *Deinococcus radiodurans* la cual contiene enzimas, y otras moléculas especializadas en reparar daños en la estructura del DNA, que pueden desenrollar, cortar, pegar y reajustar la doble hélice del DNA.

Los halófilos tienen estrategias que permiten a organismos conservar el agua que pierden por osmosis. Una adaptación en arqueas es la acumulación de

altos niveles de KCL en su citoplasma utilizando un antiportador  $K^+/Na^+$  (Gunde-Cimerman et al., 2018), de igual manera la presencia de pigmentos asociados a la membrana fotoprotectora, protegen contra el daño frecuente de los rayos UV en ambientes hipersalinos. *Haloferax volcanii* es una archaea predominante en el Mar Muerto, algunas de sus estrategias adaptativas han permitido conocer más sobre la vida en ambientes con estas características y hacer una extrapolación a otros planetas como en Marte donde la alta salinidad y la radiación están presentes.

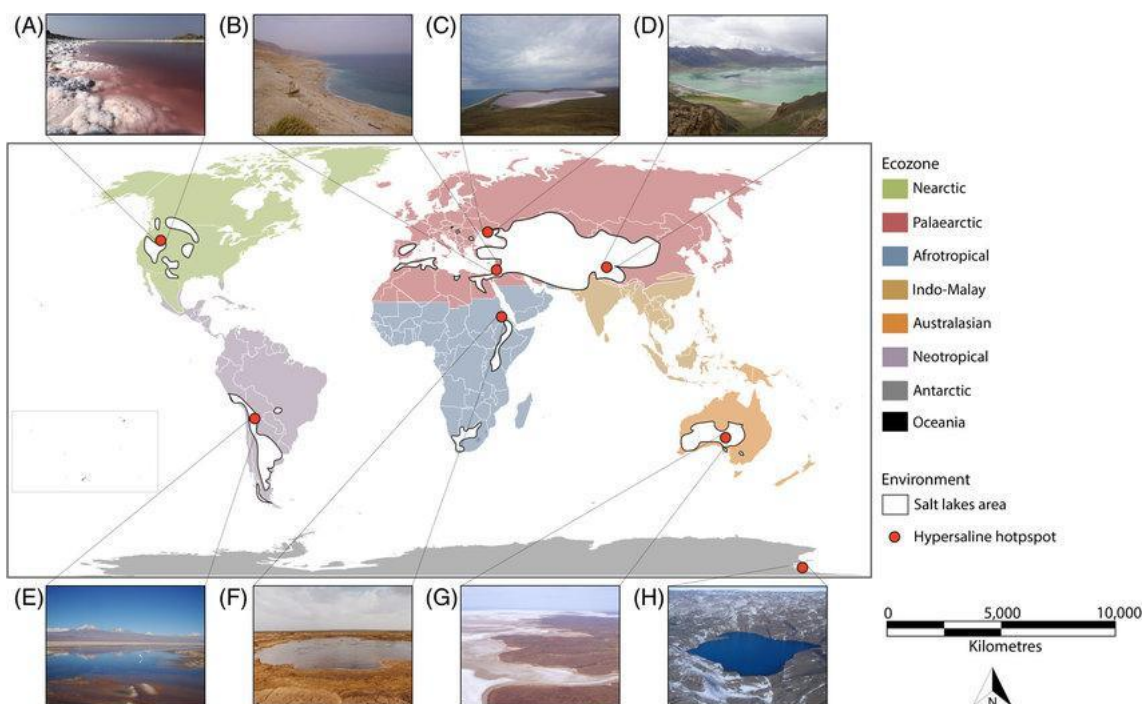
### **1.3 Ambientes hipersalinos**

Los ambientes hipersalinos son aquellos que tienen una concentración mayor de sales comparadas con el mar (35 g/l), los dos elementos principales en el agua del mar son el cloro y sodio, algunos fenómenos naturales hacen que aumente o disminuya el porcentaje de salinidad como lo son la evaporación, las erupciones volcánicas, el deshielo, entre otros. La salinidad del agua del mar varía en las diferentes latitudes de la tierra, siendo del 1- 5% p/v, en la región cercana a los trópicos donde la evaporación es mayor debido a la radiación solar, mientras que en ambos hemisferios que son regiones frías la salinidad es menor.

Los ambientes hipersalinos se forman a través de diferentes procesos naturales. Puede ser en regiones de lagos con minerales abundantes, o en zonas donde el agua del océano se evapora, o, en la superficie de minas de sal o por depósitos salinos.



Debido a su origen, los ambientes salinos se clasifican en talasohalinas, cuyo origen se da por la evaporación del agua de mar, las cuales reflejan la composición iónica del agua de mar y tienen un pH de casi neutro a ligeramente alcalino (Ma et al., 2010). Las atalasohalinas que no derivan del agua del mar, contienen altas concentraciones de iones,  $\text{Na}^+$ ,  $\text{Mg}^{2+}$  y  $\text{Ca}^{2+}$  con un pH ligeramente ácido. La composición iónica puede variar de acuerdo con las características biogeográficas.



**Figura 3** Distribución global de lagos hipersalinos (Saccò et al., 2021).

Los lagos salinos cuentan con (~44%) del volumen total del agua del planeta, y se localizan preferencialmente en el hemisferio norte (Figura 3).

La presencia de altos niveles de sal en los ambientes acuáticos puede tener un impacto significativo en la estructura y función de la membrana celular de los organismos que habitan en ellos, esto a su vez puede afectar la diversidad de

especies presentes en estos ambientes, lo que es crucial para el mantenimiento y desarrollo de los ecosistemas acuáticos.

#### **1.4.1 Microorganismo halófilos**

En estos ambientes aparentemente poco favorables para el desarrollo de la vida se encuentra una gran diversidad bacteriana. Por ejemplo, 42 de los principales filos bacterianos y 15 nuevos filos candidatos fueron reportados en un tapete microbiano en Guerrero Negro (Ley et al., 2006).

Los organismos que necesitan la presencia de sal para un desarrollo óptimo son denominados halófilos, este término tiene su origen del griego hals (sal) y filis (amante), los amantes de la sal.

Clasificación	Concentración NaCl % p/v
Halófilos débiles	0.2-0.5 M 1-3 %
Halófilos moderados	0.5-2,5M 3-15%
Halófilos extremos	2,5M-5,0M 15-30%
Halotolerantes	0.5-5,0 < 1% grandes conc. De sal

**Tabla 1.1** Clasificación de halófilos

#### **1.5 Estudio de muestras ambientales**

El estudio de organismos provenientes de ambientes extremos ha sido un desafío a distintos niveles, en primer lugar la obtención de las muestras debido al acceso a esos ambientes tan precarios, seguido por el aislamiento y cultivo de los organismos que también resulta problemático ya que es difícil emular las condiciones de cultivo óptimas para su crecimiento. Sin embargo, el

desarrollo de estrategias moleculares y bioinformáticas nos ha llevado a poder explorar las comunidades presentes en estos ambientes, evitando el paso de aislamiento y cultivo.

## **1.6 Metagenómica**

En las últimas décadas la metagenómica se ha convertido en una herramienta de gran utilidad debido a que permite recuperar genomas de microorganismos a partir de muestras ambientales, permitiendo conocer la estructura de la comunidad presente en las distintas muestras ambientales sin necesidad de cultivarlas. El desarrollo de herramientas bioinformáticas nos da la oportunidad de clasificar taxonómicamente los organismos presentes. La metagenómica por shotgun proporciona acceso a la composición de genes funcionales de las comunidades microbianas a comparación de la metagenómica por profiling realizada con el gen 16S rRNA por ejemplo (Thomas et al., 2012).

El análisis de datos metagenómicos permite identificar nuevos organismos y secuenciar genomas completos, a partir de las lecturas ensambladas de las muestras del estudio.

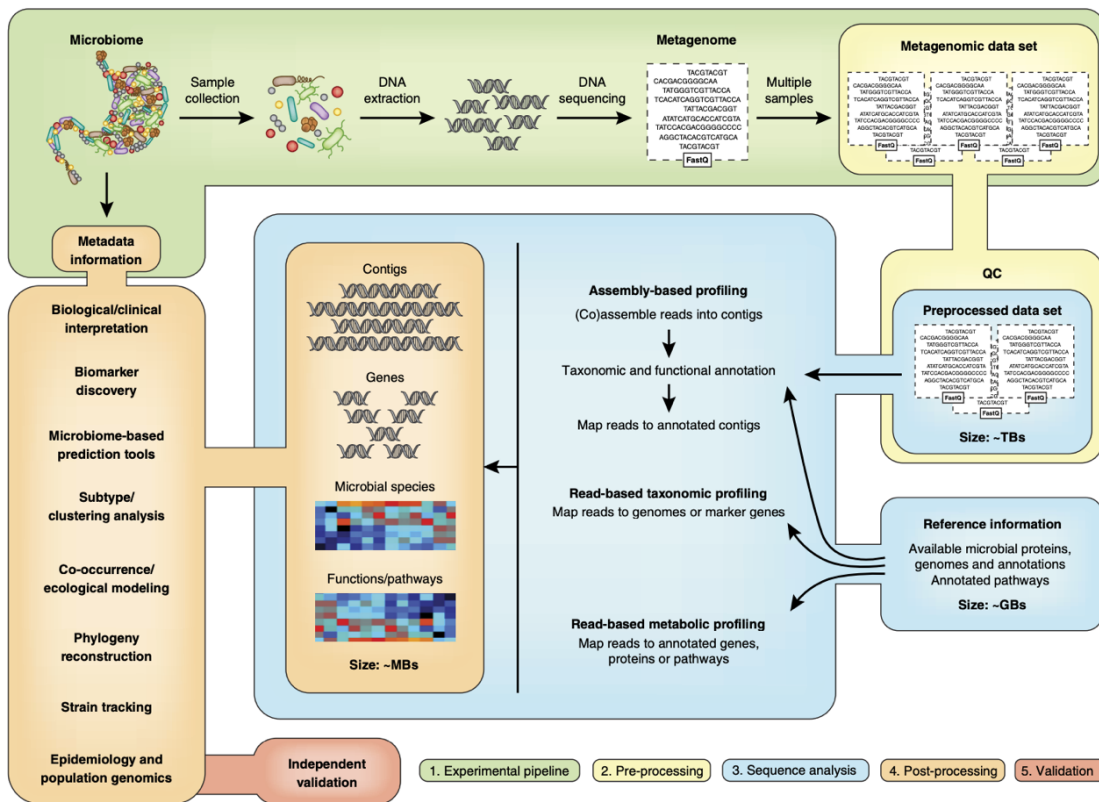
La metagenómica por shotgun es un enfoque amplio que se utiliza para investigar comunidades microbianas complejas. Difiere de la metagenómica de profiling, que solo ofrece un estudio filogenético limitado a bacterias y arqueas a partir de un solo gen, como el ARN ribosomal 16S. La metagenómica de profiling permite estudiar la abundancia relativa de organismos en muestras similares, pero no permite cuantificar la abundancia absoluta, por lo que no es suficiente para una investigación exhaustiva. La metagenómica por shotgun,

por otro lado, permite un análisis más detallado y profundo de la composición de la comunidad microbiana.

La secuenciación genómica de shotgun también utiliza técnicas de secuenciación masiva de última generación (NGS) la cual proporciona las secuencias de los genomas de todos los organismos presentes en la muestra ambiental determinada, siendo eficaz para determinar tanto la taxonomía como el potencial funcional de las comunidades microbianas.

### **1.6.1 Análisis metagenómico**

Los análisis metagenómicos tienen los pasos generales: recolección de muestras, se toman muestras de la comunidad microbiana de interés, que pueden incluir suelos, agua, aire, alimentos, entre otros. Preparación de la muestra, la muestra se somete a una serie de procesos para separar y purificar el material genético. Secuenciación del material genético utilizando tecnologías modernas, como la secuenciación por shotgun. Pre-procesamiento de los datos obtenidos de la secuenciación, incluyendo la eliminación de secuencias no deseadas, la corrección de errores y la asignación de cada secuencia a su organismo fuente. Análisis funcional para identificar las funciones y características de las especies presentes en la comunidad. Análisis filogenético permite identificar la relación evolutiva entre los organismos. Interpretación de resultados, comprensión detallada de la composición y función de la comunidad microbiana.



**Figura 4.** Descripción de flujo de trabajo de metagenómica

### 1.6.2 Metagenómica viral

Los virus son las entidades biológicas más abundantes en la tierra. La identificación de nuevos virus se ha visto limitada durante años ya que las investigaciones se sesgan a la búsqueda de patógenos, principalmente de humanos, dejando de lado los demás virus (Alavandi & Poornima, 2012)

La metagenómica viral es la aplicación de la genómica ambiental en virus, que necesita una combinación de tecnologías innovadoras como enriquecimientos virales, PCR aleatoria, secuenciación masiva (NGS) y/o secuenciación de lecturas largas.

En los últimos años por medio de la metagenómica viral la estructura de la comunidad y la biogeografía de los virus ambientales ha ampliado el conocimiento de los viomas pero poco se conoce sobre las interacciones virus-hospedero (Suttle, 2005).

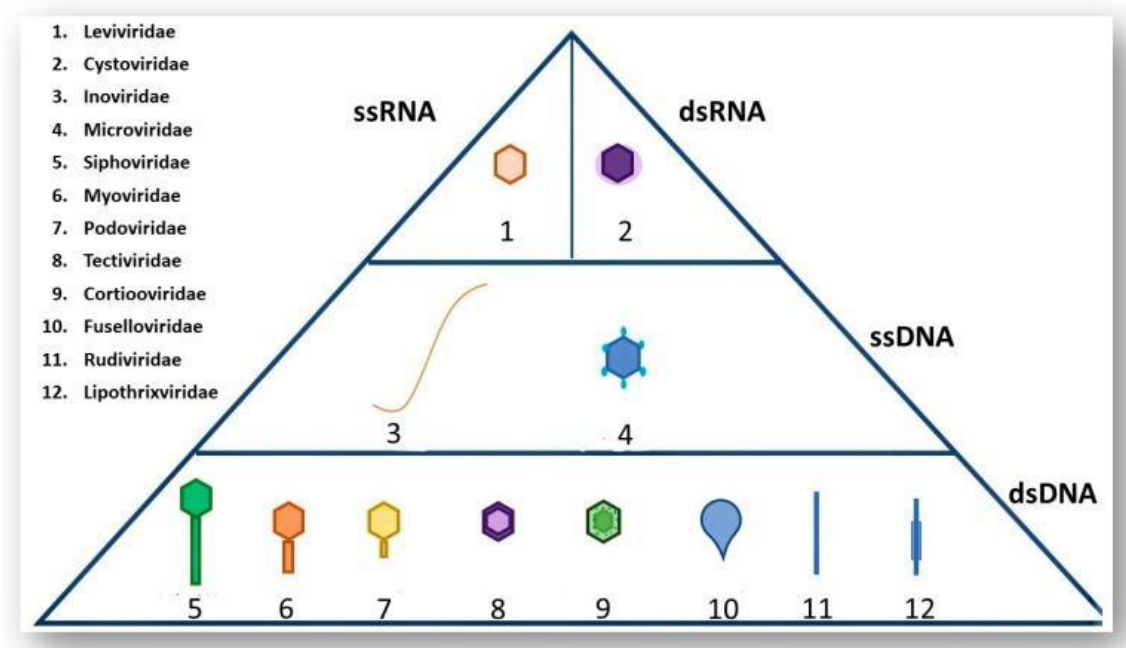
### **1.8 Virus en ambientes extremos**

Organismos extremófilos de los tres dominios de la vida, Bacteria, Archaea y Eukarya, pueden actuar como hospederos para la replicación viral. Esta interacción ayuda a mantener el equilibrio en las poblaciones microbianas de diferentes ambientes y contribuye con genes que mejoran la adaptabilidad al medio y participan en los ciclos biogeoquímicos de manera significativa.

A pesar de la dificultad que representa identificar virus, se han logrado identificar por microscopio electrónico a virus que infectan a huéspedes halófilos, los cuáles suelen ser sensibles a fuerzas iónicas (Kukkaro & Bamford, 2009).

Los virus que infectan arqueas se han clasificado en 12 familias y 1 género no asignado. Los virus con forma de botella (familia *Ampullaviridae*), fusiformes (*Bicaudaviridae*, *Fuselloviridae* y género *Salterprovirus*), espirales (*Spiraviridae*) y gotitas (*Guttaviridae*), Estos tienen cápsides icosaédricas y colas contráctiles, no contráctiles o cortas, *Myoviridae*, *Siphoviridae* y *Podoviridae* pertenecen al orden de Cuadovirales (Krupovic et al., 2018).

Los genomas de los virus aislados contienen DNA el cual puede ser de cadena sencilla o doble, en forma lineal o circular.



**Figura 5** Ejemplos de familias de fagos agrupados en función de su material genético (Stone et al., 2019).

El Comité Internacional de Taxonomía de Virus (ICTV) utiliza las características de los fagos para su clasificación, en el cual se incluye la morfología del fago, además de las propiedades fisicoquímicas del virión.

### 1.8 Estrategias para identificar hospederos de virus

El descubrimiento metagenómico de nuevos virus, lleva a estudiar a ¿quién infecta?, la bioinformática ha facilitado la interpretación de los grandes conjuntos de datos asociados a la metagenómica mediante señales computacionales para identificar relaciones bacteriófago-hospedero, las estrategias a utilizadas comprenden el conocimiento de la abundancia de perfiles, este funciona por medio de patrones de abundancia bacterianos en los metagenomas para identificar su asociación por correlación (retardada);

homología genética, en el cual se evalúa la homología genética entre fagos y bacterias de nucleótidos y proteínas pueden representar secuencias adquiridas por un fago durante una infección anterior; CRISPR, este funciona cuando bacterias colocan un fragmento de 25 a 75 pb de una secuencia de fago infectante en CRISPR en su genoma. Estas matrices pueden identificar y alinear los espaciadores con los genomas de los fagos para detectar infecciones recientes.

Las coincidencias exactas entre genomas bacterianos pueden representar sitios de integración, espaciadores CRISPR regiones de homología genética; coincidencia de matches donde se encuentran las coincidencias exactas entre genomas bacterianos pueden representar sitios de integración, espaciadores CRISPR regiones de homología genética; perfiles de oligonucleótidos siendo que los fagos llevan genes de ARNt para alterar el perfil de uso de codones típico del huésped (Edwards et al., 2016)

## **1.9 CRISPR-Cas**

Los virus pueden infectar a la mayoría de los microorganismos, sin importar en el ambiente que se localicen, estas infecciones son un riesgo epidemiológico, debido a esto es importante conocer la interacción virus – hospedero.

CRISPR Cas se identificó por primera vez en E. Coli 1987, por sus siglas en inglés que significa: Repeticiones Palindrómicas Cortas Agrupadas y Regularmente interespaciadas también conocido como sistema inmunitario adaptativo procariótico, está presente en bacterias y arqueas (Ishino et al., 1987).



El sistema CRISPR Cas es un mecanismo de resistencia altamente adaptativo y hereditario que incorpora secuencias cortas de virus y otros elementos genéticos móviles en el locus CRISPR del huésped para transcribirlos y procesarlos pequeños ARN que guían la destrucción de los ácidos nucleicos invasores (Charpentier & Marraffini, 2014).

Los sistemas CRISPR-Cas se clasifican en dos clases Clase 1 y Clase 2, a su vez esta se dividen en subtipos Clase 1 tipo I, III y IV; Clase 2 tipo II, V y VI. (Makarova et al., 2020).

Utilizando el sistema CRISPR, las bacterias colocan un fragmento corto de una secuencia genómica de un fago infectante en el ADN del fago. Esta secuencia genómica del fago infectante, normalmente de 25-75 pares de bases (pb) como espaciador en una matriz CRISPR, una región de repetición reconocible en el genoma bacteriano. Esto da lugar a una firma de secuencia fago hospedero, lo cual se ha explotado para identificar interacciones fago hospedero en diversos sistemas (Edwards et al., 2016).

## Capítulo 2. FUNDAMENTACIÓN TEÓRICA

### 2.1 Antecedentes

En el artículo de Derui Zhu y cols. del 2020, se obtuvieron muestras de ambientes hipersalinos ubicados en China: lagos y salinas de los cuales estudiaron muestras de agua y sedimentos utilizando secuenciación Illumina de alto rendimiento de genes 16S rRNA con el cual se describe la diversidad en ambientes hipersalinos.

En el artículo de Santos y cols del 2010, se describe la diversidad de la microbiota de los ambientes marinos y acuáticos, en el cuál compara ambientes hipersalinos, siendo la microbiota dominante halófilos, bacterias cerca del 15% de la comunidad, la bacteria más abundante *Salinibacter*.

Además también se ha descrito los virus presentes en arqueas (Atanasova et al., 2015); los cuales han sido denominados halovirus dentro de estos están presentes los Myoviruses, Siphoviruses, Podoviruses, virus icosaédricos, virus polimórficos, virus con forma de limón.

Se ha relacionado la interacción fago hospedero (Edwards et al., 2016), por medio de predicciones computacionales, por lo cual es posible conocer la dinámica ecológica de esta relación al identificar matrices de CRISPR.

## **2.2 JUSTIFICACIÓN**

Los ambientes salinos, como los océanos y los hábitats hipersalinos, están extendidos en todo el mundo. Incluyen agua de mar, lagos salados, salinas solares o suelos hipersalinos. Sin embargo, se sabe poco sobre la diversidad viral, el rango de huéspedes y la geografía, lo que tiene implicaciones importantes sobre cómo los virus manipulan la función y la evolución microbiana.

Actualmente se conoce de manera limitada las interacciones virus-huésped en metagenomas microbianos derivados de distintos sitios de ambientes salinos.

Por consiguiente, analizar las muestras del ecosistema hipersalino mediante técnicas bioinformáticas que permiten conocer las secuencias virales y los espaciadores de repeticiones palindrómicas cortas CRISPR, proporciona datos sobre la diversidad, estructura, la relación virus huésped y su relación con el entorno.

## **2.3 HIPÓTESIS**

Análisis de secuencias virales a través de CRISPR CAS en ambientes hipersalinos en diferentes regiones geográficas puede ayudar a establecer las relaciones entre virus y hospedero.

## **2.4 OBJETIVO GENERAL**

Analizar la diversidad de la microbiota y diversidad viral e identificación de sus posibles hospederos a través CRISPR CAS en ambientes hipersalinos de regiones geográficamente distintas.

## **OBJETIVOS PARTICULARES**

- Analizar la microbiota bacteriana en distintos ambientes hipersalinos.
- Describir la diversidad viral en ambientes hipersalinos de diferentes lugares geográficos.
- Identificar los posibles hospederos presentes en ambientes hipersalinos por medio de CRISPR.

## Capítulo 3. METODOLOGÍA

### 3.1 Obtención de muestras

Se obtuvieron seis secuencias metagenómicas de ambientes hipersalinos por medio de Sequence Read Archive (SRA) disponible en línea en: <https://www.ncbi.nlm.nih.gov/sra> la cual es una base de datos pública que almacena datos de secuenciación de alto rendimiento sin procesar, el cual permite compartir datos de secuenciación de nueva generación, las muestras metagenómicas pertenecen a un ambiente hipersalino, de distintas regiones geográficas (Figura 6), península de España en dos regiones Santa Pola e Isla Cristina ubicados en las coordenadas 38.185388°, -0.626205°; 37.199444° -7.325278° respectivamente; Salinas Maras ubicada en Perú a -72.155556°, -13.300278° y Deep Lake con una profundidades de 8.0 um y 3.0 um en la Antártida a -68.540000°, - 78.180000° (Tabla 2). Posteriormente de ser descargados por medio de las herramientas en SRA Toolkit 2.9.6 se transformaron los archivos comprimidos SRA a fastq.



**Figura 6** Ubicación geográfica de las muestras metagenómicas de ambientes hipersalinos., A) Salinas Maras en Perú, B) Salinas, Isla Cristina y Santa Pola en la península de España C) Deep Lake en la Antártida.

### 3.2 Análisis de control de calidad

A las secuencias descargadas, se le lleva a cabo el análisis de control de calidad, esto se realiza con el fin de detectar aquellas secuencias que hayan sufrido, amplificación selectiva, errores de sustitución de bases, entre otras. Esta se evalúa con el software FastQC v0.11.9 disponible en línea en <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, el cual proporciona un informe detallado, mediante representación de gráficas y tablas, del cual permite exportar los resultados.

De cada informe a cada módulo se da indicador de aprobado, advertencia o fallo, esto se toma en cuenta antes de realizar análisis posteriores.

### **3.3 Recorte de calidad de las secuencias y remoción de adaptadores**

Remoción de adaptadores por medio del comando Trimmomatic, el cual funciona con archivos FastQ, se recortaron las regiones y los adaptadores de baja calidad con Trimmomatic versión 0.39 con parámetros ILLUMINACLIP: NexteraPE-PE.fa SLIDINGWINDOW:4:20 HEADCROP:10 MINLEN:75. Estos parámetros eliminan bases del final de una lectura si están por debajo de una puntuación Phred de 20, recortan la lectura si la puntuación Phred media dentro de una ventana deslizante de 4 bases avanzada desde el extremo 5' cae por debajo de 20 y especifican una longitud mínima de lectura de 75 bases.

### **3.4 Ensamble de metagenomas**

En este estudio se realizó por medio del ensamblaje de novo, este ensamble se realiza sin una secuencia de referencia, el cuál se realizó en todas las muestras metagenómicas con Megahit-1.0.6 disponible en línea en <https://de.cyverse.org/apps> este es un ensamblador NGS el cual está optimizado para datos metagenómicos, mediante el tamaño de k mers pequeños a de gran tamaño siendo el primero útil para filtrar los bordes erróneos y mientras que el de mayor tamaño permite resolver las repeticiones, con el parámetro de mínimo de lectura de contigs en 200.

### **3.5 Identificación de genomas virales**

Por medio de la herramienta VirSorter 1.0.3 disponible en línea en: <https://de.iplantcollaborative.org/de/> la cual permite la identificación de secuencias virales de DNA, el cual se emplea una base de datos llamada

RefSeqABVir en esta se realiza la búsqueda de genomas virales en los contigs la cuál contiene 114 297 proteínas de virus que infectan bacterias o arqueas, esta base de datos incluye biomas de muestras de aguas dulce, agua de mar, intestino humano, pulmón y saliva (Enright, 2002).

### **3.6 Identificación de profagos**

Se realizó la identificación de fagos en contigs ensamblados a partir de datos metagenómicos esto se realizó por medio de PHASTER (Phage Search Tool) disponible en línea en: <https://phaster.ca/>, la cuál es una versión mejorada de PHAST, permite a través del formato fasta por medio de una interfaz web realizar búsquedas BLAST en la base de datos de profagos de NCBI (National Center for Biotechnology Information), de los genes completos, los parciales se predicen utilizando FragGeneScan, este encuentra genes fragmentados en lecturas cortas, estos profagos predichos se organizan por contig en los resultados presentados (Arndt et al., 2016).

### **3.7 Evaluación de calidad de genomas**

Se estimó la calidad del genoma y la contaminación del genoma , a partir de los contigs de VirSorter por medio del programa CheckV disponible en: <https://bitbucket.org/berkeleylab/checkv/src/master/> se utiliza debido a que este programa permite eliminar regiones virales en los bordes de los contigs lo que se logra es estimar la integridad de los fragmentos encontrados en estos genomas, calidad alta, calidad media y calidad baja, lo cual brinda la información mínima sobre un genoma de virus no Cultivado (MIUViG).



### **3.8 Predicción de genes**

Se ocupa la predicción de genes pharokka disponible en: <https://github.com/gbouras13/pharokka> este está diseñado para la identificación de bacteriofagos, por medio de homología con genes conocidos, de este se obtuvo un archivo GFF (SignalMap Gene-Finding Format) con el cuál se realiza el análisis comparativo de genes.

### **3.9 Evaluación comparativa de genes**

Los grupos de genes generalmente se conservan hasta cierto punto en todos los taxones, y el análisis comparativo puede revelar información importante sobre su evolución y las posibles diferencias en las vías que codifican, con este se obtiene la forma de realizar la visualización de datos por medio de Clinker disponible en: <https://github.com/gamcil/clinker> el cuál permite realizar la visualización de alineaciones de grupos de genes esto se realizó por medio de puntuación de identidad de secuencia acumulada.

### **3.10 Asignación taxonómica**

El genoma de virus consta de DNA o RNA en especial los que infectan a arqueas, se han descrito en mayor parte como virus de DNA, pero eso no descarta que se puedan encontrar virus de RNA, conocer su relación taxonómica es importante para la clasificación, a partir de las secuencias, se da la clasificación por medio de superposición de k-mer de una secuencia dada

con conjuntos discriminativos (conjuntos mutuamente excluyentes) de k-mers de géneros de virus conocidos. Se usó VirusTaxo programa el cual contiene información sobre 6950 genomas de virus que abarcan 129 familias y 682 géneros de virus de ADN y ARN que podrían ayudar a descubrir los taxones de virus no caracterizados que están relacionados con géneros de virus conocidos ( Raju).

### **3.11 Identificación de hospederos**

La identificación de posibles hospederos se realizó por medio de HostPhinder 1.1 disponible en línea en: <https://cge.food.dtu.dk/services/HostPhinder/>, la similitud genética se define como el número de k-mers que ocurren simultáneamente entre el fago de consulta y los fagos en la base de datos de referencia, en este se examina al genoma completo. HostPhinder podría ayudar a establecer el vínculo entre los fagos y las bacterias (Nielsen et al., 2014)

#### **3.11.1 Identificación de operones CRISPR CAS y relación de hospederos**

Los archivos obtenidos de VirSorter 1.0.3 se utilizaron para identificar a los operones de CRISPR CAS, disponible en: <https://github.com/Russel88/CRISPRCasTyper> es un software, el cual detecta matrices y genes CRISPR Cas, y su vez predice el subtipo en función de los genes Cas y la secuencia de repetición CRISPR, esto se realiza por BLASTing y un aprendizaje automático basado en kmer. Se obtuvieron el número de identificación de contigs que contienen secuencias de repetición y a estas, se

les realizó un Blastn en <https://blast.ncbi.nlm.nih.gov/Blast.cgi> para obtener a los posibles hospederos.

### **3.12 Clasificación taxonómica y comparación de genomas**

La clasificación taxonómica de las lecturas metagenómicas está relacionada con los análisis del gen marcador 16S rRNA en la forma en que asigna secuencias a grupos taxonómicos, deduciendo así un perfil comunitario. En consecuencia, dicha clasificación puede hacerse basada en referencias o sin referencias. Se realizó la clasificación taxonómica en MEGAN ([Huson et al., 2016](#)), disponible en: <https://www.computomics.com/services/megan6.html> la cuál se basa enfoque del ancestro común más bajo (LCA) y asigna la taxonomía LCA de todos los aciertos de BLAST que tienen una puntuación similar al mejor acierto para cada secuencia.

### **3.13 Abundancia relativa**

Por medio de R Studio Version: 2022.07.2+576 el cuál es un entorno de desarrollo integral que permite de una manera amigable realizar visualizaciones por medio de los paquetes ggplot2, reshape2 para los procedimientos de forma estadística y además en la creación de las gráficas, se realizó la abundancia relativa para la identificación de virus con mayor presencia en las muestras metagenómicas.

## 4. RESULTADOS Y DISCUSIÓN DE RESULTADOS

### 4.1 Clasificación taxonómica y evaluación de la diversidad

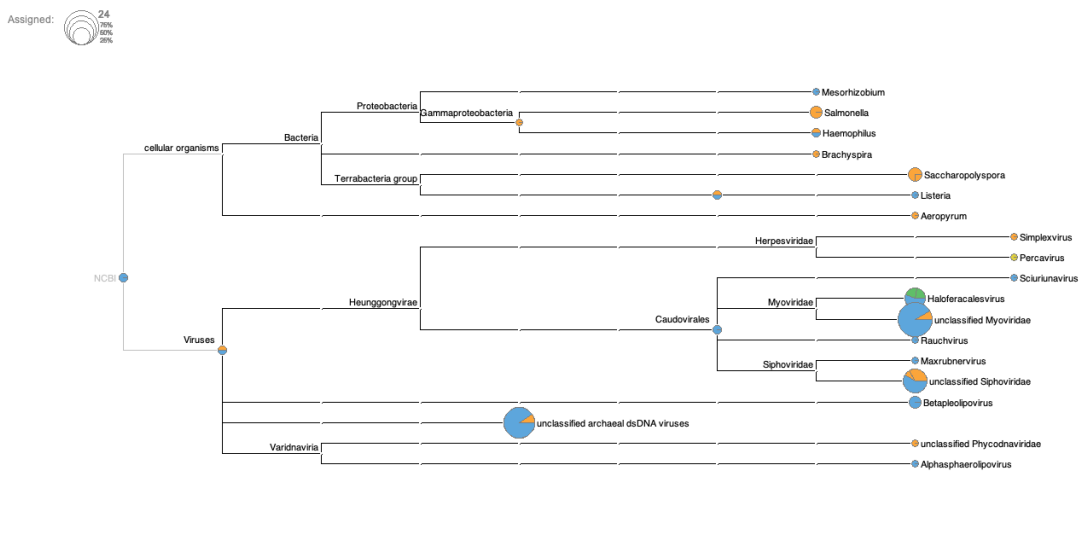
Fueron analizadas seis muestras metagenómicas de 4 regiones geográficas distintas, separadas en el rango de 9000 km a 700 km. Se puede observar la diversidad en todas las muestras (figura 7) a nivel de género, presentando diversidad de virus halófilos y bacterias, entre las presentes está *Mesorhizobium*, la cual se ha aislado en suelo con característica morfológica de Gram negativa y ejerce la función de ser fijadora del nitrógeno en el suelo, *Salmonella* es una bacteria Gram negativa con forma de bastón, entre los hábitats que se ha localizado se encuentra en el agua, el medio ambiente y los alimentos, representa contaminación fecal (Jajere, 2019), *Haemophilus* con característica morfológica de un bacilo Gram negativo, tiene un desarrollo óptimo de 7.6 pH (*Manual for the Laboratory Identification and Antimicrobial Susceptibility Testing of Bacterial Pathogens of Public Health Importance in the Developing World*, n.d.), *Brachyspira hyodysenteriae* es una bacteria Gram negativa con temperatura óptima de crecimiento 37-41°C, el cual se ha detectado también en heces fecales de cerdos (Dors et al., 2021), *Saccharopolyspora* han sido ampliamente reportados en una amplia gama de hábitats, incluidos suelo, sedimentos marinos, invertebrados marinos, plantas y muestras clínicas, además de está ampliamente distribuida en los ambientes marinos con alrededor de la mitad de sus especies (Sayed et al., 2020), *Listeria* con temperatura óptima de crecimiento 31-34°C, con un pH óptima de 7.5, además de poder desarrollarse con concentraciones de NaCl de hasta 10% (Lang Halter et al., 2013), arquea *Aeropyrum* tiene una forma de coco

clasificada como hipertermófila debido a la temperatura a la que logra desarrollarse 70-100 °C a un pH de 7, esta se ha recolectado en sedimentos del océano y ventilas hidrotermales, salinidad hasta del 7% (Gogliettino et al., 2022). Los microorganismos encontrados que pertenecen al dominio de Bacteria en su mayoría presenta la característica de crecer en concentraciones de sales superiores al 7%, y pH alcalinos siendo organismos tolerantes a estas concentraciones, esto tiene congruencia ya que estas salinas tienen además las características de tener un pH alcalino y radiación UV elevado. A nivel de especie los virus en su mayoría son halófilos pero no pertenecen a la misma familia y algunos de estos no se encuentran clasificados, en su mayoría estos virus infectan a arqueas. La composición microbiana encontrada en estos metagenomas concuerda con lo reportado en ambientes hipersalinos con concentraciones de sales similares (Rosselli et al., 2022).

Acceso	Ubicación	País	Salinidad	Plataforma de secuenciación	Fecha
SRR944625	Parque natural de las Salinas de Santa Pola	España	19%	454 GS FLX Titanium	2010
SRR979792	Parque natural de las Salinas de Santa Pola	España	15%	454 GS FLX Titanium	2010
SRR988245	Salinas Isla Cristina	España	15%	454 GS FLX Titanium	2010

SRR4027688	Antarctic Deep Lake, sample 24m 8.0um	Antártida	35%	Illumina Genome Analyzer II	2008
SRR4027687	Antarctic Deep Lake, sample 24m 0.3um	Antártida	35%	Illumina Genome Analyzer II	2008
	Salinas Maras	Perú	19%	Illumina NextSeq	2019

**Tabla 2** Metadatos utilizados en esta comparación de ambientes hipersalinos con acceso de los datos metagenómicos.



**Figura 7** Asignación taxonómica de orden jerárquico. Clasificación a nivel de género de datos metagenómicos de ambientes hipersalinos.

#### 4.1.2 Análisis de diversidad

Se evaluó la diversidad al respecto de el contenido de salinidad y lugar geográfico por medio de la diversidad alfa que describe la diversidad en la muestra, al evaluar la diversidad con el índice de Shannon, se observó menor

abundancia de diversidad en las muestras con un contenido de salinidad del 35 % las cuales son de Deep Lake, de igual manera la muestra de Santa Pola, Isla Cristina y Salinas Maras presentan una diversidad mayor respecto a las anteriores ambas comparten una salinidad del 19 % (Figura 8 A). Con el índice de Simpson se evaluó la dominancia de las especies, con resultados similares al índice Shannon, donde Isla Cristina y Salinas Maras presentan una mayor posibilidad de dominancia de una especie, mientras que Deep Lake y Santa Pola tienen menor probabilidad de presentar dominancia de una especie.

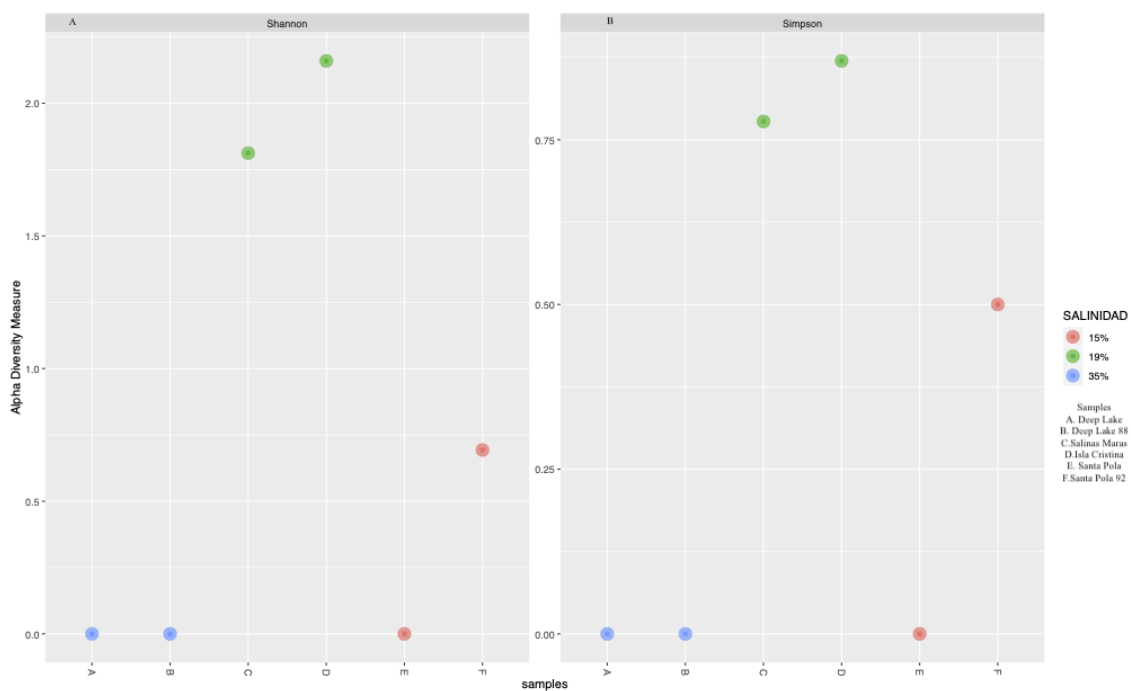


Figura 8 Análisis de diversidad. A) Índice de Shannon, B) Índice de Simpson.

#### 4.1.3 Análisis de agrupamiento

Se observaron dos agrupaciones principales de clusters 1 y 2 . Donde 1 está compuesto por las Salinas Maras C, Isla Cristina y Santa Pola 92 F agrupado

porque entre estos presentan mayor similitudes y 2 compuesto por Deep Lake 88 B, Deep Lake A y Santa Pola E (Figura 9).

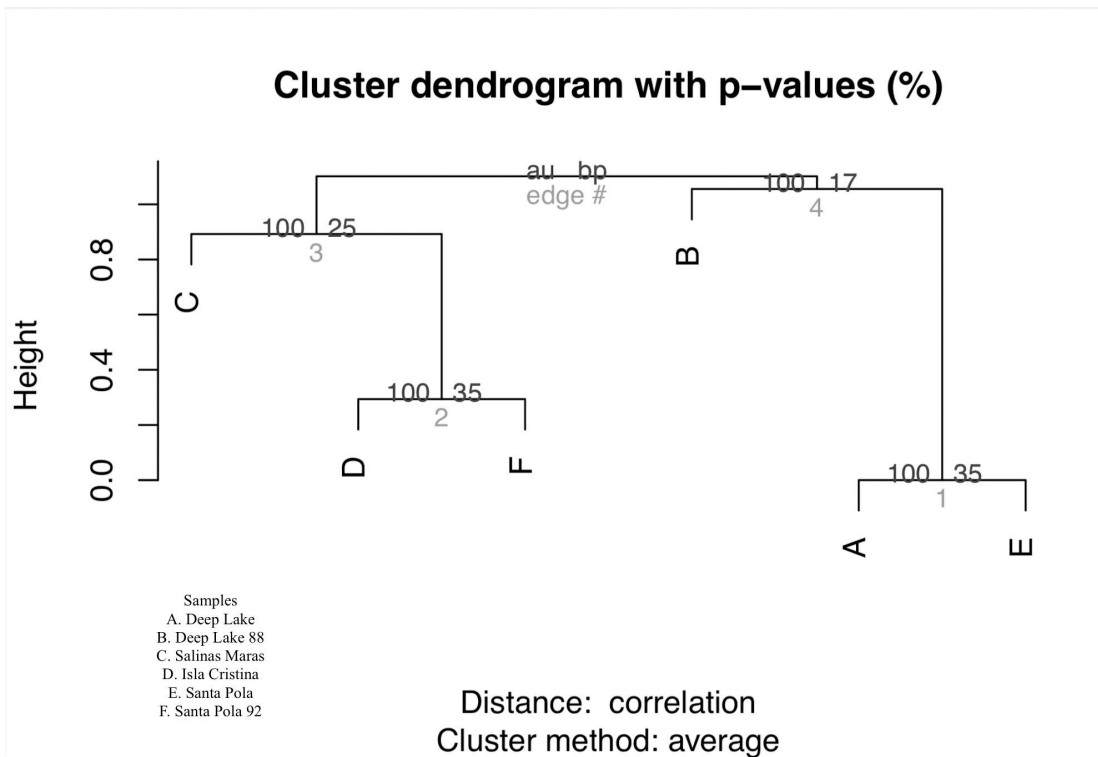


Figura 9 Análisis de clúster.

## 4.2 Comparación de datos metagenómicos

En la diversidad de este ecosistema, los virus más abundantes identificados por especie son halovirus y estos infectan a arqueas, bacterias y eucariotas (Atanasova et al., 2015). Se encontró una predominancia de virus dsDNA la cuál ya ha sido reportada en sedimentos alrededor de 95- 99 % en aguas profundas (Yoshida et al., 2013). Han sido reportados los géneros de myoviridae, siphoviridae, podoviridae en ambientes con altas concentraciones de sal (Atanasova et al., 2015). Las familias presentes en las muestras fueron



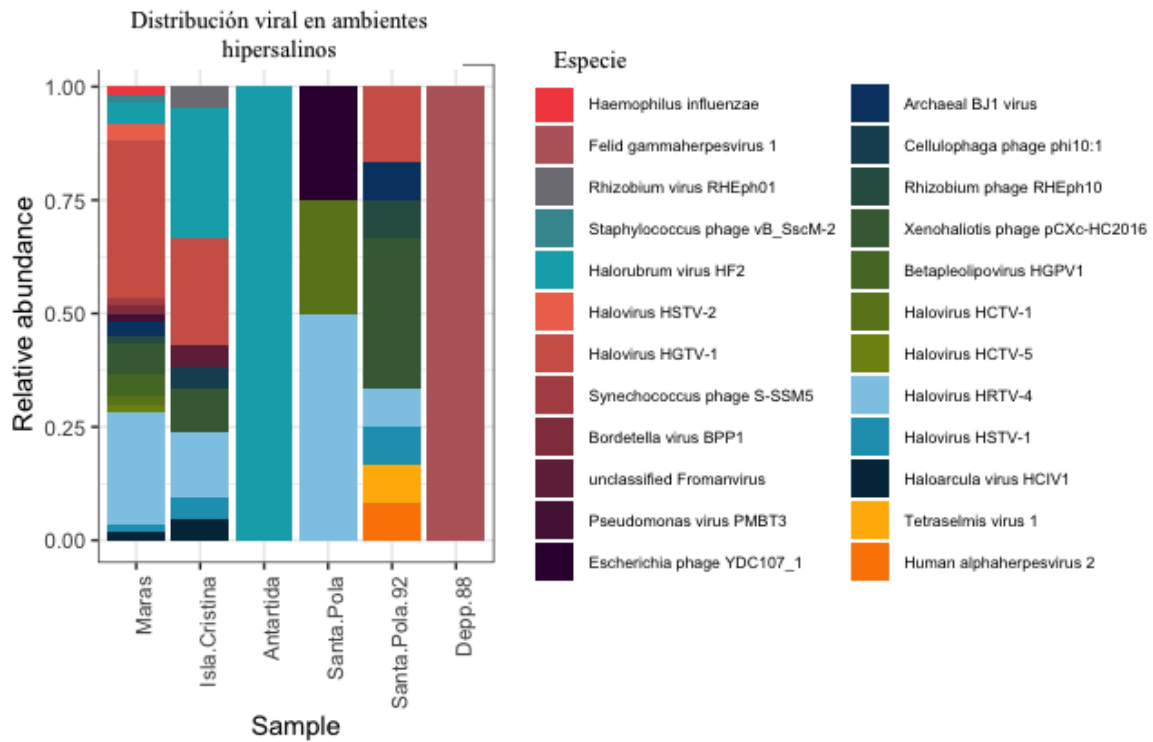
Halovirus HSTV1, Halovirus HSTV2, HSTV 1, HSTV 4, HSTV 5 (Figura 8). Estos halovirus comparten similitud en sus secuencias genómicas. La distribución viral de Santa Pola varía entre una muestra y otra, esta tiene una concentración de sales del 15 % también denominado baja salinidad, solo se comparte la presencia de un virus en el cual es Halovirus HRTV 4, esto puede ser debido a que, aunque, se obtuvieron de la misma salinera no se tomaron del mismo punto el muestreo. La distribución viral en las Salinas Maras en Perú es similar a la distribución presente en la Isla Cristina ambas cuentan con la presencia de Halovirus HSTV2 siendo el más abundante, una característica que comparten es la salinidad alrededor del 19% también denominado como salinidad intermedia.

Las salinas de Santa Pola e Isla Cristina a pesar de encontrarse en el mismo país y ambas ser de origen marino tienen un manantial diferente: el mar Mediterráneo y el océano Atlántico, respectivamente.

Los virus más abundantes son los Halovirus los cuales en su mayoría infectan al dominio Archae, los cuales se encuentran presentes en 5 de las 6 secuencias, siendo el orden de los caudovirales el de mayor presencia. Se observó un contraste significativo con las muestras de Deep Lake ubicado en la Antártida ya que solo se encuentra la presencia de un virus por muestra, Halovirus HGTV1 y Halobrum virus HF2, estos se encuentran en la presencia de una salinidad de alrededor de 31.5 % clasificado como salinidad saturada.

Los caudovirales también conocidos como bacteriófagos se han encontrado que predominan en el ambiente marino y otros ambientes acuáticos. Los

géneros en estas muestras de ambientes hipersalinos en el orden de abundancia son Haloferacalesvirus, Haloviruses, Siphoviridae y Myoviridae.



**Figura 10** Abundancia relativa de ambientes hipersalinos. Especie más abundantes Halovirus.

### 4.3 Identificación fagos y similitud de genes

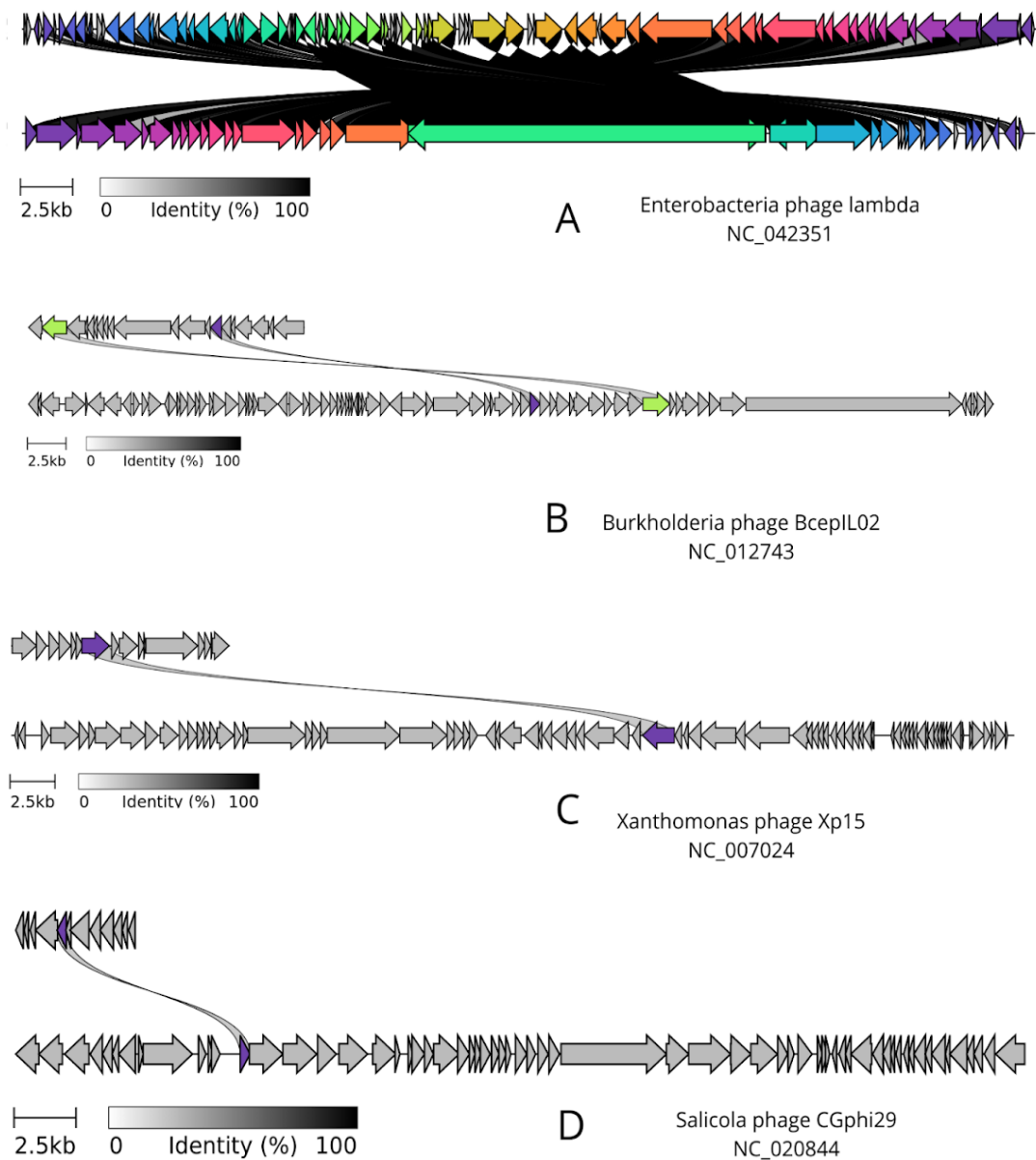
Se identificaron 11 fagos en los contigs de los ambientes hipersalinos, por medio de phaster el cuál permite conocer cuál es el fago más común en estos contigs, utilizando Checkv se evaluó la calidad de genomas virales, de los cuales se obtuvo 1 de forma intacta Enterobacteria phage lambda, 3 de calidad media Burkholderia phage BcepIL02, Xanthomonas phage Xp15 y Salicola phage CGphi29 (Tabla 3), además de la presencia de 7 fagos de calidad baja,

esto se debe a que los bacteriofagos encontrados solo estaban conformados por una parte de su secuencia.

Contig length	Completeness	Most Common Phage
48,502	100	Enterobacteria phage lambda
17921	3.069.948.806	Burkholderia phage BcepL02
12153	2.019.504.355	Xanthomonas phage Xp15
4818	1.269.630.434	Salicola phage CGphi29
4989	9.290.795.991	Natrialba phage PhiCh1
4472	739.719.181	Halorubrum tailed phage 7
4435	7.033.458.635	Natrialba phage PhiCh1
2849	3.769.005.238	Halorubrum tailed phage 8
2390	3.622.189.398	Halovirus HGTV-1
2661	3.577.972.183	Halorubrum tailed phage 7
2961	2.818.016.178	Halovirus HGTV-1

**Tabla 3.** Tabla de Completeness

Con los fagos de mayor completeness se realizó un análisis de similitud de genes, el cual la mejor sintenia conservada de manera inversa es de Enterobacteria phage lambda, la sintenia en B, C y D (figura 9) tiene un porcentaje de similitud baja, se observa la conservación de conexiones genéticas.



**Figura 11** Análisis de similitud de genes. A) Enterobacteria phage lambda B) Burkholderia phage BcepIL02 C) Xanthomonas phage Xp15 D) Salicola phage CGphi29.

#### 4.4 Identificación de posibles hospederos

Se clasificó a los hospederos en dos tipos, los hospederos en cualquier lugar y hospederos que se han reportado en ambientes extremos. De los posibles

hospederos (Tabla 4), se encuentran *Escherichia* que ha desarrollado la capacidad de desarrollarse en ambientes hipersalinos como ya han sido reportado en Kebrit Deep ubicado en el mar Rojo (Ramadan et al., 2019). *Salmonella* se ha reportado en ambientes acuáticos y algunos de los hospederos posibles como lo es *Xylella* son parte de la microbiota de las plantas.

Posibles hospederos	
<i>Escherichia</i>	<i>Caulobacter</i>
<i>Salmonella</i>	<i>Rhodococcus</i>
<i>Streptomyces</i>	<i>Saccharomonospora</i>
<i>Shigella</i>	<i>Rhodobacter</i>
<i>Burkholderia</i>	<i>Acidithiobacillus</i>
<i>Tsukamurella</i>	<i>Rhodovulum</i>
<i>Xylella</i>	<i>Halomonas</i>
<i>Actinoplanes</i>	<i>Microbacterium</i>
<i>Ralstonia</i>	

**Tabla 4** Identificación de posibles hospederos HostPhinder. Esta tabla se encuentra dividida en dos secciones la primera con los posibles hospederos y en la segunda los hospederos identificados en ambientes extremos y salinos.

#### 4.5 Identificación de operones CRISPR CAS y relación de hospederos

Se identificaron cinco secuencias de repetición de las cuales tres pertenecen al tipo 1 subtipo 1B con los genes cas asociados cas3 cas5 cas7 cas8b, se relacionaron a los posibles hospederos por Blastn, *Natribaculum longums* también se han aislado en suelos salinos con concentraciones de NaCl mayores a 15% (Liu et al., 2015), *Salarchaeum sp.* JOR-1 es una arquea halófila extrema (Anton et al., 2020), *Haloquadratum walsbyi* también presente

en salmueras con concentraciones elevadas de NaCl y MgCl<sub>2</sub>, *Salinibacter ruber* aislado en salinas cristalinas con concentraciones mayores al 15% (Anton et al., 2020)., *Natrinema sp. J7-2* es una halo arquea extrema, la cual prospera en salinas, lagos salados y depósitos de sal (Feng et al., 2012). La relación de posibles hospederos por las secuencias de repetición es congruente con los hospederos encontrados, ya que estos se han reportado en ambientes salinos.

Virus	Posible hospedero	Secuencia de repetición	Predicción	Tipo	Subtipo	Genes
Salicola phage CGphi29	Salinibacter ruber	AGGTCTATCCCCGCG TGTGCGGGGTCAACC	I-E	1	I-E	cas3 cas5 cas7 cas8e/cs e1 cas11/cs e2
Natrialba phage PhiCh1	Natribaculum longums	GTTTCAGACGAACCC TCGTGGGGTTGAAG C	I-B	1	I-B	cas3 cas5 cas7 cas8b
Halovirus HGTV-1	Natrinema sp. J7-2	GCTTCGACCCACAA GGGTCCGTCTGAAA CGGG	III-A	1	III-A	cas10 csm2 csm3 csm4 csm5
Halorubrum virus	Salarchaeum sp. JOR-1	GCTTCAACCCACGA GGGTACGTCTGAAAC	I-B	1	I-B	cas3 cas5 cas7 cas8b
Halorubrum tailed phage 7	Haloquadratum walsbyi	ACTTCAACCCAACAA GGGTTCATCTGAAAC	I-B	1	I-B	cas3 cas5 cas7 cas8b

Tabla 5 Identificación de hospederos por CRISPR Cas

## Capítulo 5. CONCLUSIONES

La metagenómica viral en ambientes hipersalinos ha mostrado que los virus más comunes en estos ambientes son los Caudovirales. Esto indica que la presencia de iones de sales en concentraciones elevadas es un factor importante en el desarrollo de la microbiota en estos ambientes. Los estudios comparativos entre diferentes lugares geográficos han demostrado similitudes en los ensamblajes virales de ambientes hipersalinos ubicados a más de 9,000 km de distancia. Esto sugiere que la ubicación geográfica no es tan importante como la concentración de iones de sales presentes en el ambiente.

La presencia de operones CRISPR Cas en los contigs de estos ambientes también puede indicar la interacción entre virus y hospederos. Los fagos son un mecanismo importante para la regulación de la diversidad genética y la evolución de los microorganismos en estos ambientes. Además, se identificaron cinco secuencias de repetición relacionadas con posibles hospederos, como *Natrabaculum longums*, *Salarchaeum* sp. JOR-1, *Haloquadratum walsbyi*, *Salinibacter ruber* y *Natrinema* sp. J7-2. Estos organismos son conocidos por su capacidad para prosperar en ambientes con altas concentraciones de iones de sales y se han aislado previamente en salinas, lagos salados y depósitos

de sal. Sin embargo, es posible que en estudios metagenómicos de estos ambientes, la presencia de virus no muestran suficiente similitud con los genomas de referencia para una estimación precisa de su integridad.

En general, la metagenómica viral en ambientes hipersalinos proporciona información valiosa sobre los virus presentes en estos ambientes y cómo interactúan con la microbiota. Además, ayuda a entender cómo los ambientes hipersalinos influyen en la diversidad y distribución de los virus. En el futuro, se espera que los estudios metagenómicos en ambientes hipersalinos proporcionen información adicional sobre la ecología y evolución de los virus en estos ambientes únicos.

#### Referencias:

Alavandi, S. V., & Poornima, M. (2012). Viral Metagenomics: A Tool for Virus



- Discovery and Diversity in Aquaculture. *Indian Journal of Virology*, 23(2), 88–98.  
<https://doi.org/10.1007/s13337-012-0075-2>
- Annamalai, N., Rajeswari, M. V., & Balasubramanian, T. (2016). Thermostable and Alkaline Cellulases from Marine Sources. In *New and Future Developments in Microbial Biotechnology and Bioengineering* (pp. 91–98). Elsevier.  
<https://doi.org/10.1016/B978-0-444-63507-5.00009-5>
- Anton, B. P., DasSarma, P., Martinez, F. L., DasSarma, S. L., Al Madadha, M., Roberts, R. J., & DasSarma, S. (2020). Genome Sequence of *Salarchaeum* sp. Strain JOR-1, an Extremely Halophilic Archaeon from the Dead Sea. *Microbiology Resource Announcements*, 9(5). <https://doi.org/10.1128/MRA.01505-19>
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 44(W1), W16–W21. <https://doi.org/10.1093/nar/gkw387>
- Atanasova, N. S., Oksanen, H. M., & Bamford, D. H. (2015). Haloviruses of archaea, bacteria, and eukaryotes. *Current Opinion in Microbiology*, 25, 40–48.  
<https://doi.org/10.1016/j.mib.2015.04.001>
- Banciu, H. L., & Muntyan, M. S. (2015). Adaptive strategies in the double-extremophilic prokaryotes inhabiting soda lakes. *Current Opinion in Microbiology*, 25, 73–79. <https://doi.org/10.1016/j.mib.2015.05.003>
- Manual for the Laboratory Identification and Antimicrobial Susceptibility Testing of Bacterial Pathogens of Public Health Importance in the Developing World.
- Charpentier, E., & Marraffini, L. A. (2014). Harnessing CRISPR-Cas9 immunity for genetic engineering. *Current Opinion in Microbiology*, 19, 114–119.  
<https://doi.org/10.1016/j.mib.2014.07.001>

- Dors, A., Czyżewska-Dors, E., & Woźniakowski, G. (2021). A survey on the occurrence of *Brachyspira pilosicoli* and *Brachyspira hyodysenteriae* in growing-finishing pigs. *F1000Research*, 8, 1702.  
<https://doi.org/10.12688/f1000research.20639.3>
- Edwards, R. A., McNair, K., Faust, K., Raes, J., & Dutilh, B. E. (2016). Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiology Reviews*, 40(2), 258–272. <https://doi.org/10.1093/femsre/fuv048>
- Enright, A. J. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, 30(7), 1575–1584.  
<https://doi.org/10.1093/nar/30.7.1575>
- Feng, J., Liu, B., Zhang, Z., Ren, Y., Li, Y., Gan, F., Huang, Y., Chen, X., Shen, P., Wang, L., Tang, B., & Tang, X.-F. (2012). The Complete Genome Sequence of *Natrinema* sp. J7-2, a Haloarchaeon Capable of Growth on Synthetic Media without Amino Acid Supplements. *PLoS ONE*, 7(7), e41621.  
<https://doi.org/10.1371/journal.pone.0041621>
- Gogliettino, M., Arciello, S., Cillo, F., Carluccio, A. V., Palmieri, G., Apone, F., Ambrosio, R. L., Anastasio, A., Gratino, L., Carola, A., & Cocca, E. (2022). Recombinant Expression of Archaeal Superoxide Dismutases in Plant Cell Cultures: A Sustainable Solution with Potential Application in the Food Industry. *Antioxidants*, 11(9), 1731. <https://doi.org/10.3390/antiox11091731>
- Gunde-Cimerman, N., Plemenitaš, A., & Oren, A. (2018). Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiology Reviews*, 42(3), 353–375. <https://doi.org/10.1093/femsre/fuy009>
- Hallsworth, J. E. (2018). Stress-free microbes lack vitality. *Fungal Biology*, 122(6),

379–385. <https://doi.org/10.1016/j.funbio.2018.04.003>

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. <https://doi.org/10.1128/jb.169.12.5429-5433.1987>

Jajere, S. M. (2019). A review of *Salmonella enterica* with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Veterinary World*, *12*(4), 504–521. <https://doi.org/10.14202/vetworld.2019.504-521>

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., & Koonin, E. V. (2018). Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research*, *244*, 181–193. <https://doi.org/10.1016/j.virusres.2017.11.025>

Kukkaro, P., & Bamford, D. H. (2009). Virus-host interactions in environments with a wide range of ionic strengths. *Environmental Microbiology Reports*, *1*(1), 71–77. <https://doi.org/10.1111/j.1758-2229.2008.00007.x>

Lang Halter, E., Neuhaus, K., & Scherer, S. (2013). *Listeria weihenstephanensis* sp. nov., isolated from the water plant *Lemna trisulca* taken from a freshwater pond. *International Journal of Systematic and Evolutionary Microbiology*, *63*(Pt\_2), 641–647. <https://doi.org/10.1099/ijs.0.036830-0>

Ley, R. E., Harris, J. K., Wilcox, J., Spear, J. R., Miller, S. R., Bebout, B. M., Maresca, J. A., Bryant, D. A., Sogin, M. L., & Pace, N. R. (2006). Unexpected Diversity and Complexity of the Guerrero Negro Hypersaline Microbial Mat. *Applied and Environmental Microbiology*, *72*(5), 3685–3695.

<https://doi.org/10.1128/AEM.72.5.3685-3695.2006>

- Liu, Q., Ren, M., & Zhang, L.-L. (2015). *Natribaculum breve* gen. nov., sp. nov. and *Natribaculum longum* sp. nov., halophilic archaea isolated from saline soil. *International Journal of Systematic and Evolutionary Microbiology*, 65(Pt\_2), 604–608. <https://doi.org/10.1099/ijms.0.060541-0>
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnyš, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., ... Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18(2), 67–83. <https://doi.org/10.1038/s41579-019-0299-x>
- Nielsen, H. B., Almeida, M., Juncker, A. S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A. G., Le Chatelier, E., Pelletier, E., Bonde, I., Nielsen, T., Manichanh, C., Arumugam, M., Batto, J.-M., Quintanilha dos Santos, M. B., Blom, N., Borrueal, N., ... Ehrlich, S. D. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, 32(8), 822–828. <https://doi.org/10.1038/nbt.2939>
- Ramadan, E., Maged, M., El Hosseiny, A., Chambergo, F. S., Setubal, J. C., & El Dorry, H. (2019). Molecular Adaptations of Bacterial Mercuric Reductase to the Hypersaline Kebrit Deep in the Red Sea. *Applied and Environmental Microbiology*, 85(4). <https://doi.org/10.1128/AEM.01431-18>
- Rampelotto, P. (2013). *Extremophiles and Extreme Environments*. <https://www.mdpi.com/Journal/Life>. <https://www.mdpi.com/journal/life>

- Rosselli, R., López-Pérez, M., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., & Bolhuis, H. (2022). Differences in gene expression patterns between cultured and natural *Haloquadratum walsbyi* ecotypes. *Frontiers in Microbiology*, *13*.  
<https://doi.org/10.3389/fmicb.2022.1044446>
- Rothschild, L. J., & Mancinelli, R. L. (2001). Life in extreme environments. *Nature*, *409*(6823), 1092–1101. <https://doi.org/10.1038/35059215>
- Saccò, M., White, N. E., Harrod, C., Salazar, G., Aguilar, P., Cubillos, C. F., Meredith, K., Baxter, B. K., Oren, A., Anufriieva, E., Shadrin, N., Marambio-Alfaro, Y., Bravo-Naranjo, V., & Allentoft, M. E. (2021). Salt to conserve: a review on the ecology and preservation of hypersaline ecosystems. *Biological Reviews*, *96*(6), 2828–2850. <https://doi.org/10.1111/brv.12780>
- Santos, F., Yarza, P., Parro, V., Briones, C., & Antón, J. (2010). The metavirome of a hypersaline environment. *Environmental Microbiology*, *12*(11), 2965–2976.  
<https://doi.org/10.1111/j.1462-2920.2010.02273.x>
- Sayed, A. M., Abdel-Wahab, N. M., Hassan, H. M., & Abdelmohsen, U. R. (2020). Saccharopolyspora : an underexplored source for bioactive natural products. *Journal of Applied Microbiology*, *128*(2), 314–329.  
<https://doi.org/10.1111/jam.14360>
- Siliakus, M. F., van der Oost, J., & Kengen, S. W. M. (2017). Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles*, *21*(4), 651–670. <https://doi.org/10.1007/s00792-017-0939-x>
- Stone, E., Campbell, K., Grant, I., & McAuliffe, O. (2019). Understanding and Exploiting Phage–Host Interactions. *Viruses*, *11*(6), 567.  
<https://doi.org/10.3390/v11060567>

- Suttle, C. A. (2005). Viruses in the sea. *Nature*, 437(7057), 356–361.  
<https://doi.org/10.1038/nature04160>
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3.  
<https://doi.org/10.1186/2042-5783-2-3>
- Williams, T. J., Allen, M. A., DeMaere, M. Z., Kyrpides, N. C., Tringe, S. G., Woyke, T., & Cavicchioli, R. (2014). Microbial ecology of an Antarctic hypersaline lake: genomic assessment of ecophysiology among dominant haloarchaea. *The ISME Journal*, 8(8), 1645–1658. <https://doi.org/10.1038/ismej.2014.18>
- Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., & Takai, K. (2013). Metagenomic Analysis of Viral Communities in (Hado)Pelagic Sediments. *PLoS ONE*, 8(2), e57271. <https://doi.org/10.1371/journal.pone.0057271>
- Zhu, Daochen, Adebisi, W. A., Ahmad, F., Sethupathy, S., Danso, B., & Sun, J. (2020). Recent Development of Extremophilic Bacteria and Their Application in Biorefinery. *Frontiers in Bioengineering and Biotechnology*, 8.  
<https://doi.org/10.3389/fbioe.2020.00483>
- Zhu, Derui, Han, R., Long, Q., Gao, X., Xing, J., Shen, G., Li, Y., & Wang, R. (2020). An evaluation of the core bacterial communities associated with hypersaline environments in the Qaidam Basin, China. *Archives of Microbiology*, 202(8), 2093–2103. <https://doi.org/10.1007/s00203-020-01927-7>



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS



## INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS



Control Escolar de Licenciatura

### VOTOS DE APROBATORIOS

Secretaría ejecutiva del Instituto de Investigación en Ciencias Básicas Aplicadas de la Universidad Autónoma del Estado de Morelos.

Presente.

Por medio de la presente le informamos que después de revisar la versión escrita de la tesis que realizó la C. **AYALA RODRÍGUEZ MARIA GUADALUPE** con número de matrícula **10018616** cuyo título es:

***Metagenómica viral de ambientes extremos e identificación de sus posibles hosperos mediante CRISPR CAS***

Consideramos que **SI** reúne los méritos que son necesarios para continuar los trámites para obtener el título de **LICENCIADO EN CIENCIAS CON AREA TERMINAL BIOQUIMICA Y BIOLOGIA MOLECULAR**.

Cuernavaca, Mor a 04 de mayo del 2023

Atentamente  
Por una universidad culta

Se adiciona página con la e-firma UAEM de los siguientes:

<b>DR. AGUSTÍN REYES PÉREZ</b>	<b>(PRESIDENTE).</b>
<b>DR. RAMÓN ALBERTO BATISTA GARCIA</b>	<b>(SECRETARIO).</b>
<b>DRA. SONIA DÁVILA RAMOS</b>	<b>(VOCAL).</b>
<b>DR. HUGO GILDARDO CASTELÁN SÁNCHEZ</b>	<b>(SUPLENTE).</b>
<b>DR. LUIS FERNANDO LOZANO AGUIRRE BELTRÁN</b>	<b>(SUPLENTE).</b>



UNIVERSIDAD AUTÓNOMA DEL  
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

### Sello electrónico

**RAMON ALBERTO BATISTA GARCIA | Fecha:2023-06-02 10:29:13 | Firmante**

Cnf/Gk5xOluOG4cvUjiiOOIJEWEx6l6OvaS5AZAgmEZXCjPg/bnkYz9y0EPNKRwAPCYDMDuTWzE/vrTLVEVPKR59HFLNjhLnsq5NTR0+2eF5k0a5A7nRGi97UBDaTQ2Py348wNbDJZxP2rpaec8V+TOiwE6awlHqm4SsV0nygU+SY7Fr3DOMqoQium+B4R+wb4MwC/2COsmjuBvOewLiykLebmrYT/VG2OjEqOVFR+1Du90ZiLrSW7dPwKpr9MKc0L4By+Io7ZwUaO2AMGISfo+JqyqMppRRDoUmxlcZxNcF142B3t/IQ9yO0Mc0r7u3Fe5GQtDeOH8ev5+SyZMSLg==

**LUIS FERNANDO LOZANO AGUIRRE BELTRÁN | Fecha:2023-06-02 10:56:09 | Firmante**

Bz5DZDWugxFaiD+gwjCwxXP2C9PvEyoTyWgEESXB+zd7c2ovam6GRwqUXJZhW30H7/yZgRIRKb1oqnNgFzzDXept4xl3OpO2+glTqn+4AhYM7wcU/MIBSvPQDg2rUnBNbRaX83FzxfnOSdRiuHglzctO6NzrzvDCs/bMxEBCTXDPCOfy+wys4Aj7L2secg4ezLqUxHgSPtnaErUMM7cHHM5QsdAeX9yILQtL7IHmPQ7Fyc3PGPcQzjFHUKqQVHR3/1llmsFYV1UjGY1qJ/CnNv8yaJawITD6cl4Hyd6MBcFhRLJVRl/0pcxyZtptwd8H3qAoKheMHfETy6zQrw==

**SONIA DAVILA RAMOS | Fecha:2023-06-02 13:32:54 | Firmante**

A0FPVw8KNQE+PCr7yZQ5b8+ecmpWBUuiZ+vb758Wl1v6NZpLyU6lpkkZ/mF412ToGIPQeJGxrNHZparLOniQq4mbl3p0KkradlFR9TxljIRD+ESp99F/N6Q9n7jHe4M6hJtNBpEOFEbXA/IZRnUQ/NckAPcKmi8N3bRD5lsmHocawd1rmetxKVTGnRuOeXzQVGVYO1iPPdaDPIUUKwuSMJHeLenpQwLhJR6CZgJlflKj/6t9KaQ33RS9GucQVnA05qfHvBzGZKnSe tpK+2bPvOH3UNMdb13k6kWiXfoxgbMjl14uVJq3MqVyh06vDGCcSxwFM5e0zrhfwQ8tjRyIVQ==

**HUGO GILDARDO CASTELÁN SÁNCHEZ | Fecha:2023-06-02 17:32:09 | Firmante**

DO9U88vZx8qo1+c4arT8aPAZq71quMqJlIfgyPZi7v1TJt0ddclq6anXRclj4cZJKqTxgGBpbRjUMW/y0wBzKFFrhxJHqko2LZrMTDDn14NrxCDcfVBAHjrSASWJm30Qh9EE3ThmWqN5sKV97+6l6V5QUARqslT7xL1rM+kmL1ztu6fmxgdvVvSWv6Mt/fkXfllkk3WWL01Dkw/x1dCVH1qXozhEli6yNg3zSmcy2iYpixG0ZD3HK8ESOjtdv3lp17vWCU4AiCHW169NzNeOoR39Plby1xj9OG7oScUqFhrb/E3Q85YwtiXHwQ2m1SAO2ZI7gN81ghX+U6MzRQw==

**AGUSTIN REYES PEREZ | Fecha:2023-06-06 08:16:35 | Firmante**

e9nGjQWnDlbugjvd3Sn6OVRkgJONKC0kgWhSOvj0mODKGX+i+j0RFUhpACqtQdME+FDJIANmjhj+qRIKdJl2u/zyRvrM+K7QRPQzzv7Eu9qFzG4umD8sP/6Z235NBjKS4v1E3Ct1OAXuGnNcLQ2FoW9IG5W2kBgFG9oTpBl6vwjWVj9rZ2Ayu7OqGQrwb+poEFhEOsSxNOhkB68cF+0rEzncoppjOIRZhfKf5XAwmdGCf9V9s2zujSmUo/5EHRjVO9mgdAYzm4MzA/8NCg1YtE+8MIDjd5a4/ZodDU3wH4Hp8ozooJU2z8q8IXmfUbPA5RwF5OPeRvKJCu8Og==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



SWKuCFRtn

<https://efirma.uaem.mx/noRepudio/pwNli561rFRYnb8QB5oeG29Ql3uBCCTo>

