UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS
CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

"ESTUDIO COMPARATIVO METAGENÓMICO VIRAL DE AMBIENTES ACUÁTICOS EXTREMOS"

TESIS

QUE PARA OBTENER EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

HUGO GILDARDO CASTELÁN SÁNCHEZ

DIRECTOR DE TESIS
Dra. Sonia Dávila Ramos

CUERNAVACA, MORELOS AGOSTO, 2021

DR. JEAN MICHEL GRÉVY MACQUART
COORDINADOR DEL POSGRADO EN CIENCIAS
PRESENTE

Atendiendo a la solicitud para emitir DICTAMEN sobre la revisión de la TESIS titulada **"Estudio comparativo metagenómico viral de ambientes acuáticos extremos",** que presenta el alumno **Hugo Gildardo Castelán Sánchez (10010090)** para obtener el título de **Doctor en Ciencias.**

Nos permitimos informarle que nuestro voto es:

NOMBRE	DICTAMEN	FIRMA
Dr. Ramón Alberto Batista García CIDC-UAEM	APROBADO	
Dr. Armando Hernández Mendoza CIDC-UAEM	APROBADO	
Dr. Lorenzo Patrick Segovia Forcella IBT-UNAM	APROBADO	
Dra. María del Rayo Sánchez Carbente CEIB-UAEM	APROBADO	
Dr. Enrique Merino Pérez IBT-UNAM	APROBADO	
Dr. Ramón Antonio Gonzalez García-Conde CIDC-UAEM	APROBADO	
Dr. Jorge Luis Folch Mallol CEIB-UAEM	APROBADO	

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

JORGE LUIS FOLCH MALLOL | Fecha:2021-03-23 09:35:54 | Firmante

IWKiyD2NHrBrDbIDApmKB5O8Ilhb3akkaeccfXyvMzDJEg0x8C+2PJxSMe0VTrIfX4gLbHLQQsY5RnovnRI4xysDgiEkOv8Cj7YSdzAk0IyEZ8pyewcHRuaTeY+VKcy0ek7vuYF+pD62NYRkiwcv7jKwJKVOUx6uJNIvxpwG1BEPTK5kjuJJ5fNcZVaIPx7Tp8ehe575P+q1dJQRJN14sNsCR5RSlYF5RsqxNAmexvR2B2VmpWXMUwwsp+0qtzDVDIF9pKB8M8M+1shTYvrGknYzlrhrer+07LlboX7VkRn1mAtpWY/hjGq5YW37v6sTrRUWV2VYVo74iwSoZeqfBQ==

RAMON ALBERTO BATISTA GARCIA | Fecha:2021-03-23 20:19:10 | Firmante

DL0JShO67t44TKm3CtoIdWmwn7XKoipSeIN2eybNKNTTIO5sUlpKqQTwalaA8v7FgsMAQuC8DWNCPW3dZV92Ai7RX/f2HJf6lSxzGeVWwFc/O/MGG5jCedTcqkcN8xjwQkkRvmRL4l6DOYobmTLNgdJxt5eeclABLVJtLleK+HcTBUZDypNiiFsWYsopoFsZbSkRkASjBHtE2GVveSzLyN2nbKjHUyVD8BN2jJbJGuiq92+1JKKT1BiQQKSV0ZAl4EF9r6VVV3GdweUYDBso2jHHmzcGGkd6KmQllnFqrQhBd7vC9C6OwkqEOYNmf8+ZR7wlbq6aNYtpbmEpenPMLQ==

LORENZO PATRICK SEGOVIA FORCELLA | Fecha:2021-03-25 08:36:54 | Firmante

cpMktRkNo16FnG3sTMidTxUZE06BwAx3QAg4o1zK6R4nqbzBlh44lAj8g//hv/8eeOrtwTCD5ZzhDaSwNvJnVi5/GQbkJJsPu7013/1YXN0t4uLUfZ+d9tvUj7HM0zkYFb9Yg3Vd7JMJKLcDCtVMtBDi0yjlj7oni+3SsfTF0jvwwV4S63TYmpHw9RRO4/eScKTQ0USiqKjKdpVNfLX/1/lSuBUl8B6F96ZBTVlUGkrvn0wVvB1LSHkzVw8lxQjHfd7ADPsKTsIoUOCsm+55mMaIFy1O8tNeIfNL2MSn65d0izfGz82M5FiLhMRV7DHc8duv5Hj3wB5gFYcULLb6UA==

ENRIQUE MERINO PÉREZ | Fecha:2021-03-25 09:37:06 | Firmante

tkx8bJ7mSBOaqLaP1n++XXqUT2iTt06ZDkUBqn3YQT8xhl+fB0VYog1rEU+PKlMcLWks1Toe0iXHo2tIlJt/+CnAwAS0vTTqPuISpr26bBJ8pnd1gQuFDFi+9hJSqpzkb464CmubuypRB1wA5B+TGYTeraqqJNDV4mxgHxurWOO4absep9oQ3Krcsr5XG6tYJ+/HPVWSUPhUtqXfENbXAF2JFUAMkG9PqxO1/Qq4MZtb8ea/QNZiGjdna9JZnfAfJ2av2KzKto8W+tasfpRpYkkA5DWHQ5DuEkVOg8m8/7tdEDVjM1LmT216gd/5NUxHD56wFelDZZwWPCvlzQ7DLQ==

RAMON ANTONIO GONZALEZ GARCIA CONDE | Fecha:2021-03-25 10:08:01 | Firmante

bTRGjFNzByOB5U6j9ogJ6cyDeYANsfsAuCNKGLMHCVACssU0TQO0adSGUBPKkVHva+yllS5vpuKnFnzZWfyx3NcJ1N5VJtskIzVMsZfHOPdjHwoODSq+NNEcWRIzU57CkuccUmcMWTlWybv8/xNiyGbQaL91kae9viRBPgptldrpnhFdImb/2zPLYq+PqjwFBC/qAYBfLiqeBhJSGd9zpyctUeGwggBR2lgfcnMDdDcUmiy62uJe1b0fa/UsKg5PrPtkgr0xeYGl4O9yzYManSwkO5u1FJuVIiUQAIx4IKgA7klri47V6Q5QEZgqxJbRfXzwzIVeZ0DFXNbPgZsMEQ==

MARIA DEL RAYO SANCHEZ CARBENTE | Fecha:2021-04-12 20:52:14 | Firmante

ru6+VoAhenXFazfsgpZQYFMDSIuvBR0fKXY9vcRxloxOjxoBheOYKjd3ZkPFf2xjxsVubeW2f1H1gS7ERnNvS6Le+Fryx10a1KPklUjQLZcogs5LFxK5537/f4MpJN9bgV93r7pjD/JeDrgfaqqSHpidYCYeeEXHl+3A5LKapk6gxHH8kLdzpKvscxMo8OhDDqquD/moBPBsVVKif0zjtBApypfdNc10m4nCY6X6ExvaGZ7EnPBsjmJv9FB8wSIAamE09QG1OmbXRcalpPcoHUnbS2xUzlubyfl6dtNw6Cseij6La6ghmkiljqMz0c0fCjQeJUfzia24+0BqhmirdA==

ARMANDO HERNANDEZ MENDOZA | Fecha:2021-04-29 18:37:35 | Firmante

D9S+08GhICwEBVYDvzBuv/N4lQTUKf8ywi27t1qoA41FlnGxql6al/d9NnDWb4hvQEuHDoxm3FGIyOXh7LhSsSnZltzeRtuW0ngQgGVQpiCRsaIZPtRM+TyxDZvASIlUtFvYvmqR1UwhCy2mfrbHEgKA720ANMV3INdoE0XavEBie51JVUd9LhrNcyYFPXl2Fn8CNZwaPwPEVRHQuZAKYTcFN2WOIwdQyMpMT6Nkp9TeIlvWV94tvZSX/nOJXUhtyOTRQSi3kp5nz+u4RDPpbRxW7/gt4MN/TXahwbUoRswCNGG0jOPOgiz0RitI89p2FUZHiMBHXBLuL57RB37AZg==

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

8Um0YH

https://efirma.uaem.mx/noRepudio/HanqeKuEQZUzHN9VCgk0ufurGF5zOvhK

LISTA DE PUBLICACIONES

1. **Castelán-Sánchez, H.,** Lopéz-Rosas, I., García-Suastegui, W., Peralta, R., Dobson, A., Batista-García, R., & Dávila-Ramos, S. (2019). **Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis.** *Marine Genomics*, *46*, 16-28. doi: 10.1016/j.margen.2019.03.001

2. Dávila-Ramos, S., **Castelán-Sánchez, H.,** Martínez-Ávila, L., Sánchez-Carbente, M., Peralta, R., & Hernández-Mendoza, A. et al. (2019). **A Review on Viral Metagenomics in Extreme Environments.** *Frontiers In Microbiology*, *10*. doi: 10.3389/fmicb.2019.02403

3. **Castelán-Sánchez, H.,** Elorrieta, P., Romoacca, P., Liñan-Torres, A., Sierra, J., & Vera, I. et al. (2019). **Intermediate-Salinity Systems at High Altitudes in the Peruvian Andes Unveil a High Diversity and Abundance of Bacteria and Viruses.** *Genes*, *10*(11), 891. doi: 10.3390/genes10110891

4. **Castelán-Sánchez, H.,** Meza-Rodríguez, P., Carrillo, E., Ríos-Vázquez, D., Liñan-Torres, A**.,** & Batista-García, R. et al. (2020). **The Microbial Composition in Circumneutral Thermal Springs from Chignahuapan, Puebla, Mexico Reveals the Presence of Particular Sulfur-Oxidizing Bacterial and Viral Communities**. Microorganisms, 8(11), 1677. doi: 10.3390/microorganisms8111677

5. **Castelán-Sánchez H.,** Liang X., Sánchez-Carbente M, & Dávila-Ramos, S. (2021). **Chapter 9 Viral metagenomics** in book Metagenomics and Microbial Ecology: Techniques and Applications. CRC Press, Taylor & Francis. In process.

RESUMEN

Los virus son entidades biológicas ubicuas con la capacidad de infectar bacterias, arqueas y eucariontes. Los virus, al igual que sus hospederos, pueden estar distribuidos en diversos ambientes, incluyendo los extremos. Dentro de este tipo de ecosistemas, un organismo extremófilo funciona como hospedero de los virus para su replicación, sugiriendo que los virus contribuyen al equilibro de la dinámica poblacional de sus hospederos.

A la fecha se tiene información sobre la ecología microbiana de los microorganismos extremófilos, sin embargo, de sus co-simbiontes virales se tiene poco conocimiento. El presente trabajo responde a esta falta de conocimiento, analizando la estructura de las comunidades virales, para poder determinar el papel de los virus en dichos entornos. Además, se evaluó el rol funcional de los virus y sus hospederos.

Inicialmente se presenta el estado del arte de la metagenómica viral, desde la toma de muestras, su procesamiento en el laboratorio hasta el análisis bioinformático, dejando claro que se carece de estándares generalizados que permitan comparaciones entre distintos estudios y evaluar posibles sesgos en los datos.

Posteriormente, se evaluó la estructura general y funcional de viromas y metagenomas en tres ambientes extremófilos acuáticos: ventilas hidrotermales profundas, sistemas acuáticos de salinidad intermedia y un sistema acuático termófilo circumneutral; cada uno con características fisicoquímicas diferentes representando una muestra de los ambientes extremos.

El análisis de la metagenómica estructural reveló que en estos tres ambientes predominan los bacteriófagos de los órdenes Caudovirales y Nucleocytoviricota, mostrando que son ubicuos en estos ecosistemas, no obstante, existe estratificación del viroma dentro de las ventilas hidrotermales. Particularmente en los sedimentos de ventilas hidrotermales, y en los ambientes hipersalinos y termófilos, existen virus característicos del entorno como Microviridae, Hapunaviruses y Herelleviridae respectivamente.

Los resultados del ensamble metagenómico mostraron que los virus de DNA de doble cadena, bacteriófagos de bacterias y arqueas, son predominantes sobre los virus de DNA de cadena sencilla. Además, se recuperaron genomas virales completos correspondientes al género Hapunaviruses en los ambientes hipersalinos.

Los análisis del metagenoma funcional revelaron que las proteínas de virus más abundantes corresponden a secuencias con la estructura del virión, además de las involucradas en procesos de replicación, transcripción y reparación del DNA. Adicionalmente se recuperaron genes virales auxiliares del metabolismo, los cuales están involucrados en el metabolismo de carbohidratos, de aminoácidos, de terpenoides, de azufre, carbono y nitrógeno. Resultados que podría indicar que complementan vías metabólicas de sus hospederos celulares.

Además, un análisis de las presiones selectivas sobre los genes auxiliares del metabolismo mostró que estos evolucionan bajo selección natural purificadora, lo que indica que están conservados y podrían ser relevantes para la adecuación de su hospedero en los ambientes extremos.

ABSTRACT

Viruses are ubiquitous biological entities with the ability to infect bacteria, archaea, and eukaryotes. Viruses, like their hosts, can be distributed in a variety of environments, including extremes. Within these types of ecosystems, an extremophilic organism functions as a host for viruses for their replication, suggesting that viruses contribute to the balance of the population dynamics of their hosts.

To date there is information on the microbial ecology of the Extremophilic microorganisms, however, little is known about their viral co-symbionts. The present work responds to this lack of knowledge, analyzing the structure of viral communities, in order to determine the role of viruses in these environments. In addition, the functional role of viruses and their hosts was evaluated.

Initially, the state of the art of viral metagenomics is presented, from the taking of samples, their processing in the laboratory to the bioinformatic analysis, making it clear that there is a lack of generalized standards that allow comparisons between different studies and evaluate possible biases in the data.

Subsequently, the general and functional structure of viromes and metagenomes was evaluated in three extremophilic aquatic environments: deep hydrothermal vents, intermediate salinity aquatic systems and a circumneutral thermophilic aquatic system; each with different physicochemical characteristics representing a sample from extreme environments.

The structural metagenomics analysis revealed that bacteriophages of the orders Caudovirales and Nucleocytoviricota predominate in these three environments, showing that they are ubiquitous in these ecosystems; however, there is stratification of the virome within hydrothermal vents. Particularly in the sediments of hydrothermal vents, and in hypersaline and thermophilic environments, there are viruses characteristic of the environment such as Microviridae, Hapunaviruses and Herelleviridae respectively.

The results of the metagenomic assembly showed that double-stranded DNA viruses, bacteriophages from bacteria and archaea, are predominant over single-stranded DNA viruses. Furthermore, complete viral genomes corresponding to the genus Hapunaviruses were recovered in hypersaline environments.

Analysis of the functional metagenome revealed that the most abundant virus proteins correspond to sequences with the virion structure, in addition to those involved in DNA replication, transcription and repair processes. Additionally, viral genes auxiliary to metabolism were recovered, which are involved in the metabolism of carbohydrates, amino acids, terpenoids, sulfur, carbon and nitrogen. Results that could indicate that they complement metabolic pathways of their cellular hosts.

Furthermore, an analysis of the selective pressures on the genes ancillary to metabolism showed that they evolve under purifying natural selection, indicating that they are conserved and could be relevant for the adequacy of their host in extreme environments.

RECONOCIMIENTOS ACADÉMICOS

Esta tesis de desarrollo bajo la tutoría de la Dra. Sonia Dávila Ramos en el Laboratorio de Metagenómica Viral y Bioinformática en el Centro de Investigación en Dinámica Celular de la Universidad Autónoma del Estado de Morelos, con apoyo de la beca No. 493622 del Consejo Nacional de Ciencia y Tecnología.

Durante el desarrollo de este trabajo se realizaron estancias cortas de investigación en:

- Universidad de San Antonio de Abad del Cusco en Perú, con la Dra. María Antonieta Quispe Ricalde
- Universidad Nacional Autónoma de México Centro de Ciencias de la Complejidad campus Yucatán, con el Dr. Ernesto Pérez Rueda
- Colegio de Postgraduados campus Campeche, con la Dra. Itzel López Rosas

Durante el desarrollo de este trabajo se recibieron becas internacionales para la asistencia a cursos:

- Travel award OSU Viromics Workshop, The Ohio State University EUA, Mayo 2019
- Travel award CABANA Workshop: NGS analysis applied to virome sequencing in agricultural systems, University of Costa Rica Marzo 2020

El comité tutoral que participó en el desarrollo de este proyecto fueron:

- Dra. Sonia Dávila Ramos CIDC-UAEM
- Dr. Ramón A. González García-Conde CIDC-UAEM
- Dr. Luis Fernando Lozano Aguirre-Beltrán CCG-UNAM

AGRADECIMIENTOS

A la Universidad Autónoma del Estado de Morelos y al Centro de Investigación en Dinámica Celular por brindarme la oportunidad de realizar mis estudios de Doctorado en esta casa de estudios.

Dra. Sonia Dávila Ramos, gracias por permitirme y darme la libertad de realizar este proyecto de doctorado y poder desarrollar mis ideas, por tu gran amistad, y apoyarme en todo momento aprecio, y mucho todo lo aprendido. ¡Gracias Sonia!

Dr. Ramón A. Gonzalez, por todos los comentarios recibios durante el desarrollo de este proyecto y el rigor científico que me ayudaron en mi formación.

Dr. Luis Lozano, por los comentarios durantes los tutorales y las herramientas bioinformáticas.

Al comité sinodal de esta tesis los Doctores María del Rayo Sanchez, Lorenzo Segovia, Enrique Merino, Jorge Folch y Armando Hernandez por leer y revisar esta tesis que enriquecieron el trabajo.

Dr. Ernesto Pérez Rueda, gracias por el apoyo bionformático, y la discusión de ideas que apoyaron mucho el desarrollo de este trabajo.

Dr. Ramón Batista, por su apoyo y por la confianza de hacerme participe de sus proyectos.

Dra. Maria Antonieta Quispe, por todo el apoyo brindado durante mi estancia en Cusco Perú, y por la confianza que me brindaste para ser parte de tus colaboradores.

Dr. Simon Roux, por el soporte en bioinformática, y apoyo recibido durante mi estancia en Ohio State University.

Dra. Valerie de Anda, por el soporte en bioinformática y apoyo en los ciclos biogeoquímicos.

Dra. Itzel Lopez Rosas, gracias por tu amistad y apoyo en el desarrollo de este proyecto.

Dra. Wendy García gracias por tu amistad y apoyo en el desarrollo de este proyecto.

Dra. Norma Rojas, gracias por el apoyo brindado durante los muestreos en Puebla y por confiar en este proyecto.

Dra. Selene Zárate, por la confianza y comentarios recibios que enriquecieron mi formación.

A los miembros del laboratorio del Dr. Gonzalez García-Conde, Virología Molecular del Centro de Investigación en Dinámica Celular por los comentarios durante los seminarios de grupo.

A Arturo Liñan, mucho de este trabajo realizado aquí fue gracias a tu apoyo, gracias por escucharme siempre. Eres un gran ser humano, científico y amigo! ¡Gracias Paps!

A Pedro Rommoaca, gracias por todo el apoyo durante los muestreos en Perú y en México, los muestreos no hubieran sido posibles sin tu experiencia. Por mostrarme tu cultura y ser un gran amigo. ¡Gracias Causa!

A mis amigos Peruanos Kevin, Paola, Ilución, David y Sharly muchas gracias por su amistad y hacer mis estancias en Perú mi agradables y mostrarme lo maravilloso de su país.

José Hernández Valle, gracias por tu amistad, por compartir tu conocimiento, por el apoyo en los momentos difíciles siempre lo recordaré, por compartir las clases durante el doctorado. ¡Gracias amigo!

A mis amigos, Enrique, Jose y Fabricio. Por los momentos compartidos en los congresos y cursos, y las reuniones.

A Chrystian Hernández, gracias por estar siempre presente apoyando a la distancia y escucharme siempre y leerme, siendo para mi un apoyo emocional importante, lo agradezco.

A Ulises Rodríguez, gracias por escucharme, por apoyarme, colaborar en los proyectos, discutir de ciencia, tus consejos, los conciertos, por la música compartida, todo este tiempo de amistad.

A todos los miembros de Solaria Biodata Sandra, Blanca, Luis, Robert G., Roberto C. y Josh, por comprender muchas veces mis ausencias, gracias por el apoyo recibido y su amistad.

A Luz Torres, muchas por tu amistad y el apoyo en todo momento.

A Annely Lujano, gracias amiga por siempre escucharme y darme excelente consejos de vida que me ayudaron mucho en esta etapa lo agradezco sinceramente.

A Jorge Rojas Vargas, gracias por enseñarme muchas cosas nuevas, por todos esos consejos, comprensión y tu apoyo en muchos momentos. ¡Gracias por todo!

A mi padres Guillermina Sánchez, Benito Castelán, de los que siempre recibí su apoyo incondicional para realizar este doctorado y darme la libertad para seguir formado académicamente y mi hermano Juan Castelán por el apoyo recibido.

ÍNDICE

INTRODUCCIÓN GENERAL

Los virus son los entes biológicos más abundantes en la biosfera, por ejemplo, se ha estimado un número de 1.2 x 10^{30}, 2.6 x 10^{30}, 3.5 x 10^{31}, y 0.2-2.5 x 10^{31} virus en océanos abiertos, suelos, superficies oceánicas y terrestres respectivamente (Mokili et al., 2012; Bolduc et al., 2012; de Cárcer et al., 2015). Lo anterior implica que una cantidad importante de virus es hasta el momento desconocida.

Los virus pueden influir en la abundancia, composición, diversidad genómica, y la evolución de las comunidades microbianas presentes en los distintos ecosistemas (de Cárcer et al., 2015; Koonin et al., 2012). Por ejemplo, dentro de los ambientes acuáticos extremos, existen casos particulares que, por sus características, se pueden hacer preguntas relevantes para entender la dinámica y evolución de las comunidades microbianas, como: ¿Las poblaciones virales están estratificadas?, ¿las barreras biogeográficas delimitan su diversidad genómica?, ¿cómo es la estructura y el papel del viroma dentro de estos ecosistemas? (López et al., 2009).

Algunas de estas preguntas han sido respondidas, en su mayoría, con el análisis en los océanos, pero poco en los ambientes extremos. Ejemplos de estos estudios en los últimos ecosistemas, son los reportados por de Cárcer et al. en 2015 en el aguas del polar Ártico, donde los virus de ssDNA (*single strand DNA*) fueron abundantes y diversos, los cuales infectan principalmente a eucariontes (de Cárcer et al., 2015); o los obtenidos por López et al. 2009, en agua dulce del Antártico, donde predominan virus de ssDNA y dsDNA (*doble strand DNA*), donde la abundancia viral, se ve modifica dependiendo de la estación del año (López et al., 2009). Ambas regiones psicrófilas geográficamente aisladas presentan una gran diversidad genética de virus de eucariontes (Mizuno et al., 2014).

La elevada diversidad genética en estos ambientes podría explicar los distintos papeles ecológicos que juegan los virus en el control de floración de algas, en la interacción y mortalidad de bacterias, en la contribución de nutrientes, así como en la regulación de ciclos biogeoquímicos (Anesio y Bellas., 2010; Ortmann y Suttle., 2005; de ; Mizuno et al., 2014; Cárcer et al., 2015). Esto hace proponer a las entidades virales como entes importantes en el desarrollo de sus hospederos en estos ambientes.

Aunque se cuenta con información de los metagenomas acuáticos de estos ecosistemas, han sido poco exploradas las estructuras de las comunidades de virus

presentes. Es por ello por lo que en esta tesis se describieron y compararon las comunidades virales en diferentes ambientes acuáticos extremos, con el objetivo de comprender cómo es la estructura del viroma, como contribuyen a sus comunidades y su potencial funcional.

En el Capítulo 1 se hace una revisión de los aspectos metodológicos para la implementación de la metagenómica viral en el laboratorio y las herramientas bioinformáticas para el análisis de secuencias virales.

El Capítulo 2 presenta los resultados de la estructura y diversidad de las comunidades virales a través de diferentes ambientes extremos alrededor del mundo, mediante la búsqueda de los virus en bases de datos.

En el Capítulo 3 se analizó la estructura de las comunidades virales dentro de diez (10) ventilas hidrotermales y el potencial funcional que tienen los virus dentro de este ecosistema.

El Capítulo 4 presenta el análisis metagenómico de ambientes acuáticos hipersalinos en los Andes Peruanos, describiendo las comunidades de bacterias y virus, además de la obtención de genomas completos particulares de ambientes hipersalinos.

Finalmente, en el capítulo 5 se analizó un ambiente acuático termófilo terrestre ubicado en Chignahuapan, Puebla, donde se encontraron comunidades de bacterias y virus particulares en dicho ambiente circunneutral termófilo.

Microorganismos extremófilos y ambientes extremos

Los organismos extremófilos están representados en los tres dominios de la vida Archaea, Bacteria y Eucariontes y son aquellos que prosperan bajo condiciones fisicoquímicas que se consideran extremas para el humano, como, por ejemplo, alta/bajas temperaturas, pHs ácidos y/o básicos, alta salinidad y presiones (Rothschild y Mancinelli., 2001; Merino et al. 2019).

El término "extremófilo", acuñado por Robert MacElroy en 1974, considera a las condiciones fisicoquímicas de los organismos mesófilos como las condiciones normales de supervivencia. Cualquier estado por fuera de esta normalidad es considerada extrema. Los organismos que se desarrollan en ambientes extremos pueden ser categorizados en: *extremófilos*, los cuales requieren una o más condiciones fisicoquímicas extremas para vivir; o *extremo-tolerantes*, aquellos que pueden tolerar una o más condiciones extremas, pero cuyo crecimiento óptimo se da en condiciones estándar de laboratorio.

Además de esta clasificación, los organismos extremófilos se pueden clasificar de acuerdo con el ambiente en el que se desarrollan (ver Tabla1 y figura 1):

- Termófilos, termófilos moderados, Hipertermófilos: Microorganismos que crecen a altas temperaturas.
- Psicrófilos: Microorganismos que crecen bien a bajas temperaturas.
- Acidófilos y Alcalófilos: Microorganismos adaptados a pH ácidos y básicos.
- Barófilos: Microorganismos que crecen mejor a altas presiones.
- Halófilos: Microorganismos que requieren una gran concentración de sales para su crecimiento.
- Xerófilos: Microorganismos que crecen en ambientes con poca disponibilidad de agua y son resistentes a la desecación.
- Endolítico: Microorganismos que crece dentro de las rocas.
- Hipólito: Microorganismos que crece en rocas y desiertos fríos.
- Metalotolerantes: Microorganimos que toleran altas concentraciones de metales pesados.
- Toxitolerante: Microorganismos que toleran altas concentraciones de agentes tóxicos (por ejemplo, disolventes orgánicos).
- Resistentes a la radiación: Organismos que pueden sobrevivir a dosis de radiación ionizante mucho mayores que las presentes en el entorno natural.

- Poliextremófilo: Organismo que tolera o prefiere diferentes condiciones fisicoquímicas extremas para su crecimiento.

Tabla 1. Nomenclatura de extremófilos y rangos de parámetros fisicoquímicos (modificada de Merino et al., 2019).

			Bajo a Alto		
pH	Hiperacidos (<pH3)	Acidófilos (<pH 5)	Neutros (pH 5-9)	Alcalinos (>pH 9)	Hiperalcalinos (> pH 11)
Temperatura		Psicrófilos (< 20ºC)	Mesofilo (20-45º)	Termófilo (42-80ºC)	Hipertermófilos (> 80ºC)
Salinidad		No halófilos (< 1.2%)	Halotolerantes (1.2-2.9%; tolerante < 14.5%)	Halófilo (>8.8%)	Halófilo extremo (> 14.6%, no puede crecer < 8.8%)
Presión			Piezotolerantes o barotolerantes (0.1-10 MPa)	Piezofilo o barofilo (10-50 MPa)	Hiperpiezofilo o hiperbarofilo (> 50MPa)
Actividad en agua			Xerófilo (aw <0.7)		
Poli Extremófilos			Tolerancia o preferencia por múltiples parámetros combinados		

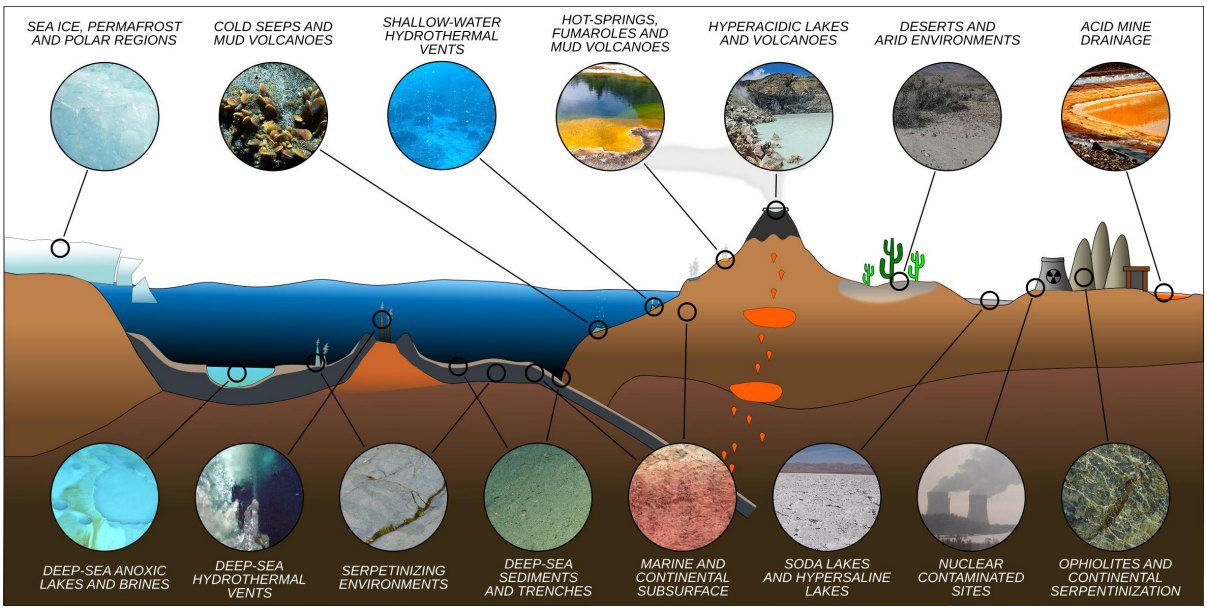


Figura 1. Representación de diferentes ambientes extremos (Tomada de Merino et al., 2019).

La diversidad filogenética de los organismos extremófilos es amplia, pues estos se encuentran distribuidos en los tres dominios de la vida. Los extremófilos que están adaptados a más de una condición extrema para vivir pueden estar representados en diversos clados en el árbol filogenético (Leal Dalmaso et al., 2015).

En general, la diversidad filogenética de los microorganismos es compleja para estudiar por su amplia distribución en el árbol de la vida, pero para ello se han utilizado típicamente el marcador molecular RNA ribosomal 16S, pero dentro estas reconstrucciones filogenéticas no se han incorporado los virus (Figura 2). A pesar de ello, se espera que los virus estén representes en los tres dominios porque estos tienen la capacidad de infectar a diversos hospederos dentro de estos dominios (Leal Dalmaso et al., 2015).

La mayoría de los microorganismos extremófilos son procariontes, de los cuales, las arqueas se consideran los principales microorganismos que prosperan en ambientes extremos, por ejemplo, el filo Crenarchaeota (que incluye hipertermófilos, acidófilos, reductores y/u oxidantes del azufre y quimiolitoheterótrofos) y el filo Euryarchaeota (que incluye microorganismos metanógenos, termoacidófilos e hiperhalófilos). Ejemplos característicos a nivel de especie están *Methanopyrus kandleri* que crece a 112ºC, *Picrophilus torridus* que crece a un pH 0.06, *Haloferax volcanii* que crece a concentraciones de saturación de Cloruro de Sodio (NaCl) y Cloruro de Potasio (KCl) y *Halobacterium* sp. NRC-1 adaptada a vivir en altos niveles de luz ultravioleta (UV) (> 100 J/ m^2) (Rampelotto, P. 2013; Rothschild y Mancinelli, 2001).

En el caso de las bacterias, éstas pueden estar adaptadas a diferentes ambientes extremos, sin embargo, proliferan mejor en condiciones fisicoquímicas estándar. Uno de los ejemplos característicos son las Cianobacterias, adaptadas a diversos ambientes extremos, desde tapetes microbianos en ambientes psicrófilos hasta ambientes termófilos (Rothschild y Mancinelli, 2001).

Respecto a los eucariontes, los hongos son los principales microorganismos presentes en los ambientes extremos, como en condiciones alcalinas, en desiertos, en océanos profundos, y en regiones hipersalinas (van Thielen y Garbary, 1999). Sin embargo, el eucarionte poliextremófilo más conocido es el tardígrado que puede sobrevivir a temperaturas de -272ºC y 151°C, condiciones de vacío, presión de 6.000 atm y exposición a rayos X y gamma (Rothschild y Mancinelli, 2001).

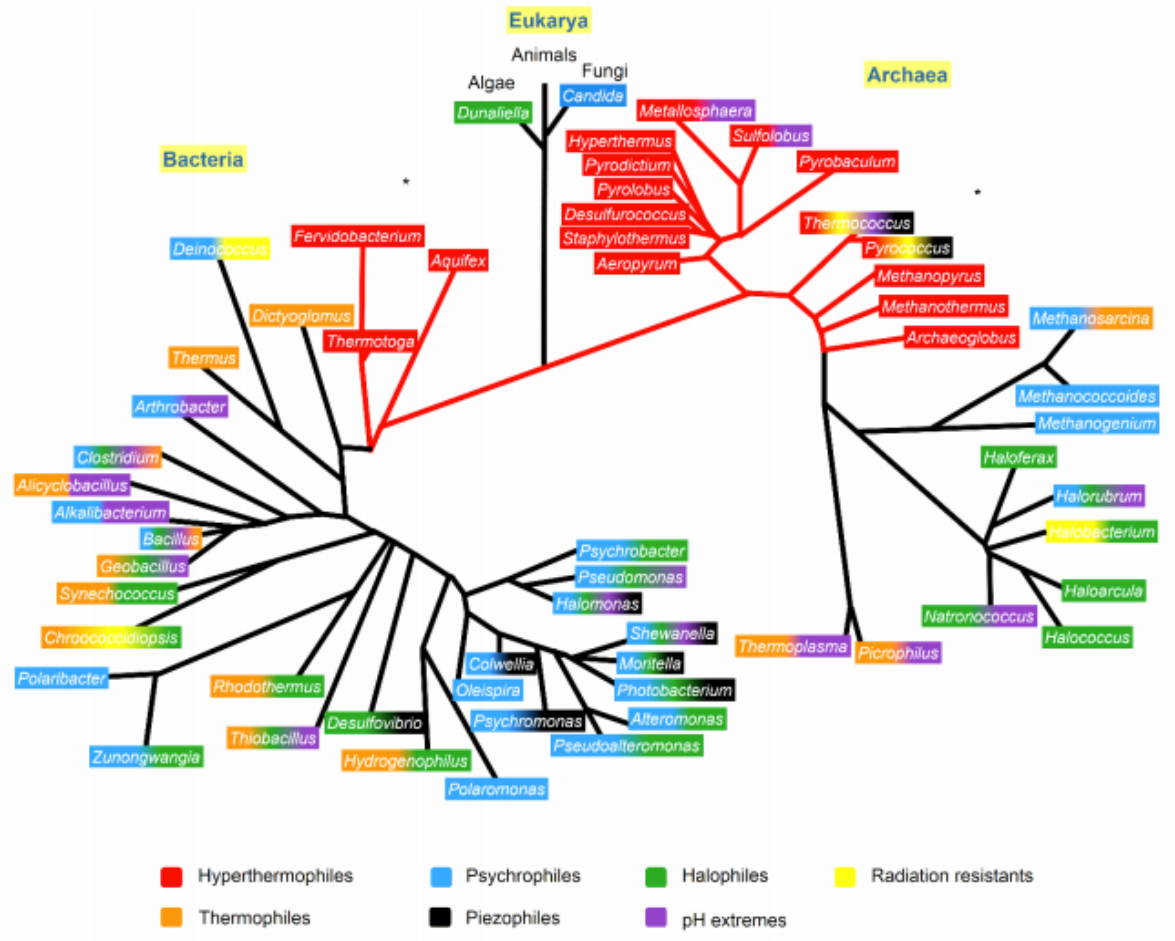


Figura 2. Árbol filogenético de microorganismos extremófilos. Muestra los principales géneros de microorganismos extremos en los tres dominios de la vida, el código de color corresponde a distintos ambientes extremos. Los virus no están incluidos (Tomado de Leal Dalmaso et al., 2015).

Por otro lado, los ambientes extremos están distribuidos en distintas regiones del mundo, sin embargo, cualquier lugar donde haya agua existe una mayor diversidad de microorganismos (Merino et al., 2019). Por lo tanto, el agua influye directamente en las condiciones fisicoquímicas de estos ecosistemas, la cual puede controlar la composición y diversidad de la comunidad microbiana. Sin embargo, los parámetros abióticos como temperatura, salinidad, pH y presión modifican la diversidad y abundancia de las comunidades microbianas (Merino et al., 2019).

Ahondando en los parámetros fisicoquímicos, se sabe que los límites teóricos actuales de temperatura para la vida son −40°C a 150°C, los cuales pueden extenderse. Sin embargo, la complejidad y diversidad de la comunidad disminuye a medida que se incrementa la temperatura (Merino et al., 2019).

Varios estudios demuestran que el pH afecta la diversidad de la comunidad microbiana más que cualquier otro parámetro (Lauber et al., 2009; Rousk et al., 2010). Esto es porque muchos microorganismos deben mantener un pH neutro en el citoplasma (7.2–8.7) para mantener una homeostasis, y cualquier incremento o disminución afecta su desarrollo considerablemente. Los microorganismos deben mantener su pH intracelular para ser metabólicamente activos, pero los extremófilos se han adaptatado para mantener su pH intracelular mediante la excreción de metabolitos orgánicos, como el ácido láctico o el ácido acético, cambiando así el pH circundante (Lauber et al., 2009; Rousk et al., 2010). Se ha demostrado que tanto en entornos naturales y como en el laboratorio los microorganismos puede alterar significativamente su pH intracelular debido a reacciones metabólicas. En general, se sabe que la composición del microbioma varía drásticamente con el cambio del pH, donde se ha sugerido que las arqueas y hongos, se ven menos afectados, mientras que las bacterias se afectan considerablemente (Merino et al., 2019).

Por otro lado, la salinidad también es considerada como otro parámetro determinante de la estructura de las comunidades dentro de los ecosistemas extremos. A medida que se incrementa la concentración de sal (NaCl), la diversidad y abundancia de microorganismos disminuye, favoreciendo la presencia de arqueas a altas concentraciones de NaCl. Generalmente los halófilos extremos crecen óptimamente a concentración de 150–300 g/L (2.5–5.2 M) de NaCl, y los halófilos moderados a una concentración de 30 a 150 g/L (0,5 a 2,5 M) de NaCl (Ventosa et al., 2015; Andreí et al., 2012).

Aunado a la temperatura y salinidad, se ha informado que la presión influye en el crecimiento microbiano, especialmente bajo temperaturas extremas. En la Tierra la presión varía de 0,1 a 112 Mpa. Se han aislado varios piezófilos y microorganismos piezotolerantes de lugares de aguas profundas , y el poseedor del récord actual es *Thermococcus piezophilus*, capaz de sobrevivir hasta 125 MPa. Los piezófilos tienen tiempos de generación más bajos a mayor presión que a presión atmosférica , y considerando que la profundidad promedio del océano es de 3.800 m, con temperaturas del fondo entre 0 y 3°C. Ya en la superficie terrestre se ha observado

que la diversidad bacteriana de reduce alrededor de 3700 m, pero otros factores como las temperaturas extremas, la radiación UV y la falta de nutrientes probablemente afectan a la comunidad microbiana de manera más significativa que los cambios de presión (Merino et al., 2019).

Por último las radiaciones son otro paramétro que puede afectar la diversidad de la comunidades. Todos los ecosistemas están expuestos a diferentes fuentes de radiación como radiación UV, rayos X, rayos gamma (Merino et al., 2019). En particular los rayos ultravioleta y gamma, pueden impactar a los microbios de forma directa e indirecta en la formación de especies reactivas de oxígeno (ROS), las cuales pueden dañar el DNA, las proteínas, los lípidos y el RNA. Los microorganismos han sufrido diferentes mecanismos de adaptación como cambios en las funciones de reparación del ADN (Byrne et al., 2014), acumulación de Mn (II) (Daly et al., 2004), y la producción de pigmentos (Mojib et al., 2013). En conjunto, todos estos paramétros influyen en la diversidad y abundancia de los ambientes extremos.

Metagenómica

Debido a las condiciones extremas donde viven los microorganismos extremófilos es complejo realizar una caracterización de estos mediante técnicas de medios de cultivo en el laboratorio, por lo que se ha optado por el uso de técnicas independientes de cultivo para su estudio, como la metagenómica.

La metagenómica es una técnica que permite recuperar los genomas de microorganismos directamente de muestras ambientales, independientemente de la naturaleza de la muestra y la abundancia de los microorganismos, sin necesidad de realizar aislamiento a través de un cultivo, ni clonación de fragmentos de ADN (Thomas et al., 2012; Bharti, y Dominik 2019).

La metagenómica puede dividirse en dos tipos, la metagenómica estructural y la metagenómica funcional. La primera describe las comunidades de microorganismos presentes en un ambiente en particular, es decir, obtiene información de cuáles y cuántos microorganismos viven en un determinando ambiente. Mientras que la segunda sirve para inferir las posibles funciones metabólicas que tienen los genomas recuperados en el ambiente, y con ello predecir el potencial metabólico de los microorganismos dentro del ecosistema (Sharpton, 2014).

Con el inicio de la secuenciación masiva se comenzó el desarrollo de la metagenómica, pero en el principio las técnicas de secuenciación podían secuenciar solo fragmentos de DNA, y se secuenciaron marcadores moleculares, lo cual se le conoce como *metaprofiling*. Con el mejoramiento del rendimiento de la secuenciación se secuenciaron genomas completos y metagenomas sin utilizar un marcador molecular en particular, lo que se conoce como secuenciación tipo *shotgun*. En la Tabla 2 se muestran algunas diferencias de los tipos de secuenciación.

Como se mencionó anteriormente las primeras aproximaciones para el estudio de los metagenomas comenzaron con los marcadores moleculares, como los genes ribosomales 16S, 18S que actualmente se siguen usando, además de la región espaciadora interna transcrita ITS (*internal transcribed spacer*) en le caso de los hongos. Este *metaprofiling* se logra a través de la clonación y secuenciación de estos marcadores, con el objetivo de describir la estructura de una comunidad (Bharti, y Dominik 2019).

En el caso del marcador molecular 16S se utilizan las regiones hipervariables (V1-V9) del 16S rRNA de procariontes. Sin embargo, una de las limitaciones de esta técnica es que generalmente no es posible distinguir dos especies que están estrechamente relacionadas o que son cepas clonales. Por ejemplo, las bacterias *Escherichia coli* O157:H7 y *E. coli* K-12, no podrían diferenciarse por un análisis de este tipo (Bharti, y Dominik 2019).

En el caso del marcador ribosomal 18S, éste es aplicado para la identificación taxonómica de eucariontes principalmente de los hongos. Dicho marcador comprende regiones hipervariables y conservadas. Finalmente, la región espaciadora interna transcrita (ITS), se encuentra entre los genes 18S y 5.8S, y cuenta con un alto grado de variación de la secuencia, lo que permite realizar una identificación taxonómica a nivel de género (Banos et al., 2018).

Dado que la secuenciación de marcadores se limita a una molécula en particular, es difícil poder inferir funciones de la comunidad a partir de dicho marcador por lo tanto se hace uso de la metagenoma tipo shogun.

La metagenómica tipo *shotgun* es la secuenciación del ADN total de una muestra, con ello se pueden recuperar genomas de nuevas especies cultivables y no cultivables. La ventaja que tiene la metagenómica tipo *shotgun*, en comparación con las técnicas anteriores, es que además de obtener la asignación taxonómica de las secuencias y de esta forma conocer la estructura de las comunidades dentro de un

ecosistema, se puede conocer el potencial funcional basándose en las secuencias completas de genes que codifican para proteínas (Quince et al., 2015).

Tabla 2. Diferencias entre la secuenciación de amplicon y *shotgun*

	Secuenciación de amplicon (*metaprofiling*)	Secuenciación tipo *shotgun*
Objetivo	- Secuenciación de amplicones de genes marcadores	- Secuenciación de todo el ADN accesible de una comunidad mixta
Formación de la biblioteca metagenómica	-Se elabora con iniciadores específicos para un marcador molecular	- Opcional utiliza iniciadores aleatorios y, por lo tanto, sufre mucho menos de sesgo de la PCR.
Usos	- Es una métrica de composición de la comunidad. - Informa sobre la abundancia relativa de géneros recuperados. - Puede rastrear los cambios en la estructura de la comunidad.	- Descripción de la composición de la comunidad (Eucariontes, Bacterias, Arqueas, Hongos y Virus). - Descripción del potencial funcional - Conocimiento sobre los genomas de microbios aún no cultivables - No hay un problema de confusión en el número de copias ni un sesgo de PCR drástico para cálculo de la abundancia.
No utilizar	- Cálculo de abundancia de organismos (o abundancia relativa de organismos). - Copias de genes recuperadas ≠ recuentos de organismos - Cálculo de número de copias de genes, éste varía según el genoma u organismo.	- Abundancia de organismos - Análisis de transcriptómica o proteómica, cada uno brinda información sobre la regulación celular en diferentes niveles.

Las aplicaciones de la metagenómica van desde la identificación de patógenos, la seguridad alimentaria, el diagnóstico en muestras clínicas y muestras ambientales. Además, brinda información de la diversidad y la abundancia de microorganismos en esos tipos de muestras.

Adicionalmente está la metatranscriptómica, que permite la caracterización y análisis funcional del microbioma, a través de la secuenciación del ARN total, y finalmente la metaproteómica que es la secuenciación total de las proteínas presentes en una muestra.

Análisis del metagenoma shotgun

Un estudio típico de metagenómica tipo *shotgun* consta de cuatro pasos, después el diseño inicial del estudio: (i) la recolección, procesamiento y secuenciación de las muestras; (ii) pre-procesamiento de las lecturas de secuenciación; (iii) análisis de secuencias para perfilar características taxonómicas, funcionales y genómicas del microbioma; (iv) análisis de post-procesamiento estadístico y biológico (Quince et al., 2015) (Figura 3).

El primer paso de pre-procesamiento de las lecturas de secuenciación consiste en tomar los archivos tipo fastq provenientes de una corrida de secuenciación masiva y realizar un control de calidad.

Anterior al control de calidad, un paso que se puede realizar es un "demultiplixing", que consiste en dividir los archivos fastq de acuerdo con sus identificadores o "barcodes".

El control de calidad es la evaluación de la calidad de las lecturas, que indica si un nucleótido fue introducido correctamente durante el proceso de secuenciación. Esta calidad viene dada en una escala de *Phred score*, que se define como una propiedad que está relacionada logarítmicamente con las probabilidades de error de las llamadas de base; con una escala de Q1 a Q60, donde 60 es la probablidad que 1 base entre 1,000,000 sea incorrecta (Dominguez et al., 2018). Siendo el límite mínimo permitido un Q20 donde es 1 en 100 o un 99% de probabilidad que esa base de haya introducido de forma correcta. Otros parámetros importantes durante el control de calidad es la revisión del contenido de GC, el cual tiene una distribución gaussiana, si existe desviaciones de la curva, podría ser indicativo de contaminante de la muestra

Posteriormente se filtran las secuencias con base en la calidad y se remueven los adaptadores. Los adaptadores son secuencias exógenas que se anclan en los extremos del DNA para para protección, además de que tiene la función de reconocimiento del instrumento de secuenciación y etiquetar a la muestra para poder realizar combinaciones. (Quince et al., 2015; Domínguez et al., 2018).

Una vez que se han eliminado los adaptadores y se filtraron las secuencias de baja calidad se puede realizar un ensamble de las secuencias, utilizando el mismo enfoque para genómica, ensambladores basados superposición de consenso (overlap-layout-consensus OLC) y grafos de Bruijn (De-Bruijn-Graph DBG) (Li et al.,

2011). Donde el algoritmo OLC es más adecuado para las lecturas largas de baja cobertura, mientras que el algoritmo DBG es más adecuado para lecturas cortas de alta cobertura. Actualmente los ensambladores más utilizados son los del algoritmo OLC para secuenciadores de tercera generación, mientras que los ensambladores DBG son más utilizados para secuencias cortas (Li et al., 2011).

Seguido de esto se puede hacer reconstrucción de genomas a partir de metagenomas o MAG (*Metagenome assembly genome*) por sus siglas en inglés. Un MAG es un ensamblaje de un solo taxón basado en uno o más metagenomas agrupados, que contiene secuencias que representan cercana relación a un genoma individual real (que podría coincidir con un aislado ya existente o representar un aislado nuevo).

Un MAG se obtiene mediante a partir los contigs resultantes de los ensambles, los cuales son agrupados de acuerdo con su composición de secuencia, es decir, son secuencias similares, las secuencia. Los cuales posteriormente tienen que ser asignados taxonómicamente para conocer su género y especie.

Una vez obtenidos los contigs se puede realizar la asignación taxonómica, pero no necesariamente se tiene que realizar directamente de los ensambles, esta se puede realizar a partir de las lecturas. Para esta tarea existen tres aproximaciones como analizar la frecuencia de k-meros (secuencia de tamaño k) en las lecturas crudas sin procesar, y compararlo con un modelo entrenado con secuencias de genomas conocidos, como se implementó en Kraken2 (Wood et al., 2019), Centrifuge (Kim et al., 2016) y Kaiju (Menzel et al., 2016).

En un segundo método, las lecturas de los datos metagenómico se mapean con una base de datos de genomas de referencia como Proyecto de Microbioma Humano o NCBI RefSeq. Generalmente, las herramientas diseñadas para realizar la comparación con base en la similitud entre las lecturas y la referencia, son BWA, Bowtie2 o STAR (Bengtsson, 2018).

El tercer método se basa en la identificación de genes marcadores filogenéticos en lecturas sin procesar para estimar la abundancia de cada taxón en el metagenoma, por ejemplo, Metaphlan2 o TIPP. Estos métodos deben considerarse perfiladores, ya que no intentan clasificar el conjunto completo de lecturas, sino que reconocen la identidad de genes marcadores particulares para inferir la composición de la comunidad a partir de estos (Bengtsson, 2018; Tamames et al., 2019).

Posterior a esto, se puede realizar una anotación funcional para asignar información biológicamente relevante a las secuencias de aminoácidos, y estas se clasifican de acuerdo con diferentes bases de datos como SEED, KEEG, Cazyome, NCBI. Esta comparación se realiza mediante la comparación con Blast de las secuencias codificantes SEED (Overbeek et al., 2013), KEGG (Kanehisa et al. 2010), Cazyome (Ospina et al., 2010).

Adicionalmente, se realizan análisis de diversidad. La diversidad se describe típicamente dentro de una muestra (diversidad alfa) o entre muestras (diversidad beta). La mayoría de estas estimaciones se pueden aplicar a genes marcadores, géneros o especies. La diversidad alfa cuantifica la diversidad dentro de la muestra y generalmente se caracteriza por riqueza de variantes la cual es estimada mediante Chao, ACE, y las métricas Shannon y Simpson estiman la riqueza de especies y la uniformidad de las especies (Bengtsson, 2018).

Para calcular la diversidad beta se calculan las distancias o diferencias al comparar dos especies, y estos se realizan sin distancias filogenéticas como Bray-Curtis o Jaccard y con distancias filogenéticas como UniFrac (Lozupone, y Knight, 2005).

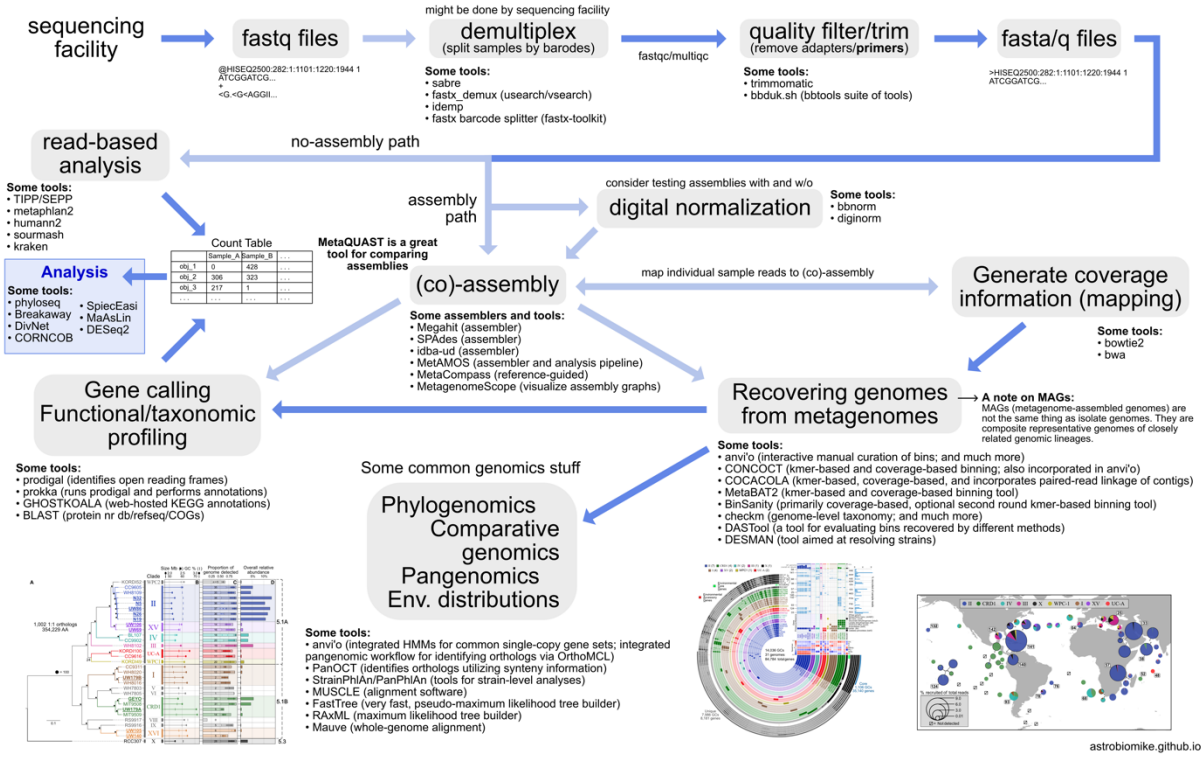


Figura 3. Flujo de trabajo de metagenómica. Flujo de trabajo general para un análisis de metagenómica estructural y funcional.

Virus en ambientes extremos

El descubrimiento de virus en los ambientes extremos se realizó mediante la investigación de las comunidades microbianas, es decir, en sus hospederos, pudiendo determinar que los virus también prosperan en dichos nichos. Aunado a esto se presume que los virus dentro de los ambientes extremos son responsables del equilibrio de la dinámica poblacional, de mejorar la adecuación (fitness) de sus hospederos, e influir en los ciclos biogeoquímicos (Dávila 2019, Gil et al., 2021).

Los virus que infectan a las arqueas tienen morfologías muy diversas cuando se observan en el microscopio electrónico, sin embargo, han sido difíciles de aislar. A pesar de ello se han logrado identificar que los virus de dsADN son los que más infectan a arqueas de ambientes extremos, los cuales han sido agrupados en 13 familias: Ampullaviridae (morfología en forma de botella), Bicaudaviridae (en forma de huso), Clavaridae (en forma de palo), Fuselloviridae (en forma de huso), Guttaviridae (en forma de gota), Globuloviridae (forma redonda), Lipothrixviridae (en forma de pelo graso), Myoviridae (forma de músculo, contráctil), Portogloboviridae (forma de globular), Rudiviridae (forma de varilla), Sphaerolipoviridae (forma de esfera), Tristromaviridae (formado por tres capas) y Turriviridae (apéndices en forma de torreta en expansión de la cápside). Mientras que los virus de ssADN se restringen a dos familias Pleolipoviridea (compuestos por muchos lípidos) y Spiravidiae (en forma de espiral) (Figura 4) (Le Romancer et al., 2006; Gil et al., 2021).

Respecto a los virus que infectan a bacterias (bacteriófagos) son los más abundantes y diversos en el planeta (Dion et al., 2020). Principalmente los virus dentro del orden de los Caudovirales como Myoviridae, Siphoviridae, Podoviridae, Herelleviridae y Ackermannviridae (en forma pleomórfica conectado por cola), son virus de dsADN. Otras familias de virus con dsADN son los Tectiviridae (en forma poliédrica), Corticoviridae (en forma poliédrica), Plasmaviridae (en forma pleomórfica) y Sphaerolipoviridae. Por otro lado, los virus de ssADN incluyen las familias Microviridae (en forma poliédrica) e Inoviridae (en forma filamentosa). También se incluyen virus de ARN de doble cadena (dsARN) como son las familias Cystoviridae y Picobirnaviridae, y ARN de cadena sencilla (ssARN) como Leviviridae (Dion et al., 2020; Gil et al., 2021).

En el caso de virus de eucariontes en ambientes extremos, se conoce aún menos, y la mayoría de estos corresponden a virus de ARN. Algunas de las familias

identificadas son la Phasmaviridae y respecto a los virus de dsADN la Herpesviridae, y los Retrovirus; así también se han reportado virus de dsADN dentro del filo Nucleocytovirocota que incluye a las familias Poxviridae, Iridoviridae, Ascoviridae, Asfarviridae, Marseilleviridae, Mimiviridae, y Phycodnaviridae, así como linajes como pithoviruses, pandoraviruses, molliviruses, y faustoviruses entre otros que aún no han sido clasificados (Gil et al., 2020).

Por otro lado, de acuerdo con los análisis metagenómicos, se ha podido determinar la diversidad y abundancia de virus en los ambientes extremos. Estos análisis han demostrado que los bacteriófagos, dentro del orden de los Caudovirales, son ubicuos, mientras que en otros virus quedan restringido a ambientes particulares y esto depende de la adaptación de su hospedero a un determinado ambiente (Dávila et al., 2019).

Por ejemplo, en los ambientes hipertermófilos ácidos, se han caracterizado principalmente la presencia por ejemplo de los Fuselloviridae, los cuales tienen morfologías únicas como forma de limón, de huso, de gota, entre otras. Así mismo los virus de ambientes hipersalinos, donde prosperan virus adaptados a las concentraciones de sal como son los Pleolipoviridae que infectan a la familia la familia Halobacteriaceae (Le Romancer et al., 2006; Atanasova et al., 2018).

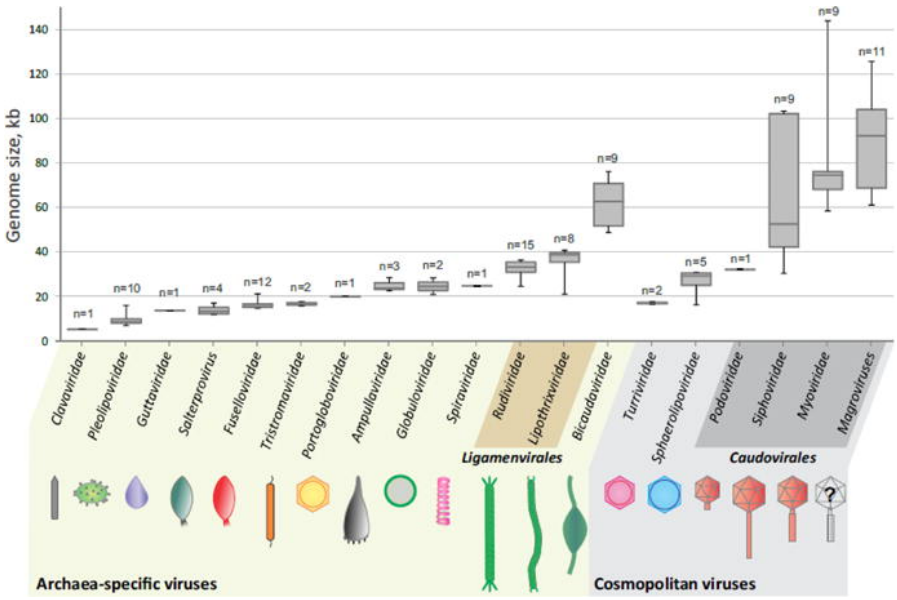


Figura 4. Morfología de los virus arqueas. Diagrama de cajas de los tamaños de genomas y diversidad en morfologías de los virones (Tomada de Krupovic et al., 2018).

CAPÍTULO 1. Metagenómica viral

La colección de todos los virus presentes en la muestra ha sido definida como viroma (Anderson et al., 2003; Zárate et al., 2019). El viroma se puede obtener mediante el metagenoma, o mediante técnicas de concentración de partículas virales en estudios recientes. Se estima que la virosfera del planeta esta compuesta a alrededor de 10^{31} (Delwart, 2007; Dion et al., 2020), no obstante, la limitante es poder aislarlos y caracterizarlos en medios de cultivo.

La metagenómica viral es una estrategia que permite estudiar la diversidad viral y describir la estructura de las comunidades virales dentro de un ecosistema particular, sin necesidad de realizar cultivos en el laboratorio.

El enfoque de metagenómica viral se puede aplicar a una amplia diversidad de muestras desde agua de mar, sedimentos cerca de la costa, heces, suero, plasma, ambientes extremos etc.; posterior a esto se realiza una concentración y extracción del material genético viral para subsecuentemente ser secuenciado y finalmente realizar el análisis bioinfomático (Delwart, 2007), del cual existen diferentes protocolos y herramientas que permiten el estudio y análisis de secuencias virales.

En ese contexto, en el presente capítulo se realiza una revisión de los principales métodos de laboratorio para la recuperación de virus a partir de diferentes muestras y se hace una descripción del estado del arte de las herramientas bioinformáticas para el análisis de secuencias virales.

1.1 ¿Qué es la metagenómica viral?

La metagenómica viral es el estudio de todos los virus presentes en una muestra ambiental, con el objetivo de determinar la estructura de la comunidad viral y sus posibles funciones dentro del ecosistema. A nivel estructural la metagenómica viral ha permitido determinar algunas de las principales familias virales dentro los ecosistemas acuáticos. Desde el punto de vista funcional, ha permitido conocer el posible intercambio de material genético entre los virus y sus hospederos a través de mecanismos de transducción de genes, y se ha demostrado la existencia de genes metabólicos auxiliares virales, los cuales son responsables de incrementar el metabolismo del hospedero, participar en la reprogramación metabólica, estar involucrados en ciclos biogeoquímicos y aumentar la producción de progenie viral

aumentando su aptitud, donde el virus ha tomado genes de su hospedador a través de la estrategia 'Saqueo y Pillaje' (Warwick-Dugdale et al., 2019).

Esta técnica ha sido útil en la vigilancia de la salud y las enfermedades. Es el caso de la identificación de virus causales de algunas enfermedades respiratorias complejas en eucariontes. Con la metagenómica viral ha sido posible identificar alrededor de 750.000 genomas de virus sin cultivar.

A diferencia de las células procariontes y eucariontes, los virus no comparten un marcador molecular universal, por lo que identificar la diversidad del material genético viral (ADN, ARN de doble cadena o de cadena sencilla) en una muestra representa un gran desafío para su estudio (Woolhouse y Adair 2013).

1.2 Viromas de ambientes acuáticos

La mayoría de los estudios sobre viromas se han realizado en el medio acuático, determinando la abundancia viral mediante microscopía con epifluorescencia y citometría de flujo, y combinado otras técnicas de viabilidad. El resultado de la abundancia viral ha variado entre 10^6 a 10^8 partículas ml^{-1} (Corinaldesi, Tangherlini, y Dell'Anno 2017).

Sin embargo, esta abundancia está sujeta a la estacionalidad y a las características fisicoquímicas del sitio de muestreo, la abundancia de los huéspedes y el ciclo de replicación en el que se encuentran los virus (lisogénico y lítico). En este tipo de ambiente, es necesario considerar el factor de estratificación causado principalmente por la luz y el oxígeno como último aceptor de electrones, que es el principal componente que diferencia la composición de las poblaciones en los ecosistemas acuáticos (Wommack y Colwell 2000).

Teniendo en cuenta tanto la abundancia y diversidad viral en estos entornos, se pueden plantear algunas hipótesis y metodologías de diseño que permitan conocer el microbioma y viroma, reduciendo en la medida de lo posible los sesgos metodológicos en la toma de muestras para obtener resultados consistentes y comparables (Nooij et al. 2018).

El desarrollo de la metagenómica, particularmente del tipo *shotgun*, permite conocer las poblaciones virales de un medio sin necesidad de tener un conocimiento previo de las secuencias que las componen. A partir de los estudios realizados, se ha

encontrado mayoritariamente a virus del orden Caudovirales, seguidos de virus de arqueas, y virus que infectan hongos y algas (Nooij et al. 2018).

Los avances y la aplicación de la virómica han impulsado enormemente el campo de la virología ambiental y se han realizado grandes esfuerzos para dilucidar la diversidad de la comunidad viral, la composición genética y las asociaciones con las poblaciones de huéspedes. Sin embargo, mucho de esto se ha hecho en los océanos y aguas dulces, y falta por explorar dicha diversidad en otros entornos.

El estudio de las comunidades virales se ha centrado en la distribución y la variabilidad de estas, tanto en el espacio como en el tiempo, con la finalidad de entender la ecología de estos organismos en diversos ambientes. Así también la ubicación geográfica, la distancia desde la costa y la profundidad eran los principales predictores de la estructura de la comunidad, y la estación, la profundidad y la concentración de oxígeno son factores importantes que dan forma a las comunidades virales en una estación oceánica específica.

Para la extracción de viromas de ambientes acuáticos, las muestras pasan por una prefiltración o centrifugación que elimina contaminantes y detritos celulares de los microorganismos que pertenecen a ese ecosistema. Esto favorece el enriquecimiento de las secuencias virales al eliminar poblaciones celulares abundantes, como las procariotas (Ramos-Barbero et al. 2019; Rice et al. 2001)

El siguiente paso es pasar la muestra por un filtro de 0,02 mm. Las muestras de volúmenes mayores a los 10 L de agua requieren de estrategias de concentración, siendo las más aceptadas la filtración de flujo tangencial y el método químico (con Cloruro férrico $FeCl_3$). Para el método de $FeCl_3$, descrito en John et al. 2011, se crean flóculos del virus con una solución de cloruro, que precipitan para luego ser recuperados en filtros de policarbonato de 1.0 µm. Luego se resuspenden los flóculos en un tampón de magnesio-EDTA-acido ascórbico, para posteriormente realizar la extracción del material genético viral con estuches especiales para material genético viral y su posterior secuenciación.

1.3 Análisis bioinformático del metagenoma viral

Existen dos tipos de metodologías para realizar los análisis bioinformáticos de un metagenoma viral. Uno de ellos, conocido como *pipeline* o flujos de trabajos, son protocolos que agrupan una serie de programas para estudiar de forma integral las

secuencias virales obtenidas de un metagenoma o por metagenómica viral. La otra metodología comprende en usar directamente programas *stand alone* para recuperar secuencias virales de un ensamble metagenómico y/o realizar la asignación taxonómica. Los programas de asignación taxonómica agrupan las secuencias virales basándose en estimaciones de similitud de secuencia (Roux et al. 2019).

Sin embargo, un aspecto importante que los usuarios de herramientas bioinformáticas deben saber es que no todos los programas sirven para buscar todos los tipos de virus. Dadas las diferencias genéticas que tienen virus, se necesita más de un programa para recuperar virus.

1.3.1 Flujos de trabajo para recuperar virus a partir de metagenoma

Los programas bioinformáticos que recuperan genomas virales están desarrollados principalmente para la detección de virus que causan enfermedades en humanos y para la búsqueda de virus en muestras ambientales, estos utilizan los datos crudos o archivos tipo fastq (secuencias de ADN con calidad asociada).

Algunos programas tipo *pipeline* o flujo de trabajo realizan más de una tarea a la vez, como eliminar secuencias duplicadas y nucleótidos indeterminados. Posteriormente, realizan un ensamblaje de las lecturas para seguir con una clasificación taxonómica según unas bases de datos como NCBI, VirusSite (Stano et al., 2016), Virus Pathogen Resource (Pickett et al., 2012) y finalmente clasifican a los virus según los criterios del Comité Internacional de Taxonomía de Virus (ICTV, *International Committee on Taxonomy of Viruses*).

No obstante, cada *pipeline* utiliza estrategias particulares para recuperar genomas de virus dentro de un metagenoma viral. Muchos están diseñados para buscar virus que causan enfermedades en humanos, por ejemplo, PathSeq (Kostic et al. 2011), CaPSID (Borozan et al. 2012), “Integrated Metagenomic Sequence Analysis” (Dimon et al. 2013), VirusHunter (Zhao et al. 2013) , “Virus Identification Pipeline” (VIP) (Zhao et al. 2013), VirusSeeker (Zhao el al 2017) , ViromeScan (Rampelli et al. 2016), "Viral Genome-Targeted Assembly Pipeline" (VirusTAP) (Yamashita, Sekizuka, and Kuroda 2016), virMine (Garretto, Hatzopoulos, and Putonti 2019), DisCVR (Maabar et al. 2019),Virmet (https://virmet.readthedocs.io/en/latest/) y VirAnnotOTU (Lefebvre et al. 2019). Estos programas usualmente mapean las

lecturas crudas contra el genoma humano, y después las lecturas filtradas son anotadas taxonómicamente.

Para la búsqueda de virus ambientales los *pipelines* más utilizados son VirAmp (Wan et al. 2015), Holovir (Laffy et al. 2016), y FastViromeExplorer (Tithi et al. 2018). Mientras, que otros programas son exclusivos para virus de RNA como "Short RNA subtraction and assembly" (SRSA), VirFind (Ho y Tzanetakis 2014), Kodoja (Baizan-Edge et al. 2019) y VirusDetect (Zheng et al. 2017). Estos últimos manejan diferentes enfoques como un análisis de k-meros y comparando contra una base de datos o ensamble de small RNA.

1.3.2 Recuperación de virus a partir del ensamble

Generalmente, los ensambladores se han diseñado para recuperar lecturas específicamente de metagenomas celulares, no para secuencias virales. Los ensambladores de lecturas cortas de metagenómica suelen utilizar el algoritmo de Bruijn basado en grafos, los cuales dividen las secuencias en k-meros. Un k-mer es una secuencia de nucleótidos de longitud específica, por ejemplo, un dinucleótido es un k-mer donde k=2. Ejemplos de estos programas son SPAdes (Bankevich et al. 2012), metaSPAdes (Nurk 2017), IDBA-UD (Peng et al. 2012), Megahit (D. Li et al. 2015), MetaVelvet (Namiki et al. 2012), y MetaVelvet-SL (Afiahayati, Sato, and Sakakibara 2015).

Estos funcionan con cadenas de una longitud particular (k-meros), que forman un grafo, y cada nodo del grafo representa una superposición entre dos lecturas a través de los k-meros. En otras palabras, las secuencias se descomponen en k-meros superpuestos, formando grafos donde los nodos representan los primeros k-1 nucleótidos de una lectura que se superponen con los k-1 nucleótidos de otra lectura (Compeau, Pevzner, y Tesler 2011)

Para determinar con cuál de los ensambladores se podría recuperar el mayor número de virus, Roux et al. 2017 hicieron una comparación de los diferentes ensambladores. Determinaron que Megahit fue el mejor para evitar la construcción de contigs quiméricos. metaSPAdes fue el mejor para ensamblar genomas de baja cobertura dentro de un solo contig y muestra menos falsos positivos al recuperar contigs circulares o genomas completos de virus.

Actualmente se ha desarrollado MetaviralSPADES (Antipov et al. 2020) enfocado directamente a genomas virales, permitiendo ensamblar genomas virales dentro de un metagenoma. Este ensamblador se centra en el ensamblaje viral a nivel de especie y, teniendo en cuenta la diversidad viral, dado que los genomas son variables elimina el procedimiento formación de burbujas (estructuras en la formación del grafo tipo burbuja formadas para la aparición de variantes y errores en la secuenciación). También dado que muchos virus son circulares, conserva grafos circulares de alta cobertura (mínima 5x) (Figura 5). De igual forma los grafos largos de alta cobertura representan supuestas secuencias virales lineales y tienen secuencias invertidos repetidos en los extremos que forman subgrafos pequeños. Por lo tanto, el programa identifica estas regiones dentro de los genomas para identificarlos como virus.

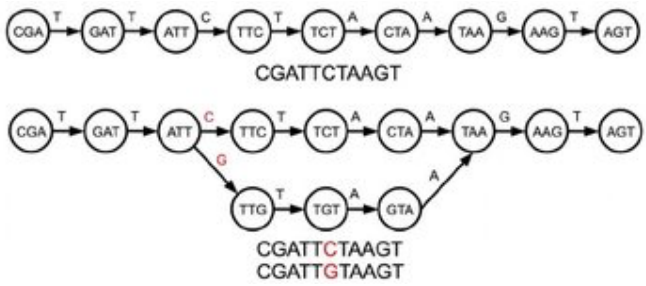


Figura 5. Formación de burbujas durante el ensamble. MetaViralSpades, conserva las burbujas pues los virus son muy variables.

Otro método utilizado para obtener secuencias virales a partir de lecturas crudas es el reclutamiento de fragmentos (Fragment Recruitment), el cual consiste en mapear lecturas metagenómicas contra una base de datos, que pueden ser secuencias de referencia virales o un conjunto de genes de interés. Las lecturas se mapean contra genomas de referencia utilizando los algoritmos Burrows-Wheeler Transform y Tablas Hash. Actualmente, existen algunas herramientas para realizar el reclutamiento de fragmentos como el FR-HIT, algunas especializadas en secuencias virales son FRAP and VR-MG-FRAP (Cobían Güemes 2019), y el programa Tanoti que permite mapear lecturas a través de una referencia guiada usando BLAST (https://bioinformatics.cvr.ac.uk/software/tanoti/).

1.3.2.1 Métodos basados en homología de genes

Estos métodos buscan virus dentro de los contigs de un metagenoma, mediante el alineamiento de las secuencias con las bases de datos de genomas virales. Algunas de las herramientas más populares se centran en la búsqueda de virus ambientales como VirSorter que busca los virus con base a características particulares de virus para poder identificarlos, como proteínas conservadas en los virus. A estas secuencias se les denomina *hallmarks* virales, los cuales pueden ser las cápsides y las polimerasas (Roux et al., 2016). Otros ejemplos de herramientas diseñadas para buscar virus ambientales son Phast (Zhou et al., 2011) y Phaster (Arndt et al., 2016), que comparan las secuencias de consulta con una base de datos de genes virales.

Una vez recuperados los contigs virales estos se pueden anotar mediante otros programas como, por ejemplo, Viral Genome Organizer (Upton et al., 2000), Viral MetaGenome Annotation Pipeline (VMGAP) (Lorenzi et al., 2011), Viral Genome Anotation (Vgas) (Zhang et al., 2019), PHANOTATE (McNair et al., 2019), Viral Genome ORF Reader (S. Wang, Sundaram, y Spiro 2010).

1.3.2.2 Arquitecturas de aprendizaje automático

Los algoritmos de aprendizaje automático, conocidos como *deep learning*, permiten identificar patrones o características dentro de las secuencias y luego hacer predicciones a través de enfoques basados en k-mers. Con el tiempo algunos de estos algoritmos han mejorado la búsqueda, usando por ejemplo las redes neuronales artificiales para conocer las características virales (Ponsero and Hurwitz 2019).

Uno de los programas más populares que utiliza este tipo de enfoque es VirFinder (Ren et al. 2017), que identifica firmas virales en contigs a través del aprendizaje automático, prediciendo un mayor número de genomas virales en comparación con VirSorter. Sin embargo, una limitación de este programa es que necesita un conjunto grande de viromas y genomas virales para tener suficientes datos para realizar el entrenamiento del algoritmo y encontrar nuevos virus.

Recientemente, se han desarrollado nuevas herramientas con diferentes algoritmos de clasificación basados en el aprendizaje automático, como por ejemplo PPR-Meta (Fang et al. 2019), el cual es una arquitectura de red neuronal capaz de identificar plásmidos y fagos en el ensamble. Otro es VIBRANT (Kieft, Zhou, and

Anantharaman 2020) que utiliza redes neuronales que se entrenan con un conjunto de una familia de proteínas virales, que es capaz de identificar virus dsDNA, ssDNA y de RNA.

También se han desarrollado herramientas integrales como What the Phage (Marquet, et al., 2020), el cual integra diferentes herramientas de identificación en paralelo para detectar y anotar fagos. Esta herramienta utiliza un total de nueve herramientas diferentes en paralelo, VirSorter v1.0.6 83 (Roux et al. 2015) Sourmash v2.0.1 (Pierce et al. 2019), Metaphinder with no release 84 version (Jurtz et al. 2016), VirFinder v1.1 (Ren et al. 2017), MARVEL v0.2 (Amgarten et al. 2018), VirNet v0.1 (Abdelkareem et al. 2018), PPR-Meta v1.1 (Fang et al. 2019), Vibrant v1.2.1 (Kieft, Zhou, and Anantharaman 2020), y DeepVirFinder v1.0 (Ren et al. 2020).

Posteriormente, para evaluar la integridad de los genomas virales recuperados, se ha desarrollado recientemente CheckV (Nayfach et al. 2020). Este programa primero elimina la contaminación del huésped identificando genes no virales utilizando modelos ocultos de Markov (HMM en inglés). Luego estima la integridad del genoma a través de la identidad de los aminoácidos promedio comparados con una base de datos. Determina la longitud del nuevo genoma basándose en los resultados arrojados por la base de datos. Finalmente evalúa las regiones de repeticiones terminales directas, las repeticiones terminales invertidas y los sitios de integración de provirus (genoma insertado) y se infieren en las secuencias para evaluar el genoma circular.

1.3.4 Herramientas de clasificación taxonómica

Para la clasificación taxonómica de los virus se han desarrollado diversas herramientas, aunque no ha sido una tarea fácil por la carencia de genes marcadores (Nooij et al. 2018). Una forma de clasificarlos es mediante la reconstrucción de una filogenia, y así poder inferir su origen. Recientemente se ha propuesto una nueva estrategia de clasificación agrupando en cuatro reinos mega taxonómicos, Riboviria, Monodnaviria, Viridnaviria y Duplodnaviria (Koonin et al. 2020). Algunas herramientas para la taxonomía de virus son Phage Proteomic VICTOR (Meier-Kolthoff y Göker 2017), ViPTree (Nishimura et al.2017), VicTree (Modha et al.2018), GRAViTy (Aiewsakun y Simmonds 2018). Los programas más utilizados para la clasificación taxonómica son MEGAN (Huson et al. 2007) y vConTACT (Bolduc et al. 2017). El

primero utiliza el algoritmo de asignación de ancestro común más bajo (LCA) para la clasificación de las secuencias. El segundo programa emplea los enfoques de agrupación en red, basados en el contenido de proteínas compartidas entre fagos.

JUSTIFICACIÓN

Los ecosistemas acuáticos como los océanos y aguas dulces presentan una amplia diversidad de virus, los cuales están implicados en la regulación de procesos biogeoquímicos, diversidad bacteriana, abundancia de nutrientes entre otros aspectos, y en algunos casos proporcionando genes que les confieren ventajas adaptativas a sus hospederos. Sin embargo, aún se desconoce el papel de los virus en ambientes acuáticos extremos.

Actualmente la metagenómica, ha permitido conocer la diversidad microbiana en diversos ambientes, así como la interacción de organismos en las comunidades, la abundancia de determinados genes en una condición especial; lo que la hace una herramienta de aplicación en diversos campos de estudio. Por lo tanto, el estudio del metagenoma de ecosistemas acuáticos de ambientes extremos, permitirá conocer la estructura del viroma de estos ambientes, y el rol funcional de los virus en ambientes extremos, así como poder determinar como los virus pueden influir a la adaptación en sus hospederos en estos ecosistemas.

Objetivo General

Analizar y caracterizar la diversidad genómica viral de muestras de aguas de tres ambientes extremos de las ventilas hidrotermales, aguas termales de Chignahuapan y sistemas hipersalinas de Cusco mediante metagenómica.

Objetivos particulares

- Determinar la diversidad de microorganismos presentes en el metagenoma de agua de ambientes extremos.
- Determinar la estructura de las comunidades virales dentro de los ambientes extremos acuáticos.
- Obtener genomas bacterianos y virales a partir del ensamble de secuencias obtenidas del metagenoma.
- Generar predicciones funcionales de las secuencias virales presentes en los metagenoma.
- Analizar la presión de selección que ha modelado la evolución de genes virales auxilares del metabolismo.

Capítulo 2. Revisión de la metagenómica viral en ambientes extremos

Los virus son las entidades biológicas más abundantes y diversas sobre el planeta (Dion et al., 2020). En recientes años se han descrito las comunidades de virus que habitan en ambientes extremos las cuales parecen ser genéticamente distintas a los virus de otros ambientes, lo que sugiere una adaptación a nichos específicos.

La observación de la diversidad morfológica que presentan los virus en los ambientes extremos ha sido posible mediante la microscopía electrónica, mientras que la diversidad genética en los genomas de los virus se ha estudiado a través de la metagenómica viral (Le Romance et al., 2007).

La metagenómica viral presenta varios retos tanto metodológicos como bioinformáticos, aunado a esto aún existe poca información acerca de la ecología de los virus en ambientes extremos; estudios iniciales de la diversidad viral estuvieron enfocados en heces humanos, mar abierto y costero, lagos de agua dulce, y finalmente estanques geotérmicos, hipersalinos e hipertermófilos, porque los protocolos iniciales para la extracción de partículas virales se enfocaron en muestras acuáticas. Sin embargo, la metagenómica ha permitido comenzar a describir la estructura de las comunidades virales en ambientes extremos (Roux et al., 2019).

Sobre la estructura de las comunidades de virales en ambientes extremos, se conoce relativamente poco, en comparación con las bacterias, por lo tanto, es importante realizar una descripción precisa del viroma, conocer la ecología viral y potencial biotecnológico que tienen los virus en los ambientes extremos. Es por lo que en este capítulo realizamos una revisión de los virus en los diferentes ambientes extremos desde los terrestres a los acuáticos, donde se observa que hay virus que se limitan a ambientes determinados, sugiriendo adaptaciones específicas en tiempo y espacio, por último, se describe el potencial biotecnológico de los virus.

Brevemente en este capitulo se llevó a cabo un análisis comparativo de la estructura de la población de viromas en entornos extremos, utilizando 17 metagenomas de virus de acceso público depositadas en MetaVir2. Las comunidades virales de diferentes ambientes extremos mostraron comunidades similares de virus, siendo la más abundante el orden Caudovirales, además de ser ubicuos. Las familias más abundantes fueron Siphoviridae, Myoviridae y Podoviridae, seguidas de Circoviridae, Phycodnaviridae, Microviridae e Inoviridae. Adicionalmente, se

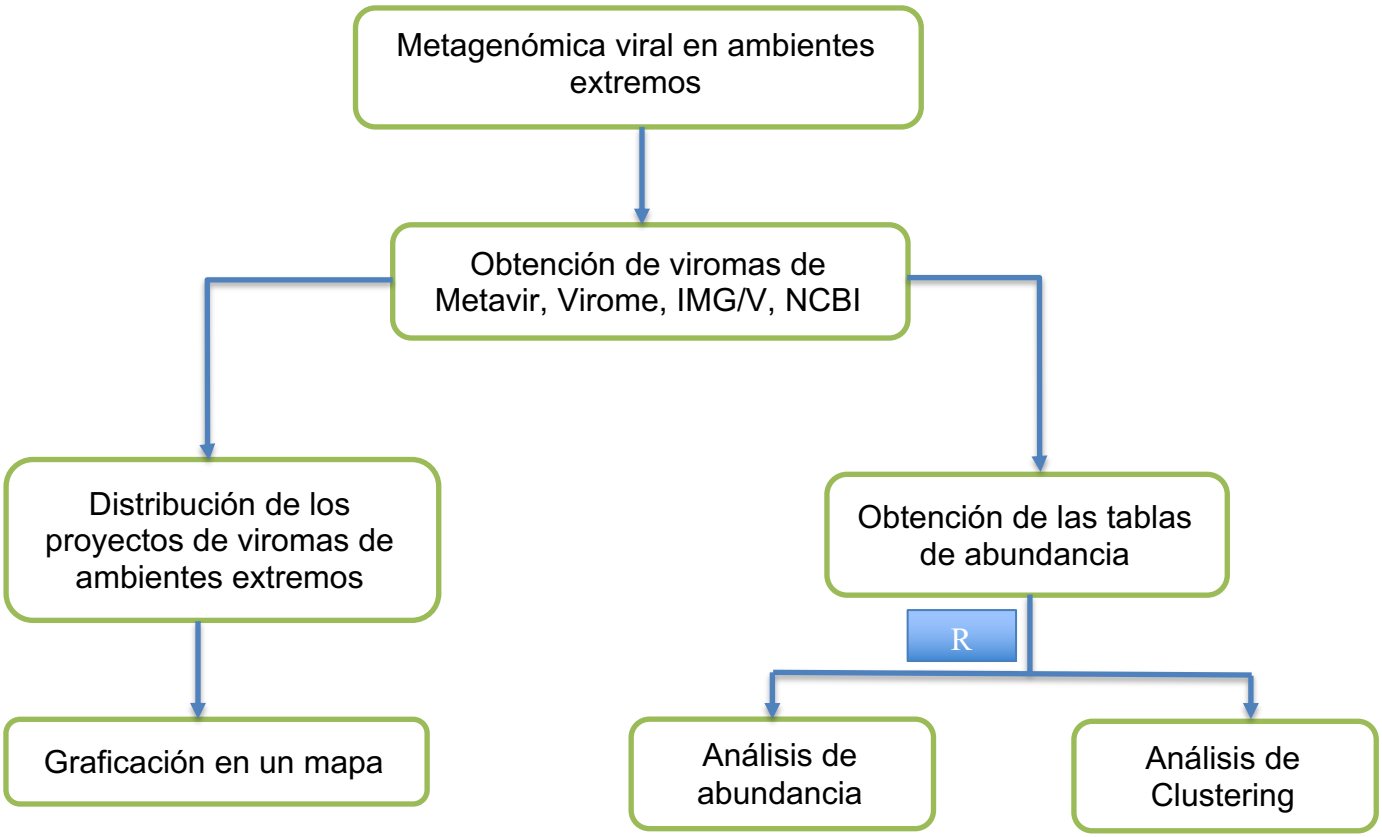
identificaron fracciones muy elevadas de virus ssDNA, en ambientes psicrófilos y sedimentos.

Mediante un análisis de clustering jerárquico de las más abundantes familias virales se pudo concluir que algunos ambientes extremos, como los hipersalinos, marinos profundos e hiperáridos, tienen grupos de virus partículares que indican similitudes en las comunidades virales presentes en estos ambientes.

Además, la prospección funcional de virus en nichos ecológicos extremos se ha limitado casi exclusivamente al cribado basado en secuencias hasta la fecha. Aunque algunas secuencias virales se han anotado y asignado a funciones específicas, muy pocas proteínas virales descubiertas usando enfoques metagenómicos se han clonado posteriormente, expresado heterólogamente y caracterizado bioquímicamente. Otros métodos basados en funciones, como los cribados basados en la actividad y los cribados basados en la PCR o la hibridación, están infra explotados actualmente como enfoques para identificar proteínas virales de viromas extremófilos.

El conocimiento actual sobre viromas asociados a ecosistemas extremos es bastante limitado, todavía no es posible apreciar plenamente el gran potencial biotecnológico que pueden representar. Por lo tanto, se deben realizar más esfuerzos para analizar metagenomas virales extremófilos en busca de nuevas proteínas y biomoléculas para lograr tener una mayor comprensión de su impacto biológico.
Los resultados de este capítulo fueron publicados en el artículo del anexo 1.

Diagrama metodológico capítulo 2



Programas utilizados

Capítulo 3 Análisis del viroma de ventilas hidrotermales profundas

Los primeros protocolos para la realización de metagenómica viral fueron diseñados para recuperar virus a partir de ambientes acuáticos, como las ventilas hidrotermales.

Las ventilas hidrotermales, chimeneas o géiseres, expulsan agua y alcanzan temperaturas tan altas como 400ºC. Se caracterizan por poseer abundantes organismos procariontes quimiosintéticos, los cuales tienen la capacidad de convertir el dióxido de carbono, agua y nitratos en sustancias orgánicas esenciales (Ortmann, et al., 2005).

Dávila et al. 2015, analizaron la diversidad microbiana presente en dos regiones hidrotermales de México, encontrando diferentes consorcios bacterianos en cada uno de ellos. La mayoría de filogrupos bacterianos encontrados fueron Actinobacteria, Bacteroidetes, Firmicutes, Spirochaetes, Thermotogae, Cyanobacteria, Chloroflexi, Aquificacae, y Proteobacteria (Alphaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, y Epsilonproteobacteria), sin embargo, la presencia o ausencia de estas familias bacterianas depende de las condiciones de reducción-oxidación de las ventilas hidrotermales analizadas (Dávila et al., 2015). Por lo tanto, se ha demostrado la presencia de diferentes linajes bacterianos en el ventilas hidrotermales, de manera interesante hay una alta abundancia de virus que regulan la morbilidad de las bacterias. Así mismo, la temperatura que se encuentra en las ventilas parece que regula la distribución de procariontes y virus (Ortmann, et al., 2005; Breitbart, M 2011).

Las primeras aproximaciones para el análisis de virus presentes dentro de las ventilas hidrotermales fue mediante métodos tradicionales, donde 6 virus de ventilas hidrotermales fueron descubiertos: el bacteriófagos “BVW1” sin clasificar dentro de una familia, los virus GVE1, GVE2, NRS-1, MPV-1 (Siphoviridae), D6E (Myoviridae) y dos virus que infectan archeas PAV1 y TPV1 (Fuselloviridae) (Lossouarn et al., 2015).

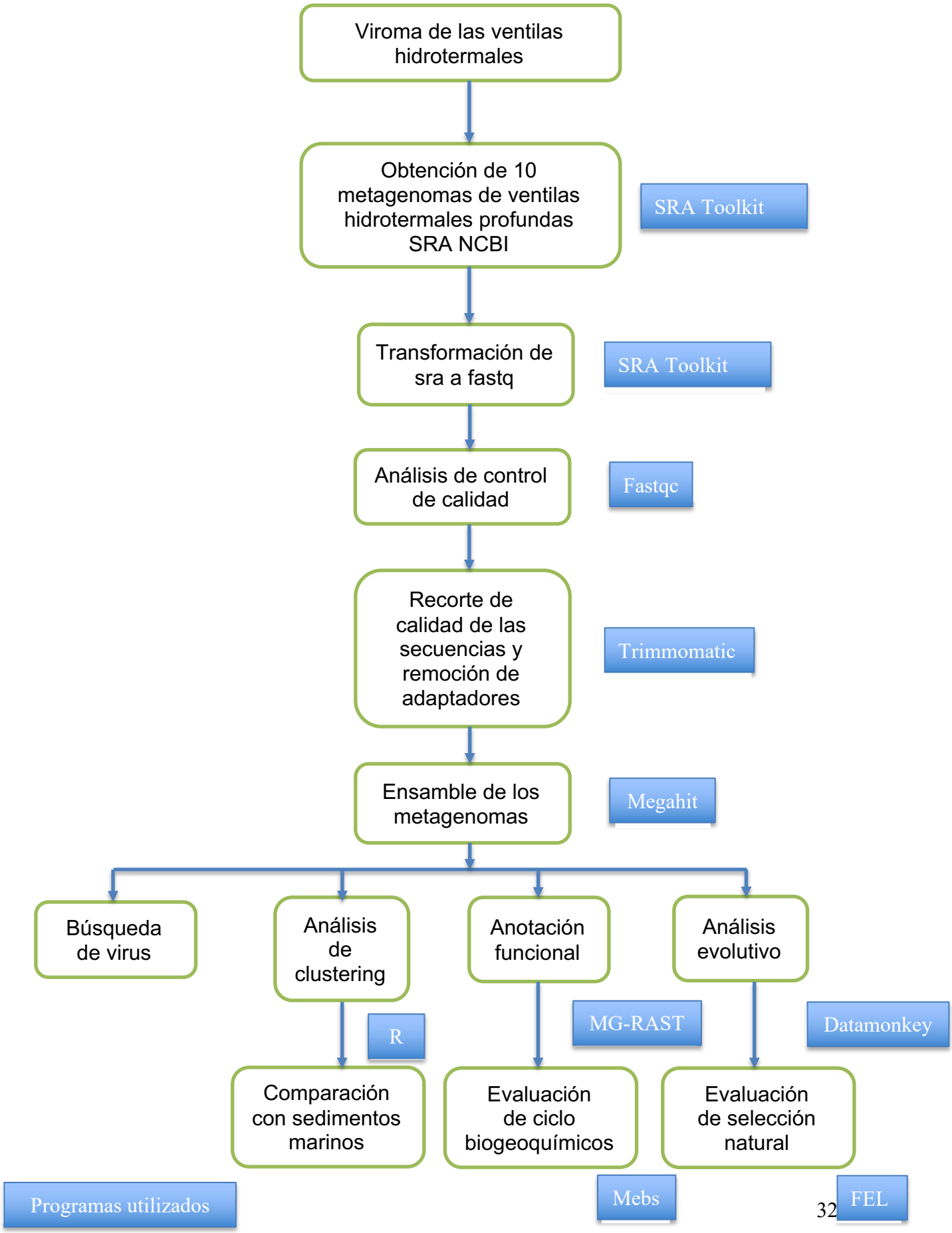
Sin embargo, el conocimiento acerca de los virus en las aguas profundas es significativamente escaso, reportándose en general una abundancia de entre un 0.6 a 60 × 10^8 virus por litro (Breitbart et al., 2007). Además, se sabe muy poco sobre la diversidad de virus en las ventilas hidrotermales; pero algunos reportes de fluidos difusos en ventilas han reportado distintas familias virales (Anderson et al., 2011).

En este capítulo se realizó un análisis de la estructura y función de las comunidades virales dentro de las ventilas hidrotermales, donde anteriormente no se ha descrito la estructural viroma. Adicionalmente los virus desempeñan papeles cruciales en la transferencia de genes, provocando la alteración del genoma del hospedero, así como en los ciclos biogeoquímicos y en general en el ecosistema acuático, por lo que se consideró que una metagenómica comparativa permitiría dilucidar el papel que juegan los virus dentro de las ventilas hidrotermales.

Durante este análisis se determinó que los virus dentro del orden de los Caudovirales se hallan en sedimentos y plumas de las ventilas hidrotermales. Sin embargo, la comunidad viral está estratificada. En las plumas de las ventilas hidrotermales se encontraron virus de dsDNA, mientras que en los sedimentos solo había virus de ssDNA como los pertenecientes a las familias Microviridae, e Inoviridae. Por otro lado, análisis funcional reveló la presencia de proteínas estructurales de los virus y de genes auxiliarles del metabolismo, como la cititidiltransferasa y ribonucleotido reductasa, los cuales se analizaron desde una perspectiva evolutiva. Dicho análisis sugirió que estos genes auxiliares del metabolismo están altamente conservados, evolucionan bajo selección purificadora y se mantiene en el genoma.

Los resultados de este capítulo se presentan en el anexo 2.

Diagrama metodológico capítulo 3



Capítulo 4. Metagenoma de sistemas salinos en los Andes Peruanos

Los ecosistemas hipersalinos se encuentran distribuidos en todo el mundo y representan una fuente natural de sal. Estos ecosistemas se caracterizan por la alcalinidad y la baja concentración de oxígeno (Grant 2004; Demergasso, et al., 2004; Ventosa et al., 2015; Pontefract et al., 2017; Naghoni et al. 2017).

Los ambientes acuáticos hipersalinos se clasifican en dos categorías principales: las *talasohalinas*, que se generan de la evaporación de agua de mar, tienen pH neutro o ligeramente alcalino y poseen una salinidad que excede la del agua de mar por un factor de cinco a 10 veces; y las *atalasohalinas*, que no se derivan de agua de mar, con altas concentraciones de iones como Mg^{2+} o Ca^{2+} y un pH ácido ligero (Grant 2004; Demergasso, et al., 2004; Daffonchio et al., 2006; Ventosa et al., 2015; Pontefract et al., 2017; Naghoni et al., 2017).

Estos sistemas representan excelentes modelos para el estudio de la ecología y la diversidad de los microorganismos halófilos (Fernández et al., 2014). Los halófilos se han clasificado según la concentración de sales donde habitan, en halófilos débiles (1-3%), halófilos moderados (3- 15% de NaCl) y halófilos extremos (más del 15% de NaCl) (Quillaguaman et al., 2006). En cuanto a los ambientes hipersalinos, se han clasificado de acuerdo con el porcentaje de sal, en hábitats marinos (3,5-3.8% de sales totales); de baja salinidad (menos del 10% de NaCl), salinidad intermedia (10-20% de NaCl) (Oren, 2008) y alta salinidad (superior a 20% de NaCl) (Ventosa A., 2014).

Con respecto a la comunidad microbiana que habita en estos ecosistemas, se ha reportado un gran diversidad de microorganismos, en particular de la familia Halobacteriaceae dentro del dominio de las Arqueas; en el caso de bacterias los géneros Haloanaerobiales, Halorhodospira, Salinibacter, Halomonas, Chromohalobacter y Salicola son abundantes; mientras que organismos eucariontes como *Artemia salina*, *Colpodella edax* y *Dunaliella salina*, se han identificado en bajas proporciones (Ventosa et al., 2015; Haferburg et al., 2017, Emerson et al., 2012; Sarwar et al., 2015).

Se ha descrito una alta diversidad de halovirus, en concentraciones $\geq 1x10^7$ por mL en agua de mar, de los cuales, y en un bajo porcentaje son cultivables (Emerson

et al., 2012). Los ambientes hipersalinos en América del Sur han sido poco estudiados, entre ellos los ubicados en la región de los Andes.

Entre estos se encuentra el sistema Hipersalino de Acos y las salmueras de Maras, dos ambientes atalasohalinos situados en el suroeste de Perú, dentro del departamento de Cusco. Acos se encuentra en el distrito de Acomayo a una altitud de 2.852 m y se compone de corrientes de agua salinas. Y Maras, ubicada en el distrito de Urubamba a 3.380 m.s.n.m, está compuesta de 3.000 pequeños estanques poco profundos que forman terrazas en toda la ladera de la montaña *Qaqawiñay* (palabra quechua que significa "roca eterna") (Maturrano et al., 2006).

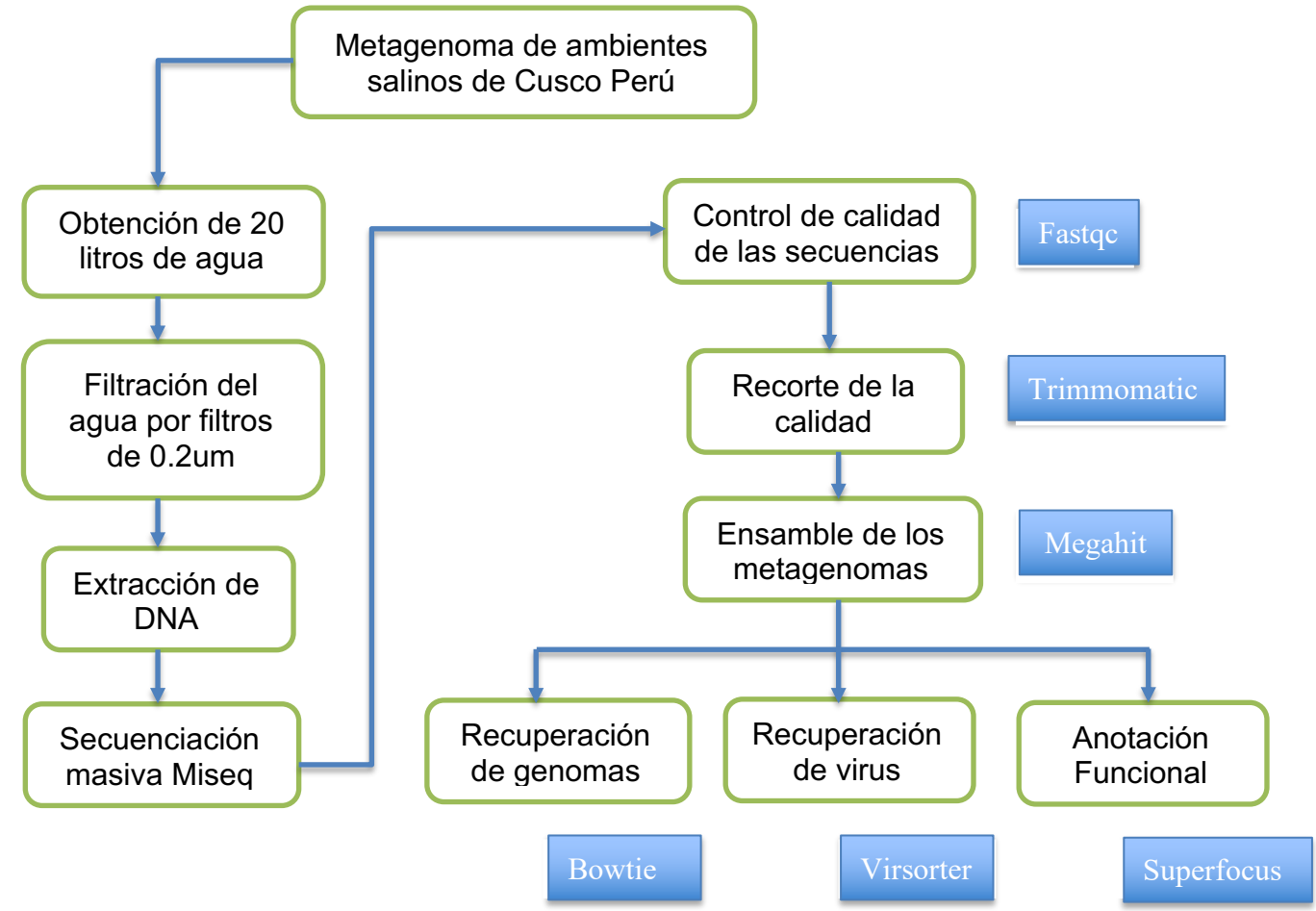
En este capítulo se determinó la diversidad de microorganismos halófilos y la diversidad funcional en estos dos ambientes de atalasohalinos, que presentan diferencias en salinidad y pH. A partir de este análisis metagenómico se evidenció una alta diversidad bacteriana en contraste con las arqueas; ambos metagenomas muestran predominancia de genes involucrados en el metabolismo de los carbohidratos y de factores de estrés, es decir, que los microorganismos en estos metagenomas tienen posiblemente una elevada expresión de genes involucrados con estas funciones para la supervivencia en este ambiente, y genes relacionados con la osmorregulación para evitar la desecación. De allí que se observara una baja abundancia de genes que codifican para pigmentos.

Por otro lado, se obtuvo el borrador de los genomas de las bacterias halófilas de las especies *Halomonas elongata* y *Idiomarina loihiensis*. Así también se obtuvieron dos genomas completos del bacteriófago *Halomonas elongata-like*, y de 27 genomas de virus nuevos que se pudieron agrupar dentro del orden de los Caudovirales.

En general, la diversidad en las comunidades dentro de estos ambientes hipersalinos moderados fue principalmente bacteriana, con la identificación de algunos virus que pueden estar relacionados con ellas, mientras que las arqueas presentaron una diversidad y abundancia menores.

Finalmente, el perfil funcional de los genes que presentan las comunidades de microorganismos en estos ambientes hipersalinos, mostró una gran abundancia de secuencias asociadas con la desintoxicación, reparación del ADN, formación de cápsulas, paredes celulares y metabolismo de nucleótidos; todas estas pertenecientes principalmente al dominio bacteria. Los resultados de este capítulo se presentan en el anexo 3.

Diagrama metodológico capítulo 3



Programas utilizados

CAPÍTULO 5 Metagenoma de aguas termales circunneutrales de Chignahuapan, Puebla, México.

En México hay diferentes regiones de aguas termales, respiraderos de vapor, solfataras, suelos calentados geotérmicamente, lagunas de lodo hirviendo, zonas geotérmicas y volcanes activos, ubicados en el cinturón volcánico transmexicano (TMVB) del centro de México; sin embargo, la diversidad de microorganismos y el potencial funcional con interés biotecnológico que se desarrolla en estos sitios han sido poco estudiados en el país.

Las aguas termales se distribuyen por todo el mundo y se han realizado algunos estudios sobre la diversidad en estos ambientes. El más estudiado es el Parque Nacional de Yellowstone en EUA, con 10.000 sitios termales, y en este se ha encontrado que los ambientes de manantiales terrestres tienen una variedad de procariotas novedosos a diferencia de los respiraderos hidrotermales en el océano profundo. Los ambientes de aguas termales se pueden clasificar de acuerdo con la temperatura en los de baja temperatura (<55ºC) y alta temperatura (> 55ºC), y de acuerdo con el pH de los manantiales ácidos (<4), de pH intermedio (~4), manantiales circunneutral o neutros (~7) y resortes alcalinos (>7) (Inskeep et al. 2013; Mardanov et al., 2017).

Además, las aguas termales pueden clasificarse según su origen en *magmáticas*, aquellas que nacen en áreas volcánicas y sus temperaturas son altas y en promedio superiores a 50°C; y en *telúricas*, la cuales pueden surgir de cualquier suelo. En cuanto a los organismos termófilos, éstos se pueden clasificar en tres categorías definidas por Wagner y Wiegel, en termófilos moderados (crecimiento óptimo, 50–60ºC), termófilos extremos (crecimiento óptimo, 60–80°C) e hipertermófilos (crecimiento óptimo, 80–110ºC) (De Castro et al., 2016).

Los factores determinantes de las comunidades en las aguas termales son la temperatura y el pH. Se ha observado que a pH bajo son más abundantes las arqueas que las bacterias.

En el caso de México, el área de aguas termales más explorada es “Los Azufres”, en el estado de Michoacán en México, que es una fuente termal ácida (pH 3.6 y 65 °C) y comprende una serie de manantiales hidrotermales, fumarolas y lodo

hirviendo, con altas concentraciones de metales. De este lugar se ha determinado la diversidad de bacterias y se demostró que las más abundantes fueron de los géneros *Rhodobacter*, *Acidithiobacillus* y *Lyzobacter* (Brito et al., 2014). También se han recuperado mediante metagenómica algunos virus y arqueas, por ejemplo, *Archaeal Fusellovirus*, *Archaeal Rudivirus* y *Sulfolobales Archaeon AZ1*.

Se han explorado otras aguas termales en México en el estado de Veracruz en la zona termal de Carrizal y en las aguas termales "Los Baños", que se caracterizan por ser manantiales circunneutrales o neutrales con baja temperatura (50ºC); de los que se ha logrado el aislamiento de bacterias. En este lugar, las bacterias más abundantes fueron de los géneros *Geobacillus*, *Anoxybacillus* y *Aeribacillus,* y dichos estudios solo se limitan a organismos cultivables (Pinzón et al., 2010).

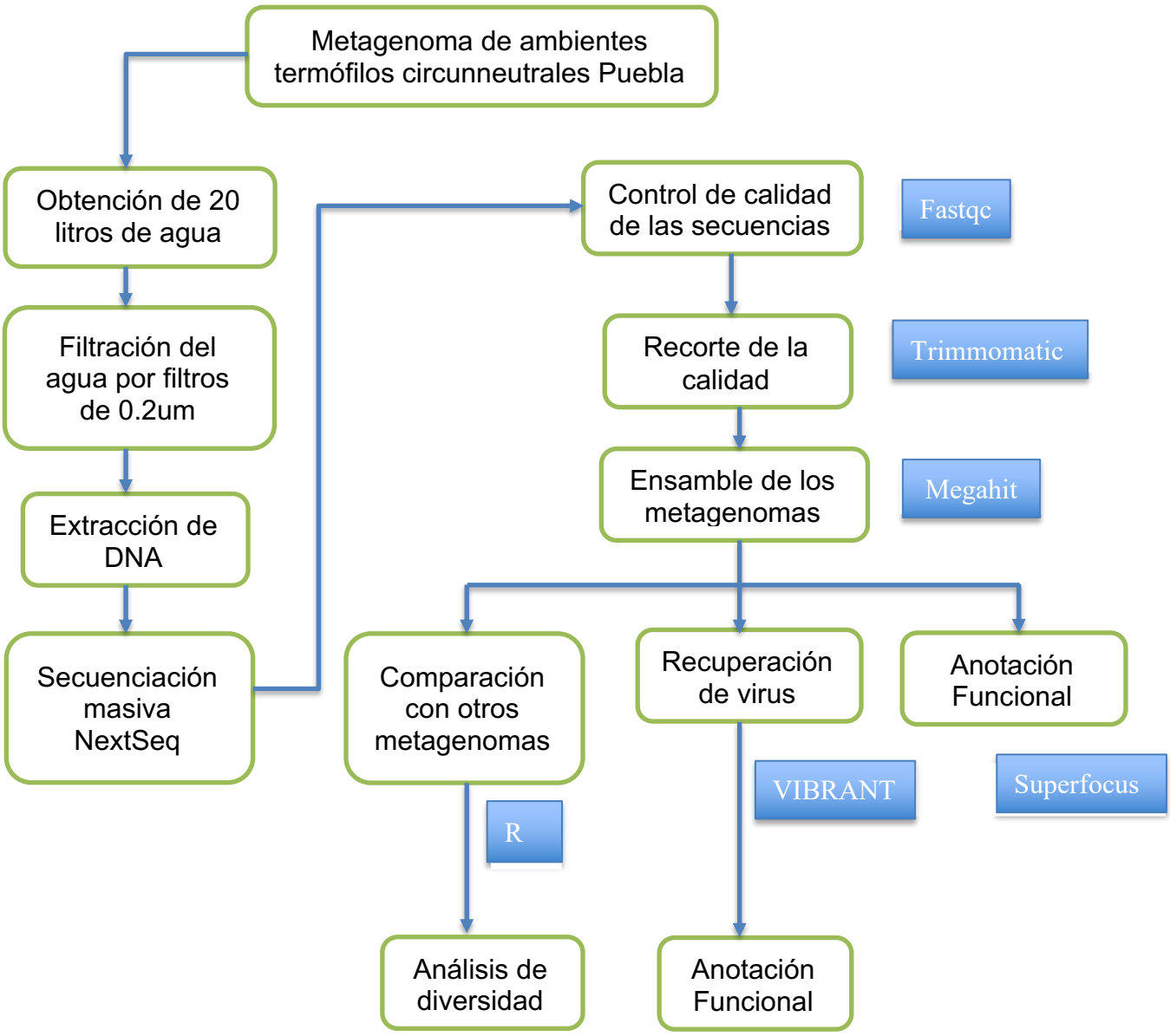
Otras regiones geotérmicas en el país se han explorado, y en muy pocas se ha descrito la diversidad de microorganismos a través de la metagenómica. Nosotros evaluamos la diversidad de microorganismos en una fuente termal circunneutral en Chignahuapan en el Estado de Puebla, con el objetivo de determinar la composición taxonómica y el potencial funcional en estas condiciones.

Chignahuapan es un poblado que pertenece a la región geotérmica Tulancingo-Acoculco, que cuenta con un diámetro de 18 km y está distribuida entre los estados de Hidalgo y Puebla. Chignahuapan está ubicado en la Sierra Norte del Estado de Puebla a 2.260 m.s.n.m. Cuenta con la presencia de aguas termales magmáticas, que se remontan a 1.7-0.9 millones de años aproximadamente. Allí se han encontrado indicios de explosiones hidrotermales recientes que sugieren un origen magmático activo con composiciones de dacitas y riolitas (López-Hernández et al., 2009).

En este capítulo se realizó un análisis metagenómico del ambiente hidrotermal de Chignahuapan, Puebla, para determinar la estructura de la comunidad bacteriana y viral. Siendo este un ambiente termal circunneutral, se observó que los microorganismos más abundantes correspondieron a Bacterias dentro de los géneros *Rhodococcus*, *Thermomonas*, *Thiomonas*, *Acinetobacter*, *Sulfurovum*, y *Bacillus*, siendo todas consideradas oxidantes de azufre. Mientras que la estructura de las comunidades virales mostró un predominio de las familias virales dentro Herelleviridae, Siphoviridae, Myoviridae, y Podoviridae, los cuales infectan a bacterias, sin embargo, se encontró que existen virus ubicuos como el *Acidithiobacillus phage AcaML1*.

Finalmente, desde el punto de vista funcional, se encontraron diversos genes auxiliares del metabolismo como cofactores vitaminas, metabolismo de carbohidratos y relacionados con el ciclo del azufre, donde bacterias y virus contribuyen en este ciclo biogeoquímico. Los resultados de este capítulo se presentan en el anexo 4.

Diagrama metodológico capítulo 5



DISCUSIÓN GENERAL

La descripción de la virosfera en nuestro planeta ha sido posible gracias el desarrollo de técnicas moleculares como la metagenómica. A través de ella ha aumentado la comprensión de la ecología microbiana dentro un determinado ambiente, además de responder las preguntas: “¿quiénes están allí?”, “¿qué están haciendo? y “¿qué funciones realizan?”. En este contexto, la metagenómica viral ha podido dilucidar la estructura de las comunidades virales en diferentes ambientes, aunque se conoce relativamente poco de los ambientes extremos.

Por lo tanto, este proyecto se llevó acabo con la finalidad de determinar la estructura y diversidad de las comunidades virales, así como de analizar la dinámica funcional de los virus dentro de ambientes extremos acuáticos, y finalmente evaluar los procesos evolutivos que actúan sobre genes virales como los genes auxiliares del metabolismo (*Auxiliary metabolic genes*, AMG).

Lo primero que evaluamos fue la composición y diversidad de comunidades virales en diferentes ambientes extremos acuáticos hasta los terrestres. Para realizar dicha evaluación, se analizaron los metagenomas y viromas disponibles en las bases de datos como Metavir y Virome.

De forma general, se encontró que la estructura de las comunidades virales depende del ambiente extremo. Cada uno de ellos puede contener virus particulares, sugiriendo adaptaciones específicas de estas entidades biológicas. No obstante, existen virus ubicuos en todos los ambientes extremos como los que pertenecen a las familias Siphoviridae, Myovridae y Podoviridae, dentro del orden de los Caudovirales. Esto es de esperarse puesto que los virus que pertenecen a estas familias infectan a más de 140 géneros bacterianos, y podrían tener mecanismos de adaptación a diferentes ambientes (Dávila et al., 2019).

Dentro de las 10 familias virales más abundantes encontradas están, Phycodnaviridae, Mimiviridae, Circoviridae, Microviridae, e Inoviridae, además de virus no clasificados en una familia. Pero la estructura de las comunidades virales no fue similar en todos los ambientes extremos, a pesar de que estas familias fueron las más abundantes.

Ejemplo, de esto son las claras diferencias en los metagenomas de aguas profundas y sedimentos marinos del Atlántico, Ártico y Pacífico Noroeste, donde los virus ssDNA están en elevada abundancia en comparación con los virus de dsDNA.

Las familias predominantes fueron Microviridae que infecta a bacterias, Circoviridae que infecta a los animales, y las Inoviridae, Nanoviridae y Geminiviridae que infectan a las plantas (Dávila et al., 2019).

Además, cabe señalar que los ambientes fríos muestran una composición similar a la de los sedimentos, incluyendo estas familias de virus ssDNA anteriormente mencionadas en elevada abundancia. Esto concuerda con un informe de Yoshida et al., quienes reportan a los virus ssDNA como predominantes en los sedimentos marinos y tienen una abundancia estimada de 1×10^8 a 3×10^9 copias del genoma por cm^3 de sedimento, más abundantes que los virus dsDNA (Yoshida et al., 2018).

Respecto a virus particulares dentro de un ambiente extremo, por ejemplo, en los ecosistemas hipersalinos, la diversidad fue menor comparado con otros ambientes, un factor determinante de la composición de la comunidad son las concentraciones de sales donde la baja actividad hídrica afecta directamente la composición de las comunidades microbianas (Le Romancer et al., 2006; Ma et al., 2010; Merino et al., 2019). Los virus que se han identificado en estos ecosistemas son halovirus y una gran cantidad de estos infectan arqueas, bacterias y eucariotas (Atanasova et al., 2018; Plominsky et al., 2018; Ramos-Barbero et al., 2019). Un total de 64 virus de arqueas de las dos filas, Crenarchaeota y Euryarchaeota, se han podido aislar de estos dos ecosistemas (Porter et al., 2007). En general la alta concentración de NaCl limita la diversidad viral debido a la escasez de agua disponible y, por lo tanto, de hospederos. Otro factor limitante de la diversidad es la dinámica de los ciclos lisogénicos y líticos, puesto que esto presenta un mecanismo de adaptación específica al alto contenido de salinidad (Roux et al., 2016).

Al igual que la concentración de sal limita la diversidad de microorganismos, la baja concentración de oxígeno modela la diversidad de virus en un ambiente en particular, donde la comunidad viral a lo largo de los gradientes verticales de oxígeno disuelto en el océano tiene fluctuaciones de diversidad disminuyendo de forma directa a medida que baja el oxígeno (Cassman et al., 2012; Parvathi et al., 2018).

Estas diferencias en las comunidades virales también se pudieron observar cuando analizamos los 10 metagenomas de ventilas hidrotermales, donde los resultados mostraron que las comunidades virales están estratificadas, predominando de igual forma los virus de ssDNA en los sedimentos, como las familias Microviridae e Inoviridae, mientras que en las plumas predominaron los virus de dsDNA (Castelán et al., 2019).

La predominancia de virus ssDNA, como se mencionó anteriormente, había sido reportada en sedimentos de subsuelos poco profundos y en ambientes terrestres (Yoshida et al., 2013). Mediante el análisis determinamos que están presentes en sedimentos profundos de océanos, ventilas hidrotermales, lugares con bajas concentraciones de oxígeno y ambientes psicrófilos (Lagos antárticos) (Castelán et al., 2019; Dávila et al., 2019).

Pocos virus de ssDNA habían sido recuperados de ambientes terrestres, por lo que resulta interesante la predominancia de estos virus en sedimentos, lo que podría indicar que éstos se depositan en el fondo del mar o su hospedero vive en los sedimentos. Sin embargo, se ha propuesto que los virus de ssDNA son más abundantes en los ambientes marinos, representando un 96.3% de partículas virales en zonas bentónicas (Yoshida et al., 2013; Yoshida et al., 2018).

Hasta la fecha se han aislado virus ssDNA que infectan a diatomeas marinas, Bacteroidetes marinos (*Celluophaga baltica*), además de haber sido recuperados a partir de microbialitas costeras (Desnues et al., 2008), corales (Wegley et al., 2007) y de agua dulce, lago Antártico (Sommers et al., 2019), y células protistas marinas. Esto podría indicar que posiblemente los hospederos de los virus ssDNA habitan en ventilas hidrotermales y sedimentos oceánicos (Holmfedt et al., 2012). No obstante, también se ha propuesto que los virus ssDNA se consideran virus alóctonos en sedimentos marinos, ya que se cree que se depositan en las zonas bentónicas por sedimentación (Hewson et al., 2012).

Estos resultados podrían indicar una estratificación dentro de los viromas en ventilas hidrotermales. Para corroborarlo, realizamos una comparación en la composición de viromas de las muestras pertenecientes a ventilas hidrotermales profundas (plumas y sedimentos), con otras muestras de sedimentos del océano. Un análisis de clustering reveló que los virus ssDNA (Circoviridae Microviridae y Geminiviridae) están presentes predominantemente en las muestras de sedimentos, los virus dsDNA (Myoviridae, Siphoviridae y Podoviridae) están en abundancia, en las muestras de agua, sugiriendo que las comunidades virales de respiraderos hidrotermales son distintas, y se correlaciona con los reportes de virus de ssDNA en sedimentos mencionados anteriormente (Yoshida et al., 2013).

Otra de las diferencias observadas en los ambientes extremos analizados fue que los ambientes hipersalinos están enriquecidos en organismos pertenecientes al dominio Archaea, principalmente Euryarcheaota y, por lo tanto, era de esperarse que

se encuentrara una mayor cantidad de virus de las familias, Pleolipoviridae, Halspiviridea, Sphaerolipoviridae, y Fuselloviridae. Estos virus son denominados Haloarchaeoviruses, los cuales se han caracterizado por tener morfologías diversas y tamaños de genoma entre 7.0 a 103.2 kb, con genomas de dsDNA lineales en su mayoría, y muy pocos de ssDNA (Roien y Oksanen 2011; Luk et al., 2014).

Esta misma especialización de virus a un determinado ambiente ocurre en los ambientes hipertermófilos, donde las familias de virus que predominan son Fuselloviridae, Lipothrixviridae, Rudiviridae, Guttaviridae, Globuloviridae, Ampullaviridae, y Bicaudaviridae (Prangishvili 2013), las cuales infectan a arqueas principalmente.

En este trabajo analizamos dos ambientes salinos y termófilos, el primero de ellos ubicado en Cuzco, Perú, y el segundo en Puebla, México. En ambos se determinó la estructura de las comunidades de microorganismos y virales, mediante un análisis metagenómico tipo shotgun. Este análisis reveló comunidades de microorganismos particulares y diversas respecto a lo reportado anteriormente. En el caso de los ambientes salinos de Perú, éstos exhibieron una alta diversidad y abundancia bacteriana de los filos Proteobacteria, Bacteroidetes, Balneolaeota y Actinobacteria; en contraste, los phyla Euryarchaeota, Thaumarchaeota y Crenarchaeota de arqueas en poca abundancia (Castelán et al., 2019). Mientras que la estructura de la comunidad viral presentó principalmente virus dentro del orden de los Caudovirales, donde se pudieron ensamblar dos genomas de los bacteriófagos de *Halomonas elongata.* Adicionalmente se recuperaron virus dsDNA asociados con eucariotas como Phycodnaviridae, Poxviridae, Mimiviridae, y Pandoravidae, junto con 27 virus nuevos. Estos resultados contrastan con lo reportado anteriormente en los ambientes hipersalinos. Esto se debe principalmente a los factores fisicoquímicos del ambiente, en el caso de los ambientes salinos de Cusco Perú la concentración de sal de 19 a 21% de salinidad, a dicha concentración, existe una mayor disponibilidad de oxígeno y agua, por lo tanto, mayor diversidad y abundancia, estos resultados correlacionan con otros ambientes circunneutrales como los de Santa Pola con salinidades de 13% a 19% (Fernández et al., 2014).

Lo mismo ocurrió cuando se analizó el metagenoma termal ubicado en Puebla, la estructura de la comunidad de microorganismos mostró una mayor abundancia de bacterias, en contraste con la abundancia de arqueas reportadas. Respecto a la población de bacterias, se encontraron los géneros Rhodococcus, Acinetobacter,

Halothobacillus, Bacillus, Thermomonas, Lysobacter, Luteiomonas, entre otras, mientras que la comunidad de virus correspondió principalmente a los Caudovirales, con una particular abundancia en la familia Herelleviridae, además de otras familias de virus como Baculoviridae, Phycodnaviridae y Mimiviridae (Castelán et al., 2020).

En estos dos ambientes extremos analizados se observaron comunidades diferentes a lo reportado en los ambientes hipersalinos e hipertermófilos. La diferencia es debida principalmente a las condiciones fisicoquímicas del entorno, incluyendo altas concentraciones de sal, temperaturas, pH, el bajo contenido de nutrientes y la disponibilidad de oxígeno.

Los ambientes analizados en este trabajo corresponden a ambientes extremos, cuyos rangos de salinidad (18-20% NaCl) y temperatura (50ºC) se consideran moderados. En el caso de las concentraciones de sal en los ambientes acuáticos, se ha reportado que a medida que se incrementa la concentración de sal, disminuye la disponibilidad de agua y oxígeno para uso de los microorganismos, limitando la diversidad de microorganismos (Ventosa et al., 2015).

La diversidad y abundancia de microorganismos en ambientes hipersalinos está en función al porcentaje de salinidad, en porcentajes cerca de la saturación con 33 a 37% de salinidad, la diversidad es baja y predominan principalmente arqueas, y cuando disminuye el porcentaje de cationes cerca del 13% al 19% (salinidad intermedia) la abundancia de arqueas decrece y aumenta la abundancia de bacterias (Ventosa et al., 2015). Por lo tanto, el gradiente de salinidad influye directamente en la estructura de la comunidad de microorganismos, mostrando diferentes abundancias en las comunidades de bacterias y arqueas.

No obstante, una característica en la fisiología de estos microorganismos halófilos interesante es que el enriquecimiento de aminoácidos ácidos en las proteínas citoplasmáticas les permite funcionar correctamente a altas salinidades. Esta especialización se conserva en microorganismos que viven tanto salinidades intermedias como en altas (Fernández et al., 2014, Ventosa et al., 2015).

Dadas estas diferencias fisicoquímicas fue posible encontrar comunidades particulares de bacterias en estos ambientes extremos moderados, los mismo que se observó en la estructura viroma donde predominan principalmente bacteriófagos. Pero independiente de los parámetros fisicoquímicos en los ambientes extremos, predominan virus de dsDNA y ssDNA, mientras que virus de RNA son escasos. Esto

sugiere que el genoma de RNA no es muy estable en las duras limitaciones de los hábitats extremos o que hay un sesgo en la obtención de los genomas virales.

De igual manera, en los ambientes termófilos estudiados se observó la misma abundancia de bacterias. Esta estructura de la comunidad también se ve determinada por los parámetros fisicoquímicos como la temperatura y pH.

En el caso del ambiente termal de Chignahuapan, la abundancia de bacterias es debida a las características circunneutrales. Particularmente se recuperaron nueve familias virales, como Siphoviridae, Myoviridae, Podoviridae, Corticoviridae y Herelleviridae. Estos virus también infectan a los invertebrados Baculoviridae, algas eucariotas Phycodnaviridae y Protozoan Mimiviridae.

Mediante una comparación de este sitio con otros circunneutrales, se reveló que las familias más abundantes son Myoviridae, Siphoviridae y Herelleviriadae, además de ser ubicuas en todos los metagenomas. Estos resultados contrastan con previos reportes en ambientes terrestres térmicos o geotérmicos, donde las principales familias de virus presentes son Fuselloviridae, Bicaudaviridae, Turriviridae, Ampullaviridae, Guttaviridae, Lipothrixviridae, Rudiviridae y Globuloviridae. No obstante, se esperaba que los virus recuperados en los metagenomas de Puebla infectaran principalmente bacterias, porque a la temperatura y el pH descritas en el ambiente existe una mayor diversidad de microorganismos En general, las comunidades virales en ambientes termófilos moderados se han reportado que infectan a las bacterias termófilas moderadas de los géneros Rhodobacter, Acidithiobacillus, Lysobacter, Rhodococcus, Halothiobacillus, Thermomonas, Bacillus, entre otras. Se evidenció que existen distintas estructuras de comunidades virales entre las fuentes termales, en comparación con fuentes termales ácidas o hipertermófilas como Yellowstone.

Por otra parte, en cuanto al análisis funcional de los metagenomas obtenidos por shotgun, en las ventilas hidrotermales se evidenció que gran parte de los genes tuvieron una función estructural, es decir, que forman parte de las cápsides de los viriones, y pocos genes involucrados en la reparación de ADN. Contrastando con los reportado por Anderson et al., 2014, donde reporta que el perfil metabólico de los virus dentro de las ventilas se han observado una alta abundancia de genes relacionados con el metabolismo energético, así como cofactores y vitaminas en la fracción viral, lo que sugiere la codificación de los genes metabólicos auxiliares en el

genoma viral, que favorece la adaptación de sus hospederos al medio (Anderson et al., 2014).

En nuestro estudio únicamente dos genes eran auxiliares del metabolismo AMG, la cititidiltransferasa y la ribonucleotido reductasa se encontraron, un número limitado de genes. Este primer análisis funcional realizado estuvo limitado por cuestiones metodológicas, pues la anotación fue obtenida por medio de MG-RAST, que hace una comparación contra la base de datos RefSeq, donde solo se encuentran genomas de referencia, sesgando la búsqueda a solo aquellos genes reportados previamente.

Los genes auxiliares del metabolismo virales (AMG) sirven de apoyo para complementar procesos metabólicos del hospedero, que favorecen la producción de fagos, por lo tanto, actúan en vías de aumento de energía, y en la replicación eficiente del fago. Estos genes codifican para proteínas que participan en la fotosíntesis, el metabolismo del azufre, carbono, nitrógeno, la biosíntesis de nucleótidos, oxidación del metano, reparación de ADN, y la tolerancia al estrés (Crummett et al., 2016; Warwick et al., 2019; Mara et al., 2020).

Los AMG encontrados dentro de las ventilas hidrotermales evolucionan bajo selección negativa o purificadora, lo que indica que son bastante conservados en los genomas virales, por que cumplen funciones vitales como la biosíntesis de nucleótidos.

La anotación funcional de los AMG actualmente se puede realizar por medio del desarrollo de nuevas herramientas bioinformáticas, como VIBRANT (Kieft et al., 2020). Dicha herramienta se utilizó en el caso de la anotación en lo metagenomas analizados en esta tesis. Cabe destacar que, para las muestras de aguas termales de Puebla, los AGM se clasificaron dentro de las categorías funcionales de metabolismo de cofactores y vitaminas (MCV), metabolismo de carbohidratos (CM), metabolismo de aminoácidos (AAM), metabolismo de terpenoides, metabolismo azufre (Castelán et al., 2020).

Estos resultados indican que los virus participan activamente en vías metabólicas en el medio ambiente. Es el caso del metagenoma de Puebla, donde la vía SOX se encontró completa en las proteobacterias de este ambiente. Esta vía involucra la oxidación de sulfuro (S^{2-}) y tiosulfato ($S_2O_3^{2-}$) a sulfato (SO_4^{2-}), utilizadas por las bacterias para producir o conservar la energía a través de la cadena de transporte de electrones unida a la membrana. Los virus presentaron dos genes *cysH*

y *sufS*. El primero codifica para una fosfoadenosina fosfosultafo reductasa, involucrada en la reducción de sulfato a sulfito. El segundo codifica para una cisteína desulfurasa, la cual moviliza el azufre de la L-cisteína para producir L-alanina, un paso esencial en el metabolismo del azufre para la biosíntesis de una variedad de biomoléculas que contienen azufre. Por lo tanto, bacterias y virus contribuyen activamente en el ciclo del azufre (Castelán et al., 2020).

En el caso de los metagenomas de salinidad intermedia, solo por mencionar un ejemplo, se encontraron AMG involucrados con el metabolismo del ADN, como el gene *phoH*, que codifica para la proteína de unión a ADN phoH, y es activa en respuesta al estrés. Así también se encontraron genes que codifican para el ribonucleótido reductasa, involucrada en la formación de desoxirribonucleótidos a partir de ribonucleótidos. De esta manera, se confirma que los AMG virales son importantes en la complementación de vías metabólicas de los hospederos virales en ambientes extremos.

Con la anotación funcional de los metagenomas tipo shotgun, se encontró que en los ambientes de salinidad intermedia y termales, existen genes relacionados con el estrés oxidativo. Su presencia era esperada, dado que podrían permitir una mejor adaptación al medio como estrategias de sobrevivencia en ambientes extremos.

En general los microorganismos extremófilos son de interés ya que sus enzimas y proteínas activas tienen que estar adaptadas a condiciones fisicoquímica extremas. Como se mencionó en el Capítulo 2, por ejemplo, los virus tienen enzimas biotecnológicamente útiles como las endolisinas, lisinas asociadas a virones, depolimerasas de polisacáridos, estas se encargan de despolimerizar el peptido glicano, polisacárido como la cápsula, el lipopolisacárido (LPS), degradación que es requerida para la inyección de material genético de fagos a la célula infectada. Estos son de interés para su uso potencial en las farmacéuticas como nuevas estrategias contra patógenos bacterianos resistentes a los antibióticos (Dávila et al., 2019). Otro ejemplo, las ADN polimerasas virales, las ribonucleótidas reductasas las cuales son termoestables, son utilizadas en biología molecular y en la biotecnología enzimas como transaminasas, anhidrasas carbónicas, deshalogenasas, esterasas y epóxido hidrolasas.

Es importante resaltar que el análisis funcional de los viromas en los ambientes extremos en este trabajo solo se limitó a una predicción bioinformática, quedando

claro que es necesario hacer un análisis posterior para corroborar la actividad funcional que tiene los virus en los metagenomas analizados.

En resumen, en este trabajo obtuvimos la estructura del viroma en diferentes ambientes extremos, como ventilas hidrotermales, salinidad intermedia y termales terrestres, donde cada uno presentó virus particulares. Adicional a esto se recuperaron genomas completos bacterianos y virales de los metagenomas y finalmente dichos metagenomas fueron anotados funcionalmente donde interesantemente los virus están involucrados en diferentes vías metabólicas contribuyendo activamente en los ambientes extremos.

CONCLUSIONES

En conclusión, este trabajo hace una aportación importante en el estudio de los virus en ambientes extremos, describiendo cual es la estructura del viroma en estos ambientes, donde los Caudovirales son ubicuos, mientras que en otros ambientes extremos hay virus particulares, adicional a esto el análisis funcional de los metagenomas revelo que los virus participan en vías metabólicas importantes y el evolutivo de los genes auxiliares del metabolismo mostró que dichos genes son conversados y evolucionan bajo selección natural negativa.

A continuación, se enlistan las conclusiones más relevantes de este trabajo.

- Los caudovirales son virus ubicuos en diferentes ambientes extremos y mostrando aparentemente una adaptación poliextremofíla.
- Los ambientes hipersalinos, marinos profundos e hiperáridos, tiene similares en familias virales.
- Los ambientes hipersalinos y hipertermófilos tienen familias de virus que son exclusivas de estos ambientes como Fuselloviridae, Lipothrixviridae, Rudiviridae, Guttaviridae, Globuloviridae, Pleolipoviridae, Halspiviridea, y Sphaerolipoviridae.
- Los ambientes extremos tienen poblaciones de virus predominantes de dsDNA y ssDNA, los virus de RNA son escasos.
- La estructura del viroma en las ventilas hidrotermales tiene comunidad viral estratificada donde poblaciones específicas de virus de ssDNA, como Circoviridae, Microviridae, Geminiviridae son exclusivas de sedimentos.
- Existe una mayor abundancia de virus dsDNA en las muestras de plumas de ventilas hidrotermales en comparación con los sedimentos.
- Los genes auxiliares del metabolismo cititidiltransferasa y ribonucleotido reductasa, evolucionan bajo selección purificadora y, por lo tanto, se mantienen en el genoma viral.
- El análisis funcional de los viromas de las ventilas hidrotermales, mostró principalmente secuencias de proteínas estructurales de los viriones.
- Los ambientes de salinidad intermedia mostraron una gran diversidad y abundancia de bacterias, en comparación de otros ambientes hipersalinos donde predominan las arqueas.

- En los ambientes salinos los fila de las proteobacterias fueron las más abundantes, seguido de Bacteroidetes, Balneolaeota, y Actinobacteria.
- Dos borradores de genomas (draft genome) fueron ensamblados de los metagenomas de salinos de Perú, estos fueron las bacterias *H. elongata* e *I. loihiensis*, las cuales se observó que tienen diferentes mecanismos de adaptación a ambientes hipersalinos, a través de la síntesis de novo de los transportadores ectoína y *natAB*.
- La estructura del viroma de las comunidades de salinidad intermedia reveló una abundancia de virus del orden de los Caudovirales.
- Se obtuvieron dos nuevos fagos completos de *Halomonas eloganta* con una logitud de 13 kb.
- El análisis funcional de los ambientes salinos intermedios indicó que los microorganismos contribuyen a los ciclos biogeoquímicos principalemente en el ciclo de carbono y nitrógeno como fuente de energía.
- El análisis funcional mostros una abundancia importante de genes relacionados con el estrés oxidativo y la reparación del ADN, en los ambientes salinos.
- El viroma de los ambientes salinos intermedios presentó genes de proteínas reparadoras de dsDNA exclusivos de eucariotas y bacterias.
- La composición de microorganismos en las aguas termales de Puebla predominó los géneros: Rhodococcus, Acinetobacter, Thermomonas, Tepidimonas, y Azotobacter, en comparación con otros ambientes circumneutrales en el mundo estas bacterias son oxidantes de azufre.
- El análisis funcional de las aguas termales predijo que las vías metabólicas del ciclo de Calvin-Benson son las principales vías para contribuir a la fijación de carbono.
- El viroma de las aguas termales mostró que la familia *Herelleviridae* fue principalmente abundante.
- El análisis de los genes auxiliares reveló que los virus también contribuyen a las vías metabólicas del ciclo del azufre.

PERSPECTIVAS

La metagenómica viral a permitido dilucidar la estructura del viroma en a los ambientes extremos, sin embargo, aún no es suficiente pues existen una diversidad de ambientes con diferentes condiciones fisicoquímicas que aún no se han explorado las comunidades virales, por ejemplo, en este trabajo, se describieron dos ambientes con condiciones fisicoquímicas de ambiente extremos moderados, revelando que existe una mayor diversidad de microorganismos, comparado con otros ambientes.

Uno de los retos más importantes es realizar la caracterización funcional de las secuencias virales, para ello se tienen que realizar la amplificación, clonación y asilamiento de los genes auxiliares del metabolismo de los virus para poder afirmar las funciones que se predicen con la bioinformática.

Mucho de los descrito en el trabajo reporta la abundancia de virus de DNA en ambientes extremos, no obstante, aún falta realizar la exploración con más detalle de los virus de RNA que están en los ambientes extremos, para lo cual no hay una metodología estándar para su estudio.

El estudio de los viromas en ambientes extremos aún tiene limitaciones por no existe una estrategia metodológica unificada, por lo tanto, es una limitante para poder comprender el fenómeno biológico.

BIBLIOGRAFÍA

1. Afiahayati, Kengo Sato, and Yasubumi Sakakibara. 2015. “MetaVelvet-SL: An Extension of the Velvet Assembler to a de Novo Metagenomic Assembler Utilizing Supervised Learning.” DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes 22 (1): 69–77. doi:10.1093/dnares/dsu041.
2. Aguirre de Cárcer, Daniel, Alberto López-Bueno, David A. Pearce, and Antonio Alcamí. 2015. "Biodiversity And Distribution Of Polar Freshwater DNA Viruses". Science Advances 1 (5): e1400127. doi:10.1126/sciadv.1400127.
3. Aiewsakun, Pakorn, and Peter Simmonds. 2018. “The Genomic Underpinnings of Eukaryotic Virus Taxonomy: Creating a Sequence-Based Framework for Family-Level Virus Classification.” Microbiome 6 (1): 38. doi:10.1186/s40168-018-0422-7.
4. Anderson, Norman G., John L. Gerin, and N. Leigh Anderson. 2003. “Global Screening for Human Viral Pathogens.” Emerging Infectious Diseases 9 (7): 768–73. doi:10.3201/eid0907.030004.
5. Anderson, Rika E., Mitchell L. Sogin, and John A. Baross. 2014. "Evolutionary Strategies Of Viruses, Bacteria And Archaea In Hydrothermal Vent Ecosystems Revealed Through Metagenomics". Plos ONE 9 (10): e109696. doi:10.1371/journal.pone.0109696.
6. Anderson, Rika E., William J. Brazelton, and John A. Baross. 2011. "Is The Genetic Landscape Of The Deep Subsurface Biosphere Affected By Viruses?". Frontiers In Microbiology 2. doi:10.3389/fmicb.2011.00219.
7. Andrei, Adrian-Ştefan, Horia Leonard Banciu, and Aharon Oren. 2012. "Living With Salt: Metabolic And Phylogenetic Diversity Of Archaea Inhabiting Saline Ecosystems". *FEMS Microbiology Letters* 330 (1): 1-9. doi:10.1111/j.1574-6968.2012.02526.x.
8. Anesio, Alexandre M., and Christopher M. Bellas. 2011. "Are Low Temperature Habitats Hot Spots Of Microbial Evolution Driven By Viruses?". Trends In Microbiology 19 (2): 52-57. doi:10.1016/j.tim.2010.11.002.
9. Antipov, Dmitry, Mikhail Raiko, Alla Lapidus, and Pavel A. Pevzner. 2020. “MetaviralSPAdes: Assembly of Viruses from Metagenomic Data.” Bioinformatics 36 (14). Oxford Academic: 4126–29. doi:10.1093/bioinformatics/btaa490.
10. Arndt, David, Jason R. Grant, Ana Marcu, Tanvir Sajed, Allison Pon, Yongjie Liang, and David S. Wishart. 2016. “PHASTER: A Better, Faster Version of the PHAST Phage Search Tool.” Nucleic Acids Research 44 (W1): W16-21. doi:10.1093/nar/gkw387.
11. Atanasova, N., Demina, T., Krishnam Rajan Shanthi, S., Oksanen, H., & Bamford, D. (2018). Extremely halophilic pleomorphic archaeal virus HRPV9 extends the diversity of pleolipoviruses with integrases. *Research In Microbiology*, *169*(9), 500-504. doi: 10.1016/j.resmic.2018.04.004
12. Baizan-Edge, Amanda, Peter Cock, Stuart MacFarlane, Wendy McGavin, Lesley Torrance, and Susan Jones. 2019. “Kodoja: A Workflow for Virus Detection in Plants Using k-Mer Analysis of RNA-Sequencing Data.” Journal of General Virology 100 (3): 533–42. doi:10.1099/jgv.0.001210.
13. Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” Journal of Computational Biology 19 (5): 455–77. doi:10.1089/cmb.2012.0021.
14. Banos, Stefanos, Guillaume Lentendu, Anna Kopf, Tesfaye Wubet, Frank Oliver Glöckner, and Marlis Reich. 2018. "A Comprehensive Fungi-Specific 18S Rrna Gene

Sequence Primer Toolkit Suited For Diverse Research Issues And Sequencing Platforms". BMC Microbiology 18 (1). doi:10.1186/s12866-018-1331-4.
15. Bharti, Richa, and Dominik G Grimm. 2019. "Current Challenges And Best-Practice Protocols For Microbiome Analysis". Briefings In Bioinformatics 22 (1): 178-193. doi:10.1093/bib/bbz155.
16. Bolduc, Benjamin, Ho Bin Jang, Guilhem Doulcier, Zhi-Qiang You, Simon Roux, and Matthew B. Sullivan. 2017. “VConTACT: An IVirus Tool to Classify Double-Stranded DNA Viruses That Infect Archaea and Bacteria.” PeerJ 5 (May). doi:10.7717/peerj.3243.
17. Borozan, Ivan, Shane Wilson, Paola Blanchette, Philippe Laflamme, Stuart N. Watt, Paul M. Krzyzanowski, Fabrice Sircoulomb, Robert Rottapel, Philip E. Branton, and Vincent Ferretti. 2012. “CaPSID: A Bioinformatics Platform for Computational Pathogen Sequence Identification in Human Genomes and Transcriptomes.” BMC Bioinformatics 13 (August): 206. doi:10.1186/1471-2105-13-206.
18. Breitbart, Mya, Luke Thompson, Curtis Suttle, and Matthew Sullivan. 2007. “Exploring the Vast Diversity of Marine Viruses.” Oceanography 20 (2): 135–39. doi:10.5670/oceanog.2007.58.
19. Brito, Elcia M. S., Norberto Villegas-Negrete, Irene A. Sotelo-González, César A. Caretta, Marisol Goñi-Urriza, Claire Gassie, and Florence Hakil et al. 2014. "Microbial Diversity In Los Azufres Geothermal Field (Michoacán, Mexico) And Isolation Of Representative Sulfate And Sulfur Reducers". Extremophiles 18 (2): 385-398. doi:10.1007/s00792-013-0624-7.
20. Castelán-Sánchez, Hugo G., Itzel Lopéz-Rosas, Wendy A. García-Suastegui, Raúl Peralta, Alan D.W. Dobson, Ramón Alberto Batista-García, and Sonia Dávila-Ramos. 2019. "Extremophile Deep-Sea Viral Communities From Hydrothermal Vents: Structural And Functional Analysis". Marine Genomics 46: 16-28. doi:10.1016/j.margen.2019.03.001.
21. Castelán-Sánchez, Hugo G., Itzel Lopéz-Rosas, Wendy A. García-Suastegui, Raúl Peralta, Alan D.W. Dobson, Ramón Alberto Batista-García, and Sonia Dávila-Ramos. 2019. "Extremophile Deep-Sea Viral Communities From Hydrothermal Vents: Structural And Functional Analysis". Marine Genomics 46: 16-28. doi:10.1016/j.margen.2019.03.001.
22. Castelán-Sánchez, Hugo Gildardo, Pablo M. Meza-Rodríguez, Erika Carrillo, David I. Ríos-Vázquez, Arturo Liñan-Torres, Ramón Alberto Batista-García, Ernesto Pérez-Rueda, Norma Elena Rojas-Ruíz, and Sonia Dávila-Ramos. 2020. "The Microbial Composition In Circumneutral Thermal Springs From Chignahuapan, Puebla, Mexico Reveals The Presence Of Particular Sulfur-Oxidizing Bacterial And Viral Communities". Microorganisms 8 (11): 1677. doi:10.3390/microorganisms8111677.
23. Castelán-Sánchez, Hugo Gildardo, Paola Elorrieta, Pedro Romoacca, Arturo Liñan-Torres, José Luis Sierra, Ingrid Vera, and Ramón Alberto Batista-García et al. 2019. "Intermediate-Salinity Systems At High Altitudes In The Peruvian Andes Unveil A High Diversity And Abundance Of Bacteria And Viruses". Genes 10 (11): 891. doi:10.3390/genes10110891.
24. Compeau, Phillip E. C., Pavel A. Pevzner, and Glenn Tesler. 2011. “How to Apply de Bruijn Graphs to Genome Assembly.” Nature Biotechnology 29 (11). Nature Publishing Group: 987–91. doi:10.1038/nbt.2023.
25. Corinaldesi, Cinzia, Michael Tangherlini, and Antonio Dell’Anno. 2017. “From Virus Isolation to Metagenoma Generation for Investigating Viral Diversity in Deep-Sea Sediments.” Scientific Reports 7 (1). Nature Publishing Group: 8355. doi:10.1038/s41598-017-08783-4.
26. Crummett, Lisa T., Richard J. Puxty, Claudia Weihe, Marcia F. Marston, and Jennifer B.H. Martiny. 2016. "The Genomic Content And Context Of Auxiliary Metabolic Genes In Marine Cyanomyoviruses". Virology 499: 219-229. doi:10.1016/j.virol.2016.09.016.
27. Daffonchio, Daniele, Sara Borin, Tullio Brusa, Lorenzo Brusetti, Paul W. J. J. van der Wielen, Henk Bolhuis, and Michail M. Yakimov et al. 2006. "Stratified Prokaryote

Network In The Oxic–Anoxic Transition Of A Deep-Sea Halocline". Nature 440 (7081): 203-207. doi:10.1038/nature04418.
28. Dalmaso, Gabriel, Davis Ferreira, and Alane Vermelho. 2015. "Marine Extremophiles: A Source Of Hydrolases For Biotechnological Applications". Marine Drugs 13 (4): 1925-1965. doi:10.3390/md13041925.
29. Dávila-Ramos, Sonia, A. Estradas-Romero, R. Prol-ledesma, and K. Juárez-López. 2015. "Bacterial Populations (First Record) At Two Shallow Hydrothermal Vents Of The Mexican Pacific West Coast".
30. Dávila-Ramos, Sonia, Hugo G. Castelán-Sánchez, Liliana Martínez-Ávila, María del Rayo Sánchez-Carbente, Raúl Peralta, Armando Hernández-Mendoza, Alan D. W. Dobson, Ramón A. Gonzalez, Nina Pastor, and Ramón Alberto Batista-García. 2019. “A Review on Viral Metagenomics in Extreme Environments.” Frontiers in Microbiology 10. Frontiers. doi:10.3389/fmicb.2019.02403.
31. Dávila-Ramos, Sonia, Hugo G. Castelán-Sánchez, Liliana Martínez-Ávila, María del Rayo Sánchez-Carbente, Raúl Peralta, Armando Hernández-Mendoza, Alan D. W. Dobson, Ramón A. Gonzalez, Nina Pastor, and Ramón Alberto Batista-García. 2019. "A Review On Viral Metagenomics In Extreme Environments". Frontiers In Microbiology 10. doi:10.3389/fmicb.2019.02403.
32. Dávila-Ramos, Sonia, Hugo G. Castelán-Sánchez, Liliana Martínez-Ávila, María del Rayo Sánchez-Carbente, Raúl Peralta, Armando Hernández-Mendoza, Alan D. W. Dobson, Ramón A. Gonzalez, Nina Pastor, and Ramón Alberto Batista-García. 2019. "A Review On Viral Metagenomics In Extreme Environments". Frontiers In Microbiology 10. doi:10.3389/fmicb.2019.02403.
33. DeCastro, María-Eugenia, Esther Rodríguez-Belmonte, and María-Isabel González-Siso. 2016. "Metagenomics Of Thermophiles With A Focus On Discovery Of Novel Thermozymes". Frontiers In Microbiology 7. doi:10.3389/fmicb.2016.01521.
34. Delwart, Eric L. 2007. "Viral Metagenomics". Reviews In Medical Virology 17 (2): 115-131. doi:10.1002/rmv.532.
35. Demergasso, Cecilia, Emilio O Casamayor, Guillermo Chong, Pedro Galleguillos, Lorena Escudero, and Carlos PedrÃ³s-AliÃ³. 2004. "Distribution Of Prokaryotic Genetic Diversity In Athalassohaline Lakes Of The Atacama Desert, Northern Chile". FEMS Microbiology Ecology 48 (1): 57-69.
36. Desnues, Christelle, Beltran Rodriguez-Brito, Steve Rayhawk, Scott Kelley, Tuong Tran, Matthew Haynes, and Hong Liu et al. 2008. "Biodiversity And Biogeography Of Phages In Modern Stromatolites And Thrombolites". Nature 452 (7185): 340-343. doi:10.1038/nature06735.
37. Dimon , Michelle T., Henry M. Wood, Pamela H. Rabbitts, and Sarah T. Arron. 2013. “IMSA: Integrated Metagenomic Sequence Analysis for Identification of Exogenous Reads in a Host Genomic Background.” PLOS ONE 8 (5). Public Library of Science: e64546. doi:10.1371/journal.pone.0064546.
38. Dion, Moïra B., Frank Oechslin, and Sylvain Moineau. 2020. "Phage Diversity, Genomics And Phylogeny". Nature Reviews Microbiology 18 (3): 125-138. doi:10.1038/s41579-019-0311-5.
39. Emerson, Joanne B., Brian C. Thomas, Karen Andrade, Eric E. Allen, Karla B. Heidelberg, and Jillian F. Banfield. 2012. "Dynamic Viral Populations In Hypersaline Systems As Revealed By Metagenomic Assembly". Applied And Environmental Microbiology 78 (17): 6309-6320. doi:10.1128/aem.01212-12.
40. environments. Extremophiles, 18(5), 811-824
41. Fang, Zhencheng, Jie Tan, Shufang Wu, Mo Li, Congmin Xu, Zhongjie Xie, and Huaiqiu Zhu. 2019. “PPR-Meta: A Tool for Identifying Phages and Plasmids from Metagenomic Fragments Using Deep Learning.” GigaScience 8 (6). Oxford Academic. doi:10.1093/gigascience/giz066.
42. Fernández, Ana B., Rohit Ghai, Ana-Belen Martin-Cuadrado, Cristina Sánchez-Porro, Francisco Rodriguez-Valera, and Antonio Ventosa. 2014. "Prokaryotic Taxonomic And Metabolic Diversity Of An Intermediate Salinity Hypersaline Habitat Assessed By

Metagenomics". FEMS Microbiology Ecology 88 (3): 623-635. doi:10.1111/1574-6941.12329.
43. Garretto, Andrea, Thomas Hatzopoulos, and Catherine Putonti. 2019. “VirMine: Automated Detection of Viral Sequences from Complex Metagenomic Samples.” PeerJ 7 (April). PeerJ Inc.: e6695. doi:10.7717/peerj.6695.
44. Gil, Jose F., Victoria Mesa, Natalia Estrada-Ortiz, Mauricio Lopez-Obando, Andrés Gómez, and Jersson Plácido. 2021. "Viruses In Extreme Environments, Current Overview, And Biotechnological Potential". Viruses 13 (1): 81. doi:10.3390/v13010081.
45. Grant, W. D. 2004. "Life At Low Water Activity". Philosophical Transactions Of The Royal Society Of London. Series B: Biological Sciences 359 (1448): 1249-1267. doi:10.1098/rstb.2004.1502.
46. Haferburg, Götz, Janosch A.D. Gröning, Nadja Schmidt, Nicolai-Alexeji Kummer, Juan Carlos Erquicia, and Michael Schlömann. 2017. "Microbial Diversity Of The Hypersaline And Lithium-Rich Salar De Uyuni, Bolivia". Microbiological Research 199: 19-28. doi:10.1016/j.micres.2017.02.007.
47. Ho, Thien, and Ioannis E. Tzanetakis. 2014. “Development of a Virus Detection and Discovery Pipeline Using next Generation Sequencing.” Virology 471–473 (December): 54–60. doi:10.1016/j.virol.2014.09.019.
48. Inskeep, William. 2013. "The YNP Metagenoma Project: Environmental Parameters Responsible For Microbial Distribution In The Yellowstone Geothermal Ecosystem". Frontiers In Microbiology 4. doi:10.3389/fmicb.2013.00067.
49. John, Seth G, Carolina B Mendez, Li Deng, Bonnie Poulos, Anne Kathryn M Kauffman, Suzanne Kern, Jennifer Brum, Martin F Polz, Edward A Boyle, and Matthew B Sullivan. 2011. “A Simple and Efficient Method for Concentration of Ocean Viruses by Chemical Flocculation.” Environmental Microbiology Reports 3 (2): 195–202. doi:10.1111/j.1758-2229.2010.00208.x.
50. Jurtz, Vanessa Isabell, Julia Villarroel, Ole Lund, Mette Voldby Larsen, and Morten Nielsen. 2016. “MetaPhinder—Identifying Bacteriophage Sequences in Metagenomic Data Sets.” PLOS ONE 11 (9). Public Library of Science: e0163111. doi:10.1371/journal.pone.0163111.
51. Kieft, Kristopher, Zhichao Zhou, and Karthik Anantharaman. 2020. “VIBRANT: Automated Recovery, Annotation and Curation of Microbial Viruses, and Evaluation of Viral Community Function from Genomic Sequences.” Microbiome 8 (1): 90. doi:10.1186/s40168-020-00867-0.
52. Koonin, Eugene V., and Yuri I. Wolf. 2012. "Evolution Of Microbes And Viruses: A Paradigm Shift In Evolutionary Biology?". Frontiers In Cellular And Infection Microbiology 2. doi:10.3389/fcimb.2012.00119.
53. Koonin, Eugene V., Valerian V. Dolja, Mart Krupovic, Arvind Varsani, Yuri I. Wolf, Natalya Yutin, F. Murilo Zerbini, and Jens H. Kuhn. 2020. “Global Organization and Proposed Megataxonomy of the Virus World.” Microbiology and Molecular Biology Reviews 84 (2). American Society for Microbiology. doi:10.1128/MMBR.00061-19.
54. Kostic, Aleksandar D., Akinyemi I. Ojesina, Chandra Sekhar Pedamallu, Joonil Jung, Roel G. W. Verhaak, Gad Getz, and Matthew Meyerson. 2011. “PathSeq: Software to Identify or Discover Microbes by Deep Sequencing of Human Tissue.” Nature Biotechnology 29 (5): 393–96. doi:10.1038/nbt.1868.
55. Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., & Koonin, E. (2018). Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research*, *244*, 181-193. doi: 10.1016/j.virusres.2017.11.025
56. Laffy, Patrick W., Elisha M. Wood-Charlson, Dmitrij Turaev, Karen D. Weynberg, Emmanuelle S. Botté, Madeleine J. H. van Oppen, Nicole S. Webster, and Thomas Rattei. 2016. “HoloVir: A Workflow for Investigating the Diversity and Function of Viruses in Invertebrate Holobionts.” Frontiers in Microbiology 7 (June). doi:10.3389/fmicb.2016.00822.

57. Lefebvre, Marie, Sébastien Theil, Yuxin Ma, and Thierry Candresse. 2019. “The VirAnnot Pipeline: A Resource for Automated Viral Diversity Estimation and Operational Taxonomy Units Assignation for Virome Sequencing Data.” Phytobiomes Journal 3 (4). Scientific Societies: 256–59. doi:10.1094/PBIOMES-07-19-0037-A.
58. Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. “MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph.” Bioinformatics 31 (10): 1674–76. doi:10.1093/bioinformatics/btv033.
59. Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., & Zhang, H. et al. (2011). Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Briefings In Functional Genomics*, *11*(1), 25-37. doi: 10.1093/bfgp/elr035
60. Lopez-Bueno, A., J. Tamames, D. Velazquez, A. Moya, A. Quesada, and A. Alcami. 2009. "High Diversity Of The Viral Community From An Antarctic Lake". Science 326 (5954): 858-861. doi:10.1126/science.1179287.
61. López-Hernández, Aída, Gerardo García-Estrada, Gerardo Aguirre-Díaz, Eduardo González-Partida, Hugo Palma-Guzmán, and José L. Quijano-León. 2009. "Hydrothermal Activity In The Tulancingo–Acoculco Caldera Complex, Central Mexico: Exploratory Studies". Geothermics 38 (3): 279-293. doi:10.1016/j.geothermics.2009.05.001.
62. Lorenzi, Hernan A., Jeff Hoover, Jason Inman, Todd Safford, Sean Murphy, Leonid Kagan, and Shannon J. Williamson. 2011. “TheViral MetaGenome Annotation Pipeline(VMGAP):An Automated Tool for the Functional Annotation of Viral Metagenomic Shotgun Sequencing Data.” Standards in Genomic Sciences 4 (3): 418–29. doi:10.4056/sigs.1694706.
63. Lossouarn, Julien, Samuel Dupont, Aurore Gorlas, Coraline Mercier, Nadege Bienvenu, Evelyne Marguet, Patrick Forterre, and Claire Geslin. 2015. "An Abyssal Mobilome: Viruses, Plasmids And Vesicles From Deep-Sea Hydrothermal Vents". Research In Microbiology 166 (10): 742-752. doi:10.1016/j.resmic.2015.04.001.
64. Lozupone, C., & Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied And Environmental Microbiology*, *71*(12), 8228-8235. doi: 10.1128/aem.71.12.8228-8235.2005
65. Luk, Alison, Timothy Williams, Susanne Erdmann, R. Papke, and Ricardo Cavicchioli. 2014. "Viruses Of Haloarchaea". Life 4 (4): 681-715. doi:10.3390/life4040681.
66. Maabar, Maha, Andrew J. Davison, Matej Vučak, Fiona Thorburn, Pablo R. Murcia, Rory Gunson, Massimo Palmarini, and Joseph Hughes. 2019. “DisCVR: Rapid Viral Diagnosis from High-Throughput Sequencing Data.” Virus Evolution 5 (2). Oxford Academic. doi:10.1093/ve/vez033.
67. Mara, Paraskevi, Dean Vik, Maria G. Pachiadaki, Elizabeth A. Suter, Bonnie Poulos, Gordon T. Taylor, Matthew B. Sullivan, and Virginia P. Edgcomb. 2020. "Viral Elements And Their Potential Influence On Microbial Processes Along The Permanently Stratified Cariaco Basin Redoxcline". The ISME Journal 14 (12): 3079-3092. doi:10.1038/s41396-020-00739-3.
68. Mardanov, Andrey V., Alexey V. Beletsky, Denis A. Ivasenko, Olga V. Karnachuk, and Nikolai V. Ravin. 2017. "Metagenoma Sequence Of A Microbial Community From The Gold Mine Tailings In The Kuzbass Area, Russia". Genome Announcements 5 (49). doi:10.1128/genomea.01355-17.
69. Marquet, M., Hölzer, M., Pletz, M., Viehweger, A., Makarewicz, O., Ehricht, R., & Brandt, C. (2020). What the Phage: A scalable workflow for the identification and analysis of phage sequences. doi: 10.1101/2020.07.24.219899
70. Maturrano, Lenin, Fernando Santos, Ramon Rosselló-Mora, and Josefa Antón. 2006. "Microbial Diversity In Maras Salterns, A Hypersaline Environment In The Peruvian

Andes". Applied And Environmental Microbiology 72 (6): 3887-3895. doi:10.1128/aem.02214-05.
71. McNair, Katelyn, Carol Zhou, Elizabeth A. Dinsdale, Brian Souza, and Robert A. Edwards. 2019. “PHANOTATE: A Novel Approach to Gene Identification in Phage Genomes.” Bioinformatics 35 (22). Oxford Academic: 4537–42. doi:10.1093/bioinformatics/btz265.
72. Meier-Kolthoff, Jan P, and Markus Göker. 2017. “VICTOR: Genome-Based Phylogeny and Classification of Prokaryotic Viruses.” Edited by Janet Kelso. Bioinformatics 33 (21): 3396–3404. doi:10.1093/bioinformatics/btx440.
73. Merino, Nancy, Heidi S. Aronson, Diana P. Bojanova, Jayme Feyhl-Buska, Michael L. Wong, Shu Zhang, and Donato Giovannelli. 2019. "Living At The Extremes: Extremophiles And The Limits Of Life In A Planetary Context". Frontiers In Microbiology 10. doi:10.3389/fmicb.2019.00780.
74. Mizuno, Carolina Megumi, Francisco Rodriguez-Valera, Nikole E. Kimes, and Rohit Ghai. 2013. "Expanding The Marine Virosphere Using Metagenomics". Plos Genetics 9 (12): e1003987. doi:10.1371/journal.pgen.1003987.
75. Modha, Sejal, Anil S. Thanki, Susan F. Cotmore, Andrew J. Davison, and Joseph Hughes. 2018. “ViCTree: An Automated Framework for Taxonomic Classification from Protein Sequences.” Bioinformatics (Oxford, England) 34 (13): 2195–2200. doi:10.1093/bioinformatics/bty099.
76. Mokili, John L, Forest Rohwer, and Bas E Dutilh. 2012. "Metagenomics And Future Perspectives In Virus Discovery". Current Opinion In Virology 2 (1): 63-77. doi:10.1016/j.coviro.2011.12.004.
77. Naghoni, Ali, Giti Emtiazi, Mohammad Ali Amoozegar, Mariana Silvia Cretoiu, Lucas J. Stal, Zahra Etemadifar, Seyed Abolhassan Shahzadeh Fazeli, and Henk Bolhuis. 2017. "Microbial Diversity In The Hypersaline Lake Meyghan, Iran". Scientific Reports 7 (1). doi:10.1038/s41598-017-11585-3.
78. Namiki, Toshiaki, Tsuyoshi Hachiya, Hideaki Tanaka, and Yasubumi Sakakibara. 2012. “MetaVelvet: An Extension of Velvet Assembler to de Novo Metagenoma Assembly from Short Sequence Reads.” Nucleic Acids Research 40 (20): e155. doi:10.1093/nar/gks678.
79. Nayfach, Stephen, Antonio Pedro Camargo, Emiley Eloe-Fadrosh, Simon Roux, and Nikos Kyrpides. 2020. “CheckV: Assessing the Quality of Metagenome-Assembled Viral Genomes.” Preprint. Bioinformatics. doi:10.1101/2020.05.06.081778.
80. Nishimura, Yosuke, Takashi Yoshida, Megumi Kuronishi, Hideya Uehara, Hiroyuki Ogata, and Susumu Goto. 2017. “ViPTree: The Viral Proteomic Tree Server.” Bioinformatics 33 (15). Oxford Academic: 2379–80. doi:10.1093/bioinformatics/btx157.
81. Nooij, Sam, Dennis Schmitz, Harry Vennema, Annelies Kroneman, and Marion P. G. Koopmans. 2018. “Overview of Virus Metagenomic Classification Methods and Their Biological Applications.” Frontiers in Microbiology 9. Frontiers. doi:10.3389/fmicb.2018.00749.
82. Nooij, Sam, Dennis Schmitz, Harry Vennema, Annelies Kroneman, and Marion P. G. Koopmans. 2018. “Overview of Virus Metagenomic Classification Methods and Their Biological Applications.” Frontiers in Microbiology 9. Frontiers. doi:10.3389/fmicb.2018.00749.
83. Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824-834. doi: 10.1101/gr.213959.116
84. Oren, Aharon. 2008. "Microbial Life At High Salt Concentrations: Phylogenetic And Metabolic Diversity". Saline Systems 4 (1): 2. doi:10.1186/1746-1448-4-2.
85. Ortmann, Alice C., and Curtis A. Suttle. 2005. "High Abundances Of Viruses In A Deep-Sea Hydrothermal Vent System Indicates Viral Mediated Microbial Mortality". Deep

Sea Research Part I: Oceanographic Research Papers 52 (8): 1515-1527. doi:10.1016/j.dsr.2005.04.002.

86. Ospina-Giraldo, M., Griffith, J., Laird, E., & Mingora, C. (2010). The CAZyome of Phytophthora spp.: A comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus Phytophthora. *BMC Genomics*, *11*(1), 525. doi: 10.1186/1471-2164-11-525
87. Overbeek, R., Olson, R., Pusch, G., Olsen, G., Davis, J., & Disz, T. et al. (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), D206-D214. doi: 10.1093/nar/gkt1226
88. Peng, Y., H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin. 2012. “IDBA-UD: A de Novo Assembler for Single-Cell and Metagenomic Sequencing Data with Highly Uneven Depth.” Bioinformatics 28 (11): 1420–28. doi:10.1093/bioinformatics/bts174.
89. Pickett, B., Greer, D., Zhang, Y., Stewart, L., Zhou, L., & Sun, G. et al. (2012). Virus Pathogen Database and Analysis Resource (ViPR): A Comprehensive Bioinformatics Database and Analysis Resource for the Coronavirus Research Community. *Viruses*, *4*(11), 3209-3226. doi: 10.3390/v4113209
90. Pierce, N. Tessa, Luiz Irber, Taylor Reiter, Phillip Brooks, and C. Titus Brown. 2019. “Large-Scale Sequence Comparisons with Sourmash.” F1000Research 8 (July). doi:10.12688/f1000research.19675.1.
91. Pinzón-Martínez, D.L., C. Rodríguez-Gómez, D. Miñana-Galbis, J.A. Carrillo-Chávez, G. Valerio-Alfaro, and R. Oliart-Ros. 2010. "Thermophilic Bacteria From Mexican Thermal Environments: Isolation And Potential Applications". Environmental Technology 31 (8-9): 957-966. doi:10.1080/09593331003758797.
92. Ponsero, Alise J., and Bonnie L. Hurwitz. 2019. “The Promises and Pitfalls of Machine Learning for Detecting Viruses in Aquatic Metagenomes.” Frontiers in Microbiology 10. Frontiers. doi:10.3389/fmicb.2019.00806.
93. Pontefract, Alexandra, Ting F. Zhu, Virginia K. Walker, Holli Hepburn, Clarissa Lui, Maria T. Zuber, Gary Ruvkun, and Christopher E. Carr. 2017. "Microbial Diversity In A Hypersaline Sulfate Lake: A Terrestrial Analog Of Ancient Mars". Frontiers In Microbiology 8. doi:10.3389/fmicb.2017.01819.
94. Prangishvili, David. 2013. "The Wonderful World Of Archaeal Viruses". Annual Review Of Microbiology 67 (1): 565-585. doi:10.1146/annurev-micro-092412-155633.
95. Quillaguamán, J., S. Hashim, F. Bento, B. Mattiasson, and R. Hatti-Kaul. 2005. "Poly(B-Hydroxybutyrate) Production By A Moderate Halophile,Halomonas BoliviensisIc1 Using Starch Hydrolysate As Substrate". Journal Of Applied Microbiology 99 (1): 151-157. doi:10.1111/j.1365-2672.2005.02589.x.
96. Quince, Christopher, Alan W Walker, Jared T Simpson, Nicholas J Loman, and Nicola Segata. 2017. "Shotgun Metagenomics, From Sampling To Analysis". *Nature Biotechnology* 35 (9): 833-844. doi:10.1038/nbt.3935.
97. Ramos-Barbero, María Dolores, José M. Martínez, Cristina Almansa, Nuria Rodríguez, Judith Villamor, María Gomariz, Cristina Escudero, et al. 2019. “Prokaryotic and Viral Community Structure in the Singular Chaotropic Salt Lake Salar de Uyuni.” Environmental Microbiology 21 (6): 2029–42. doi:https://doi.org/10.1111/1462-2920.14549.
98. Rampelli, Simone, Matteo Soverini, Silvia Turroni, Sara Quercia, Elena Biagi, Patrizia Brigidi, and Marco Candela. 2016. “ViromeScan: A New Tool for Metagenomic Viral Community Profiling.” BMC Genomics 17 (148). BioMed Central Ltd.

99. Rampelotto, Pabulo. 2013. "Extremophiles And Extreme Environments". Life 3 (3): 482-485. doi:10.3390/life3030482.
100. Ren, Jie, Kai Song, Chao Deng, Nathan A. Ahlgren, Jed A. Fuhrman, Yi Li, Xiaohui Xie, Ryan Poplin, and Fengzhu Sun. 2020. “Identifying Viruses from Metagenomic Data Using Deep Learning.” Quantitative Biology 8 (1): 64–77. doi:10.1007/s40484-019-0187-4.
101. Ren, Jie, Nathan A. Ahlgren, Yang Young Lu, Jed A. Fuhrman, and Fengzhu Sun. 2017. “VirFinder: A Novel k-Mer Based Tool for Identifying Viral Sequences from Assembled Metagenomic Data.” Microbiome 5 (1): 69. doi:10.1186/s40168-017-0283-5.
102. Rice, George, Kenneth Stedman, Jamie Snyder, Blake Wiedenheft, Debbie Willits, Susan Brumfield, Timothy McDermott, and Mark J. Young. 2001. “Viruses from Extreme Thermal Environments.” Proceedings of the National Academy of Sciences 98 (23). National Academy of Sciences: 13341–45. doi:10.1073/pnas.231170198.
103. Roine, Elina, and Hanna Oksanen. 2011. "Viruses From The Hypersaline Environment". Springer, Berlin, Heidelberg, 153-172. https://link.springer.com/chapter/10.1007/978-3-642-20198-1_8.
104. Rothschild, Lynn J., and Rocco L. Mancinelli. 2001. "Life In Extreme Environments". Nature 409 (6823): 1092-1101. doi:10.1038/35059215.
105. Roux, Simon, Evelien M. Adriaenssens, Bas E. Dutilh, Eugene V. Koonin, Andrew M. Kropinski, Mart Krupovic, Jens H. Kuhn, et al. 2019. “Minimum Information about an Uncultivated Virus Genome (MIUViG).” Nature Biotechnology 37 (1). Nature Publishing Group: 29–37. doi:10.1038/nbt.4306.
106. Roux, Simon, Francois Enault, Bonnie L. Hurwitz, and Matthew B. Sullivan. 2015. “VirSorter: Mining Viral Signal from Microbial Genomic Data.” PeerJ 3 (May). PeerJ Inc.: e985. doi:10.7717/peerj.985.
107. Roux, Simon, Jennifer R. Brum, Bas E. Dutilh, Shinichi Sunagawa, Melissa B. Duhaime, Alexander Loy, Bonnie T. Poulos, et al. 2016. “Ecogenomics and Potential Biogeochemical Impacts of Globally Abundant Ocean Viruses.” Nature 537 (7622). Nature Publishing Group: 689–93. doi:10.1038/nature19366.
108. Roux, Simon, Joanne B. Emerson, Emiley A. Eloe-Fadrosh, and Matthew B. Sullivan. 2017. “Benchmarking Viromics: An in Silico Evaluation of Metagenome-Enabled Estimates of Viral Community Composition and Diversity.” PeerJ 5 (September). PeerJ Inc.: e3817. doi:10.7717/peerj.3817.
109. Roux, Simon. 2019. “A Viral Ecogenomics Framework To Uncover the Secrets of Nature’s ‘Microbe Whisperers.’” MSystems 4 (3). American Society for Microbiology Journals. doi:10.1128/mSystems.00111-19.
110. Sarwar Muhammmad Kaleem Iqra Azam, Tahir Iqbal. 2015. “Biology and Applications of Halophilic Bacteria and Archaea: A Review” Electronic Journal of Biology, 2015, Vol.11(3): 98-103
111. Sharpton, Thomas J. 2014. "An Introduction To The Analysis Of Shotgun Metagenomic Data". Frontiers In Plant Science 5. doi:10.3389/fpls.2014.00209.
112. Sommers, Pacifica, Rafaela S. Fontenele, Tayele Kringen, Simona Kraberger, Dorota L. Porazinska, John L. Darcy, Steven K. Schmidt, and Arvind Varsani. 2019. "Single-Stranded DNA Viruses In Antarctic Cryoconite Holes". Viruses 11 (11): 1022. doi:10.3390/v11111022.
113. Stano, M., Beke, G., & Klucar, L. (2016). viruSITE—integrated database for viral genomics. *Database*, *2016*, baw162. doi: 10.1093/database/baw162

114. Tamames, J., Cobo-Simón, M., & Puente-Sánchez, F. (2019). Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics*, *20*(1). doi: 10.1186/s12864-019-6289-6
115. Thomas, Torsten, Jack Gilbert, and Folker Meyer. 2012. "Metagenomics - A Guide From Sampling To Data Analysis". Microbial Informatics And Experimentation 2 (1). doi:10.1186/2042-5783-2-3.
116. Tithi, Saima Sultana, Frank O. Aylward, Roderick V. Jensen, and Liqing Zhang. 2018. “FastViromeExplorer: A Pipeline for Virus and Phage Identification and Abundance Profiling in Metagenomics Data.” PeerJ 6 (January): e4227. doi:10.7717/peerj.4227.
117. Upton, Chris, Duncan Hogg, David Perrin, Matthew Boone, and Nomi L Harris. 2000. “Viral Genome Organizer: A System for Analyzing Complete Viral Genomes.” Virus Research 70 (1): 55–64. doi:10.1016/S0168-1702(00)00210-0.
118. Ventosa, A., Fernández, A. B., León, M. J., Sánchez-Porro, C., & Rodriguez-Valera, F.
119. Ventosa, Antonio, Ana Beatriz Fernández, María José León, Cristina Sánchez-Porro, and Francisco Rodriguez-Valera. 2014. "The Santa Pola Saltern As A Model For Studying The Microbiota Of Hypersaline Environments". *Extremophiles* 18 (5): 811-824. doi:10.1007/s00792-014-0681-6.
120. Ventosa, Antonio, Rafael R de la Haba, Cristina Sánchez-Porro, and R Thane Papke. 2015. "Microbial Diversity Of Hypersaline Environments: A Metagenomic Approach". Current Opinion In Microbiology 25: 80-87. doi:10.1016/j.mib.2015.05.002.
121. Wan, Yinan, Daniel W. Renner, Istvan Albert, and Moriah L. Szpara. 2015. “VirAmp: A Galaxy-Based Viral Genome Assembly Pipeline.” GigaScience 4: 19. doi:10.1186/s13742-015-0060-y.
122. Wang, Shiliang, Jaideep P. Sundaram, and David Spiro. 2010. “VIGOR, an Annotation Program for Small Viral Genomes.” BMC Bioinformatics 11 (1): 451. doi:10.1186/1471-2105-11-451.
123. Warwick-Dugdale, Joanna, Holger H. Buchholz, Michael J. Allen, and Ben Temperton. 2019. “Host-Hijacking and Planktonic Piracy: How Phages Command the Microbial High Seas.” Virology Journal 16 (1): 15. doi:10.1186/s12985-019-1120-1.
124. Warwick-Dugdale, Joanna, Holger H. Buchholz, Michael J. Allen, and Ben Temperton. 2019. "Host-Hijacking And Planktonic Piracy: How Phages Command The Microbial High Seas". Virology Journal 16 (1). doi:10.1186/s12985-019-1120-1.
125. Wegley, Linda, Robert Edwards, Beltran Rodriguez-Brito, Hong Liu, and Forest Rohwer. 2007. "Metagenomic Analysis Of The Microbial Community Associated With The Coral Porites Astreoides". Environmental Microbiology 9 (11): 2707-2719. doi:10.1111/j.1462-2920.2007.01383.x.
126. Wommack, K. Eric, and Rita R. Colwell. 2000. “Virioplankton: Viruses in Aquatic Ecosystems.” Microbiology and Molecular Biology Reviews 64 (1). American Society for Microbiology: 69–114. doi:10.1128/MMBR.64.1.69-114.2000.
127. Woolhouse, Mark E.J., and Kyle Adair. 2013. “The Diversity of Human RNA Viruses.” Future Virology 8 (2): 159–71. doi:10.2217/fvl.12.129.
128. Yamashita, Akifumi, Tsuyoshi Sekizuka, and Makoto Kuroda. 2016. “VirusTAP: Viral Genome-Targeted Assembly Pipeline.” Frontiers in Microbiology 7. Frontiers. doi:10.3389/fmicb.2016.00032.
129. Yoshida, Mitsuhiro, Tomohiro Mochizuki, Syun-Ichi Urayama, Yukari Yoshida-Takashima, Shinro Nishi, Miho Hirai, Hidetaka Nomaki, Yoshihiro Takaki, Takuro Nunoura, and Ken Takai. 2018. "Quantitative Viral Community DNA Analysis Reveals The Dominance Of Single-Stranded DNA Viruses In Offshore Upper Bathyal Sediment From Tohoku, Japan". Frontiers In Microbiology 9. doi:10.3389/fmicb.2018.00075.

130. Yoshida, Mitsuhiro, Yoshihiro Takaki, Masamitsu Eitoku, Takuro Nunoura, and Ken Takai. 2013. "Metagenomic Analysis Of Viral Communities In (Hado)Pelagic Sediments". Plos ONE 8 (2): e57271. doi:10.1371/journal.pone.0057271.
131. Zárate, Selene, Blanca Taboada, Martha Yocupicio-Monroy, and Carlos F. Arias. 2017. “Human Virome.” Archives of Medical Research 48 (8): 701–16. doi:10.1016/j.arcmed.2018.01.005.
132. Zhang, Kai-Yue, Yi-Zhou Gao, Meng-Ze Du, Shuo Liu, Chuan Dong, and Feng-Biao Guo. 2019. “Vgas: A Viral Genome Annotation System.” Frontiers in Microbiology 10 (February). doi:10.3389/fmicb.2019.00184.
133. Zhao, Guoyan, Guang Wu, Efrem S. Lim, Lindsay Droit, Siddharth Krishnamurthy, Dan H. Barouch, Herbert W. Virgin, and David Wang. 2017. “VirusSeeker, a Computational Pipeline for Virus Discovery and Virome Composition Analysis.” Virology 503 (March): 21–30. doi:10.1016/j.virol.2017.01.005.
134. Zhao, Guoyan, Siddharth Krishnamurthy, Zhengqiu Cai, Vsevolod L Popov, Hilda Guzman, Song Cao, Herbert W Virgin, Robert B Tesh, and David Wang. 2013. “Identification of Novel Viruses Using VirusHunter -- an Automated Data Analysis Pipeline.” PLOS ONE 8 (10): 11.
135. Zheng, Yi, Shan Gao, Chellappan Padmanabhan, Rugang Li, Marco Galvez, Dina Gutierrez, Segundo Fuentes, Kai-Shu Ling, Jan Kreuze, and Zhangjun Fei. 2017. “VirusDetect: An Automated Pipeline for Efficient Virus Discovery Using Deep Sequencing of Small RNAs.” Virology 500 (January): 130–38. doi:10.1016/j.virol.2016.10.017.
136. Zhou, You, Yongjie Liang, Karlene H. Lynch, Jonathan J. Dennis, and David S. Wishart. 2011. “PHAST: A Fast Phage Search Tool.” Nucleic Acids Research 39 (Web Server issue): W347-352. doi:10.1093/nar/gkr485.

Anexo de protocolos utilizados

Protocolos de wetlab utilizados

1.- https://github.com/hugocastelan/protocols_wetlab

Programas bioinformáticos desarrollados

2.- https://github.com/hugocastelan/Scripts_metagenomics-

3.- https://github.com/hugocastelan/qiime

Artículos de colaboración publicados durante el doctorado.

1. Pineda-Mora, Daniel, Ana Laura Juárez-López, Jeiry Toribio-Jiménez, Maria Teresa Leal-Ascencio, Jesús Carlos Ruvalcaba-Ledezma, **Hugo Gildardo Castelán-Sánchez,** José Luis Aguirre-Noyola, and Paul Alexander Arp. 2020. "Diversity And Functions Of Epilithic Riverine Biofilms". *Water, Air, & Soil Pollution* 231 (8). doi:10.1007/s11270-020-04692-x.
2. Villacís, José E., Jorge A. Reyes**, Hugo G. Castelán-Sánchez,** Sonia Dávila-Ramos, Miguel Angel Lazo, Ahmad Wali, and Luis A. Bodero et al. 2020. "OXA-48 Carbapenemase In Klebsiella Pneumoniae Sequence Type 307 In Ecuador". *Microorganisms* 8 (3): 435. doi:10.3390/microorganisms8030435.
3. Gómez-Silva, Benito, Claudia Vilo-Muñoz, Alexandra Galetović, Qunfeng Dong**, Hugo G. Castelán-Sánchez,** Yordanis Pérez-Llano, and María del Rayo Sánchez-Carbente et al. 2019. "Metagenomics Of Atacama Lithobiontic Extremophile Life Unveils Highlights On Fungal Communities, Biogeochemical Cycles And Carbohydrate-Active Enzymes". *Microorganisms* 7 (12): 619. doi:10.3390/microorganisms7120619**.**
4. Mamani, J. Ilucion, Kevin B. Pacheco, Paola Elorrieta, Pedro Romoacca, **Hugo Castelan**, Sonia Davila, Jose L. Sierra, and Maria A. Quispe-Ricalde. 2019. "Draft Genome Sequence Of Halomonas Elongata MH25661 Isolated From A Saline Creek In The Andes Of Peru". *Microbiology Resource Announcements* 8 (1). doi:10.1128/mra.00934
5. González-Durán, Elizabeth, Mauricio Vázquez-Pichardo, Jesús Miguel Torres-Flores, Fabiola Garcés-Ayala, Alfonso Méndez-Tenorio, Everardo Curiel-Quesada, Joanna María Ortiz-Alcántara, **Hugo Gildardo Castelán Sánchez**, et al.,2018. "Genotypic Variability Analysis Of DENV-1 In Mexico Reveals The Presence Of A Novel Mexican Lineage". *Archives Of Virology* 163 (6): 1643-1647. doi:10.1007/s00705-018-3759-0.
6. Negrete-Abascal, Erasmo, Fernando Montes-Garcia, Sergio Vaca-Pacheco, Abraham M. Leyto-Gil, Edgar Fragoso-Garcia, Roberto Carvente-Garcia, Sandra Perez-Agueros, **Hugo Castelán Sánchez** et al. 2018. "Genome Sequence Of Actinobacillus Seminis Strain ATCC 15768, A Reference Strain Of Ovine Pathogens That Causes Infections In Reproductive Organs". *Genome Announcements* 6 (2). doi:10.1128/genomea.01453-17.

ARTICULO 1

Marine Genomics 46 (2019) 16–28

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Method paper

Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis

Hugo G. Castelán-Sánchez[a], Itzel Lopéz-Rosas[b], Wendy A. García-Suastegui[c], Raúl Peralta[a], Alan D.W. Dobson[d,e], Ramón Alberto Batista-García[a], Sonia Dávila-Ramos[a,*]

[a] Centro de Investigación en Dinámica Celular, Instituto de Investigaciones en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Morelos. Av. Universidad 1001. Col. Chamilpa. Cuernavca, Morelos. C.P, Cuernavaca 62209, Mexico
[b] CONACyT Research fellow-Colegio de Postgraduados Campus Campeche, Carretera Haltunchén - Edzná Km 17.5. Colonia Sihochac. Champotón, Campeche 24450, Mexico
[c] Laboratorio de Toxicología Molecular, Departamento de Biología y Toxicología de la Reproducción, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Puebla C.P., 72570, Mexico
[d] School of Microbiology, University College Cork. Cork, Ireland
[e] Environmental Research Institute, University College, Cork, Ireland

ARTICLE INFO

Keywords:
Deep-sea virome
Natural selection
Auxiliary metabolic gene
Cytitidyltransferase
Ribonucleotide reductase

ABSTRACT

Ten publicly available metagenomic data sets from hydrothermal vents were analyzed to determine the taxonomic structure of the viral communities present, as well as their potential metabolic functions. The type of natural selection on two auxiliary metabolic genes was also analyzed. The structure of the virome in the hydrothermal vents was quite different in comparison with the viruses present in sediments, with specific populations being present in greater abundance in the plume samples when compared with the sediment samples. ssDNA genomes such as *Circoviridae* and *Microviridae* were predominantly present in the sediment samples, with *Caudovirales* which are dsDNA being present in the vent samples. Genes potentially encoding enzymes that participate in carbon, nitrogen and sulfur metabolic pathways were found in greater abundance, than those involved in the oxygen cycle, in the hydrothermal vents. Functional profiling of the viromes, resulted in the discovery of genes encoding proteins involved in bacteriophage capsids, DNA synthesis, nucleotide synthesis, DNA repair, as well as viral auxiliary metabolic genes such as cytitidyltransferase and ribonucleotide reductase. These auxiliary metabolic genes participate in the synthesis of phospholipids and nucleotides respectively and are likely to contribute to enhancing the fitness of their bacterial hosts within the hydrothermal vent communities. Finally, evolutionary analysis suggested that these auxiliary metabolic genes are highly conserved and evolve under purifying selection, and are thus maintained in their genome.

1. Introduction

Hydrothermal vents are cracks or fissures in the seafloor from which geothermally heated water emerges in a column form as the seawater meets the magma (Ledesma, 2011; Tarasov et al., 2005). Despite hydrothermal vents having temperatures of up to 400 °C and a highly reducing chemical nature (Kelley et al., 2005; Martin et al., 2008), they are a source of bacteria and archaea with a high level of biodiversity; which has been investigated using both culture dependent (Cary et al., 1997; Harmsen et al., 1997; Jeanthon, 2000) and independent approaches (Xie et al., 2010; Anderson et al., 2011a; Anantharaman et al., 2015; Zhang et al., 2016a; Pjevac et al., 2018; Cerqueira et al., 2018) A recent report shows an inverse relationship between the abundance and diversity levels in the microbial populations inhabiting hydrothermal vents, suggesting the presence of specific microbial groups which are very well established in these hyperthermophilic environments (Anderson et al., 2017). Prokaryotes have to date been the best studied microorganisms in vents (Huber et al., 2007; Dick et al., 2013; Sheik et al., 2015; Poli et al., 2017; Dávila-Ramos et al., 2014), with reports showing that bacteria and archaea communities residing in hydrothermal plumes are quite different from those present in sediments primarily due to the fact that the plume is much colder and much more strongly influenced by the background seawater (Dick et al., 2013; Ding et al., 2017; Christakis et al., 2018).

In the case of the virome in these polyextremophilic ecosystems, little is known about their taxonomic structure, their metabolism or

* Corresponding author at: Ave. Universidad 1001. Col. Chamilpa, Cuernavaca, 62209, Mexico.
E-mail address: sonia.davila@uaem.mx (S. Dávila-Ramos).

https://doi.org/10.1016/j.margen.2019.03.001
Received 6 December 2018; Received in revised form 25 January 2019; Accepted 1 March 2019
Available online 09 March 2019

Table 1
Metagenome data considered in the structural and functional analysis.

Name of site	Location	Ocean	Deep (m)	ID No. SRA	Sample/Collection date	Sequencing platform	Insert size (bp)	Sequenced bases (Gb)	Reference
Lau Basin Tahi Moana	S 20 40. 927,843 W 176 11.001806	Pacific Ocean	800	SRR 1217461	Plume water/ June 2009	Illumina HiSeq 2500 (paired)	N/A	18.3	Anantharaman et al. (2014)
Mid Cayman 1	N/A	Atlantic Ocean	2238	SRR 2046236	Plume water/ January 2012	Illumina HiSeq 2000 (paired)	N/A	30.7	Li et al. (2015)
Mid Cayman 2	N/A	Atlantic Ocean	4869	SRR 2046238	Plume water/ January 2012	Illumina HiSeq 2000 (paired)	202	36.1	Li et al. (2015)
Axial Seamount	N 27.025 W 111.400	Pacific Ocean	1500	ERR 2021511	Plume water/ August 2015	Illumina NextSeq 500 (paired)	275	7.38	Fortunato et al. (2018)
Guaymas Mexico	N 27.025 W 111.400	Pacific Ocean	1993	SRR 3577362	Plume water/ July 2004	Illumina HiSeq 2000 (paired)	N/A	23.5	Anantharaman et al. (2014)
Menez Gwen	N/A	Atlantic Ocean	828	ERR 1078302	Fluid/ October 2009	Illumina MiSeq (paired)	N/A	5.3	Meier et al. (2016)
South Mid Atlantic	S 15.16 W 13.350	Atlantic Ocean	2500	SRR 4028170	Sediment/ August 2012	Illumina HiSeq 2000 (paired)	N/A	4.07	N/A
Southwest Indian Ridge 1	S 37.55 E 51.00	Indian Ocean	2400	SRR 3136143	Sediment/ March 2014	Illumina HiSeq 2000 (paired)	N/A	5.34	He et al. (2017)
Southwest Indian Ridge 2	S 27.85 E 63.94	Indian Ocean	2400	SRR 3133481	Sediment/ December 2013	Illumina HiSeq 2000 (paired)	N/A	7.30	He et al. (2017)
Southern Mariana Japan	N 12.93 E 143.62	Pacific Ocean	3024	DRR 093004	Sediment/ June 2010	Illumina MiSeq (paired)	300	9.08	Kato et al. (2018)

N/A: No apply.

their overall ecology (Sime-Ngando, 2014). Given the important of viruses as microbial predators that are known to influence global biogeochemical cycles and to impact microbial evolution (Rohwer et al., 2009), it is likely that they play an important role in the ecological relationships with these unique microbial communities inhabiting deep-sea hydrothermal vents. Viral mediated horizontal gene transfer is known to occur on a widespread basis in the oceans (McDaniel et al., 2010). Viruses are known to encode auxiliary metabolic genes (AMGs), which play a crucial role in promoting biochemical and metabolic processes (Beiko et al., 2005; Breitbart et al., 2007).

The viral abundance in active vents has been estimated to be 3.5×10^6 and 2.94×10^6 viruses per ml^{-1} from plume and sediment samples, respectively (Ortmann and Suttle, 2005; Manini et al., 2007). In spite of this, few viruses have to date been isolated from hydrothermal vents using classical techniques; those that have include; the bacteriophages *Bacillus virus* W1 (BVW1), *Geobacillus virus* E1 (GVE1), *Geobacillus virus* E2 (GVE2), *Nitratiruptor phage* (NRS-1) and *Marinitoga piezophila virus* (MPV-1), TPV1 (*Thermococcus prieurii virus* 1) (Gorlas et al., 2012; Romancer et al., 2006; Prangishvili, 2003; Lossouarn et al., 2015). With advances in next generation sequencing based approaches, it is clear that metagenomics will allow us gain a greater appreciation of the virome in these hydrothermal vents. Metagenomic studies have already shown that *Siphoviridae*, *Myoviridae* and *Podoviridae* are the predominant viral families present in these ecosystems (Breitbart et al., 2007; Millard et al., 2014; Anderson et al., 2017; Strazzulli et al., 2017), and that viruses which infect archaea are present in high abundance (Rice et al., 2001; Prangishvili, 2003; Geslin et al., 2003). Despite this, further efforts are needed to increase our knowledge relating to the virosphere that is present in hydrothermal vents; particularly with a view to determining the potential function that these viruses may play in these environments.

As previously mentioned viruses are known to play an essential role in biogeochemical cycles (Rohwer et al., 2009; Weitz and Wilhelm, 2012; Mizuno et al., 2016). Viral AMGs can complement metabolic pathways that are present in bacteria, and following acquisition can remain in the prokaryotic genomes by natural selection, and consequently enhance the fitness of bacterial strains that host these viral genes (Anderson et al., 2011b; Anderson et al., 2014; He et al., 2017). However there is little knowledge about the kind of natural selection that is important in the evolution of these auxiliary metabolic genes. In addition, lysogenic viruses can also have a significant impact on their bacterial hosts, by inducing cell lysis process within the host; and thereby modifying the microbial food web and energy transfer to higher trophic levels (Williamson et al., 2008; Rastelli et al., 2017).

In the last few years a good deal of metagenomic sequence data has been generated from the microbiota of hydrothermal vents (Zhang et al., 2016b). However, generating detailed analysis that allows us to exploit all the information that is available is currently a bottleneck for scientists. With this in mind, we focused on ten publically available metagenomic data sets deposited in the National Center Biotechnology Information (NCBI), which we investigated to analyze the taxonomic structure of the viromes in hydrothermal vents from which they data had been obtained. We found that ssDNA viruses predominate in sediments, whereas dsDNA are more abundant in plumes. In addition we analyze these metagenomes for potential metabolic functions and for the presence and integrity of metabolic pathways potentially involved biogeochemical cycles, and uncovered some specific and complete pathways involved in nitrogen, sulfur, carbon and iron biogeochemical cycles. Following the functional analysis of the viral sequences the most abundant were capsid sequences of bacteriophages, and few carbohydrates and AMGs. Finally the type of natural selection was analyzed on capsids and on two AMGs, which appear to generally evolve under purifying selection, but with a few sites in the AMGs appearing to be under episodic selection.

2. Materials and methods

2.1. Metagenomic data from hydrothermal vents

Ten publically available metagenomic data sets obtained from sequence-based metagenomic studies in hydrothermal vents were collected from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (Leinonen et al., 2010). Table 1 summarizes details of these metagenomic sequences. These datasets were chosen because they were derived from shotgun metagenomic projects in hydrothermal vents and were generated from the Illumina Next-Generation Sequencing platform, that typically produces homogeneous and high quality sequences.

2.2. Sequence quality analysis

Sequence quality control (QC) analysis was performed using the FastQC program (Bioinformatics, 2011). The quality trimming threshold was set to a 30 Q score (corresponds to an 0.001 error rate). Adapters were removed using Trimmomatic software (Bolger et al., 2014). Subsequently, duplicated sequences were eliminated using CD-HIT-DUP software with a maximum mismatch number of -e 0.03 (Huang et al., 2010) since during amplification processes some sequences are artificially produced (Gomez-Alvarez et al., 2009). After sequence quality control a sequence set was obtained and used in the subsequent analysis.

2.3. Viral taxonomic analysis of metagenomes

Sequences were assembled and contigs were obtained using MegaHit software, which uses an assembly algorithm based on Bruijn graphs, using paired-end mode, k min = 21, k max = 131, k step = 10 (Li et al., 2016). The assembled sequences were compared with a previously constructed database that contains approximately 6000 virus genomes available in NCBI (non-redundant nucleotide database). Briefly, to generate the local database, the virus genomes collected from NCBI were indexed using standalone BLASTn.

For the identification of the viral sequences from all the assembled sequences, a comparison with the local database was performed using the following parameters: number of aligments = 20, e-value = 0.0001 and word size = 11. The best twenty scoring BLAST hits were parsed and taxonomically assigned using MEGAN 5.10.6 software (Huson et al., 2007). For the virus taxonomic classification, the method of the lowest common ancestor (LCA) was used using the following parameters: minimum support = 2, minimum score = 70, top percent = 10; this reduced the risk of obtaining false positive or false negative taxonomic assignments (Huson et al., 2016). When taxonomic profiles were obtained, matrix abundances were generated and later processed in R software (version 3.2.3). Finally, plots were also done in R software with the libraries ggplot2 and RColorBrewer 175 (Team, 2013) (www.ColorBrewer.org).

Hierarchical clustering analysis was addressed to compare virome relative abundance of the data with other viromes deposited in Metavir2 (Roux et al., 2014). The hierarchical clusters were obtained using the heatmap2 package in R based in the hclust library, which evaluates the dissimilarity between the relative abundances of the virus families using distance matrix methods (Becker et al., 1988). To conduct this hierarchical clustering analysis, 21 viromes previously deposited in Metavir2 were collected (Table 2). Overall, the clustering analysis included the virus relative abundance obtained from the ten hydrothermal vent metagenomes (previously mentioned in section 2.1) and these 21 viromes.

2.4. Functional analysis of metagenomic sequences

The assembled contigs were uploaded to the MG-RAST server (Glass and Meyer, 2011) and the functional annotation was obtained in the classification of subsystem technology platform (SEED), which is a categorize system that includes five hierarchical levels of functional annotation. For the viral functional annotation on MG-RAST server, only the viral domain sequences in the RefSeq database were selected to avoid the inclusion of contigs from microbial sequences in the functional analysis. The viral functions were obtained according to the classification at level 3 of the subsystems database with an e-value threshold of 1-e^{-5}. The AMGs were identified using the same viral functional analysis.

Multigenomic entropy based score (MEBS) software (Anda et al., 2017) (FDR 0.0001) was used to identify the completeness of metabolic pathway involved in the biogeochemical cycles within the metagenomic datasets, all contigs from each sample were used in this analysis.

2.5. Natural selection analysis

Using only viral sequences obtained from the MG-RAST server we evaluated the natural selection sites by different methods. To obtain the virus sequences from the metagenomics contigs, the RefSeq database was used.

Fixed-effects likelihood (FEL) and random effects likelihood (REL) were used to conduct the natural selection analysis. These algorithms use the principles of maximum likelihood to estimate the proportion of synonymous and non-synonymous rates of each nucleotide site (Pond and Frost, 2005). These methods detect natural selection in a coding gene, identifying higher non-synonymous substitution rates (dN when an amino acid changes) in relation to synonymous substitution rates (dS, silent mutations) that are considered neutral. This relation is represented as:

$$\omega = dS/dN \quad (1)$$

while mixed effect model of evolution (MEME) allows the distribution of ω to vary site by site (fixed effects) and also branch by branch in a site. In addition, the method identifies the two types of episodic and constant natural selection (Murrell et al., 2012). The presence of natural selection was also evaluated, which makes a global comparison of dS/dN rates with "evolutionary fingerprint" software in Datamonkey server. This software is based on certain sites in genes evolving rapidly or resisting the change of natural election. These sites are typically called an "evolutionary fingerprint" (Pond et al., 2009).

3. Results and discussion

3.1. Viral communities in sediments and deep-water from hydrothermal vents

In an attempt to gain additional knowledge on the viral communities present in hydrothermal vents, for which there is currently quite limited information available; this study focused on comparing the viral populations in ten virospheres from metagenomics datasets available for ten different hydrothermal vents located in different geographical zones. The structure of the viral populations in each of the ten locations is sampled divided in plumes and sediments (Fig. 1). The most representative viral communities in both sample types belong to the *Caudovirales* order, with 70–80% of the assembled contigs being classified into three main families namely the *Siphoviridae*, *Podoviridae* and *Myoviridae*. Bacteriophages (*Myoviridae* family) showed the higher relative abundance in samples ranging from 50 to 60% in the plumes and 30–70% in the sediments.

It is well established that bacteriophages are the most abundant viruses found in environmental samples from soil, freshwater and marine ecosystems, and are known to actively regulate the ecological dynamics of the bacteria populations within these environments (Dick et al., 2013; Sepulveda et al., 2016; Hayes et al., 2017; Tetz and Tetz, 2018). The results here appear to indicate that based on the levels of

Table 2
Summary of samples collection from Metavir 2.

Name of site	Location	Ocean	Deep (m)	ID No. Metavir 2	Sample/ Collection date	Sequencing Platform	Insert size (bp)	Sequences number	Reference
Atlantic	41°43′51.2394″ N − 10°40′56.568″ E	Atlantic Ocean	3530	Atl_Vir 1157	Sediment/ October 2008	454 Roche Pyrosequencing	N/A	26,826	Corinaldesi et al. (2017)
Atlantic	41°43′51.2394″ N − 10°40′56.568″ E	Atlantic Ocean	3530	Atlantic-Extra 2125	Sediments/ August 2005	454 Roche Pyrosequencing	N/A	107,090	Corinaldesi et al. (2017)
Arctic	79°8′0.5994″ N 2°50′32.2794″ E	Arctic Ocean	5571	Arct_Vir 1159	Sediment/ August 2005	454 Roche Pyrosequencing	N/A	63,869	Corinaldesi et al. (2017)
Black Sea	42°59′54.204″ N 31°30′58.644″ E	Black Sea	1970	Black Sea 1155	Sediment/ September 2006	454 Roche Pyrosequencing	N/A	78,436	Corinaldesi et al. (2017)
Mediterranean	39°31′04.1880″ N 6°10′32.4012″ E	Mediterranean	2850	Mediterranean 1161	Sediment/ April 2002	454 Roche Pyrosequencing	N/A	65,340	Corinaldesi et al. (2017)
Atlantic	39°30′24.18″ N − 9°50′0.604″ E	Atlantic Ocean	3530	NE_Atlantic_2 1156	Sediment/ August 2005	454 Roche Pyrosequencing	N/A	165,517	Corinaldesi et al. (2017)
Arctic	79°8′0.5994″ N 2°50′32.2794″ E	Arctic Ocean	5571	Arct_Ocean 1158	Sediment/ September 2006	454 Roche Pyrosequencing	N/A	79,646	Corinaldesi et al. (2017)
Izu-Ogasawara Trench	29°09′ N 142°49′ E	Pacific Ocean	9760	Izu-Ogasawara Trench 164	Sediment/ December 2007	454 Roche Pyrosequencing	N/A	46,458	Yoshida et al. (2013)
Mariana Trench	11°22′ N 42°42′ E	Pacific Ocean	10,332	Mariana Trench 165	Sediment/ May 2008	454 Roche Pyrosequencing	N/A	49,584	Yoshida et al. (2013)
Shimokita Peninsula	41°10′ N 142°12′ E	Pacific Ocean	1181	Shimokita Peninsula 166	Sediment/ January 2006	454 Roche Pyrosequencing	N/A	76,498	Yoshida et al. (2013)
Brazos Trinity Basin	27°18.0809′ N 94°,23.2537′ W	Gulf of Mexico	1470	Brazos-Trinity 8mbsf 3813	Sediment/ N/A	454 Roche Pyrosequencing	N/A	270,730	Biddle et al. (2011)
LineP transect, ocean station	48°58′8” N -130°,40′12” W	Pacific Ocean	1000	LJ12O 1394	Water/ June 2009	454 Roche Pyrosequencing	N/A	122,565	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°58′8” N -130°,40′12” W	Pacific Ocean	2000	LJ12D 1395	Water/ June 2010	454 Roche Pyrosequencing	N/A	49,914	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LA26O 1398	Water/ August 2009	454 Roche Pyrosequencing	N/A	70,596	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N, −144°,59′56” W	Pacific Ocean	2000	LA26D 1400	Water/ August 2009	454 Roche Pyrosequencing	N/A	68,516	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LF26O 1403	Water/ Febrary 2009	454 Roche Pyrosequencing	N/A	147,537	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LF26D 1404	Water/ Febrary 2009	454 Roche Pyrosequencing	N/A	125,896	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	1000	LJ26O 1407	Water/ June 2009	454 Roche Pyrosequencing	N/A	101,179	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N -144°,59′56” W	Pacific Ocean	2000	LJ26D 1408	Water/ June 2009	454 Roche Pyrosequencing	N/A	55,332	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°38′58” N -126°39′52” W	Pacific Ocean	1000	LJ4O 1414	Water/ June 2009	454 Roche Pyrosequencing	N/A	97.126	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°38′58” N -126°39′52” W	Pacific Ocean	1300	LJ4D 1415	Water/ June 2009	454 Roche Pyrosequencing	N/A	98,478	Hurwitz and Sullivan (2013)
LineP transect, ocean station	33°17′13” N -129°25′42” W	Pacific Ocean	1000	M6O1K 1432	Water/ Octuber 2009	454 Roche Pyrosequencing	N/A	225,833	Hurwitz and Sullivan, 2013
LineP transect, ocean station	33°17′13” N -129°25′42” W	Pacific Ocean	4300	M7O4K 1433	Water/ Octuber 2009	454 Roche Pyrosequencing	N/A	144,588	Hurwitz and Sullivan, 2013
LineP transect, ocean station	50°6” N, −144°,59′56” W	Pacific Ocean	500	LF26A 1402	Water/ Febrary 2009	454 Roche Pyrosequencing	N/A	167,616	Hurwitz and Sullivan (2013)
LineP transect, ocean station	50°6” N, −144°,59′56” W	Pacific Ocean	500	LA26A 1397	Water/ August 2009	454 Roche Pyrosequencing	N/A	42,118	Hurwitz and Sullivan (2013)
LineP transect, ocean station	48°58′8” N -130°,40′12” W	Pacific Ocean	500	LJ12A 1393	Water/ June 2010	454 Roche Pyrosequencing	N/A	58,108	Hurwitz and Sullivan (2013)

(*continued on next page*)

Table 2 (*continued*)

Name of site	Location	Ocean	Deep (m)	ID No. Metavir 2	Sample/ Collection date	Sequencing Platform	Insert size (bp)	Sequences number	Reference
LineP transect, ocean station	48°38′58″ N -126°39′52″ W	Pacific Ocean	10	LJ4S 1409	Water/ June 2009	454 Roche Pyrosequencing	N/A	92,415	Hurwitz and Sullivan (2013)

N/A: No apply.

bacteriophage present that they are also likely to play a significant ecological contribution in hydrothermal vent ecosystems. The second cluster of viral families with the highest representation in the virome profiles, were the *Mimiviridae, Phycodnaviridae* and *Poxviridae* (Fig. 1). This is not surprising perhaps given that *Mimiviridae* and *Phycodnaviridae* have previously been reported in aquatic environments, with some members of these families being discovered in hot spring environments such as in Yellowstone (Zhang et al., 2015).

Single strand DNA (ssDNA) viruses were exclusively detected in the sediment samples, with *Microviridae* being the family with highest abundance. This finding is consistent with reports of ssDNA viruses being found in freshwater sediments (Hewson et al., 2012; Roux et al., 2012a) and in deep sea samples (Yoshida et al., 2018). ssDNA viruses are considered as allochthonous viruses in marine sediments since it is believed that they are deposited in the benthic zones through sedimentation (Hewson et al., 2012). dsDNA viruses tend to be preferentially detected in environmental samples due to a methodological *bias* in the multiple displacement amplification technique employed, which results in a preferential amplification of dsDNA viruses (Anderson et al., 2014). In the two metagenome data sets from the Southwest Indian Ocean (SRR3136143 and S3133481) included in this work, this type of amplification was not used, and consequently ssDNA viruses were detected. In the Indian samples, *Gokushovirus* was one of the ssDNA virus subfamily which was found, that correlates with the 16S ribosomal gene analysis carried out by Anderson et al. (2014) where *Chlamydiales*, the natural host of *Gokushovirus* (Roux et al., 2012b; Labonté and Suttle, 2013) were shown to be present (Anderson et al., 2014).

The way in which the presence of viral genomes has been calculated has often involved staining the viral capsids with SYBR green, which allows a direct quantification of the amount of viral particles in any given sample. However recent studies have shown that quantification of viral particles with a ssDNA genome has been underestimated in different environments, particularly in marine sediments (Yoshida et al., 2018). It has been estimated that there are between 1×10^8 to 3×10^9 copies of viral genomes per cm^3 in sediments, an amount which is higher than for dsDNA viruses. Given that ssDNA viruses are likely to play an important role in regulating bacterial mortality levels and in ecological succession occurring in prokaryotic communities in deep-sea sediments, further efforts should be made to study the taxonomic structure of these viral populations without methodological biases.

There is currently a lack of knowledge about the structure and ecology of ssDNA viruses in these deep-sea environments. However, it could be inferred that the viral communities within the hydrothermal vents are likely to be stratified in the same way that has been reported for bacterial communities (Dick et al., 2013).

There is however little data currently available regarding virosphere stratification in deep sea hydrothermal vents. To address this, we performed a comparison in virome composition between samples belonging to deep-hydrothermal vents (plumes and sediments), with other samples from deep-sea water (water and sediments) (Fig. 2). The clustering analysis included 21 new datasets from the Metavir2 database, and ten metagenomics datasets from the analysis of deep hydrothermal vents, and aimed to establish whether stratification of the virome is maintained independently of their hydrothermal vent origin. The clustering clearly indicates different viral families in the deep sea. While ssDNA viruses (*Circoviridae Microviridae* and *Geminiviridae*) are predominantly present in the sediment samples, the dsDNA viruses (*Myoviridae, Siphoviridae* and *Podoviridae*) are present in high abundance in the deep water samples (Fig. 2).

Regardless of the clustering of sediments and water samples, there are differences in the viromes that allow the samples that come from hydrothermal vents to be distinguished from other deep sea samples. This suggests that the viral communities of hydrothermal vents are distinct and represent unique extremophilic systems from which novel viruses may be discovered. In addition, we observed an absence of

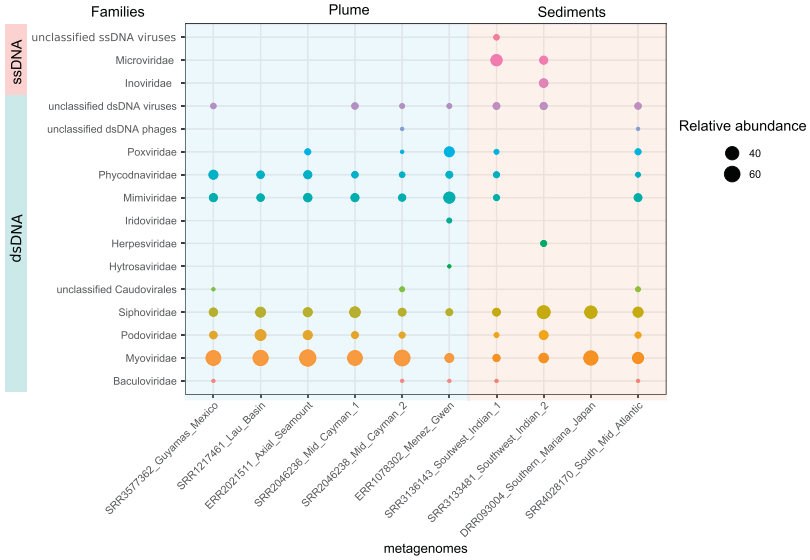


Fig. 1. Viral taxonomic composition in ten samples, which were obtained from SRA database, where correspond to six of plumes and four of sediments of deep hydrothermal vents. Relative abundance (number of contigs) of viral families dsDNA and ssDNA viruses are shown.

archaeal viruses (*Fuselloviridae*) in the metagenome data sets from the plume, sediment and deep-water samples; which have previously been reported in terrestrial high temperature environments such as the hot springs in the Yellowstone National Park (Munson-McGee et al., 2018). Therefore, it appears likely that hydrothermal vents may harbor a different viral structure that terrestrial hydrothermal ecosystems.

3.2. Biogeochemical cycles in hydrothermal vents

Coupled with the lack of information regarding the structure and composition of the virome in hydrothermal vents, the functional analysis of these communities has to date also received very little attention. However, it is clear that the functional analysis of these communities is likely to uncover a vast array of novel genes from viruses inhabiting hyperthermophilic environments such as hydrothermal vents, which may have important biotechnological applications, such as biocatalysts that are active at elevated temperatures (Frock and Kelly, 2012). In recent years, a number of genes has been identified and characterized from deep-sea viromes, including the (*PsbA*) gene encoding the D1 protein in photosystem II and the *NarG, NarH,* and *NarJ* viral nitrate reductase genes involved in the biogenesis of respiratory nitrate from hydrothermal vents (He et al., 2017; Garin-Fernandez et al., 2018). Thus, functional bioprospecting of these environments could provide an opportunity to discovery truly novel proteins.

The metagenomics data sets were firstly analyzed for the completeness of metabolic pathways which are involved in biogeochemical cycles involving carbon, nitrogen, iron and sulfur (Fig. 3), given that as previously mentioned viruses as known to play an important role in the natural recycling of these chemical elements (Weitz and Wilhelm, 2012). This analyses allowed identification of the main pathways used by the microbial community in hydrothermal vents in both the plumes and the sediments samples. This is likely to be directly related to the metabolic activities of microbial communities which are present, which mainly consist of populations involved in carbon fixation and, those involved in redox reactions of nitrogen, sulfur and iron (Eecke et al., 2012; Dick et al., 2013).

This analysis revealed 30 metabolic pathways mainly for nitrogen and sulfur with a completeness of between 80 and 100% in all the metagenomes analyzed (Fig. 3) indicating that the relevant biochemical pathways were present in both the plume water and sediment samples. Amongst these are pathways that are likely to play key ecological roles in the degradation of sulfured compounds in nature, such as: dimethylsulfoniopropionate (DMSP) oxidation, sulfoacetate oxidation, dimethylsulfone oxidation, cysteate oxidation, alkanesulfonate degradation, tetrathionate oxidation and carbon disulfide oxidation. Interestingly, others sulfur-related pathways such as those involved in sulfoquinovosyl diaglycerol (SQDG) biosynthesis and homotaurine degradation were also present in the metagenomes from both the sediments and the plumes. Genes involved in the pathway for sulfite oxidation was also present with a completeness of between 60 and 70% in all the samples with the exception of the sediment sample from Southwest India (SRR2133481).

Regarding the pathways involved in nitrogen metabolism a completeness of between 80 and 100% was observed in all metagenomes, with pathways involved in nitrate reduction (I-X), ammonia oxidation II, nitrate reduction, nitrate reductase (*nirBD*), and the superpathway ammonia being present. Genes involved in methanogenesis pathways were also present in all the metagenomics datasets including those for methanogenesis energy conversion, methanopterin (MTP) methanogenesis, dimethylsulfide (DMS) methanogenesis. However, other complete methanogenesis pathways were poorly represented in all datasets samples with only the 30% of the genes distinguishable for these pathways, while the completeness of the biochemical pathway involved in the conversion of CO_2 into methane was in the range of 80–90% of the pathway being present.

These results demonstrate that nitrogen, carbon and sulfur biogeochemical cycles participate in metabolic processes with a high ecological significance in microbial communities present in hydrothermal vents since the bacterial genomes contain a high percentage all the enzymes that participate as biocatalysts in these catabolic and anabolic

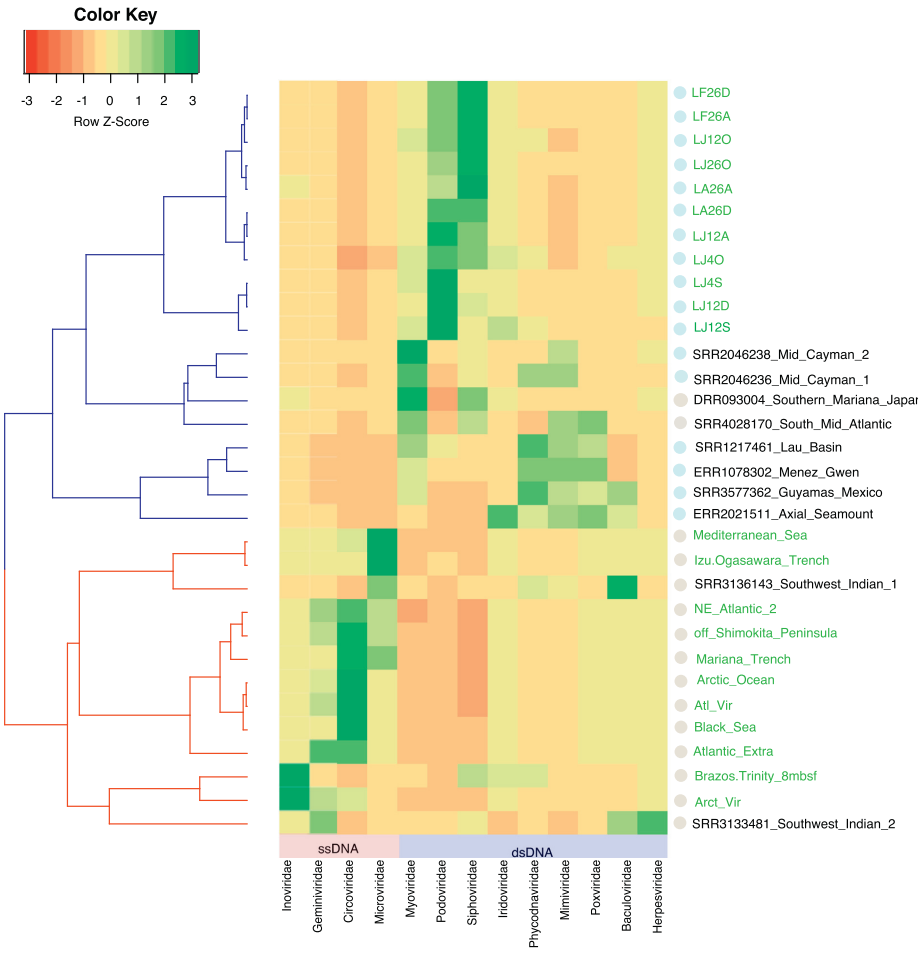


Fig. 2. Heat map of the viral communities, in deep-hydrothermal vents, and samples from deep sea. This clustering reveals two clades, one corresponding to sediments (red) and the other clade correspond to water (blue). Samples with IDs in green font are from Metavir2 and samples with IDs in black font are from the SRA database. Circles in light blue are samples from water, while circles in brown are samples from sediments. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

processes (Anantharaman et al., 2015).

3.3. Functional analysis viral in plumes and sediments from hydrothermal vents

A functional analysis using MG-RAST was then performed to specifically examine the metabolic profiles in the viral communities. In general, the prophage (capsid) was the more abundant category observed (*e*-value of $1e^{-5}$), which is to be expected since these protein structures are abundant in viruses (Brum et al., 2016). This metabolic function for phase capsid synthesis was dominant in both, plume and sediment samples with 3750 and 230 sequences being present, respectively.

Genes encoding other functions including those involved in lytic and lysogenic viral cycles and, those involved in DNA repair and replication, such as the Rlt-like protein and genes involved in phosphate metabolism were also commonly found (*e*-value of $1e^{-5}$) (Fig. 4). There was a lesser diversity in the metabolic functions in the viral genomes recovered from the sediment samples.

Not only were the metabolic functions less diverse, but the number of contigs (genes) associated with a specific metabolic function was lower. While in the plume samples there were 1496 contigs associated with virion structure, only 214 were identified in the sediment samples. The same was true for DNA metabolism, with 883 and 347 coding sequences being identified in plume and sediment samples, respectively; with a *e*-value of $1e^{-5}$.

Finally, some biochemical functions were exclusively observed in viral communities from plumes, such as phosphorus uptake, folate biosynthesis, macromolecular synthesis, ribonucleotide reductase and Type II ATP dependent DNA topoisomerases, amongst others. All these

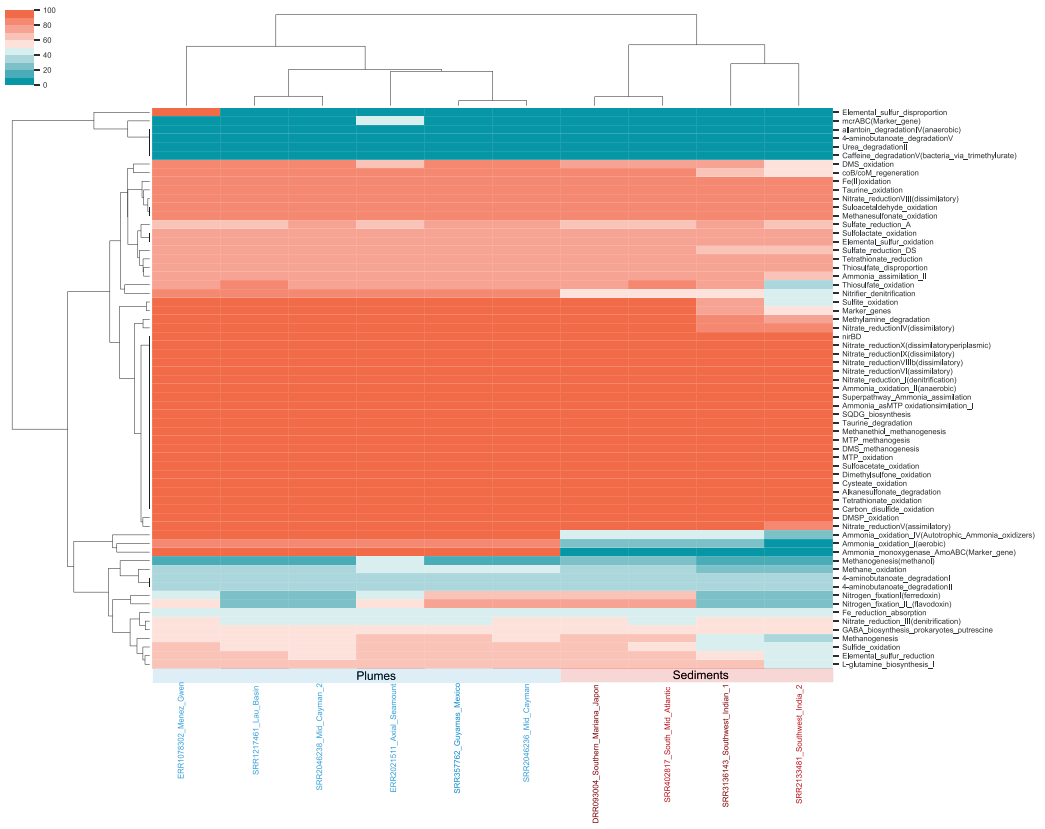


Fig. 3. Clustering hierarchical of pathways of the carbon and sulfur metabolism. Hierarchical heat map is shown, where the most pathway in biogeochemical cycles is marked in red colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are relevant for DNA replication in viral particles.

Genes involved in folate biosynthesis (thymidylate synthase thyX, (TS) was another predominant function observed in the plume datasets. This enzyme is necessary to catalyze the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), and participates in the folate cycle. It is also essential in the synthesis of methionine, together with methylation reactions (Graziani et al., 2004; Leduc et al., 2007). Thymidylate synthases have previously been reported in viruses such as *Phycodnaviridae* (Graziani et al., 2004), *Herpesviridae* and *Caudovirales* (Stern et al., 2010).

Furthermore while proteins involved in carbohydrate metabolism were not abundant; we did however identify a GDP-L-fucose synthase from *Prochlorococcus* phage P-SSM2, which is known to be involved in the biochemical synthesis of oligosaccharides (Han et al., 2012). These enzymes have been reported in *Caudovirales*, *Herpesviruses*, *Poxviruses*, *Baculoviruses* and *Phycodnaviruses* (Markine-Goriaynoff et al., 2004; Graves et al., 2001); but have not been reported from extreme environments as the virosphere from hydrothermal vents has to date not been extensively studied from a biotechnological standpoint.

This functional analysis also allowed the identification in high abundance of genes involved in auxiliary metabolic functions, with cytitidyltransferase and ribonucleotide reductase genes being the most abundant within the class Clustering-based subsystem. The former encode for nucleotidyl transferases which are typically involved in the transfer of phosphorus-containing groups, and have been reported in *Prochlorococcus* phage P-SSM2 (Sullivan et al., 2005; Sullivan et al., 2010; Aylward et al., 2017). Ribonucleotide reductases are involved in nucleotide biosynthesis and have previously been reported in many viral genomes (Sakowski et al., 2014). Moreover, *phoH* genes which are related to the acquisition of phosphate (Goldsmith et al., 2011) were observed but at lower abundances. These genes have also been previously widely reported in bacteriophages (Lindell et al., 2004).

3.4. Analysis of natural selection on auxiliary metabolic genes

In general, in the functional analysis of viruses the most abundant genes that were identified corresponded to those encoding for structural parts of the virion together with those involved in some metabolic auxiliary functions. There is particular interest in the latter sets of genes since it is known that they encode for AMGs, which are known to be involved in promoting biochemical processes and in doing so improving the fitness of their bacterial hosts; by potentially facilitating adaptation within these bacteria/archaea due to the adverse conditions present in the hydrothermal vents ecosystems (Anderson et al., 2011b; He et al., 2017). An example of the role that AMGs play has been reported in cyanobacteria, where cyanophage express their host's photosynthetic

Fig. 4. Viral functional annotation from plumes and sediments of hydrothermal vents. The most abundant correspond to capsids of phages.

genes during the lytic cycle of infection (Clokie et al., 2006), to potentially either support the host or to redirect the cyanobacterial host metabolism to support phage DNA biosynthesis (Thompson et al., 2011). Interestingly AMGs were found in high abundance during our functional analysis and given that it has been suggested that these genes may have evolved under purifying selection in hydrothermal vents (Anderson et al., 2014); a selection analysis was performed to determine which evolutionary processes dominate in the evolution of AMGs in extreme deep-sea environment. There have been very few reports of this in the literature and thus there is a deficit in our knowledge in relation to the evolution of viral genes in deep-sea ecological niches.

Concerning the AMGs, two genes with the higher relative abundance in the viral genomes were chosen, namely the cytitidyltransferase and ribonucleotide reductase genes. Fifty-one and sixty three genes encoding for these enzymes respectively were found and subsequently subjected to potential natural selection analysis. While no potential sites were identified using the REL software; however the use of MEME resulted in the detection of three and six potential sites respectively in the cytitidyltransferase genes. All these sites were predicted with

Table 3
Sites under positive selection in Cytitidiltranferase with FEL p = 0.05.

Codon	dS	dN	dN/dS	*p*-value
52	0	0.7687	Infinite	0.03576
114	0	1.2061	Infinite	0.03568
407	0	1.1755	Infinite	0.00374

Table 4
Sites under episodic selection in Cytitidiltransferese with MEME p = 0.05.

Codon	*α*	β	Pr [β = β+]	β+	Pr [β = β+]	p-value
125	0.0507	0	0.5965	2.03945	0.403492	0.014053
126	0.0907	0.0285	0.7905	7.03581	0.209416	0.0303779
407	0	0	1.0001e-09	0.380733	1	0.00920052
659	0.1596	0.1022	0.82767	139.081	0.17233	0.0326676
753	0.2383	0	0.636865	79.5711	0.363135	0.00641633
965	0	0	0.644192	40.2061	0.355808	0.00698377
1213	0	0	0.605042	1.19307	0.394958	0.0426682

statistical significance levels ($p < 0.05$). The three codons that were identified that may be under positive selection (pervasive selection) using FEL were located in positions 52, 114 and 407 (Table 3); while with MEME (episodic selection), six positions were identified namely; 125, 126, 407, 659, 753 and 1213 (Table 4). In the case of the ribonucleotide reductase genes, no evidence of positive selection was found using MEME, REL and FEL software.

Given the difficult in identifying complete genes in the viral genomes, as well as considering the limited number of sequences present, the replacement rates using evolutionary fingerprint was compared with the data of those sequences encoding viral capsid proteins and cytitidyltransferase. This analysis revealed that the substitution rates *dN/dS* evolved under a purifying selection (negative selection), as predicted by FEL, REL and MEME. However, in the sequences analyzed some changes in the non-synonymous replacement rates were observed, but these variations were subtle without exceeding the value of neutrality (Fig. 5).

In the case of AMGs, evolution under negative selection was generally also observed. This is similar to previous reports indicating that viral genes in hydrothermal system are subject to purifying selection (Anderson et al., 2011a). However, it has also been reported that the AMGs when they are transferred from the virus to their host, can evolve under positive selection (Anderson et al., 2014).

Hydrothermal vents are dynamic and fluent ecosystems, but only a small number of positions in the genes analyzed were identified as having evolved under episodic selection. This indicates that there are periods where alternating conservative selection acts and, periods of change which favor the accumulation of non-synonymous mutations thereby allowing certain adaptive advantages in those genes. This is the first analysis confirming that some genes evolve under episodic selection, and that the frequency of non-synonymous substitution indicates episodes of rapid evolution.

4. Conclusions

The structure of the virome in the hydrothermal vents allows us to distinguish specific populations and those that were present in greater abundance in the plume samples when compared with those of the sediments. The main difference in the structure appears to be due to the presence of ssDNA genomes such as *Circoviridae* and *Microviridae* in the sediment samples. In addition the viromes of the vents are very similar to other samples that have previously been analyzed from deep waters, where *Caudovirales* are ubiquitous. Genes that participate in metabolic pathways that contribute to the production of carbon, nitrogen and sulfur in the hydrothermal vents, were found in greater abundance, when compared with those involved in the oxygen cycle; indicating the types of viral populations that may be participating directly or indirectly in these cycles. On the other hand, in the functional profile of the viromes, we found that the most represented genes were those encoding for proteins involved in bacteriophage capsid synthesis, phage packaging machinery, DNA synthesis, nucleotide synthesis, DNA repair, as well as auxiliary metabolic functions. The AMGs (cytitidyltransferase and ribonucleotide reductase from viruses) which participate in the synthesis of phospholipids and are essential for the synthesis of

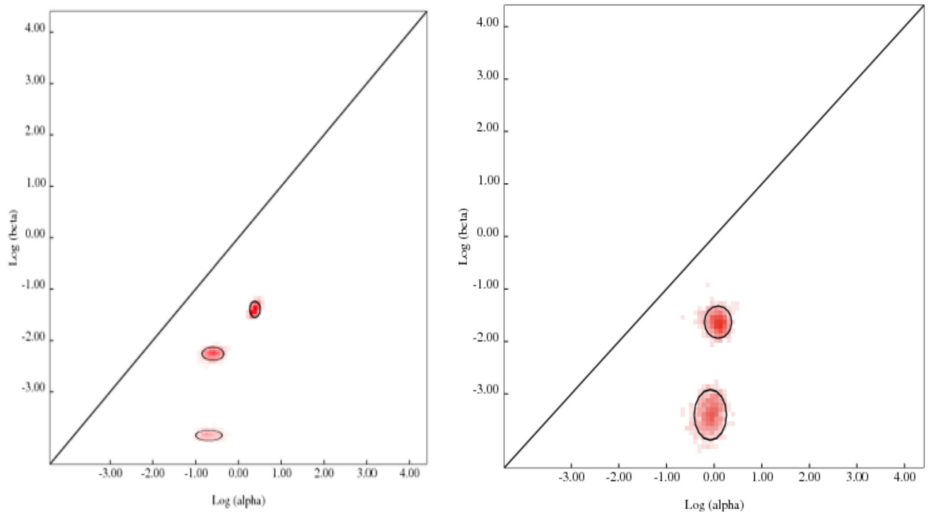


Fig. 5. Comparison of the evolutionary fingerprint *dS/dN* rates. The estimated distribution of the *dS/dN* values is shown. The diagonal indicates the neutrality state. Values located above the diagonal indicative positive selection, while values located below the diagonal indicate negative selection. A) Cytitidyltransferase genes B) Ribonucleotide reductase genes. All *dN/dS* rates are under purifying selection (negative selection).

nucleotides respectively, are likely to contribute to enhancing the fitness of their hosts within the hydrothermal vents as previously proposed (Anderson et al., 2011b; Anderson et al., 2014). The evolutionary analysis suggests that these AMGs are highly conserved and evolve under purifying selection, and are thus maintained in their genome.

Author contributions

HGCS and SDR designed and performed the analysis. RP, ILR, WAGS participated in the analysis. HGCS, SDR, ADWD and RABG analyzed and interpreted the data. All authors contributed in the preparation of the manuscript. All authors read and approved the final manuscript.

Competing interest/Conflict of interest statement

Authors declare that the research was conducted in the absence of any commercial or financial relationships. Thus, any conflict of interest exists.

Acknowledgements

Authors grateful to the Unidad de Secuenciación Masiva y Bioinformática (Instituto de Biotecnología-UNAM) and Centro de Investigación en Dinámica Celular for the access to its computer cluster. HGCS has a scholarship from CONACyT (Mexico). Authors especially grateful to the researchers that obtained the metagenomic sequences used in this manuscript, which are totally public and are deposited in NCBI web site.

References

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K., Toner, B.M., Dick, G.J., 2014. Sulfur oxidation genes in diverse deep-sea viruses. Science 344, 757–760. https://doi.org/10.1126/science.1252229.

Anantharaman, K., Breier, J.A., Dick, G.J., 2015. Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J. 10, 225–239. https://doi.org/10.1038/ismej.2015.81.

Anda, V.D., Zapata-Penasco, I., Hernandez, A.C., Fruns, L.E., Moreira, B.C., Souza, V., 2017. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: Unraveling the sulfur cycle. Gigascience 6, 1–17. https://doi.org/10.1101/191288.

Anderson, R.E., Brazelton, W.J., Baross, J.A., 2011a. Is the genetic landscape of the deep subsurface biosphere affected by viruses? Front. Microbiol. 2, 219. https://doi.org/10.3389/fmicb.2011.00219.

Anderson, R.E., Brazelton, W.J., Baross, J.A., 2011b. Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. FEMS Microbiol. Ecol. 77, 120–133. https://doi.org/10.1111/j.1574-6941.2011.01090.x.

Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., ... Huber, J.A., 2017. Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. Nat. Commun. 8 (1), 1114.

Anderson, R.E., Sogin, M.L., Baross, J.A., 2014. Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. PLoS One 9, e109696. https://doi.org/10.1371/journal.pone.0109696.

Aylward, F.O., Boeuf, D., Mende, D.R., Wood-Charlson, E.M., Vislova, A., Eppley, J.M., Delong, E.F., 2017. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. Proc. Natl. Acad. Sci. 114, 11446–11451. https://doi.org/10.1073/pnas.1714821114.

Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The New S Language. Wadsworth Brooks/Cole (S version.).

Beiko, R.G., Harlow, T.J., Ragan, M.A., 2005. Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 14332–14337. https://doi.org/10.1073/pnas.0504068102.

Biddle, J.F., White, J.R., Teske, A.P., House, C.H., 2011. Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. ISME J. 5, 1038. https://doi.org/10.1038/ismej.2010.199.

Bioinformatics, B., 2011. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Institute. Biology, Cambridge, UK. https://doi.org/10.1371/journal.pbio.0030144.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Breitbart, M., Thompson, L., Suttle, C., Sullivan, M., 2007. Exploring the vast diversity of marine viruses. Oceanography 20, 135–139. https://doi.org/10.5670/oceanog.2007.58.

Brum, J.R., Ignacio-Espinoza, J.C., Kim, E.H., Trubl, G., Jones, R.M., Roux, S., Sullivan, M.B., 2016. Illuminating structural proteins in viral "dark matter" with metaproteomics. Proc. Natl. Acad. Sci. 113, 2436–2441. https://doi.org/10.1073/pnas.1525139113.

Cary, S.C., Cottrell, M.T., Stein, J.L., Camacho, F., Desbruyeres, D., 1997. Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *Alvinella pompejana*. Appl. Environ. Microbiol. 63, 1124–1130.

Cerqueira, T., Barroso, C., Froufe, H., Egas, C., Bettencourt, R., 2018. Metagenomic signatures of microbial communities in deep-sea hydrothermal sediments of Azores vent fields. Microb. Ecol. 76, 1–17. https://doi.org/10.1007/s00248-018-1144-x.

Christakis, C.A., Polymenakou, P.N., Mandalakis, M., Nomikou, P., Kristoffersen, J.B., Lampridou, D., ... Magoulas, A., 2018. Correction to: microbial community differentiation between active and inactive sulfide chimneys of the Kolumbo submarine volcano, Hellenic Volcanic Arc. Extremophiles 22 (5). https://doi.org/10.1007/s00792-018-1017-8. (825-825).

Clokie, M.R.J., Shan, J., Bailey, S., Jia, Y., Krisch, H.M., West, S., Mann, N.H., 2006. Transcription of a "photosynthetic" T4-type phage during infection of a marine cyanobacterium. Environ. Microbiol. 8, 827–835. https://doi.org/10.1111/j.1462-2920.2005.00969.x.

Corinaldesi, C., Tangherlini, M., Dell'Anno, A., 2017. From virus isolation to metagenome generation for investigating viral diversity in deep-sea sediments. Sci. Rep. 7, 8355. https://doi.org/10.1038/s41598-017-08783-4.

Dávila-Ramos, S., Estradas-Romero, A., Prol-Ledesma, R.M., Juárez-López, K., 2014. Bacterial populations (first record) at two shallow hydrothermal vents of the Mexican Pacific West coast. Geomicrobiol J. 32, 657–665. https://doi.org/10.1080/01490451.2014.980526.

Dick, G.J., Anantharaman, K., Baker, B.J., Li, M., Reed, D.C., Sheik, C.S., 2013. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. Front. Microbiol. 4, 124. https://doi.org/10.3389/fmicb.2013.00124.

Ding, J., Zhang, Y., Wang, H., Jian, H., Leng, H., Xiao, X., 2017. Microbial community structure of deep-sea hydrothermal vents on the ultraslow spreading southwest Indian ridge. Front. Microbiol. 8, 1012. https://doi.org/10.3389/fmicb.2017.01012.

Eecke, H.C., Butterfield, D.A., Huber, J.A., Lilley, M.D., Olson, E.J., Roe, K.K., Evans, L.J., Merkel, A.Y., Cantin, H.B., Holden, J.F., 2012. Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. Proc. Natl. Acad. Sci. 109, 13674–13679. https://doi.org/10.1073/pnas.1206632109.

Fortunato, C.S., Larson, B., Butterfield, D.A., Huber, J.A., 2018. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ. Microbiol. 20 (2), 769–784.

Frock, A.D., Kelly, R.M., 2012. Extreme thermophiles: moving beyond single-enzyme catalysis. Curr. Opin. Chem. Eng. 1, 1–10. https://doi.org/10.1016/j.coche.2012.07.003.

Garin-Fernandez, A., Pereira-Flores, E., Glöckner, F., Wichels, A., 2018. The North Sea goes viral: occurrence and distribution of North Sea bacteriophages. Mar. Genomics 41, 31–41. https://doi.org/10.1016/j.margen.2018.05.004.

Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G., Prieur, D., 2003. Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. Res. Microbiol. 154, 303–307. https://doi.org/10.1016/S0923-2508(03)00075-5.

Glass, E.M., Meyer, F., 2011. The Metagenomics RAST Server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. In: Handbook of Molecular Microbial Ecology I, pp. 325–331. https://doi.org/10.1002/9781118010518.ch37.

Goldsmith, D.B., Crosti, G., Dwivedi, B., Mcdaniel, L.D., Varsani, A., Suttle, C.A., Breitbart, M., 2011. Development of pho has a novel signature gene for assessing marine phage diversity. Appl. Environ. Microbiol. 77, 7730–7739. https://doi.org/10.1128/aem.05531-11.

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317. https://doi.org/10.1038/ismej.2009.72.

Gorlas, A., Koonin, E.V., Bienvenu, N., Prieur, D., Geslin, C., 2012. TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. Environ. Microbiol. 14, 503–516. https://doi.org/10.1111/j.1462-2920.2011.02662.x.

Graves, M.V., Bernadt, C.T., Cerny, R., Etten, J.L., 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. Virology 285, 332–345. https://doi.org/10.1006/viro.2001.0937.

Graziani, S., Xia, Y., Gurnon, J.R., Van Etten, J.L., Leduc, D., Skouloubris, S., Liebl, U., 2004. Functional analysis of FAD-dependent thymidylate synthase ThyX from *Paramecium bursaria Chlorella* virus-1. J. Biol. Chem. 279, 54340–54347. https://doi.org/10.1074/jbc.M409121200.

Han, N.S., Kim, T.J., Park, Y.C., Kim, J., Seo, J.H., 2012. Biotechnological production of human milk oligosaccharides. Biotechnol. Adv. 30, 1268–1278. https://doi.org/10.1016/j.biotechadv.2011.11.003.

Harmsen, H., Prieur, D., Jeanthon, C., 1997. Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment culture of thermophilic subpopulations. Appl. Environ. Microbiol. 63, 2876–2883.

Hayes, S., Mahony, J., Nauta, A., Sinderen, D.V., 2017. Metagenomic approaches to assess bacteriophages in various environmental niches. Viruses 9, 127. https://doi.org/10.3390/v9060127.

He, T., Li, H., Zhang, X., 2017. Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. MBio 8, e00893-17. https://doi.org/10.1128/mbio.00893-17.

Hewson, I., Barbosa, J.G., Brown, J.M., Donelan, R.P., Eaglesham, J.B., Eggleston, E.M.,

nucleotides respectively, are likely to contribute to enhancing the fitness of their hosts within the hydrothermal vents as previously proposed (Anderson et al., 2011b; Anderson et al., 2014). The evolutionary analysis suggests that these AMGs are highly conserved and evolve under purifying selection, and are thus maintained in their genome.

Author contributions

HGCS and SDR designed and performed the analysis. RP, ILR, WAGS participated in the analysis. HGCS, SDR, ADWD and RABG analyzed and interpreted the data. All authors contributed in the preparation of the manuscript. All authors read and approved the final manuscript.

Competing interest/Conflict of interest statement

Authors declare that the research was conducted in the absence of any commercial or financial relationships. Thus, any conflict of interest exists.

Acknowledgements

Authors grateful to the Unidad de Secuenciación Masiva y Bioinformática (Instituto de Biotecnología-UNAM) and Centro de Investigación en Dinámica Celular for the access to its computer cluster. HGCS has a scholarship from CONACyT (Mexico). Authors especially grateful to the researchers that obtained the metagenomic sequences used in this manuscript, which are totally public and are deposited in NCBI web site.

References

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K., Toner, B.M., Dick, G.J., 2014. Sulfur oxidation genes in diverse deep-sea viruses. Science 344, 757–760. https://doi.org/10.1126/science.1252229.

Anantharaman, K., Breier, J.A., Dick, G.J., 2015. Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J. 10, 225–239. https://doi.org/10.1038/ismej.2015.81.

Anda, V.D., Zapata-Penasco, I., Hernandez, A.C., Fruns, L.E., Moreira, B.C., Souza, V., 2017. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: Unraveling the sulfur cycle. Gigascience 6, 1–17. https://doi.org/10.1101/191288.

Anderson, R.E., Brazelton, W.J., Baross, J.A., 2011a. Is the genetic landscape of the deep subsurface biosphere affected by viruses? Front. Microbiol. 2, 219. https://doi.org/10.3389/fmicb.2011.00219.

Anderson, R.E., Brazelton, W.J., Baross, J.A., 2011b. Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. FEMS Microbiol. Ecol. 77, 120–133. https://doi.org/10.1111/j.1574-6941.2011.01090.x.

Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., ... Huber, J.A., 2017. Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. Nat. Commun. 8 (1), 1114.

Anderson, R.E., Sogin, M.L., Baross, J.A., 2014. Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics. PLoS One 9, e109696. https://doi.org/10.1371/journal.pone.0109696.

Aylward, F.O., Boeuf, D., Mende, D.R., Wood-Charlson, E.M., Vislova, A., Eppley, J.M., Delong, E.F., 2017. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. Proc. Natl. Acad. Sci. 114, 11446–11451. https://doi.org/10.1073/pnas.1714821114.

Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The New S Language. Wadsworth Brooks/Cole (S version.).

Beiko, R.G., Harlow, T.J., Ragan, M.A., 2005. Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 14332–14337. https://doi.org/10.1073/pnas.0504068102.

Biddle, J.F., White, J.R., Teske, A.P., House, C.H., 2011. Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. ISME J. 5, 1038. https://doi.org/10.1038/ismej.2010.199.

Bioinformatics, B., 2011. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Institute. Biology, Cambridge, UK. https://doi.org/10.1371/journal.pbio.0030144.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Breitbart, M., Thompson, L., Suttle, C., Sullivan, M., 2007. Exploring the vast diversity of marine viruses. Oceanography 20, 135–139. https://doi.org/10.5670/oceanog.2007.58.

Brum, J.R., Ignacio-Espinoza, J.C., Kim, E.H., Trubl, G., Jones, R.M., Roux, S., Sullivan, M.B., 2016. Illuminating structural proteins in viral "dark matter" with metaproteomics. Proc. Natl. Acad. Sci. 113, 2436–2441. https://doi.org/10.1073/pnas.1525139113.

Cary, S.C., Cottrell, M.T., Stein, J.L., Camacho, F., Desbruyeres, D., 1997. Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *Alvinella pompejana*. Appl. Environ. Microbiol. 63, 1124–1130.

Cerqueira, T., Barroso, C., Froufe, H., Egas, C., Bettencourt, R., 2018. Metagenomic signatures of microbial communities in deep-sea hydrothermal sediments of Azores vent fields. Microb. Ecol. 76, 1–17. https://doi.org/10.1007/s00248-018-1144-x.

Christakis, C.A., Polymenakou, P.N., Mandalakis, M., Nomikou, P., Kristoffersen, J.B., Lampridou, D., ... Magoulas, A., 2018. Correction to: microbial community differentiation between active and inactive sulfide chimneys of the Kolumbo submarine volcano, Hellenic Volcanic Arc. Extremophiles 22 (5). https://doi.org/10.1007/s00792-018-1017-8. (825-825).

Clokie, M.R.J., Shan, J., Bailey, S., Jia, Y., Krisch, H.M., West, S., Mann, N.H., 2006. Transcription of a "photosynthetic" T4-type phage during infection of a marine cyanobacterium. Environ. Microbiol. 8, 827–835. https://doi.org/10.1111/j.1462-2920.2005.00969.x.

Corinaldesi, C., Tangherlini, M., Dell'Anno, A., 2017. From virus isolation to metagenome generation for investigating viral diversity in deep-sea sediments. Sci. Rep. 7, 8355. https://doi.org/10.1038/s41598-017-08783-4.

Dávila-Ramos, S., Estradas-Romero, A., Prol-Ledesma, R.M., Juárez-López, K., 2014. Bacterial populations (first record) at two shallow hydrothermal vents of the Mexican Pacific West coast. Geomicrobiol J. 32, 657–665. https://doi.org/10.1080/01490451.2014.980526.

Dick, G.J., Anantharaman, K., Baker, B.J., Li, M., Reed, D.C., Sheik, C.S., 2013. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. Front. Microbiol. 4, 124. https://doi.org/10.3389/fmicb.2013.00124.

Ding, J., Zhang, Y., Wang, H., Jian, H., Leng, H., Xiao, X., 2017. Microbial community structure of deep-sea hydrothermal vents on the ultraslow spreading southwest Indian ridge. Front. Microbiol. 8, 1012. https://doi.org/10.3389/fmicb.2017.01012.

Eecke, H.C., Butterfield, D.A., Huber, J.A., Lilley, M.D., Olson, E.J., Roe, K.K., Evans, L.J., Merkel, A.Y., Cantin, H.B., Holden, J.F., 2012. Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. Proc. Natl. Acad. Sci. 109, 13674–13679. https://doi.org/10.1073/pnas.1206632109.

Fortunato, C.S., Larson, B., Butterfield, D.A., Huber, J.A., 2018. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ. Microbiol. 20 (2), 769–784.

Frock, A.D., Kelly, R.M., 2012. Extreme thermophiles: moving beyond single-enzyme catalysis. Curr. Opin. Chem. Eng. 1, 1–10. https://doi.org/10.1016/j.coche.2012.07.003.

Garin-Fernandez, A., Pereira-Flores, E., Glöckner, F., Wichels, A., 2018. The North Sea goes viral: occurrence and distribution of North Sea bacteriophages. Mar. Genomics 41, 31–41. https://doi.org/10.1016/j.margen.2018.05.004.

Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G., Prieur, D., 2003. Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. Res. Microbiol. 154, 303–307. https://doi.org/10.1016/S0923-2508(03)00075-5.

Glass, E.M., Meyer, F., 2011. The Metagenomics RAST Server: A public resource for the automatic phylogenetic and functional analysis of metagenomes. In: Handbook of Molecular Microbial Ecology I, pp. 325–331. https://doi.org/10.1002/9781118010518.ch37.

Goldsmith, D.B., Crosti, G., Dwivedi, B., Mcdaniel, L.D., Varsani, A., Suttle, C.A., Breitbart, M., 2011. Development of pho has a novel signature gene for assessing marine phage diversity. Appl. Environ. Microbiol. 77, 7730–7739. https://doi.org/10.1128/aem.05531-11.

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J. 3, 1314–1317. https://doi.org/10.1038/ismej.2009.72.

Gorlas, A., Koonin, E.V., Bienvenu, N., Prieur, D., Geslin, C., 2012. TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. Environ. Microbiol. 14, 503–516. https://doi.org/10.1111/j.1462-2920.2011.02662.x.

Graves, M.V., Bernadt, C.T., Cerny, R., Etten, J.L., 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. Virology 285, 332–345. https://doi.org/10.1006/viro.2001.0937.

Graziani, S., Xia, Y., Gurnon, J.R., Van Etten, J.L., Leduc, D., Skouloubris, S., Liebl, U., 2004. Functional analysis of FAD-dependent thymidylate synthase ThyX from *Paramecium bursaria Chlorella* virus-1. J. Biol. Chem. 279, 54340–54347. https://doi.org/10.1074/jbc.M409121200.

Han, N.S., Kim, T.J., Park, Y.C., Kim, J., Seo, J.H., 2012. Biotechnological production of human milk oligosaccharides. Biotechnol. Adv. 30, 1268–1278. https://doi.org/10.1016/j.biotechadv.2011.11.003.

Harmsen, H., Prieur, D., Jeanthon, C., 1997. Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment culture of thermophilic subpopulations. Appl. Environ. Microbiol. 63, 2876–2883.

Hayes, S., Mahony, J., Nauta, A., Sinderen, D.V., 2017. Metagenomic approaches to assess bacteriophages in various environmental niches. Viruses 9, 127. https://doi.org/10.3390/v9060127.

He, T., Li, H., Zhang, X., 2017. Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. MBio 8, e00893-17. https://doi.org/10.1128/mbio.00893-17.

Hewson, I., Barbosa, J.G., Brown, J.M., Donelan, R.P., Eaglesham, J.B., Eggleston, E.M.,

Labarre, B.A., 2012. Temporal dynamics and decay of putatively allochthonous and autochthonous viral genotypes in contrasting freshwater lakes. Appl. Environ. Microbiol. 78, AE-01705. https://doi.org/10.1128/aem.01705-12.
Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/bioinformatics/btq003.
Huber, J.A., Welch, D.B.M., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., Sogin, M.L., 2007. Microbial population structures in the deep marine biosphere. Science 318, 97–100. https://doi.org/10.1126/science.1146689.
Hurwitz, B.L., Sullivan, M.B., 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS One 8, e57355. https://doi.org/10.1371/journal.pone.0057355.
Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. Genome Res. 17, 377–386. https://doi.org/10.1101/gr.5969107.
Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Tappu, R., 2016. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput. Biol. 12, e1004957. https://doi.org/10.1371/journal.pcbi.1004957.
Jeanthon, C., 2000. Molecular ecology of hydrothermal vent microbial communities. Antonie Van Leeuwenhoek 77, 117–133. https://doi.org/10.1023/A:1002463825025.
Kato, S., Shibuya, T., Takaki, Y., Hirai, M., Nunoura, T., Suzuki, K., 2018. Genome-enabled metabolic reconstruction of dominant chemosynthetic colonizers in deep-sea massive sulfide deposits. Environ. Microbiol. 20, 862–867. https://doi.org/10.1111/1462-2920.14032.
Kelley, D.S., Karson, J.A., Früh-Green, G.L., Yoerger, D.R., Shank, T.M., Butterfield, D.A., Jakuba, M., 2005. A serpentinite-hosted ecosystem: the Lost City hydrothermal field. Science 307, 1428–1434. https://doi.org/10.1126/science.1102556.
Labonté, J.M., Suttle, C.A., 2013. Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. Front. Microbiol. 4, 404. https://doi.org/10.3389/fmicb.2013.00404.
Ledesma, R.M., 2011. Nuevos ambientes submarinos: Ventilas hidrotermales. UNAM. México, pp. 54.
Leduc, D., Escartin, F., Nijhout, H.F., Reed, M.C., Liebl, U., Skouloubris, S., Myllykallio, H., 2007. Flavin-dependent thymidylate synthase ThyX Activity: implications for the folate cycle in bacteria. J. Bacteriol. 189, 8537–8545. https://doi.org/10.1128/jb.01380-07.
Leinonen, R., Sugawara, H., Shumway, M., 2010. The sequence read archive. Nucleic Acids Res. 39, D19–D21. https://doi.org/10.1093/nar/gkq1019.
Li, M., Baker, B.J., Anantharaman, K., Jain, S., Breier, J.A., Dick, G.J., 2015. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat. Commun. 6, 8933. https://doi.org/10.1038/ncomms9933.
Li, D., Luo, R., Liu, C., Leung, C., Ting, H., Sadakane, K., Lam, T., 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.
Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., Chisholm, S.W., 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc. Natl. Acad. Sci. 101, 11013–11018. https://doi.org/10.1073/pnas.0401526101.
Lossouarn, J., Dupont, S., Gorlas, A., Mercier, C., Bienvenu, N., Marguet, E., Geslin, C., 2015. An abyssal mobilome: viruses, plasmids and vesicles from deep-sea hydrothermal vents. Res. Microbiol. 166, 742–752. https://doi.org/10.1016/j.resmic.2015.04.001.
Manini, E., Luna, G.M., Corinaldesi, C., Zeppilli, D., Bortoluzzi, G., Caramanna, G., Raffa, F., Danovaro, R., 2007. Prokaryote diversity and virus abundance in shallow hydrothermal vents of the Mediterranean Sea (Panarea Island) and the Pacific Ocean (North Sulawesi-Indonesia). Microb. Ecol. 55, 626–639. https://doi.org/10.1007/s00248-007-9306-2.
Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., Vanderplasschen, A., 2004. Glycosyltransferases encoded by viruses. J. Gen. Virol. 85, 2741–2754. https://doi.org/10.1099/vir.0.80320-0.
Martin, W., Baross, J., Kelley, D., Russell, M.J., 2008. Hydrothermal vents and the origin of life. Nat. Rev. Microbiol. 6, 805–814. https://doi.org/10.1038/nrmicro1991.
McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Richie, K.B., Paul, J.H., 2010. High frequency of horizontal gene transfer in the oceans. Science 330, 50. https://doi.org/10.1126/science.1192243.
Meier, D.V., Bach, W., Girguis, P.R., Gruber-Vodicka, H.R., Reeves, E.P., Richter, M., Vidoudez, C., Amann, R., Meyerdierks, A., 2016. Heterotrophic Proteobacteria in the vicinity of diffuse hydrothermal venting. Environ. Microbiol. 18, 4348–4368. https://doi.org/10.1111/1462-2920.13304.
Millard, A.D., Hands-Portman, I., Zwirglmaier, K., 2014. Morphotypes of virus-like particles in two hydrothermal vent fields on the East Scotia Ridge, Antarctica. Bacteriophage 4, e20732. https://doi.org/10.4161/bact.28732.
Mizuno, C.M., Ghai, R., Saghaï, A., López-García, P., Rodriguez-Valera, F., 2016. Genomes of abundant and widespread viruses from the deep ocean. MBio. 7, e00805–e00816. https://doi.org/10.1128/mBio.00805-16.
Munson-McGee, J.H., Snyder, J.C., Young, M.J., 2018. Archaeal viruses from high-temperature environments. Genes 9, 128. https://doi.org/10.3390/genes9030128.
Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., Pond, S.L., 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8, 1002762. https://doi.org/10.1371/journal.pgen.1002764.
Ortmann, A.C., Suttle, C.A., 2005. High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. Deep-Sea Res. I Oceanogr. Res. Pap. 52, 1515–1527. https://doi.org/10.1016/j.dsr.2005.04.002.
Pjevac, P., Meier, D.V., Markert, S., Hentschker, C., Schweder, T., Becher, D., Gruber-Vodicka, H.R., Richter, M., Bach, W., Amann, R., Meyerdierks, A., 2018. Metaproteogenomic profiling of microbial communities colonizing actively venting hydrothermal chimneys. Front. Microbiol. 9, 680. https://doi.org/10.3389/fmicb.2018.00680.
Poli, A., Finore, I., Romano, I., Gioiello, A., Lama, L., Nicolaus, B., 2017. Microbial diversity in extreme marine habitats and their biomolecules. Microorganisms 5, 25. https://doi.org/10.3390/microorganisms5020025.
Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533. https://doi.org/10.1093/bioinformatics/bti320.
Pond, S.L., Scheffler, K., Gravenor, M.B., Poon, A.F., Frost, S.D., 2009. Evolutionary fingerprinting of genes. Mol. Biol. Evol. 27, 520–536. https://doi.org/10.1093/molbev/msp260.
Prangishvili, D., 2003. Evolutionary insights from studies on viruses of hyperthermophilic archaea. Res. Microbiol. 154, 289–294. https://doi.org/10.1016/s0923-2508(03)00073-1.
Rastelli, E., Corinaldesi, C., Dell'Anno, A., Tangherlini, M., Martorelli, E., Ingrassia, M., Chiocci, F., Lo Martire, M., Danovaro, R., 2017. High potential for temperate viruses to drive carbon cycling in chemoautotrophy-dominated shallow-water hydrothermal vents. Environ. Microbiol. 19, 4432–4446. https://doi.org/10.1111/1462-2920.13890.
Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T., Young, M.J., 2001. Viruses from extreme thermal environments. Proc. Natl. Acad. Sci. 98, 13341–13345. https://doi.org/10.1073/pnas.231170198.
Rohwer, F., Prangishvili, D., Lindell, D., 2009. Roles of viruses in the environment. Environ. Microbiol. 11, 2771–2774. https://doi.org/10.1111/j.1462-2920.2009.02101.x.
Romancer, M.L., Gaillard, M., Geslin, C., Prieur, D., 2006. Viruses in extreme environments. Life Extreme Environ. 99–113. https://doi.org/10.1007/978-1-4020-6285-8_6.
Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., Colombet, J., Sime-Ngando, T., Debroas, D., 2012a. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One 7, e33641. https://doi.org/10.1371/journal.pone.0033641.
Roux, S., Krupovic, M., Poulet, A., Debroas, D., Enault, F., 2012b. Evolution and diversity of the microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One 7, e40418. https://doi.org/10.1371/journal.pone.0040418.
Roux, S., Tournayre, J., Mahul, A., Debroas, D., Enault, F., 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinform. 15, 76. https://doi.org/10.1186/1471-2105-15-76.
Sakowski, E.G., Munsell, E.V., Hyatt, M., Kress, W., Williamson, S.J., Nasko, D.J., Polson, S.W., Wommack, K.E., 2014. Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. Proc. Natl. Acad. Sci. 111, 15786–15791. https://doi.org/10.1073/pnas.1401322111.
Sepulveda, B.P., Redgwell, T., Rihtman, B., Pitt, F., Scanlan, D.J., Millard, A., 2016. Marine phage genomics: the tip of the iceberg. FEMS Microbiol. Lett. 363, fnw158. https://doi.org/10.1093/femsle/fnw158.
Sheik, C.S., Anantharaman, K., Breier, J.A., Sylvan, J.B., Edwards, K.J., Dick, G.J., 2015. Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back-arc basin. ISME J. 9, 1434–1445. https://doi.org/10.1038/ismej.2014.228.
Sime-Ngando, T., 2014. Environmental bacteriophages: viruses of microbes in aquatic ecosystems. Front. Microbiol. 5, 355. https://doi.org/10.3389/fmicb.2014.00355.
Stern, A., Mayrose, I., Penn, O., Shaul, S., Gophna, U., Pupko, T., 2010. An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. Syst. Biol. 59, 212–225. https://doi.org/10.1093/sysbio/syp104.
Strazzulli, A., Fusco, S., Cobucci-Ponzano, B., Moracci, M., Contursi, P., 2017. Metagenomics of microbial and viral life in terrestrial geothermal environments. Rev. Environ. Sci. Biotechnol. 16, 425–454. https://doi.org/10.1007/s11157-017-9435-0.
Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., Chisholm, S.W., 2005. Three *Prochlorococcus Cyanophage* genomes: signature features and ecological interpretations. PLoS Biol. 3, e144. https://doi.org/10.1371/journal.pbio.0030144.
Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P.R., DeFrancesco, A.S., Kern, S.E., Thompson, L.R., Young, S., Yandava, C., 2010. Genomic analysis of oceanic cyanobacterias myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ. Microbiol. 12, 3035–3056. https://doi.org/10.1111/j.1462-2920.2010.02280.x.
Tarasov, V., Gebruk, A., Mironov, A., Moskalev, L., 2005. Deep-sea and shallow-water hydrothermal vent communities: two different phenomena? Chem. Geol. 224, 5–39. https://doi.org/10.1016/j.chemgeo.2005.07.021.
Team, R.C., 2013. R: A Language and Environment for Statistical Computing.
Tetz, G., Tetz, V., 2018. Bacteriophages as new human viral pathogens. Microorganisms 6, 54. https://doi.org/10.3390/microorganisms6020054.
Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc. Natl. Acad. Sci. U. S. A. 108, E757–E764. https://doi.org/10.1073/pnas.1102164108.
Weitz, J., Wilhelm, S., 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol. Rep. 4, 17. https://doi.org/10.3410/b4-17.
Williamson, S.J., Cary, S.C., Williamson, K.E., Helton, R.R., Bench, S.R., Winget, D., Wommack, K.E., 2008. Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. ISME J. 2, 1112. https://doi.org/10.1038/ismej.2008.73.
Xie, W., Wang, F., Guo, L., Chen, Z., Sievert, S.M., Meng, J., Huang, G., Li, Y., Yan, Q., Wu, S., Wang, X., Chen, S., He, G., Xiao, X., Xu, A., 2010. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with

contrasting chemistries. ISME J. 5, 414–426. https://doi.org/10.1038/ismej.2010.144.
Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., Takai, K., 2013. Metagenomic analysis of viral communities in (hado) pelagic sediments. PLoS One 8, e57271. https://doi.org/10.1371/journal.pone.0057271.
Yoshida, M., Mochizuki, T., Urayama, S., Yoshida-Takashima, Y., Nishi, S., Hirai, M., Nomaki, H., Takaki, Y., Nunoura, T., Takai, K., 2018. Quantitative viral community DNA analysis reveals the dominance of single-stranded dna viruses in Offshore Upper Bathyal sediment from Tohoku, Japan. Front. Microbiol. 9, 75. https://doi.org/10.3389/fmicb.2018.00075.
Zhang, W., Zhou, J., Liu, T., Yu, Y., Pan, Y., Yan, S., Wang, Y., 2015. Four novel algal virus genomes discovered from Yellowstone Lake metagenomes. Sci. Rep. 5, 15131. https://doi.org/10.1038/srep15131.
Zhang, L., Kang, M., Xu, J., Xu, J., Shuai, Y., Zhou, X., Ma, K., 2016a. Bacterial and archaeal communities in the deep-sea sediments of inactive hydrothermal vents in the Southwest India Ridge. Sci. Rep. 6, 25982. https://doi.org/10.1038/srep25982.
Zhang, W., Sun, J., Cao, H., Tian, R., Cai, L., Ding, W., Qian, P., 2016b. Post-translational modifications are enriched within protein functional groups important to bacterial adaptation within a deep-sea hydrothermal vent environment. Microbiome 4, 49. https://doi.org/10.1186/s40168-016-0194-x.

ARTICULO 2

frontiers in Microbiology

REVIEW
published: 18 October 2019
doi: 10.3389/fmicb.2019.02403

A Review on Viral Metagenomics in Extreme Environments

*Sonia Dávila-Ramos[1], Hugo G. Castelán-Sánchez[1], Liliana Martínez-Ávila[1], María del Rayo Sánchez-Carbente[2], Raúl Peralta[1], Armando Hernández-Mendoza[1], Alan D. W. Dobson[3,4], Ramón A. Gonzalez[1], Nina Pastor[1] and Ramón Alberto Batista-García[1]**

[1] Centro de Investigación en Dinámica Celular, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, [2] Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos, Cuernavaca, Mexico, [3] School of Microbiology, University College Cork, Cork, Ireland, [4] Environmental Research Institute, University College Cork, Cork, Ireland

OPEN ACCESS

Edited by:
Manuel Martínez García, University of Alicante, Spain

Reviewed by:
Alon Philosof, California Institute of Technology, United States
Mario López-Pérez, Universidad Miguel Hernández de Elche, Spain
Tiong Gim Aw, Tulane University, United States

***Correspondence:**
Ramón Alberto Batista-García
rabg@uaem.mx;
rbatista25@yahoo.com

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *14 May 2019*
Accepted: *04 October 2019*
Published: *18 October 2019*

Citation:
Dávila-Ramos S, Castelán-Sánchez HG, Martínez-Ávila L, Sánchez-Carbente MdR, Peralta R, Hernández-Mendoza A, Dobson ADW, Gonzalez RA, Pastor N and Batista-García RA (2019) A Review on Viral Metagenomics in Extreme Environments. Front. Microbiol. 10:2403. doi: 10.3389/fmicb.2019.02403

Viruses are the most abundant biological entities in the biosphere, and have the ability to infect Bacteria, Archaea, and Eukaryotes. The virome is estimated to be at least ten times more abundant than the microbiome with 10^7 viruses per milliliter and 10^9 viral particles per gram in marine waters and sediments or soils, respectively. Viruses represent a largely unexplored genetic diversity, having an important role in the genomic plasticity of their hosts. Moreover, they also play a significant role in the dynamics of microbial populations. In recent years, metagenomic approaches have gained increasing popularity in the study of environmental viromes, offering the possibility of extending our knowledge related to both virus diversity and their functional characterization. Extreme environments represent an interesting source of both microbiota and their virome due to their particular physicochemical conditions, such as very high or very low temperatures and >1 atm hydrostatic pressures, among others. Despite the fact that some progress has been made in our understanding of the ecology of the microbiota in these habitats, few metagenomic studies have described the viromes present in extreme ecosystems. Thus, limited advances have been made in our understanding of the virus community structure in extremophilic ecosystems, as well as in their biotechnological potential. In this review, we critically analyze recent progress in metagenomic based approaches to explore the viromes in extreme environments and we discuss the potential for new discoveries, as well as methodological challenges and perspectives.

Keywords: metagenomic, virosphere, extreme environment, viral gene bioprospection, extremophile virome

INTRODUCTION

Viruses are the most abundant biological entities in the planet, from the world oceans to the most extreme environments found in the biosphere (Zhang et al., 2018; Graham et al., 2019). Historically, the study of viral communities has been carried out by co-culture of viruses and their cellular hosts (Tennant et al., 2018), and more recently by viral metagenomic-based approaches (Nooij et al., 2018; Graham et al., 2019).

The exploration of viral populations in extreme environments has uncovered considerable genetic complexity and diversity. The biological organisms that inhabit extreme environments are

termed extremophiles, and are found in all three domains of life (Merino et al., 2019). Like all other organisms, extremophiles serve as hosts for viral replication (Castelán-Sánchez et al., 2019). As viruses depend on a cellular host for replication, the interactions with their hosts affect microbial diversity, population interactions and dynamics, and even the genomes of these hosts (Le Romancer et al., 2006; Zhang et al., 2018; Castelán-Sánchez et al., 2019). In extreme environments their impact extends from influencing microbial evolution to playing an indirect but significant role in the earth's biogeochemical cycles (Weitz and Wilhelm, 2012; Munson-McGee et al., 2018). However, despite their relevance, little is currently known about their ubiquity and diversity in extreme ecosystems (Paez-Espino et al., 2016; Berliner et al., 2018).

Nowadays the study of viruses can be carried out using metagenomic-based strategies that do not depend on cell culture approaches (Rose et al., 2016; Nooij et al., 2018; Zhang et al., 2018; Graham et al., 2019). Metagenomics represent a unique opportunity to describe the composition of viral communities in extreme environments, as well as to analyze the viral genetic reservoirs to characterize novel proteins and bioactive compounds of potential biotechnological utility.

Metagenomic studies are providing new sequences that in many cases do not share homology with sequences deposited in the reference databases (Hayes et al., 2017; Cantalupo and Pipas, 2019; Kinsella et al., 2019). It is evident that metagenomics from extreme habitats could be a powerful method to drastically increase the number of virus reported to date. It is surprising that despite the recent rapid advances in high-throughput sequencing techniques, there are still quite a limited number of studies describing the viromes of extreme environments in the literature. Here, we critically analyze recent progress in metagenomic-based approaches to explore the viromes in extreme environments, as well as methodological challenges and perspectives.

PERSPECTIVES ON SAMPLING AND PROCESSING: METHODOLOGICAL CHALLENGES FOR VIRAL METAGENOMICS IN EXTREME ENVIRONMENTS

Viral metagenomic studies are dependent on the ability to obtain sufficient amounts of nucleic acids from complex mixtures (Roux et al., 2019), particularly in extreme environmental samples that are as diverse as hot springs (Schoenfeld et al., 2008; Zablocki et al., 2017b), deep seawater, marine-sediments and oceanic basement (Breitbart et al., 2004; Hurwitz et al., 2013; Nigro et al., 2017), Antarctic and desert soils (Zablocki et al., 2014, 2017a), among others, to facilitate either the construction of metagenomic libraries or to perform direct sequencing.

The number of viral particles estimated to be present in a liter of water or kilogram of soil is in the order of 10^9–10^{11}, while the world's oceans are estimated to contain up to 10^6 viral particles per ml (Hara et al., 1991; Mokili et al., 2012). In contrast, the viral abundance from Octopus hot spring water from Yellowstone National Park or oceanic basement samples are at the lower range of $\sim 10^4$ and $\sim 10^5$ viral particles per ml respectively, compared with non-extreme aquatic environments (Schoenfeld et al., 2008; Nigro et al., 2017). In spite of viral particle abundance, during purification only subnanograms of viral DNA or RNA are typically recovered (Van Etten et al., 2010). Considering that a phage contains $\sim 10^{-17}$ g of DNA per particle, obtaining the amount of 1–5 μg of DNA required for standard pyrosequencing would implicate that $\sim 3 \times 10^{11}$ viral particles should be recovered (Thurber et al., 2009), and even for third generation PacBio technology 10 μg of DNA is required (Faino et al., 2015). With the development of new sequencing technologies it is likely that lower amounts of nucleic acids will be needed. For example, only 50 ng of DNA was required to sequence bacteriophages and archaeal viruses from hypersaline environments (Motlagh et al., 2017; Liu et al., 2019) and surprisingly only 1 ng of DNA was needed to explore the metavirome from deep sea and the chaotropic salt lake Salar de Uyuni using KAPA Hyper Prep Kit and Nextera XT kits with Illumina platforms (Hirai et al., 2017; Ramos-Barbero et al., 2019).

Filtration and Concentration of Viral Particles

To overcome the problem of obtaining sufficient viral nucleic acid amounts in extreme habitats, viral-particles must be concentrated by ultracentrifugation, flocculation, or filtration while minimizing contamination from prokaryotic or eukaryotic nucleic acids (Mokili et al., 2012; Liu et al., 2019; Ramos-Barbero et al., 2019; Roux et al., 2019). The use of classical size-selective ultrafiltration methods is not widely used, as the filters can often become blocked by impurities during concentration of the samples. Instead Tangential Flow Filtration (TFF) and/or ultracentrifugation were preferentially used in samples from hot springs and chaotropic salt lake Salar de Uyuni (Diemer and Stedman, 2012; Zablocki et al., 2017b; Ramos-Barbero et al., 2019). An excellent review by Lawrence and Steward (2010) on centrifugation highlights the efficiency of the methodology to sediment even the smallest viruses, where centrifugal separations can be divided in differential pelleting and zonal separations. The former has been successfully used to remove cell debris from samples obtained from enrichment cultures of archaeal viruses from an acidic hot spring Umi Jigoku in Beppu (Japan) (Liu et al., 2019) before concentrating and purifying viral particles and the latter has been tested using different gradient materials such as glycerol, OptiPrep and sucrose to isolate virus from a boreal lake in Finland (Laanto et al., 2017).

Due to the limitation of the volume size of aquatic samples, John et al. (2011) introduced the use of $FeCl_3$ flocculation to concentrate viruses from seawater that results in the recovery of 92–95% of viruses, which has been favorably used in samples from glacier waters and deep-sea (Bellas et al., 2015; Poulos et al., 2018). This compares favorably with traditional centrifugation or TFF, which results in recovery levels of 18–26% and 62–93% respectively (Furtak et al., 2016), as evaluated by SYBR Gold

staining, meaning that the use of $FeCl_3$ increases the efficiency of viral particle recovery from extreme environment samples to around 30–60%.

A variation of the ultracentrifugation technique uses certain compounds to precipitate viruses like PolyEthylene Glycol (PEG) and ethanol, followed by purification through CsCl gradient ultracentrifugation (Fancello et al., 2013; Kleiner et al., 2015; Rastrojo and Alcamí, 2017; Chatterjee et al., 2019). These strategies have been implemented during viral recovery from Artic and Antarctic polar habitats, hot spring lakes and hypersaline lakes (Diemer and Stedman, 2012; Motlagh et al., 2017; Rastrojo and Alcamí, 2017; Mizuno et al., 2019).

Filters in the range of 0.1–0.22 μm have been used to enrich samples from South African hot springs and igneous crust of the seafloor (Nigro et al., 2017; Zablocki et al., 2017b). However, in recent years giant viruses –*giruses*- (particle size of ∼720 nm) have been discovered (Thurber et al., 2009; Van Etten et al., 2010) leading some groups to use 0.45 μm filters that are effective in recovering these larger viral-particles (Van Etten et al., 2010; Hurwitz et al., 2013; Sangwan et al., 2015). Up to now, limited knowledge about giant viruses in extreme niches has been produced. For example, a new large DNA virus named Medusavirus, was isolated from hot spring water in Japan using a filter of 1.2 μm (Yoshikawa et al., 2019). In addition, 64 members of the *Mimiviridae* family were recently identified in Antarctic marine water (Andrade et al., 2018). Thus, selective filtration strategies should be considered to recover extreme giant viruses.

Despite the use of a variety of different approaches to enrich viral particles from extreme environments, systematic studies comparing these different concentration methods (TFF, FeCl3, PEG, commercial concentrators) are still lacking and likely the methods employed for viral enrichment may need to be adapted considering the nature of the sample.

Nuclease Treatment, Concentration and Viral Nucleic Acid Purification

Viral samples are usually treated with DNAse I, to avoid contamination with cellular genomic DNA that would, following sequence-based analysis, result in a large number of spurious DNA sequences from sources other than the virome. This treatment was used to obtain viral metagenomic DNA from Boiling Springs Lake (United States) and Great Salt Lake (thermophile and hypersaline ecosystems, respectively) (United States) (Diemer and Stedman, 2012; Motlagh et al., 2017). However, there are examples, such as in deep-sea ocean viral metagenomes (Hurwitz et al., 2013), desert perennial ponds (Fancello et al., 2013), hot springs (Zablocki et al., 2017b) among others, where despite the use of DNase treatment prior to viral genome purification, it was not possible to eliminate completely the cellular genome.

After DNase treatment, concentration steps are recommended using CsCl, sucrose or Cs_2SO_4 gradients by ultracentrifugation (Thurber et al., 2009; Fancello et al., 2013; Bellas et al., 2015). Additionally, once the viral particles have been concentrated and purified, the capsids have to be broken to release the viral genomes. The classical method is the use of formamide (Breitbart et al., 2002; Thurber et al., 2009; Fancello et al., 2013) followed by phenol:chloroform:isoamylic alcohol extraction (Diemer and Stedman, 2012; Nigro et al., 2017) or alternatively, through thermal shock (Bellas et al., 2015; Roux et al., 2016; Motlagh et al., 2017). However, in samples from hypersaline ponds, thermal shock may not be fully efficient to denude enveloped viral DNA, which may be the reason why in some studies the majority of viral DNA has been recovered from non-enveloped tailed viruses (Roux et al., 2016). Some single-stranded DNA (ssDNA) extreme viruses (e.g., HaloRubrum Pleomorphic ssDNA Virus 1, Haloarcula Hispanica Pleomorphic Virus 3, Aeropyrum Coil-shaped Virus) infecting hyperhalophile or hyperthermophile archaeal hosts, present a lipid envelope and multiprotein complexes or two criss-crossed halves of a circular nucleoprotein (Pietilä et al., 2010; Mochizuki et al., 2012; Demina et al., 2016) that could confer resistance to capsid disassembly. Thus, the capsid composition of extremophile viruses is a relevant feature of unknown viruses, as well as to access their genetic material, and consequently limits the identification of unusual extreme morphotypes.

Retrotranscription or Amplification Steps

The identification of viral genomes to date has mainly focused on ssDNA or double-stranded DNA (dsDNA) viruses and only small RNA genomes of 5–10 kb have been assembled from extreme metaviromes. For example, RNA viruses infecting archaea were discovered from an acidic hot spring in Yellowstone (United States) (Bolduc et al., 2012; Wang et al., 2015), as well as from alkaline hot springs (Schoenfeld et al., 2008). In addition, RNA cyanophages have been recently reported from Porcelana hot spring in Chilean Patagonia (Guajardo-Leiva et al., 2018).

Andrews-Pfannkoch et al. (2010) have implemented the use of hydroxyapatite chromatography to efficiently fractionate dsDNA, ssDNA, dsRNA, and ssRNA genomes of known bacteriophages from samples of marine environments. This methodology has been employed to study ssDNA viruses from deep-sea sediments, alkaline siliceous hot springs and Artic shelf seafloor (Yoshida et al., 2013; Nguyen and Landfald, 2015; Schoenfeld et al., 2015), but to our knowledge it has not been applied to study RNA viruses from extreme ecosystems.

When working with RNA viruses, a retro-transcription step is required previous to library preparation, and if the efficiency of the nucleic acid recovery is low, amplification strategies are required. Among these, phi29 polymerase-based multiple displacement amplification and random PCR using modified versions of Sequence Independent Single-Primer Amplification (SISPA) have been useful in the virome amplification in samples from hot acidic lakes, hot springs and polar aquatic environments (Diemer and Stedman, 2012; Mead et al., 2017; Yau and Seth-Pasricha, 2019). When the viral genome material is RNA, a Random-Priming SISPA (RP-SISPA) method is frequently used (Miranda et al., 2016). This approach was successfully conducted in the isolation of RNA viruses from seawater (Steward et al., 2013) and Antarctic virioplankton (Miranda et al., 2016).

Other strategies for viral amplification are also used when extremophile metaviromes are studied. While the implementation of the Linker Amplified Shotgun Library

methodology (LASL) is suggested to amplify dsDNA, Multiple Displacement Amplification (MDA) is employed to enrich ssDNA preferentially (Fancello et al., 2013). LASL has been performed to analyze viral metegenomes from Yellowstone hot springs and Antarctic virioplankton (Schoenfeld and Mead, 2015; Miranda et al., 2016), while MDA has been used to describe the virome present in deep-sea samples from Antarctica (Gong et al., 2018).

Zablocki et al. (2016) have argued that although viral amplification is commonly used in metavirome studies, especially for samples collected from extreme habitats such as hyperarid desert soils, this step should be avoided because it prevents the determination of viral particle abundance and diversity, and may promote a biased amplification of certain virus groups.

Thus, it is clear that further comparative methodological studies using samples from extreme environments are required to evaluate if purification, concentration and amplification methods have any impact in the virome structure obtained from metagenome analysis.

DATABASE AND BIOINFORMATIC ANALYSIS: GENERAL REMARKS

Up to now viral sequence search is conducted essentially on the NCBI database GenBank or RefSeq, according to their viral sequence classification criteria. The RefSeq database excludes some categories of data such as those that incorporate too much information that cannot be processed readily, such as metagenomes or genomes that have significant mismatch or indel variation compared to other closely related genomes. In addition, not all sequences have a taxonomic classification in the International Committee on Taxonomy of Viruses (ICTV) (O'Leary et al., 2015). The number of viral sequences reported in GenBank, reached almost two million by December 2018, of which only 3,279 were registered as genomes in RefSeq and of these, only 1,800 have a classification at the species level in the ICTV (Kang and Kim, 2018). The classification of viruses in ICTV has been based on the characteristics that can be used to distinguish one virus from another, such as the genome composition, the capsid structure, the gene expression program during viral replication, host range and pathogenicity, among others. Comparisons of both pairwise sequence similarity and phylogenetic relationships have become the primary guidelines used to define virus taxa (Simmonds, 2015). However, without the incorporation of metagenomic data in both the RefSeq and the ICTV databases, the comparison of sequences and their allocation is limited (Simmonds et al., 2017). Alternative viral sequence similarity search strategies, such as VirSorter (Roux et al., 2015) and VirFinder (Ren et al., 2017) have been developed. The former is designed to search protein-coding genes and the latter works with k-mer composition, both attempting to identify viral sequences in prokaryotic genomes. Integration of such strategies should reduce the number of unidentified sequences and the comparisons of viromes should then help to formulate more robust theories about their biological roles within a given community, thereby increasing the possibility of gaining a fuller understanding of the viromes in any environment (Simmonds et al., 2017).

Virus databases have been developed, such as Virome with 73 projects from viromes or metagenomes currently containing data from 270 libraries (Wommack et al., 2012); EBI metagenomics, a virome dataset and pipelines for analysis of metagenomes (Hunter et al., 2014); Metavir2, which is a web server for analysis of environmental viromes (Roux et al., 2014); and more recently, IMG/VR v.2.0 (Paez-Espino et al., 2018) that includes >600 extreme environmental metaviromes; Gut Virome Database (GVD) with 648 viral or microbial metagenomes (Gregory et al., 2019); iVirus (based on vConTACT as the main classification tool) which contains a dataset from 1,866 samples and 73 ocean expedition projects (Bolduc et al., 2017) (**Figure 1**). Some of them include viral sequences obtained through strategies such as the construction of fosmid libraries (Mizuno et al., 2013), cellular fraction of metagenomes (López-Pérez et al., 2017) or single viral genomics (Martinez-Hernandez et al., 2017), which enrich the virome sequences further. However, none of the above databases is particularly dedicated to viromes from extreme environments. Despite the limitations described above a comparative analysis of the population structure of viromes in extreme environments was carried out here, using publicly accessible virus metagenomic libraries as an attempt to exemplify the results that may be obtained using available tools and information. The data deposited in MetaVir2 until 2016 were selected because its user-friendly interface, which allows access to raw data or contig metagenome samples that contain well-classified metadata. We could select the 17 studies in MetaVir2 database that contain 66 viral metagenomes collected from most representative extreme environments: deep-sea (24), oxygen minimum zones (OMZ) (4), arid habitats (9), saline niches (23), cold environments (3) and hyperthermophile regions (3). The bioinformatic pipeline used was common to all data, so the comparison between environments relied on the same criteria. MetaVir2 followed two strategies to search the contigs in each sample: BLAST search in the RefSeq Virus database with the best-hit selection, and search for k-mer composition using di, tri or tetra nucleotides comparisons (Willner et al., 2009).

VIROMES IN EXTREME ENVIRONMENTS

Comparison of Viromes Between Extreme Habitats

The relative abundances from these data were analyzed comparing the similarities between environments. Their metadata are summarized in **Supplementary Table S1**.

The structure of the viral population in all the metagenomes analyzed by Metavir2 was compared (**Figures 2**, **3**). The 10 most abundant families are represented in **Figure 2** and the rest in **Figure 3** to visualize the differences in abundance of each family. Some families of the order *Caudovirales* are ubiquitous and the most abundant were the *Siphoviridae*, *Myoviridae,* and *Podoviridae* as expected, since the viruses that belong to these families infect a wide range of bacterial hosts from more than 140 prokaryotic genera (Konstantinidis et al., 2009;

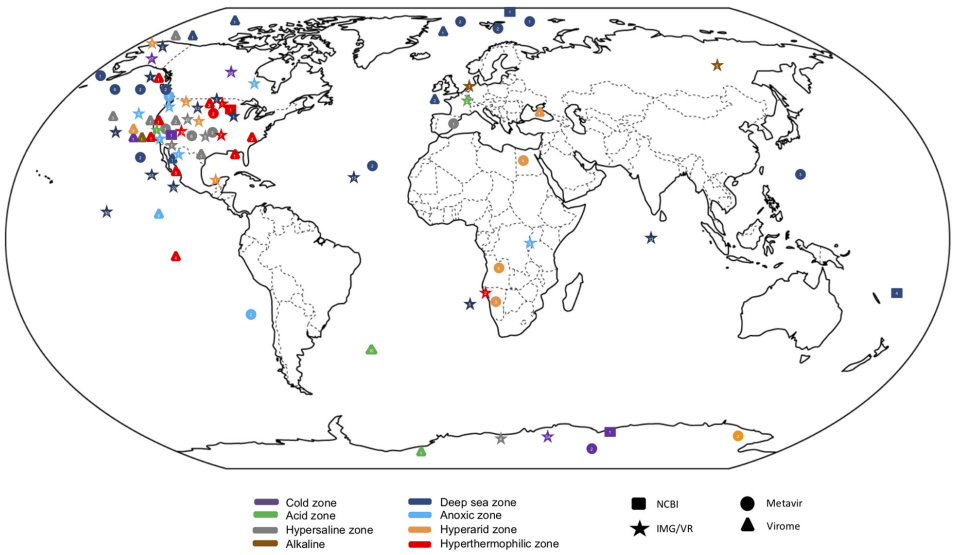


FIGURE 1 | Global distribution of metagenomics studies from extreme environments from public databases. Circles represent metagenomes deposited in Metavir database, stars in IMG/VR, triangles in Virome and squares in NCBI.

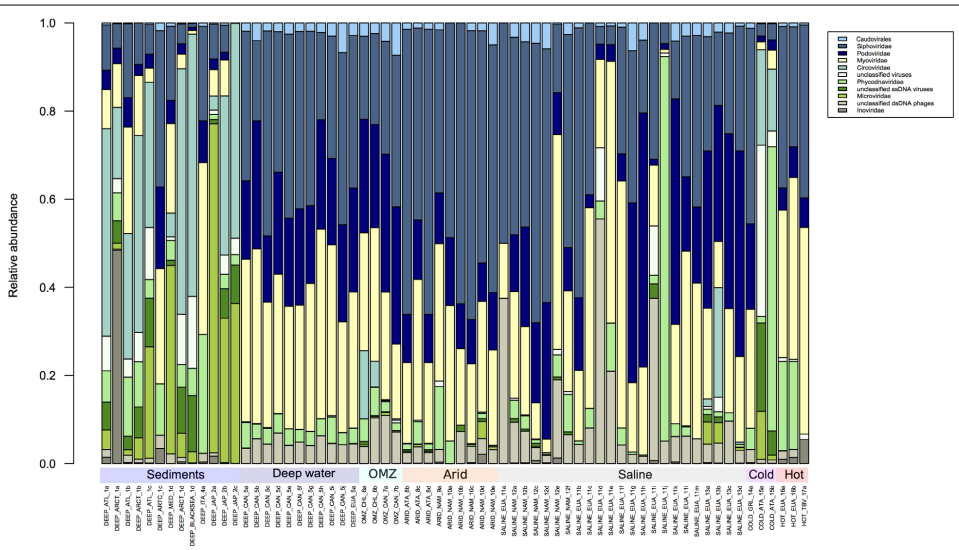


FIGURE 2 | Comparison of the 10 most abundant virus families according to the Metavir database. The taxonomic composition is expressed in relative abundance at the virus family level. The families *Siphoviridae*, *Podoviridae,* and *Myoviridae* are ubiquitous in extreme environments. The figure was constructed from an abundance matrix, using the number of sequences reported in the Metavir database, from which the relative abundance was obtained; using the R program.

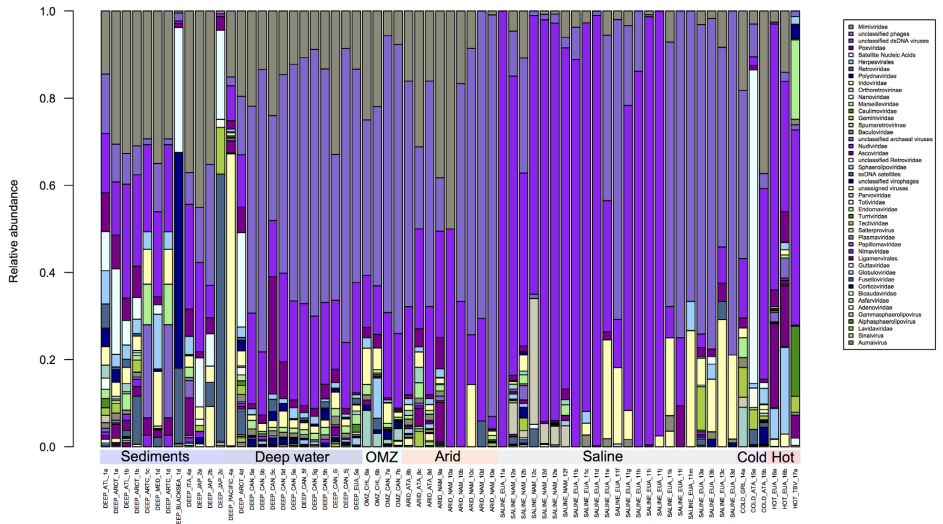


FIGURE 3 | Comparison of families with lower abundance in the Metavir database. Unclassified phages predominate in all environments, compared with other virus families. However, greater diversity is observed in the sediments, hyperthermophile and hypersaline environments compared to deep waters, Oxygen Minimum Zones (OMZ) or saline environments.

Hurwitz et al., 2013; Danovaro et al., 2016; Graham et al., 2019). It has been considered that the information from whole metagenomic analysis can give clues of potential model microorganisms to host virus replication, through the analysis of the Clusters of Regularly Interspaced Short Palindromic Repeats (CRISPR) loci from the cellular fraction of the metagenomes, that have been isolated from extreme environments (Gudbergsdóttir et al., 2016; Sharma et al., 2018; Liu et al., 2019; Martin-Cuadrado et al., 2019).

Next in relative abundance were *Circoviridae,* followed by *Phycodnaviridae, Microviridae,* and *Inoviridae* and high fractions of unidentified ssDNA and dsDNA viruses and phages. *Circoviridae* infect vertebrates and were particularly abundant in sediment samples (Dennis et al., 2018, 2019; Blanc-Mathieu et al., 2019). Members of *Phycodnaviridae* have been found at high levels in deep water samples (Mizuno et al., 2016; Gong et al., 2018; Blanc-Mathieu et al., 2019), which is curious given that they preferentially infect eukaryotic algae which require light to grow. It is possible therefore that these dsDNA viruses may infect other as yet unknown marine hosts in deep waters (Van Etten et al., 2010; Blanc-Mathieu et al., 2019). However, the predominance of the above families present obvious exceptions, such as in two samples from cold environments, a sample from a saline environment and most of the samples from deep marine sediments.

Within the metagenomes that correspond to deep sea environments (depths greater than 1,000 m), where the absence of light, oligotrophic conditions, low-oxygen concentrations, low temperatures and high hydrostatic pressure dominate (Le Romancer et al., 2006; Liang et al., 2019), two categories were considered based on the origin of the samples: sediments and deep water (**Figure 2**). It is clear that in metagenomes from sediments of the Atlantic, Arctic and Pacific Northwest, ssDNA viruses like *Circoviridae*, *Microviridae* and *Inoviridae* are more abundant than dsDNA viruses (**Figure 2**). This characteristic seems to be exclusive to samples from this environment. It should be noted that the two samples from cold environments show a similar composition to that of sediments, and all others include ssDNA viruses in low abundance (**Figure 2**). This is in agreement with a recent report by Yoshida and coworkers who reported that ssDNA viruses predominate in marine sediments and have an estimated abundance of 1×10^8 to 3×10^9 genome copies per cm^3 of sediment, clearly more abundant than dsDNA viruses which range from 3×10^6 to 5×10^6 genome copies per cm^3 (Yoshida et al., 2018).

In **Figure 3**, where the remaining viral families are shown, two general points can be highlighted: *Mimiviridae* are present in almost all environments, which is not surprising since some of their hosts are known polyextremophiles (Claverie et al., 2018; Yau and Seth-Pasricha, 2019). The second point is the abundance of unclassified sequences, which do not allow any conclusion to be made about the diversity observed by environment since these sequences could come from one or more than one family. A large part of the sequences obtained from different environments, except for sediments, have no similarity in the databases, an issue that should change with

the inclusion of additional metagenomic-derived sequences in databases (**Figure 3**). Overall analysis of the composition of viral families present in each extreme environment could at the very least allow a description of the families that are shared or that are exclusive to each environment.

Some environments are characterized by low-oxygen concentrations; these include those with high concentrations of greenhouse gases, which directly affect the biodiversity in those environments (Kiehl and Shields, 2005; Resplandy et al., 2018). There are three central oceanic regions which are considered to be Oxygen Minimum Zones (OMZ), namely the Eastern Tropical North Pacific (ETNP), the Eastern Tropical South Pacific (ETSP) and the Arabian Sea, within which the activity of anaerobic microorganisms is highly significant (Paulmier and Ruiz-Pino, 2009; Thamdrup, 2012). As expected, the viral population diversity closely reflects the microbial diversity in these environments (Cassman et al., 2012; Parvathi et al., 2018; Fuchsman et al., 2019), with the virome composition in OMZ being commonly composed of the *Myoviridae* and *Siphoviridae* families, followed by *Phycodnaviridae* (**Figure 2**).

OMZ were sampled at 200 m depths in Chile and Canada and virus composition was analyzed using the MDA (Genomiphi and GenomePlex) protocol. While the ssDNA *Circoviridae* family was predominantly observed in samples from Chile, this virus family was not observed in samples from the Canadian OMZ. In addition, in the samples from Canada (Chow et al., 2015) *Parvoviridae* (ssDNA) were highly abundant, but were totally absent in samples from Chile (**Figure 3**). In previous studies it was observed that the viral community along the vertical dissolved oxygen gradients was characterized by abundance taxa and diversity fluctuations. These differences could be related to changes in the viral replication strategy from lytic to lysogenic. It seems that oxygen reduction concurs with a decrease in viral abundance (Cassman et al., 2012; Parvathi et al., 2018). It should be noted that a large proportion of sequences obtained from these regions do not find similarity with other viruses in the databases, but those sequences could be from viruses that infect little known prokaryotic hosts, like ammonia-oxidizing archaea and anaerobic ammonia-oxidizing (anammox) bacteria which predominate in this environment (Parvathi et al., 2018).

Hyperarid environments exhibit conditions that are considered to be limiting for life, such as lack of water, high levels of UV radiation and extreme temperatures. However, both prokaryotic and eukaryotic organisms have adapted to live in these environments (Merino et al., 2019). Although low diversity might be expected in these environments, metagenomic studies performed with hypolithic communities have shown this not to be the case, with a high level of diversity being reported; particularly in bacterial communities from Antarctica (cold desert) and Namibia (desert), which are mainly *Actinobacteria*, *Proteobacteria,* and *Cyanobacteria* (Vikram et al., 2016). In the hypolithic viral communities from the Namibian desert and the Antarctic, metagenomic data has revealed the presence of *Caudovirales* which do not correlate with phages that infect *Cyanobacteria* species (Adriaenssens et al., 2015). The samples from the Antarctic hyperarid region displayed a greater diversity of unique viruses such as *Bicaudaviridae*, *Asfarviridae*, *Lavidaviridae*, *Tectiviridae,* and *Sphaerolipoviridae* when compared with the families found in the Namibian desert (**Figure 3**). Zablocki and coworkers have previously reported a higher viral diversity in the Arctic when compared with the Namibian desert, and it has been observed that Antarctic desert soils contain higher proportions of free extracellular virus-like particles compared to hot hyperarid desert soils, where a lysogenic lifestyle seems to prevail (Zablocki et al., 2016).

In **Figure 3** the variability in the composition of viral families in hypersaline habitats is evident. Such environments are widely distributed throughout the world and are present in salt lakes, salt flats and salt deposits. In these environments, the low water activity directly affects the composition of the microbial communities (Le Romancer et al., 2006; Ma et al., 2010; Merino et al., 2019). Viruses that have been identified in these ecosystems are haloviruses and a large number of these infect Archaea, Bacteria and Eukaryotes (Atanasova et al., 2018; Plominsky et al., 2018; Ramos-Barbero et al., 2019). About 64 archaeal viruses have been isolated from the two kingdoms, *Crenarchaeota* and *Euryarchaeota* (Porter et al., 2007). These samples are also those that have a greater abundance in unassigned or not classified viruses, which prevents determination of the real diversity of that group of archaea viruses, probably because they are the least studied and have low representation in the databases (Atanasova et al., 2018; Ramos-Barbero et al., 2019).

In addition, unclassified dsDNA viruses have also been observed (**Figure 3**), while haloviruses such as HGV-1, HTVAV-4 and HSTV-1 have also been identified at high levels. On the other hand, ssDNA viruses which mostly infect eukaryotes such as colpodellids, nematodes, arthropods, chlorophytes, among others, are present at low levels in hypersaline habitats (Feazel et al., 2008; Heidelberg et al., 2013).

The thermophile environments are characterized by high temperatures, where thermophilic microorganisms thrive at 65–80°C as their optimal growth temperature, and >80°C for hyperthermophiles (Merino et al., 2019). Viruses that infect bacteria and archaea are abundant in these hyperthermophilic habitats (Schoenfeld et al., 2008; Strazzulli et al., 2017; Liu et al., 2019). The virome of hyperthermophile environments is composed of viruses that infect all three domains of life, with members of the *Turriviridae, Fuselloviridae, Bicaudaviridae,* and *Globuloviridae* families that infect Archaea (Krupovic et al., 2018). Moreover, the *Nudiviridae, Phycodnaviridae,* and *Poxviridae* families that infect eukaryotes are also present (**Figure 3**).

Within the hyperthermophile metagenomes analyzed here, the presence of ssDNA or RNA viruses was not observed, but in other studies from these environments the presence of picornavirus-like, alphavirus-like, and flavivirus-like RNA viruses has been reported (Bolduc et al., 2012). It is possible that ssDNA and RNA viruses were not detected in the samples we analyzed due to differences in sample processing (**Figures 2**, **3**). Thus, as previously mentioned, if comparative viral metagenomic studies are to be undertaken to allow an accurate comparison between viromes from different ecosystems and to potentially identify novel viral clusters, then standardized methodologies will need to be developed and employed.

The polar regions of the Earth are dominated by the polar ice caps, with the microbial diversity present in these regions being much higher than might be expected. It is well established that viruses play an important role in controlling microbial mortality in these habitats (López-Bueno et al., 2009; Cárcer et al., 2015; Yau and Seth-Pasricha, 2019). While it has been reported that different lakes located in the Arctic and Antarctic share similar virome compositions, marked differences have been found.

Although at this taxonomic level it is possible to differentiate some of the particularities described above, in terms of the virus composition in each environment studied, very little information is revealed at the genus or species level that would allow a better understanding of the virus–host relationship and its influence in the environment.

Therefore, two environments, OMZ and deep-sediments, which at the family level have a very similar structure (**Figures 2**, **3**) were selected in an attempt to determine if it is possible to obtain biologically meaningful information on the differences or similarities in virus-host interactions at the genus level. The genus composition of two well-known families were analyzed: *Podoviridae* that infect bacteria and

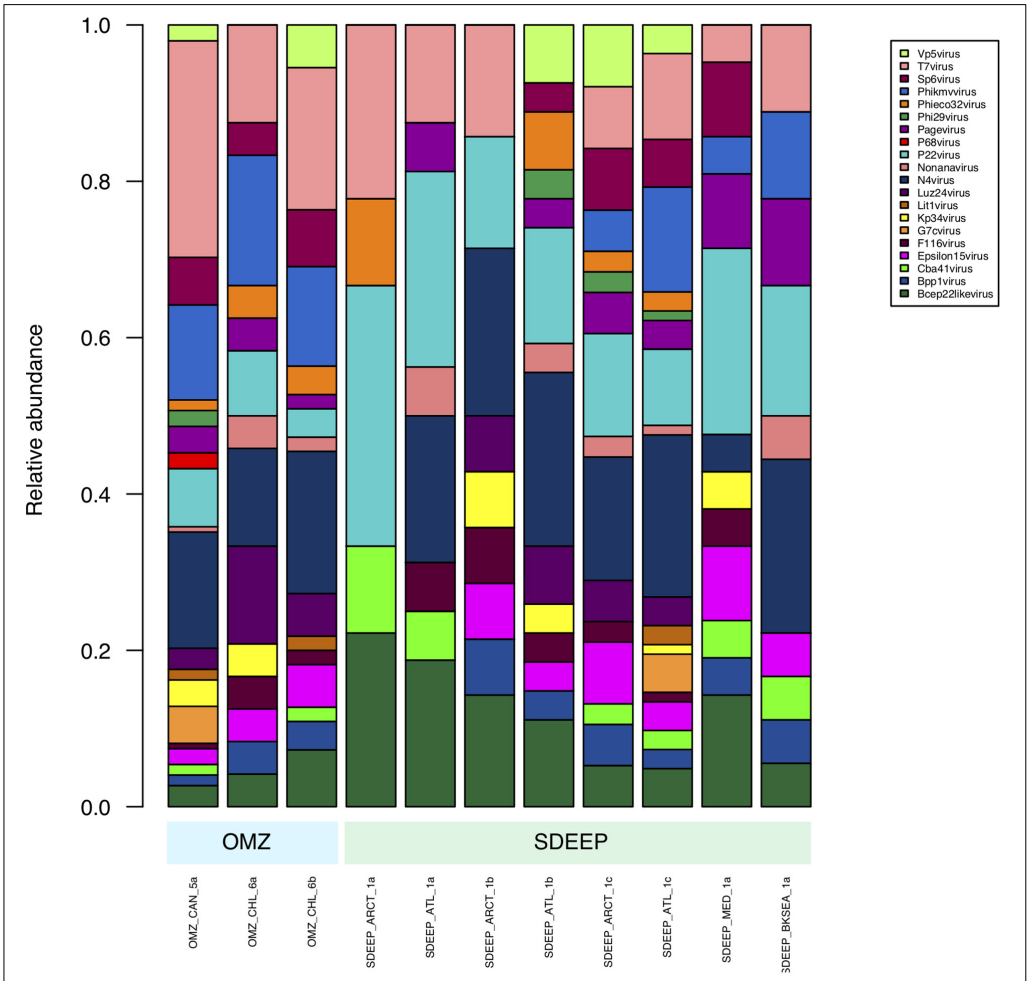


FIGURE 4 | Composition of *Podoviridae* family at genus level. Metagenomes from OMZ and deep-sediments were considered in the analysis.

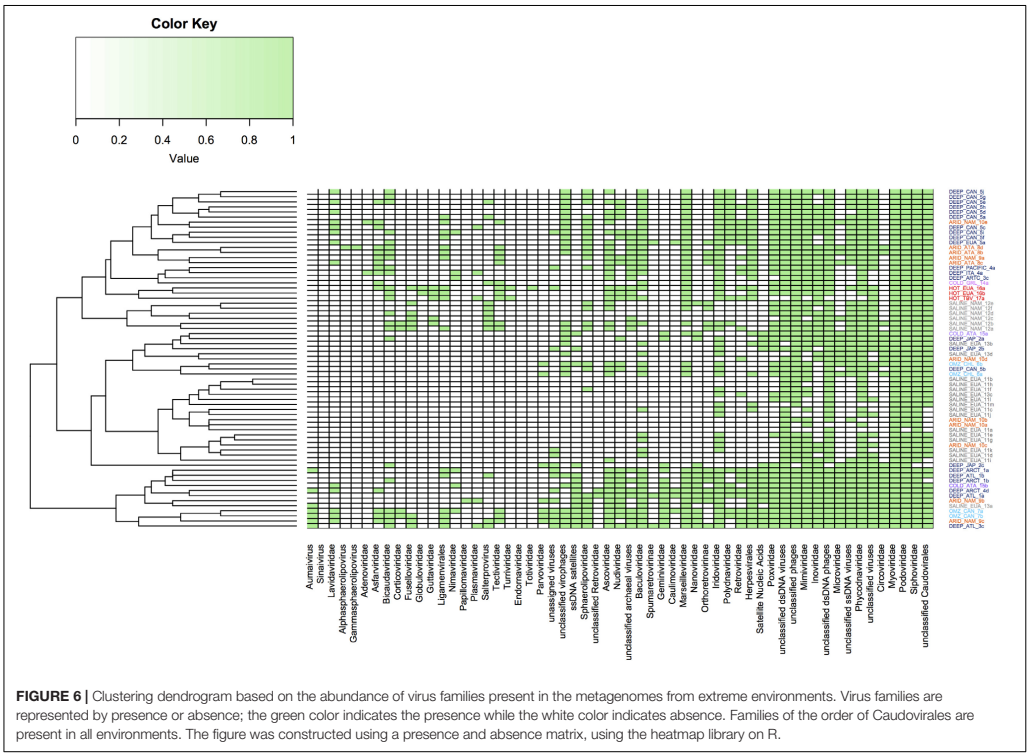


FIGURE 6 | Clustering dendrogram based on the abundance of virus families present in the metagenomes from extreme environments. Virus families are represented by presence or absence; the green color indicates the presence while the white color indicates absence. Families of the order of Caudovirales are present in all environments. The figure was constructed using a presence and absence matrix, using the heatmap library on R.

in these environments. This was particularly evident for some viromes obtained from hypersaline, deep-sea and hyperarid environments, while it was less evident in other extreme ecosystems which did not appear to show clustering, such as cold environments. However, this analysis again shows that some viral families are ubiquitous in all extreme environments, while the ssDNA viruses appear to predominate in sediments from deep-sea and cold environments.

In general, the virome structure from hypersaline samples reveals low levels of diversity, even in samples from different geographical areas (**Figure 6**). The high concentration of NaCl might limit viral diversity due to the shortage of prokaryotic hosts, since *Haloquadratum walsbyi*, *Salinibacter ruber,* and nanohaloarchaeas are the predominant organisms in these environments, with more than 90% of the contigs annotated to these taxa (Ventosa et al., 2015). Another factor that could determine the virome diversity that is observed in hypersaline environments is the dynamic switch between lytic and lysogenic replication cycles, since this represents a significant adaptation mechanism in environments with high salinity content (Roux et al., 2016).

This notwithstanding, from **Figure 6** it is clear that the virus family composition is quite similar in these environments, which could provide significant information related not only to viral evolution but also to physiological adaptation of microorganisms in response to high temperatures (Schoenfeld et al., 2008; Biddle et al., 2011).

Figure 7 shows the degree of overall similarity between the viral metagenomes in relation to the extreme environment from which the viromes were isolated. As previously described, some viral families belonging to the *Caudovirales* order are ubiquitous and display polyextremophilic adaptation. The hypersaline environments present a consistent clustering depending on the viral diversity, as well as the relative viral family abundance, which suggests that NaCl enriched environments provide strong constraints for the development of life that may restrict ecosystem diversity. Some viromes from hyperarid, deep-sea and saline environments are closely clustered (**Figure 7**) suggesting that the organisms and therefore the viral composition is partially shared, at least between these environments. Regarding deep-sea environments, **Figure 7** shows two clustered metagenome populations derived from the deep-sea, where those viromes

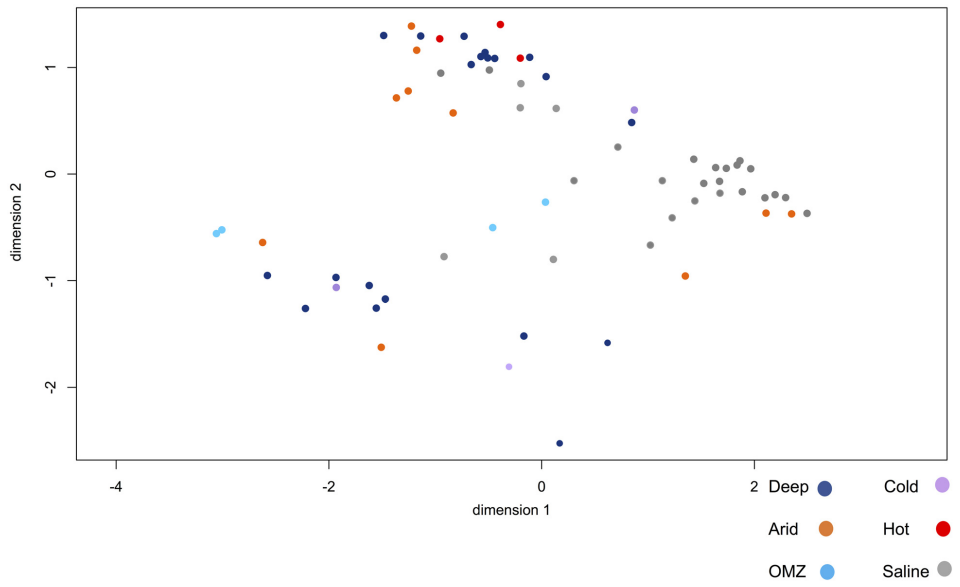


FIGURE 7 | Multidimensional scaling (NMDS) to visualize the degree of overall similarity between the metagenomes. Metagenomes from different habitats are shown: deep-sea (pink), OMZ (green), hyperarid (red), hypersaline (blue), psychrophile (orange), hyperthermophile (yellow). Hypersaline metagenomes have a strong clustering compared to other metagenomes. The figure was constructed from an abundance matrix, where the multidimensional scaling algorithm was applied to observe the general similarity of metagenomes, using the R software.

obtained from deep water are closely clustered, as well as those from sediments.

An interesting hypothesis to be investigated using metagenomics studies conducted in different geographical areas is the possibility of identifying specific viral clusters associated with a particular extreme environment. The large numbers of unclassified sequences in the databases is an important issue to consider with studies on viromes from extreme environments. The limitations of the bioinformatic pipelines to assign a taxonomic identity to a majority of the viral sequences, together with our limited understanding about viruses in extreme environments, has resulted in a lack of progress in our knowledge of extremophilic viromes. This has also negatively impacted our understanding in terms of evolution, gene horizontal transfer, ecology and virus-host interactions.

To meaningfully compare viromes from different environments it is necessary to at least partially answer the previous questions and provide new information about how viromes are potentially limited by extreme physicochemical characteristics, geographical area, or other artificial circumstances such as sampling methods, enrichment techniques and other technical biases. It should be possible to determine whether some viral populations could be closely related to a specific type of extreme ecosystem and consequently obtain more information about viral evolution (Simmonds, 2015).

FUNCTIONAL METAGENOMICS IN EXTREME ENVIRONMENTS: METHODOLOGICAL CHALLENGES, DISCOVERIES AND OPPORTUNITIES

Functional viral metagenomics focuses on exploring viral diversity to discover novel genes. Extreme environments harbor an enormous diversity of unknown viruses (Desnues et al., 2008; Dinsdale et al., 2008; Williamson et al., 2008; Rosario and Breitbart, 2011; Kristensen et al., 2012; Atanasova et al., 2016; Gudbergsdóttir et al., 2016; Nigro et al., 2017; Zablocki et al., 2017a,b; Sharma et al., 2018; Liu et al., 2019; Martin-Cuadrado et al., 2019; Mizuno et al., 2019; Roux et al., 2019) and, consequently, a potentially large number of unknown viral proteins. Functional viral metagenomics in these niches show a limited progress, with few reported recent advances (Schoenfeld et al., 2009; Schmitz et al., 2010; Moser et al., 2012; Heller et al., 2019) (**Figure 8**).

Viral-host systems (*in vitro* screening), sequence-based screening, activity-based screening (heterologous expression of

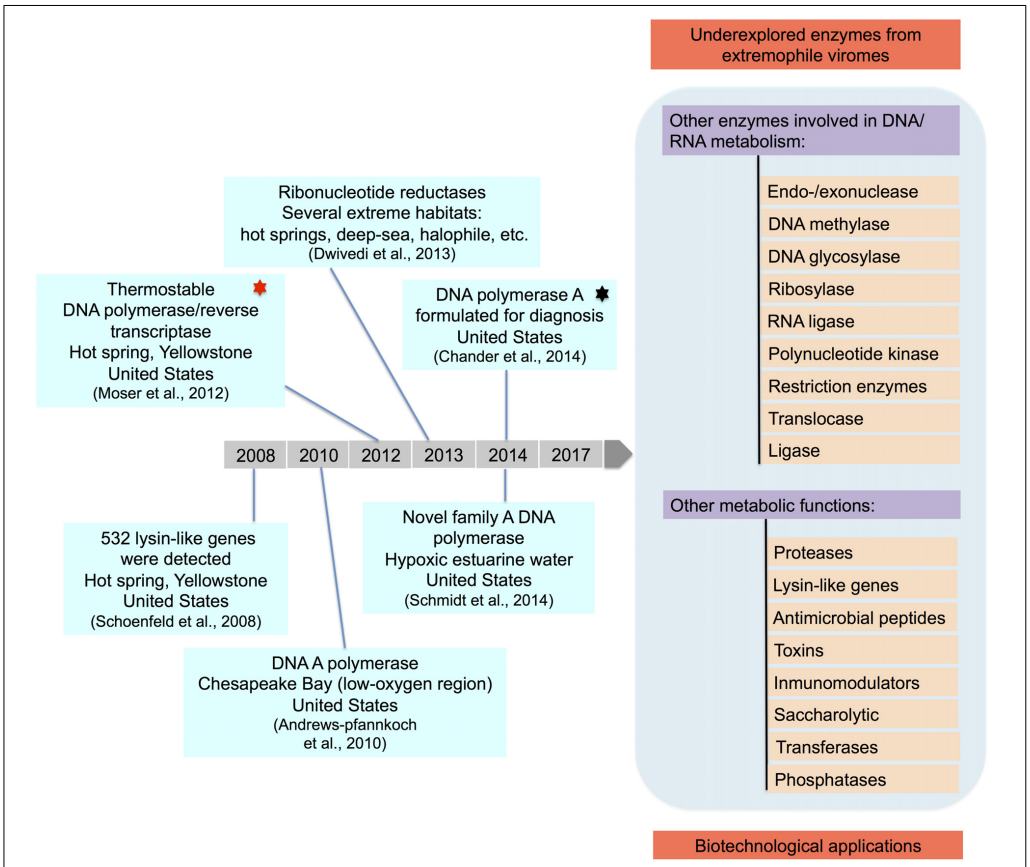


FIGURE 8 | Progress in function-based metagenomics. Functional bioprospections from extreme environments performed in the last years are summarized. Underexplored enzymes from extremophile viromes are listed.

viral proteins), and PCR- and hybridization-based screening, could be implemented for functional analysis from extreme viromes (Moser et al., 2012; Bzhalava and Dillner, 2013; Fancello et al., 2013; Heller et al., 2019). While sequence-based screenings have been responsible for the discovery of the majority of new viral enzymes (at least as annotated proteins) from extreme environments, PCR- and hybridization-driven methods have not been employed to date to functionally explore extremophile viromes.

In vitro screening from extremophile viromes is a challenge with respect to the co-cultivation of both hosts and viruses, and in particular in trying to mimic the conditions which are present in these habitats, thereby ensuring a better success rate regarding viral replication and viral protein expression. Thus, to help overcome this bottleneck, it will be important to develop new host systems (for prokaryotic and eukaryotic viruses) which grow under extreme pH, temperature, salinity, pressure and radiation (Schoenfeld et al., 2009).

Activity-based screening could be a very useful approach to identify novel enzymes. This method demands an efficient heterologous expression system for viral proteins. Thus, there are problems with the expression levels of many viral proteins in foreign host systems, particularly in genes isolated from extremophile viromes, which are dominated by rare genes, with issues such as codon usage together with promoter regulation/activation negatively impacting on enzyme production in different heterologous systems (Kristensen et al., 2012).

While the well-established *Escherichia coli* heterologous expression system is available, it is clear that additional systems

with a particular focus on extremophile bacteria and fungi will need to be developed to increase the chances of producing sufficient levels of viral extremoenzymes to allow their detection in function-based screens. These screens usually employ activity-based assays which involve colorimetric changes, typically following utilization of a substrate. However, these types of screenings are not particularly useful with viromes, since viral genes encoding enzymes involved in the metabolism of different substrates are quite rare. Due to this, there are no reports to date of the detection of viral enzymes from viromes through activity-based screening.

Despite the aforementioned disadvantages related to the heterologous expression of viral proteins, the activity-driven screenings allow functional gene annotation through an *in vitro* phenotypic-based test. This has an important advantage over sequencing-driven screening where a high number of genes are annotated as "*unknown function,*" because the gene repertoires of the extremophile viromes are currently undersampled.

A number of viral enzymes with utility in scientific, diagnosis and therapeutic applications have been identified using sequencing-based screens from extremophile viromes (Schoenfeld et al., 2008, 2009; Moser et al., 2012; Dwivedi et al., 2013; Schmidt et al., 2014; Mead et al., 2017). Also, the genomes of extremophile viruses are likely to be a source of novel antimicrobial peptides that may have applications in the biopharmaceutical and molecular diagnostics areas (Rice et al., 2001; Le Romancer et al., 2006; Schoenfeld et al., 2009).

For example, a sequencing-driven metagenomic study from two mildly alkaline hot springs in Yellowstone, allowed identification of 532 lysin-like genes (Schoenfeld et al., 2008). In recent years, these lytic enzymes have gained increasing importance due to their potential use in biomedical science applications (Schmitz et al., 2010); however, no lysin-like genes from extreme environments have to date been experimentally characterized.

DNA polymerases (502 sequences) have been also detected in extreme metaviromes, particularly from hypoxic estuarine waters obtained in the Gulf of Maine, Dry Tortugas National Park and the Chesapeake Bay (Andrews-Pfannkoch et al., 2010; Schmidt et al., 2014). These shotgun metagenomic studies revealed a novel DNA polymerase A family in marine virioplankton, since some sequences were distantly grouped in a phylogeny comprising DNA polymerase A from virus and bacteria (Schmidt et al., 2014).

Ribonucleotide reductases (RNR) have also been found from viromes obtained from hypersaline, psychrophile and thermophile niches (Dwivedi et al., 2013). For example, a bioinformatics analysis demonstrated that viruses isolated from hot springs contained a high abundance of RNR. However, some habitats such as hydrothermal vents from the East Pacific Rise, a solar saltern pond and salterns from Alicante (Spain) were found to have fewer ($\leq$5) identifiable RNR viral homologs (Dwivedi et al., 2013).

Recent efforts to characterize new viral DNA polymerases from extreme environments have resulted in the identification of a thermostable polymerase in a viral metagenomic DNA library from a near-boiling thermal pool in a hot spring in Yellowstone (Moser et al., 2012; Heller et al., 2019). This was the first report describing the isolation of a polymerase from a viral metagenomic library. In this study 59 complete polymerase clones were identified as possessing thermostable DNA polymerase activity following a functional screen. One of these polymerases, namely PyroPhage 3173 Pol, also has 5′-3′ exonuclease activity, as well an innate reverse transcriptase activity. It was subsequently tested in high fidelity reverse transcription PCR (RT-PCR) reactions and compared with some commercially available enzyme systems (Moser et al., 2012; Heller et al., 2019). The PyroPhage 3173 Pol-based RT-PCR enzyme was found to have a higher specificity and sensitivity that the other enzymes. While the PyroPhage 3173 DNA polymerase shares amino acid identity (∼32%) with another bacterial polymerase, no significant similarity was found with other viral proteins (Moser et al., 2012). This highlights the potential diversity of enzymes that may be present in extremophile viromes. The enzyme has subsequently been characterized and shown to be effective in the molecular detection of certain viral and bacterial pathogens by loop-mediated isothermal amplification (Chander et al., 2014).

STRUCTURAL BIOLOGY OF VIROMES

Specific molecular-level adaptations to extreme environments can only be appreciated once the detailed molecular structures are known. In order to explore the available structural information of proteins belonging to extremophile viruses, we carried out a manual search in the Protein Data Bank[1] (Berman et al., 2000) of all the viral families and genera identified in the metagenomes analyzed in this review, and kept only those whose hosts were either marine viruses or frank extremophiles. The resulting proteins were then classified according to their annotated function, and are discussed below.

In general, all these structures are valuable from a biochemical and biotechnological perspective, as they contain the molecular representation of the required adaptation to the particular extreme environment favored by the virus host. For example, viruses that infect *Acidianus* or *Sulfolobus* archaea are subject to the combination of high temperature and acidic pH; their proteins tend to have many charged residues, in particular, acidic ones (see structure 3DF6, an orphan protein). They also tend to have compact folds with structured termini, short loops with prolines in specific positions to stabilize them (see structure 2BBD, a major capsid protein), and the absence of cavities. Despite being DNA-binding proteins, and therefore cytoplasmic, some of them also include disulfide bridges, intramolecular (see structure 2VQC) or intermolecular (in structure 2CO5), that can impart up to 14 degrees in thermal stability for the protein. The formation of these disulfide bonds requires the existence of a sulfhydryl oxidase, either encoded by the host or by the virus itself. When a mesophilic homolog exists, a direct comparison of the structural features of the proteins can guide protein engineering to improve stability and/or function.

[1] rcsb.org

Also, as some viruses have space limitations in their capsids, resulting in compact genomes, viral homologs in these cases tend to be the minimal possible version of the protein family, allowing for the identification of the critical residues that stabilize both structure and function. A nice example of this is the minimal catalytic integrase domain of *Sulfolobus* spindle-shaped virus 1 (structures 3VCF, 4DKS and 3UXU). On the other hand, for viruses that have less space limitations, viral proteins can have surprising combinations of domains, suggesting ways to engineer multidomain proteins. This is particularly notorious in *Mimivirus*, where identifiable catalytic domains can be linked to domains with no sequence or structure homology to any known protein, as in the sulfhydryl oxidase in structure 3TD7. Less dramatic examples are basic modules known to function as transcription factors, such as the ribbon-helix-helix domain, with an extra helix added as an embellishment that increases thermal stability, as in structure 4AAI from *Sulfolobus* virus Ragged Hills.

The analysis of the conservation of proteins amongst viruses of the same or different classes is instructive, and can help in establishing families and/or events of horizontal gene transfer. This conservation has been historically one of the criteria used to choose which proteins to study structurally from a particular virus. The wealth of information derived from identifying open reading frames in the data from sequencing endeavors can certainly be a source of novel activities, as described in the previous section. This functional annotation requires sequence homology to known proteins, something that does not happen frequently with extremophile viruses. As structure diverges more slowly than sequence, protein structural analysis allows for the inference of function when sequence homology is weak. In **Supplementary Table S2** we list viral protein structures that were obtained in this spirit, sometimes as part of Structural Genomics Initiatives (Oke et al., 2010). As can be seen from **Supplementary Table S2**, the goal of assigning function is not always achieved, as on occasion novel folds are found (see, for example, structures 4ART and 3DF6 discussed above), precluding the transfer of function. In other, happier cases, the structure instructs the experiments needed to functionally annotate the sequence (see structure 3O27, with functional DNA binding activity).

Most of the structures we found were obtained with a previous inkling of the function of the protein. For example, **Supplementary Table S3** lists structural proteins, such as capsids and tail spikes. The full capsid structures are interesting, for instance, as scaffolds for drug delivery, and as models to study capsid formation, propose infection mechanisms, and study the interactions with nucleic acids and membranes. In this regard, structure 5W7G proposes a model for the membrane envelope of *Acidianus* filamentous virus 1, composed of flexible tetraether lipids that are organized as horseshoes, including a mechanism for enrichment of the viral membrane with this particular lipid of low abundance in the host. Another important interaction is that of capsid proteins with DNA, and surprisingly, it appears that rod-shaped viruses (such as *Sulfolobus islandicus* rod-shaped virus 2 in structure 3J9X) organize their DNA in the A form, stabilized by alpha helices from the major capsid proteins. This is in stark contrast to icosahedral viruses, which pack their DNA in the B form. Turrets, tails and spikes are important to understand interactions with the hosts, as part of the ecological role that these viruses play.

Supplementary Table S4 lists proteins that bind either DNA or histones; the latter come from viruses that infect either fish or shrimp and are interesting because one of them is a DNA mimic (see structure 2ZUG). The architecture of these DNA-binding proteins is sometimes reminiscent of known bacterial classes (see structure 2CO5, a winged helix-turn-helix protein with an intramolecular disulfide bond, discussed above), or is a novel fold (see structure 2J85). Finally, **Supplementary Table S5** lists enzymes found in extremophile viruses. The range of activities is wide, going from DNA, protein and sugar metabolism, to reactive oxygen species management (see, for example, structure 4U4I, a superoxide dismutase that does not require chaperones to capture copper or oxidize its disulfide bridges). There is also interest in auxiliary metabolic genes that support more efficient phage replication, and are normally related to photosynthesis; in this class we find structure 5HI8, a phycobiliprotein lyase, and 3UWA, a peptide deformylase particularly selective for the D1 protein of photosystem II.

The relevance and utility of all these structures is multiple: as crystallographic-amenable homologs of difficult targets (see structure 3VK7, a DNA glycosylase) given their stability, as inspiration to improve mesophilic orthologs in their resistance to high temperature and low pH, as examples on how to trim these orthologs to minimal yet functional versions, in the identification of novel quaternary structures (see for example structure 5Y5O, a dUTPase with novel packing), and as examples on how to adapt new modules to them (see structure 3TD7, the sulfhydryl oxidase with a novel domain attached at the C-terminus). The field of structural biology of extremophile viruses is still young, and there is plenty of room for the exploration of orphan ORFs and for viruses subject to other extreme environments.

FINAL REMARKS

Metagenomics is a powerful approach to study the virome structure of extreme environments and its potential biotechnological applications in a number of fields. However, despite its potential few studies have been undertaken to characterize viral communities in these environments. Some methodological challenges need to be overcome to ensure that samples enriched in viral particles can be obtained, as well as increasing the yields of viral nucleic acids that can be isolated.

A comparative analysis of the population structure of viromes in extreme environments was carried out here, using the 17 publicly accessible virus metagenomic libraries deposited in MetaVir2. Viral communities from different extreme environments showed quite high levels of overall similarity, with viral families belonging to the *Caudovirales* order being ubiquitous and displaying seemingly polyextremophilic adaptation. The most abundant families of *Caudovirales* were *Siphoviridae*, *Myoviridae,* and *Podoviridae*, followed by *Circoviridae, Phycodnaviridae, Microviridae,* and *Inoviridae*. However, very high fractions of unidentified ssDNA and dsDNA viruses and phages were identified. Considering the large number

of unclassified viral sequences from extremophilic viromes, it is currently not possible to definitively identify novel virus families which are uniquely present in different extreme environments. Nor is it possible to correlate the presence of specific virus families or genera with any given environment.

Attempts to further explore specific virus–host relationships in the *Podoviridae* and *Poxviridae* present in OMZ and deep-sediments resulted in the identification of viruses whose known hosts are highly unlikely to reside in these environments. Although many more sequences were obtained when a similar comparison was made using newer, more up to date databases (RefSeq from NCBI, or IMG/VR), the viruses identified often correspond to those for which more sequences are available. Therefore, although MetaVir2 is no longer kept up to date, and richer, more recent databases exist, such as IMG/VR, a similar comparative analysis using such databases produced similar results (not shown), indicating that more accurate taxonomic assignments are required and ideally they should be in a common repository where all the viral metadata are collected. Also new tools should be developed to automatize sequence classification so that viral species assignment can be obtained.

Hierarchical clustering analysis was performed from abundant viral families previously mentioned and it was possible to conclude that some extreme environments, such as hypersaline, deep-sea and hyperarid niches, have groups that indicate similarities in the viral communities present in these environments, although as above, with the available data this comparison could only be evaluated at the family level. An important challenge from viral metagenomics is to establish specific viral clusters associated with particular extreme environments and describe their role in different extreme ecosystems. Although taxonomic allocation at the level of the genus or species in viruses is a challenge, new strategies for the classification of viruses are still in development from the use of genomic sequences without previous information or clustering of their coding sequences, that allow a more efficient classification process, that is scalable and user friendly.

In addition, the functional prospecting for viruses in extreme ecological niches has been almost exclusively limited to sequence-based screening to date. While some viral sequences have been annotated and assigned to specific functions, very few viral proteins discovered using metagenomics approaches have been subsequently cloned, heterologously expressed and biochemically characterized. Other functional-based methods such as activity-based screenings and PCR -or hybridization- based screenings are currently underexploited as approaches to identify viral proteins from extremophilic viromes, which may have utility in biotechnological applications. Considering that our current knowledge of viromes associated with extreme ecosystems is quite limited, we still cannot fully appreciate the great biotechnological potential that they may represent. Thus, further efforts should be made to screen extremophile viral metagenomes for novel proteins and biomolecules if we are to advance our understanding of their biological impact and to capitalize on the unique viral diversity that is present within these novel ecosystems.

AUTHOR CONTRIBUTIONS

SD-R, RG, AD, and RB-G designed and wrote the manuscript. MS-C prepared the Section "Perspectives on Sampling and Processing: Methodological Challenges for Viral Metagenomics in Extreme Environments." HC-S, SD-R, RP, and AH collected and processed the metagenomic data and conducted the analysis of viral communities. RB-G and LM-Á prepared Section "Functional Metagenomics in Extreme Environments: Methodological Challenges, Discoveries and Opportunities" related to functional bioprospection in extreme viromes, while NP prepared Section "Structural Biology of Viromes."

FUNDING

We acknowledge the graduate fellowships that HC-S and LM-Á received from CONACyT-Mexico. The authors appreciate the support received from the Unidad de Secuenciación Masiva y Bioinformática (Instituto de Biotecnología UNAM-Mexico) for the access to computer facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02403/full#supplementary-material

REFERENCES

Adriaenssens, E. M., Van Zyl, L., De Maayer, P., Rubagotti, E., Rybicki, E., Tuffin, M., et al. (2015). Metagenomic analysis of the viral community in NAMIB DESERT hypoliths. *Environ. Microbiol.* 17, 480–495. doi: 10.1111/1462-2920.12528

Andrade, A. C., Arantes, T. S., Rodrigues, R. A., Machado, T. B., Dornas, F. P., Landell, M. F., et al. (2018). Ubiquitous giants: a plethora of giant viruses found in Brazil and Antarctica. *Virol. J.* 15:22. doi: 10.1186/s12985-018-0930-x

Andrews-Pfannkoch, C., Fadrosh, D. W., Thorpe, J., and Williamson, S. J. (2010). Hydroxyapatite-mediated separation of double-stranded DNA, single-stranded DNA, and RNA genomes from natural viral assemblages. *Appl. Environ. Microbiol.* 76, 5039–5045. doi: 10.1128/AEM.00204-10

Atanasova, N. S., Bamford, D. H., and Oksanen, H. M. (2016). Virus-host interplay in high salt environments. *Environ. Microbiol. Rep.* 8, 431–444. doi: 10.1111/1758-2229.12385

Atanasova, N. S., Demina, T. A., Shanthi, S. N. K. R., Oksanen, H. M., and Bamford, D. H. (2018). Extremely halophilic pleomorphic archaeal virus HRPV9 extends the diversity of pleolipoviruses with integrases. *Res. Microbiol.* 169, 500–504. doi: 10.1016/j.resmic.2018.04.004

Bellas, C. M., Anesio, A. M., Barker, G., and Pearce, D. A. (2015). Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. *Front. Microbiol.* 6:656. doi: 10.3389/fmicb.2015.00656

Berliner, A. J., Mochizuki, T., and Stedman, K. M. (2018). Astrovirology: viruses at large in the universe. *Astrobiology* 18, 207–223. doi: 10.1089/ast.2017.1649

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242.

Biddle, J. F., White, J. R., Teske, A. P., and House, C. H. (2011). Metagenomics of the subsurface brazos-trinity basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J.* 5, 1038–1047. doi: 10.1038/ismej.2010.199

Blanc-Mathieu, R., Kaneko, H., Endo, H., Chaffron, S., Hernández-Velázquez, R., Nguyen, C. H., et al. (2019). *Viruses of the Eukaryotic Plankton are Predicted to Increase Carbon Export Efficiency in the Global Sunlit Ocean. biorxiv*. [preprint]. doi: 10.1101/710228

Bolduc, B., Shaughnessy, D. P., Wolf, Y. I., Koonin, E. V., Roberto, F. F., and Young, M. (2012). Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. *J. Virol.* 86, 5562–5573. doi: 10.1128/JVI.07196-11

Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B. L., and Sullivan, M. B. (2017). iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J.* 11, 7–14. doi: 10.1038/ismej.2016.89

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., et al. (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc. R. Soc. Lond. B* 271, 565–574. doi: 10.1098/rspb.2003.2628

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *PNAS* 99, 14250–14255. doi: 10.1073 pnas.202488399

Bzhalava, D., and Dillner, J. (2013). Data mining in genomics proteomics. *Data Min. Genom. Proteom.* 4, 3–7. doi: 10.4172/2153-0602.1000134

Cantalupo, P. G., and Pipas, J. M. (2019). Detecting viral sequences in NGS data. *Curr. Opin. Virol.* 39, 41–48. doi: 10.1016/j.coviro.2019.07.010

Cárcer, D. A., De, López-bueno, A., Pearce, D. A., and Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Sci. Adv.* 2, 1–9. doi: 10.1126/sciadv.1400127

Cassman, N., Prieto-davó, A., Walsh, K., Silva, G. G. Z., Angly, F., Akhter, S., et al. (2012). Oxygen minimum zones harbour novel viral communities with low diversity. *Environ. Microbiol.* 14, 3043–3065. doi: 10.1111/j.1462-2920.2012.02891.x

Castelán-Sánchez, H. G., Lopéz-Rosas, I., García-Suastegui, W. A., Peralta, R., Dobson, A. D., Batista-García, R. A., et al. (2019). Extremophile deep-sea viral communities from hydrothermal vents: structural and functional analysis. *Mar. Genom.* 46, 16–28. doi: 10.1016/j.margen.2019.03.001

Chander, Y., Koelbl, J., Puckett, J., Moser, M. J., Klingele, A. J., Liles, M. R., et al. (2014). A novel thermostable polymerase for RNA and DNA loop-mediated isothermal amplification (AMP). *Front. Microbiol.* 5:395. doi: 10.3389/fmicb.2014.00395

Chatterjee, A., Sicheritz-Pontén, T., Yadav, R., and Kondabagil, K. (2019). Genomic and metagenomic signatures of giant viruses are ubiquitous in water samples from sewage, inland lake, waste water treatment plant, and municipal water supply in Mumbai. *India. Sci. Rep.* 9:3690. doi: 10.1038/s41598-019-40171-y

Chow, C. E. T., Winget, D. M., White, R. A. III, Hallam, S. J., and Suttle, C. A. (2015). Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Front. Microbiol.* 6:265. doi: 10.3389/fmicb.2015.00265

Claverie, J. M., Abergel, C., and Legendre, M. (2018). Giant viruses that create their own genes. *Med. Sci.* 34, 1087–1091. doi: 10.1051/medsci/2018300

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Rastelli, E., Cavicchioli, R., Krupovic, M., et al. (2016). Virus-mediated archaeal hecatomb in the deep seafloor. *Sci. Adv.* 2:e1600492. doi: 10.1126/sciadv.1600492

Demina, T. A., Atanasova, N. S., Pietilä, M. K., Oksanen, H. M., and Bamford, D. H. (2016). Vesicle-like virion of Haloarcula hispanica pleomorphic virus 3 preserves high infectivity in saturated salt. *Virology* 499, 40–51. doi: 10.1016/j.virol.2016.09.002

Dennis, T. P., de Souza, W. M., Marsile-Medun, S., Singer, J. B., Wilson, S. J., and Gifford, R. J. (2019). The evolution, distribution and diversity of endogenous circoviral elements in vertebrate genomes. *Virus Res.* 262, 15–23. doi: 10.1016/j.virusres.2018.03.014

Dennis, T. P., Flynn, P. J., De Souza, W. M., Singer, J. B., Moreau, C. S., Wilson, S. J., et al. (2018). Insights into circovirus host range from the genomic fossil record. *J. Virol.* 92, e145–e118. doi: 10.1128/JVI.00145-18

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 452, 340–343. doi: 10.1038/nature06735

Diemer, G. S., and Stedman, K. M. (2012). A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol. Dir.* 7:13. doi: 10.1186/1745-6150-7-13

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., et al. (2008). Functional metagenomic profiling of nine biomes. *Nat. Lett.* 452, 629–633. doi: 10.1038/nature06810

Dwivedi, B., Xue, B., Lundin, D., Edwards, R. A., and Breitbart, M. (2013). A bioinformatic analysis of ribonucleotide reductase genes in phage genomes andmetagenomes. *BMC Evol. Biol.* 13:1–17. doi: 10.1186/1471-2148-13-33

Faino, L., Seidl, M. F., Datema, E., van den Berg, G. C., Janssen, A., Wittenberg, A. H., et al. (2015). *Single-Molecule Real-Time Sequencing Combined With Optical Mapping Yields Completely Finished Fungal Genome. MBio*. [preprint]. doi: 10.1128/mBio.00936-15

Fancello, L., Trape, S., Catherine, R., Mickael, B., Nikolay, P., Raoult, D., et al. (2013). Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J.* 7, 359–369. doi: 10.1038/ismej.2012.101

Feazel, L. M., Spear, J. R., Berger, A. B., Harris, J. K., Frank, D. N., Ley, R. E., et al. (2008). Eucaryotic diversity in a hypersaline microbial mat. *Appl. Environ. Microbiol.* 74, 329–332. doi: 10.1128/AEM.01448-07

Fuchsman, C. A., Palevsky, H. I., Widner, B., Duffy, M., Carlson, M. C., Neibauer, J. A., et al. (2019). Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. *ISME J.* 1, 1–13. doi: 10.1038/s41396-019-0452-6

Furtak, V., Roivainen, M., Mirochnichenko, O., Zagorodnyaya, T., Laassri, M., Zaidi, S., et al. (2016). Environmental surveillance of viruses by tangential flow filtration and metagenomic reconstruction. *Eurosurveillance* 21:30193. doi: 10.2807/1560-7917.ES.2016.21.15.30193

Gong, Z., Liang, Y., Wang, M., Jiang, Y., Yang, Q., Xia, J., et al. (2018). Viral diversity and its relationship with environmental factors at the surface and deep sea of Prydz Bay. Antarctica. *Front. Microbiol.* 9:2981. doi: 10.3389/fmicb.2018.02981

Graham, E. B., Paez-Espino, D., Brislawn, C., Neches, R. Y., Hofmockel, K. S., Wu, R., et al. (2019). *Untapped Viral Diversity in Global Soil Metagenomes. biorxiv*. [preprint]. doi: 10.1101/583997

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., et al. (2019). Marine DNA viral macro-and microdiversity from pole to pole. *Cell* 177, 1109–1123. doi: 10.1016/j.cell.2019.03.040

Guajardo-Leiva, S., Pedrós-Alió, C., Salgado, O., Pinto, F., and Díez, B. (2018). Active crossfire between *Cyanobacteria* and Cyanophages in phototrophic mat communities within hot springs. *Front. Microbiol.* 9:2039. doi: 10.3389/fmicb.2018.02039

Gudbergsdóttir, S. R., Menzel, P., Krogh, A., Young, M., and Peng, X. (2016). Novel viral genomes identified from six metagenomes reveal wide distribution of archaeal viruses and high viral diversity in terrestrial hot springs. *Environ. Microbiol.* 18, 863–874. doi: 10.1111/1462-2920.13079

Hara, S., Terauchi, K., and Koike, I. (1991). Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy. *Appl. Environ. Microbiol.* 57, 2731–2734.

Hayes, S., Mahony, J., Nauta, A., and Van Sinderen, D. (2017). Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* 9:E127. doi: 10.3390/v9060127

Heidelberg, K. B., Nelson, W. C., Holm, J. B., Eisenkolb, N., Andrade, K., Emerson, J., et al. (2013). Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell. Australia. *Front. Microbiol.* 4:115. doi: 10.3389/fmicb.2013.00115

Heller, R., Chung, S., Crissy, K., Dumas, K., Schuster, D., and Schoenfeld, T. (2019). Engineering of a thermostable viral polymerase using metagenome-derived diversity for highly sensitive and specific RT-PCR. *Nucleic Acids Res.* 1, 1–12. doi: 10.1093/nar/gkz104

Hirai, M., Nishi, S., Tsuda, M., Sunamura, M., Takaki, Y., and Nunoura, T. (2017). Library construction from subnanogram DNA for pelagic sea water and deep-sea sediments. *Microb. Environ.* 4, 336–343. doi: 10.1264/jsme2.ME17132

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., et al. (2014). EBI metagenomics—a new resource for the analysis and archiving

of metagenomic data. *Nucleic Acids Res.* 42, D600–D606. doi: 10.1093/nar/gkt961

Hurwitz, B. L., Deng, L., Poulos, B. T., and Sullivan, M. B. (2013). Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environ. Microbiol.* 15, 1428–1440. doi: 10.1111/j.1462-2920.2012.02836.x

John, S. G., Mendez, C. B., Deng, L., Poulos, B., Kauffman, A. K. M., Kern, S., et al. (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ. Microbiol. Rep.* 3, 195–202. doi: 10.1111/j.1758-2229.2010.00208.x

Kang, S., and Kim, Y. C. (2018). Identification of Viral Taxon-Specific Genes (VTSG): application to Caliciviridae. *Genom. Inform.* 16:e23. doi: 10.5808/GI.2018.16.4.e23

Kiehl, J. T., and Shields, C. A. (2005). Climate simulation of the latest Permian: implications for mass extinction. *Geology* 33, 757–760. doi: 10.1130/G21654.1

Kinsella, C. M., Deijs, M., and van der Hoek, L. (2019). Enhanced bioinformatic profiling of VIDISCA libraries for virus detection and discovery. *Vir. Res.* 263, 21–26. doi: 10.1016/j.virusres.2018.12.010

Kleiner, M., Hooper, L. V., and Duerkop, B. A. (2015). Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics* 16:7. doi: 10.1186/s12864-014-1207-4

Konstantinidis, K. T., Braff, J., Karl, D. M., and DeLong, E. F. (2009). Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl. Environ. Microbiol.* 75, 5345–5355. doi: 10.1128/AEM.00473-09

Kristensen, D. M., Mushegian, A. R., Dolja, V. V., and Koonin, E. V. (2012). New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11–19. doi: 10.1016/j.tim.2009.11.003.New

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., and Koonin, E. V. (2018). Viruses of archaea: structural, functional, environmental and evolutionary genomics. *Vir. Res.* 244, 181–193. doi: 10.1016/j.virusres.2017.11.025

Laanto, E., Hoikkala, V., Ravantti, J., and Sundberg, L. (2017). Long-term genomic coevolution of host-parasite interaction in the natural environment. *Nat. Com.* 8, 2–8. doi: 10.1038/s41467-017-00158-7

Lawrence, J. E., and Steward, G. F. (2010). "Purification of viruses by centrifugation," in *Manual of Aquatic Viral Ecology*, eds S. W. Wilhelm, M. G. Weinbauer, and C. A. Suttle, (Waco, TX: ASLO), 166–181. doi: 10.4319/mave.2010.978-0-9845591-0-7.166

Le Romancer, M., Gaillard, M., Geslin, C., and Prieur, D. (2006). Viruses in extreme environments. *Rev. Environ. Sci. Biotechnol.* 6, 17–31. doi: 10.1007/s11157-006-0011-2

Liang, Y., Wang, L., Wang, Z., Zhao, J., Yang, Q., Wang, M., et al. (2019). Metagenomic analysis of the diversity of DNA viruses in the surface and deep-sea of the South China Sea. *Front. Microbiol.* 10:1951. doi: 10.3389/fmicb.2019.01951

Liu, Y., Brandt, D., Ishino, S., Ishino, Y., Koonin, E. V., Kalinowski, J., et al. (2019). New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. *Environ. Microbiol.* 21, 2002–2014. doi: 10.1111/1462-2920.14479

López-Bueno, A., Tamames, J., Velázquez, D., Moya, A., Quesada, A., and Alcamí, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* 80, 858–861. doi: 10.1126/science.1179287

López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M., and Rodriguez-Valera, F. (2017). Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genetics* 13:e1007018. doi: 10.1371/journal.pgen.1007018

Ma, Y., Galinski, E. A., Grant, W. D., Oren, A., and Ventosa, A. (2010). Halophiles 2010: life in Saline Environments. *Appl. Environ. Microbiol.* 76, 6971–6981. doi: 10.1128/AEM.01868-10

Martinez-Hernandez, F., Fornas, O., Gomez, M. L., Bolduc, B., de La Cruz Peña, M. J., Martínez, J. M., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* 8:15892. doi: 10.1038/ncomms15892

Martin-Cuadrado, A. B., Senel, E., Martínez-García, M., Cifuentes, A., Santos, F., Almansa, C., et al. (2019). Prokaryotic and viral community of the sulfate-rich crust from Peñahueca ephemeral lake, an astrobiology analogue. *Environ. Microbiol.* 21, 3577–3600. doi: 10.1111/1462-2920.14680

Mead, D. A., Monsma, S., Mei, B., Gowda, K., Lodes, M., Schoenfeld, T. W., et al. (2017). "Functional metagenomics of a replicase from a novel hyperthermophilic aquificales virus," in *Functional Metagenomics: Tools and Applications*, eds T. C. Charles, and M. R. Liles, (Cham: Springer), 217–242. doi: 10.1007/978-3-319-61510-3_13

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780

Miranda, J. A., Culley, A. I., Schvarcz, C. R., and Steward, G. F. (2016). RNA viruses as major contributors to Antarctic virioplankton. *Environ. Microbiol.* 18, 3714–3727. doi: 10.1111/1462-2920.13291

Mizuno, C. M., Ghai, R., Saghai, A., Lopez-Garcia, P., and Rodriguez-Valera, F. (2016). *Genomes of Abundant and Widespread Viruses from the Deep Ocean. MBio*. [preprint]. doi: 10.1128/mBio.00805-16.

Mizuno, C. M., Prajapati, B., Lucas-Staat, S., Sime-Ngando, T., Forterre, P., Bamford, D. H., et al. (2019). Novel haloarchaeal viruses from lake retba infecting haloferax and halorubrum species. *Environ. Microbiol.* 21, 2129–2147. doi: 10.1111/1462-2920.14604

Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., and Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genetics* 9:e1003987. doi: 10.1371/journal.pgen.1003987

Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., Sako, Y., Forterre, P., and Prangishvili, D. (2012). Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc. Natl. Acad. Sci.* 109, 13386–13391. doi: 10.1073/pnas.1203668109

Mokili, J. L., Rohwer, F., and Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2, 63–77. doi: 10.1016/j.coviro.2011.12.004

Moser, M. J., Difrancesco, R. A., Gowda, K., Klingele, A. J., Sugar, D. R., Stocki, S., et al. (2012). Thermostable DNA Polymerase from a Viral Metagenome Is a Potent RT-PCR Enzyme. *PLoS One* 7:e38371. doi: 10.1371/journal.pone.0038371

Motlagh, A. M., Bhattacharjee, A. S., Coutinho, F. H., Dutilh, B. E., Casjens, S. R., and Goel, R. K. (2017). Insights of phage-host interaction in hypersaline ecosystem through metagenomics analyses. *Front. Microbiol.* 8:352. doi: 10.3389/fmicb.2017.00352

Munson-McGee, J., Snyder, J., and Young, M. (2018). Archaeal viruses from high-temperature environments. *Genes* 9:E128. doi: 10.3390/genes9030128

Nguyen, T. T., and Landfald, B. (2015). Polar front associated variation in prokaryotic community structure in Arctic shelf seafloor. *Front. Microbiol.* 6:17. doi: 10.3389/fmicb.2015.00017

Nigro, O. D., Jungbluth, S. P., Lin, H. T., Hsieh, C. C., Miranda, J. A., Schvarcz, C. R., et al. (2017). *Viruses in the Oceanic Basement MBio*. [preprint]. doi: 10.1128/mBio.02129-16

Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M. P. (2018). Overview of virus metagenomic classification methods and their biological applications. *Front. Microbiol.* 9:749. doi: 10.3389/fmicb.2018.00749

Oke, M., Carter, L. G., Johnson, K. A., Liu, H., McMahon, S. A., Yan, X., et al. (2010). The scottish structural proteomics facility: targets. methods and outputs. *J. Struct. Funct. Gen.* 11, 167–180. doi: 10.1007/s10969-010-9090

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., et al. (2016). Uncovering Earth's virome. *Nature* 536, 425–430. doi: 10.1038/nature19094

Paez-Espino, D., Roux, S., Chen, I., Palaniappan, K., Ratner, A., Chu, K., et al. (2018). IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.* 47, D678–D686. doi: 10.1093/nar/gky1127

Parvathi, A., Jasna, V., Aparna, S., Pradeep Ram, A., Aswathy, V., Balachandran, K., et al. (2018). High incidence of lysogeny in the oxygen minimum zones of the arabian sea (Southwest Coast of India). *Viruses* 10:588. doi: 10.3390/v10110588

Paulmier, A., and Ruiz-Pino, D. (2009). Progress in oceanography oxygen minimum zones (OMZs) in the modern ocean. *Prog. Oceanogr.* 80, 113–128. doi: 10.1016/j.pocean.2008.08.001

Pietilä, M. K., Laurinavičius, S., Sund, J., Roine, E., and Bamford, D. H. (2010). The single-stranded DNA genome of novel archaeal virus Halorubrum pleomorphic virus 1 is enclosed in the envelope decorated with glycoprotein spikes. *J. Virol.* 8, 788–798. doi: 10.1128/JVI.01347-09

Plominsky, A., Henríquez, C. A., Delherbe, N., Podell, S., Ramirez, S., Ugalde, J. A., et al. (2018). Distinctive archaeal composition of an artisanal crystallizer pond and functional insights into salt-saturated hypersaline environment adaptation. *Front. Microbiol.* 9:1800. doi: 10.3389/fmicb.2018.01800

Porter, K., Russ, B. E., and Dyall-smith, M. L. (2007). Virus–host interactions in salt lakes. *Curr. Opin. Microbiol.* 10, 418–424. doi: 10.1016/j.mib.2007.05.017

Poulos, B. T., John, S. G., and Sullivan, M. B. (2018). "Iron Chloride Flocculation of Bacteriophages from Seawater," in *Bacteriophages. Methods in Molecular Biology*, eds M. Clokie, A. Kropinski, and R. Lavigne, (New York, NY: Humana Press), 1681.

Ramos-Barbero, M. D., Martínez, J. M., Almansa, C., Rodríguez, N., Villamor, J., Gomariz, M., et al. (2019). Prokaryotic and viral community structure in the singular chaotropic salt lake salar de uyuni. *Environ. Microbiol.* 21, 2029–2042. doi: 10.1111/1462-2920.14549

Rastrojo, A., and Alcamí, A. (2017). Aquatic viral metagenomics: Lights and shadows. *Virus Res.* 239, 87–96. doi: 10.1016/j.virusres.2016.11.021

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., and Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 5:69. doi: 10.1186/s40168-017-0283-5

Resplandy, L., Keeling, R. F., Eddebbar, Y., Brooks, M. K., Wang, R., Bopp, L., et al. (2018). Quantification of ocean heat uptake from changes in atmospheric O_2 and CO_2 composition. *Nature* 563:105. doi: 10.1038/s41586-018-0651-8

Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., et al. (2001). Viruses from extreme thermal environments. *PNAS* 98, 13341–13345. doi: 10.1073 pnas.231170198

Rosario, K., and Breitbart, M. (2011). Exploring the viral world through metagenomics karyna rosario and mya breitbart. *Curr. Opin. Virol.* 1, 289–297. doi: 10.1016/j.coviro.2011.06.004

Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L., and Prosperi, M. (2016). Challenges in the analysis of viral metagenomes. *Virus Evol.* 2:vew0022. doi: 10.1093/ve/vew022

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., et al. (2019). Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* 37:29. doi: 10.1038/nbt.4306

Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. doi: 10.7717/peerj.985

Roux, S., Enault, F., Ravet, V., Colombet, J., Bettarel, Y., Auguet, J., et al. (2016). Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environ. Microbiol.* 18, 889–903. doi: 10.1111/1462-2920.13084

Roux, S., Tournayre, J., Mahul, A., Debroas, D., and Enault, F. (2014). Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* 15:76. doi: 10.1186/1471-2105-15-76

Sangwan, N., Lambert, C., Sharma, A., Gupta, V., Khurana, P., Khurana, J. P., et al. (2015). Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator–prey genotypes. *Environ. Microbiol. Rep.* 7, 812–823. doi: 10.1111/1758-2229.12297

Schmidt, H. F., Sakowski, E. G., Williamson, S. J., Polson, S. W., and Wommack, K. E. (2014). Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. *ISME J.* 8, 103–114. doi: 10.1038/ismej.2013.124

Schmitz, J. E., Schuch, R., and Fischetti, V. A. (2010). Identifying active phage lysins through functional. *Appl. Energy* 76, 7181–7187. doi: 10.1128/AEM.00732-10

Schoenfeld, T., Liles, M., Wommack, K. E., Polson, S. W., Godiska, R., and Mead, D. (2009). Functional viral metagenomics and the next generation of molecular tools. *Cell Press* 18, 20–30. doi: 10.1016/j.tim.2009.10.001

Schoenfeld, T. W., and Mead, D. (2015). "Use of viral metagenomes from yellowstone hot springs to study phylogenetic relationships and evolution," in *Encyclopedia of Metagenomics*, ed. K. E. Nelson (New York, NY: Springer), 681–700.

Schoenfeld, T. W., Moser, M. J., and Mead, D. (2015). "Functional viral metagenomics and the development of new enzymes for DNA and RNA amplification and sequencing," in *Encyclopedia of Metagenomics*, ed. K. E. Nelson (Boston, MA: Springer), 198–218.

Schoenfeld, T., Patterson, M., Paul, M., Wommack, K. E., and Young, M. (2008). Assembly of viral metagenomes from Yellowstone Hot Springs. *Appl. Enrion. Microbiol.* 74, 4164–4174. doi: 10.1128/AEM.02598-07

Sharma, A., Schmidt, M., Kiesel, B., Mahato, N. K., Cralle, L., Singh, Y., et al. (2018). Bacterial and archaeal viruses of Himalayan Hot Springs At Manikaran Modulate Host Genomes. *Front. Microbiol.* 9:3095. doi: 10.3389/fmicb.2018.03095

Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. *J. Gen. Virol.* 96, 1193–1206. doi: 10.1099/vir.0.000016

Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15:161. doi: 10.1038/nrmicro.2016.177

Steward, G. F., Culley, A. I., Mueller, J. A., Wood-Charlson, E. M., Belcaid, M., and Poisson, G. (2013). Are we missing half of the viruses in the ocean? *ISME J.* 7, 672–679. doi: 10.1038/ismej.2012.121

Strazzulli, A., Fusco, S., Moracci, M., and Contursi, P. (2017). Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev. Environ. Sci. Bio/Techno.* 16, 425–454. doi: 10.1007/s11157-017-9435-0

Tennant, P., Fermin, G., and Foster, J. E. (eds) (2018). *Viruses: Molecular Biology, Host Interactions, and Applications to Biotechnology*. Cambridge, MA: Academic Press.

Thamdrup, B. (2012). New pathways and processes in the global nitrogen cycle. *Annu. Rev. Ecol. Evol. Syst.* 43, 407–428. doi: 10.1146/annurev-ecolsys-102710-145048

Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470–483. doi: 10.1038/nprot.2009.10

Van Etten, J. L., Lane, L. C., and Dunigan, D. D. (2010). DNA viruses: the really big ones (giruses). *Annu. Rev. Microbiol.* 64, 83–99. doi: 10.1146/annurev.micro.112408.134338

Ventosa, A., de la Haba, R. R., Sánchez-Porro, C., and Papke, R. T. (2015). Microbial diversity of hypersaline environments: a metagenomic approach. *Curr. Opin. Microb.* 25, 80–87. doi: 10.1016/j.mib.2015.05.002

Vikram, S., Guerrero, L. D., Makhalanyane, T. P., Le, P. T., Seely, M., and Cowan, D. A. (2016). Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environ. Microbiol.* 18, 1875–1888. doi: 10.1111/1462-2920.13088

Wang, H., Yu, Y., Liu, T., Pan, Y., Yan, S., and Wang, Y. (2015). Diversity of putative archaeal RNA viruses in metagenomic datasets of a yellowstone acidic hot spring. *Springer Plus* 18 4:189. doi: 10.1186/s40064-015-0973-z

Weitz, J. S., and Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* 4, 17. doi: 10.3410/B4-17

Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., et al. (2008). The sorcerer ii global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* 3:e1456. doi: 10.1371/journal.pone.0001456

Willner, D., Thurber, R. V., and Rohwer, F. (2009). Metagenomic signatures of 86 microbial and viral metagenomes. *Environ. Microbiol.* 11, 1752–1766. doi: 10.1111/j.1462-2920.2009.01901.x

Wommack, K. E., Bhavsar, J., Polson, S. W., Chen, J., Dumas, M., Srinivasiah, S., et al. (2012). VIROME: a standard operating procedure for analysis of viral metagenome sequences. *Stand. Genomic Sci.* 6, 421–433. doi: 10.4056/sigs.2945050

Yau, S., and Seth-Pasricha, M. (2019). Viruses of polar aquatic environments. *Viruses* 11:189. doi: 10.3390/v11020189

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., and Takai, K. (2013). Metagenomic analysis of viral communities in (Hado) pelagic sediments. *PLoS One* 8:e57271. doi: 10.1371/journal.pone.0057271

Yoshida, M., Mochizuki, T., Urayama, S. I., Yoshida-Takashima, Y., Nishi, S., et al. (2018). Quantitative viral community dna analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku. Japan. *Front. Microbiol.* 9:75. doi: 10.3389/fmicb.2018.00075

Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., et al. (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *J. Virol.* 93, e2130–e2118. doi: 10.1111/1462-2920.14549

Zablocki, O., Adriaenssens, E. M., and Cowan, D. (2016). Diversity and ecology of viruses in hyperarid desert soils. *Appl. Environ. Microbiol.* 82, 770–777. doi: 10.1128/AEM.02651-15

Zablocki, O., Adriaenssens, E. M., Frossard, A., and Seely, M. (2017a). Metaviromes of extracellular soil viruses along a Namib desert aridity gradient. *Genome Announc.* 5, 4–5. doi: 10.1128/genomeA.01470-16

Zablocki, O., van Zyl, L., Kirby, B., and Trindade, M. (2017b). Diversity of dsDNA viruses in a South African hot spring assessed by metagenomics and microscopy. *Viruses* 9:348. doi: 10.3390/v9110348

Zablocki, O., Zyl, L. Van, Adriaenssens, E. M., Rubagotti, E., Tuffin, M., and Cary, C. (2014). High-Level diversity of tailed phages, eukaryote-associated viruses,and virophage-like elements in the metaviromes of Antarctic soils. *Appl. Environ. Microbiol.* 80, 6888–6897. doi: 10.1128/AEM.01525-14

Zhang, Y. Z., Shi, M., and Holmes, E. C. (2018). Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172. doi: 10.1016/j.cell.2018.02.043

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ARTICULO 3

genes

MDPI

Article

Intermediate-Salinity Systems at High Altitudes in the Peruvian Andes Unveil a High Diversity and Abundance of Bacteria and Viruses

Hugo Gildardo Castelán-Sánchez [1], Paola Elorrieta [2], Pedro Romoacca [3], Arturo Liñan-Torres [1], José Luis Sierra [4], Ingrid Vera [3], Ramón Alberto Batista-García [1], Silvia Tenorio-Salgado [5], Gabriel Lizama-Uc [5], Ernesto Pérez-Rueda [6,7], María Antonieta Quispe-Ricalde [2,*] and Sonia Dávila-Ramos [1,*]

1 Centro de Investigación en Dinámica Celular, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos C.P. 62209, Mexico; hcastelans@gmail.com (H.G.C.-S.); arturo21lt@gmail.com (A.L.-T.); rabg@uaem.mx (R.A.B.-G.)
2 Departamento de Biología, Facultad de Ciencias, Universidad Nacional de San Antonio Abad del Cusco, Cusco C.P. 0800, Peru; pao91827364@gmail.com
3 Departamento de Farmacia y Bioquímica, Facultad de Ciencias de la Salud, Universidad Nacional de San Antonio Abad del Cusco, Cusco C.P. 0800, Peru; ordep.rh1211@gmail.com (P.R.); iveraf@yahoo.es (I.V.)
4 Escuela de Postgrado, Universidad Nacional de San Antonio Abad del Cusco, Cusco C.P. 0800, Peru; lsierrah77@gmail.com
5 Tecnológico Nacional de México, Instituto Tecnológico de Mérida, Mérida, Yucatán C.P. 97000, Mexico; s.tenorio.salgado@gmail.com (S.T.-S.); gabriel.lizama29@gmail.com (G.L.-U.)
6 Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Sede Mérida, Universidad Nacional Autónoma de México, Unidad Académica de Ciencias y Tecnología, Mérida, Yucatán C.P. 97302, Mexico; eprueda@gmail.com
7 Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, Providencia, Santiago C.P. 7500000, Chile
* Correspondence: antonieta.quispe@unsaac.edu.pe (M.A.Q.-R.); sonia.davila@uaem.mx (S.D.-R.); Tel.: +52-(777)-3297000 (ext. 3276) (S.D.-R.)

Received: 11 September 2019; Accepted: 26 October 2019; Published: 5 November 2019

check for updates

Abstract: Intermediate-salinity environments are distributed around the world. Here, we present a snapshot characterization of two Peruvian thalassohaline environments at high altitude, Maras and Acos, which provide an excellent opportunity to increase our understanding of these ecosystems. The main goal of this study was to assess the structure and functional diversity of the communities of microorganisms in an intermediate-salinity environment, and we used a metagenomic shotgun approach for this analysis. These Andean hypersaline systems exhibited high bacterial diversity and abundance of the phyla *Proteobacteria*, *Bacteroidetes*, *Balneolaeota*, and *Actinobacteria*; in contrast, *Archaea* from the phyla *Euryarchaeota*, *Thaumarchaeota*, and *Crenarchaeota* were identified in low abundance. Acos harbored a more diverse prokaryotic community and a higher number of unique species compared with Maras. In addition, we obtained the draft genomes of two bacteria, *Halomonas elongata* and *Idiomarina loihiensis*, as well as the viral genomes of *Enterobacteria* lambda-like phage and *Halomonas elongata*-like phage and 27 partial novel viral halophilic genomes. The functional metagenome annotation showed a high abundance of sequences associated with detoxification, DNA repair, cell wall and capsule formation, and nucleotide metabolism; sequences for these functions were overexpressed mainly in bacteria and also in some archaea and viruses. Thus, their metabolic profiles afford a decrease in oxidative stress as well as the assimilation of nitrogen, a critical energy source for survival. Our work represents the first microbial characterization of a community structure in samples collected from Peruvian hypersaline systems.

Keywords: Peruvian Andes; metagenomics; intermediate salinity; microbiome; virus

1. Introduction

Millions of years ago (80–110 million years), the ocean covered the central region of Peru; during the formation of the Andes mountains, these marine waters remained inland and, by evaporation, formed deposits of salt in ponds. Different hypersaline water systems are distributed throughout Peru, such as the salterns of the Acos system and the brines from Maras, two thalassohaline environments located in the Andes mountains in southeast Peru. These two systems have not received much study. Acos is located in the district of Acomayo (southeast Peru) at an altitude of 2852 m above sea level, while Maras is located in the district of Urubamba at an altitude of 3380 m and is composed of 3000 small shallow ponds that form terraces on the slope of the mountain Qaqawiñay (a Quechua word meaning eternal rock) [1,2].

The hypersaline ecosystems are characterized by alkalinity and low oxygen concentrations [3–6]. Hypersaline aquatic environments are classified into two main categories: (1) thalassohaline environments, which result from the evaporation of seawater and contain a high concentration of NaCl, neutral or slightly alkaline pH, and a salinity exceeding that of seawater by a factor of 5–10; and (2) athalassohaline environments, which are not derived from seawater and contain high concentrations of ions such as Mg^{2+} or Ca^{2+} and a slightly acidic pH [3–6].

Aquatic hypersaline systems represent excellent models for the study of the ecology and diversity of microorganisms. Most saline systems are composed of ponds with different salinity gradients [7]. Microorganisms identified in hypersaline environments have been classified according to the concentration of salts in the environments they inhabit: weak halophiles (1–3% NaCl), moderate halophiles (3–15% NaCl), and extreme halophiles (more than 15% NaCl) [8]. In contrast, there is no generalized classification for saline environments, but they can be divided into low salinity (less than 10% NaCl), intermediate salinity (10–20% NaCl) [9], and high salinity (higher than 20% NaCl) [10].

Regarding microbial communities that live in these ecosystems, a great diversity of microorganisms has been reported, in particular of the *Halobacteriaceae* family within the Archaea domain. For bacteria, the *Halorhodospira*, *Salinibacter*, *Halomonas*, *Chromohalobacter*, and *Salicola* genera are abundant; and eukaryotic organisms such as *Artemia salina*, *Colpodella edax*, and *Dunaliella salina* have been identified in low proportions [5,11–13]. In addition, a high diversity of haloviruses has been identified, at concentrations of $\geq 1 \times 10^7$ per mL in seawater, among which a few are cultivable [12].

In this work, the diversity of halophilic microorganisms and functional diversity were determined in two thalassohaline environments, Acos and Maras, that have physicochemical differences in salinity and pH. We expected that these intermediate-salinity environments would contain a greater microbial diversity than high-salinity environments and with a particular microbial community structure given the high altitude. Thus, we consider that this analysis opens diverse opportunities to describe the microbial diversity and functional profile within the Peruvian hypersaline systems and will contribute to knowledge in these environments. This is the first characterization of a microbial community structure of intermediate salinity in samples collected from Peruvian high-altitude salterns.

2. Materials and Methods

2.1. Sampling, DNA Extraction, and Sequencing

Water (20 liters) was collected during the rainy season (January 2018) from two points where the water emerges in the mountain in two hypersaline systems located in Cusco, Peru. The first is in Maras (13°57′59.3″ S, 71°05′65″ W), and the second is in Acos (11°16′25″ S, 72°9′15″ W). The samples were obtained with sterilized tools and containers, and salinity and pH were measured in situ using a hand refractometer (Spectronic Instruments Inc., Rochester, NY, USA) and pH potentiometer (HANNA

Instruments, Portugal), respectively. All samples were transported to the laboratory under refrigerated conditions, where liters of water were filtered through 0.22-µm Millipore filters. The DNA was purified from the filters by using ZymoBIOMICS DNA kits (MoBio, West Carlsbad, CA, USA). The DNA concentration was determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific), and fluorometry was measured using a Qubit 4 fluorometer (Invitrogen). The DNA was sequenced using the Illumina NextSeq 500 platform with the Nextera V2.0 kit (150 bp, 2 × 75 bases) at the Instituto de Biotecnología of Universidad Nacional Autónoma de México.

2.2. Quality Control and Assembly

The quality control of sequences was performed by FASTQC v0.11.4 software [14], and duplicated sequences were removed using CD-HIT-DUP v4.7 [15] with a maximum mismatch number of 0.03. Reads were assembled in contigs using MEGAHIT v1.1.2 [16] under default parameters in paired-end mode, and contigs of a minimum length of 1000 bp were considered for further analysis.

2.3. Microbial Community Taxonomic Assignments

Taxonomic assignments were performed with software Kaiju v16.0. In addition, we used MetaGenome Rapid Annotation Subsystems Technology (MG-RAST v4.03) [17], which compares the assembly sequences with a comprehensive non-redundant database sourced from the National Center for Biotechnology Information (NCBI) databases, and SEED, which categorizes gene function into five levels of resolution. An expected value (E) cutoff of 10^{-5} was employed for taxonomic classifications. Raw data of Metagenomes have been deposited in MG-RAST with accession numbers: mgm4810306.3, mgm4808260.3, and mgm4810472.3.

For virus classification, the viral contigs were achieved with VirSorter v2 [18], and these contigs were classified with MEGAN v5.10.6. For fungi classification, the sequences were compared against a constructed database comprised of 35,296 complete and draft fungi genomes from NCBI. For both viruses and fungi, the best-scoring BLAST results with an E-value of 10^{-6} were parsed, and the taxonomic assignment was determined using MEGAN software [19]. The lowest common ancestor (LCA) method in MEGAN was used for taxonomic assignment, with the following parameters: minimum support of 2; minimum score of 50; top percent of 10.

2.4. Diversity Index

The taxonomic profiles at the species level were used to calculate the diversity indices from all data, and different alpha diversity descriptors were obtained using the Phyloseq function in R v3.3.3 [20]. The beta diversity was determined by Bray-Curtis dissimilarity, and the sampling effort was evaluated through the rarefaction curves using a Vegan library implemented in R [21].

2.5. Genome Reconstruction

The reconstruction of the bacterial genome was directed to those species that had the highest abundance according to the taxonomic classification. The genomes were retrieved using the strategy fragment recruitments within Bowtie2 v2.2.6 [22]. The coverage was evaluated using BBmap v38.25 [23], and the consensus sequence was inferred using UGENE v1.31.1 [24]. For the reconstructed genome, the presence of contamination was evaluated using One Codex [25] and Genome Peek. Briefly, One Codex assigns an unknown nucleotide sequence for the identification of k-mers of fixed size k-31 in comparison with its own database. Genome Peek extracts the 16S gene and *radA*/*recA*, *rpoB*, and *groEL*, the principal molecular markers, from a genome for taxonomic identification. The annotation was achieved using Prokka v1.12 [26] and visualized with Genome Atlas.

For viral sequences, identification was achieved by VirSorter [18] and was based on viral hallmark genes annotated as "major capsid protein," "portal," "terminase large subunit," "spike," "tail," "virion formation", and "coat," among others. The entire contig was considered viral if more than 80% of predicted genes on a contig had a viral signal. This software finds new viruses at different confidence

levels, with scores of categories 1 to 4, with 4 being the highest confidence level. Viral sequences identified within category 1 by VirSorter were visualized with the easyfig v2.2.2 tool and also assessed with the PHAge search tool (PHAST) [27].

Finally, contigs with lengths of ≥10 kbp within category 2 ("quite sure") in VirSorter were translated into protein sequences and classified taxonomically using the vConTACT v2 software [28] with default parameters (https://bitbucket.org/MAVERICLab/vcontact), with the aim of classifying these possible new viruses.

2.6. Binning for Putative Genomes

Assembled contigs were clustered into bins or metagenome-assembled genomes (MAGs), using MaxBin v2.2.4 [29]. Briefly, MaxBin performs genome reconstruction from metagenomes based on two genomic characteristics, tetranucleotide frequencies and the level of bin coverage, using single-copy marker genes. The two metagenomes from Acos were used to recover the MAGs, which were later annotated with Prokka [26].

From the annotation of MAGs, the ribosomal sequences were extracted in single copy (L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, and S19), and then these sequences were aligned with those reported by Hung et al. [30] by using MAFFT v7.005 for taxonomic identification [31]. The phylogenetic analysis was performed using FastTree v2.1.7 [32], which considers an approximate maximum likelihood with 100 bootstrap replicates. Finally, the phylogenetic tree was displayed using ITOL [33].

2.7. Functional Analysis and Biogeochemical Cycles

Prodigal v2.6.3 [34] was used for predicting protein-coding genes in the assembled contigs by using the metagenomic mode, and the functional assignment was achieved using SUPERFOCUS [35], which contains the SEED database with an E-value of 10^{-5}. From functional abundance tables, a heatmap using the ggplot2 library [36] and RColorBrewer library in R (www.ColorBrewer.org) was generated. Finally, microbial metabolic pathways involved in the biogeochemical cycles for carbon, sulfur, nitrogen, hydrogen, iron, and oxygen were identified using the Multigenomic Entropy-Based Score pipeline (MEBS v1) with a false-discovery rate of 0.0001 [37].

3. Results and Discussion

3.1. Site Characterization and Field Sampling

The water samples were collected from two locations in the district of Cusco, Peru. The first sample was collected from Maras; its pH was 7 and its salinity concentration was 23% NaCl (Figure 1). This concentration was slightly lower than previously reported (25% NaCl) in emergent water, whereas in the crystallizer ponds the concentration was higher (30% NaCl) [1].

The second and third samples were collected from Acos, with a pH of 7.9 and 19% salinity (Table 1). The salinities of the thalossohaline water samples from Maras and Acos [1,2] were similar to levels in other solar salterns with intermediate salinity, such as Marine Saltern in Santa Pola, Spain (13–19% NaCl) [7,38] and Saltern in Isla Cristina, Spain (21% NaCl) [38,39]. In this regard, salterns exhibiting an intermediate salinity have been found to contain a greater diversity of microorganisms than salterns with higher salt concentration [38]; the concentration of NaCl defines the diversity and structure of the microbiome in these environments [40].

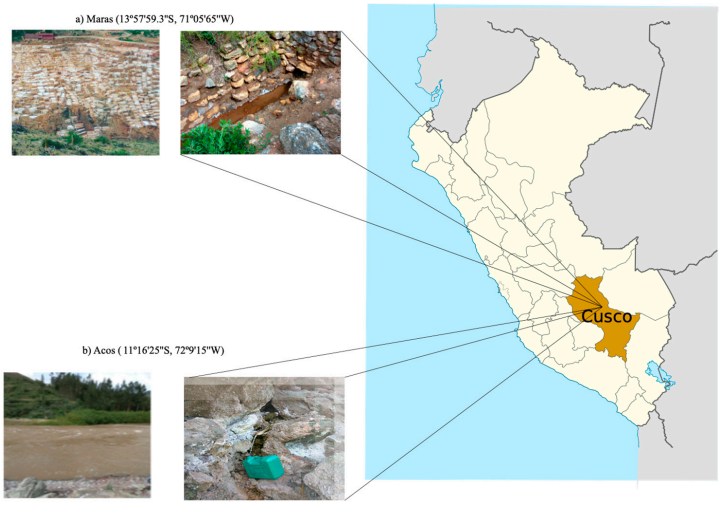


Figure 1. Location of the field sites in Cusco Peru. (**a**) Maras with 3000 shallow ponds. (**b**) Acos at the origin of the water.

Table 1. Sequence features of hypersaline metagenomes from Cusco, Peru.

Data Set	Salinity	pH	Number of Paired-End Reads	Number of Contigs Assembled	Sequences		Taxonomical Classification			
					Classified	Unclassified	Bacteria	Archaea	Eukarya	Viruses
Acos 1	19%	7.9	63,387,998	257,314	71%	29%	57%	14%	2%	0.2%
Acos 2	19%	7.9	79,304,621	256,430	71%	29%	57%	16%	2%	0.2%
Maras	23%	7.0	56,086,809	2650	70%	30%	56%	11%	1%	1.32%

3.2. Community Structures of Intermediate Hypersaline Systems

In order to analyze the diversity, abundance, and genes involved in metabolic profiles of samples from Maras and Acos, shotgun metagenomic sequencing was performed. Maras and Acos salterns can be considered environments at high altitude with intermediate salinity (according to the determined percentage of salt) (Table 1). However, salinity is not the only parameter that modifies the abundance and diversity of microorganisms present in these ecosystems; biogeographic patterns that may also have a role include altitude, remoteness of these environments, oxygen availability, alkalinity, altitude, and UV irradiation [41–43].

With the metagenomes obtained from the two locations, the general structure of the microbiome was determined. To this end, the sequences were classified with Kaiju (Table 1), and the results showed a high abundance of bacterial organisms (~57% of the sequences), followed by Archaea (~16%). These results contrast with the abundance reported in crystallizer ponds in Maras, where the salinity of >30% NaCl showed a microbiota dominated by Archaea (80–86% of total counts) with much lower percentages of Bacteria (10–13%) [1].

The enrichment analysis of species and diversity in these sites, evaluated with Chao, Shannon, and Simpson indexes, revealed that Acos samples had a greater richness than Maras samples (Supplementary Table S1). These results correlated with the rarefaction curves, i.e., in Acos samples, the asymptotic distribution was reached, which indicates a greater diversity showing correlation to the other diversity indexes, whereas in the Maras sample the asymptote was not reached, since most of the contigs were assigned to *Cutibacterium acnes*, which is highly unlikely to reside in this environment and was considered a contaminant and was therefore eliminated from diversity curves and subsequent analyses.

However, the remaining organisms present in this sample are halophilic, but as shown in the diversity curve it is necessary to perform new sampling to know the diversity in Maras (Supplementary Figure S1). In addition, the Bray-Curtis dissimilarity index was performed to evaluate the beta diversity, showing an index equal to 1, which indicates a different species composition between Maras and Acos. In contrast, the index value between the two samples from Acos was close to zero, suggesting that these samples contained the same species (Supplementary Figure S2).

These results correlate with findings reported for other saltern ponds with intermediate salinity, such as those in Santa Pola, Spain, with 13–19% NaCl, where high abundance levels of bacteria (~73 and ~54%) and archaeal organisms (~27% and ~46%, respectively) were found [38,44]. The same was found when the Chao index was compared for these metagenomes [45]. In contrast, in the saltern pond located in Isla Cristina, Spain (21% NaCl), Archaea were predominant (~84%), followed by Bacteria (~16%) [38]; although the structure at the phylum level is equivalent, important differences at the genus level are attributed to particular local ecological conditions [38].

These results suggest that in environments with higher salt concentrations there is less diversity and species richness, probably because there is lower availability of nutrients and oxygen, in contrast to intermediate-salinity environments, where there is a greater availability of nutrients and oxygen. Therefore, salt concentration is an important factor that shapes the structure of the microbial community in hypersaline environments and determines its diversity and abundance.

3.3. Bacterial and Archaeal Community Composition

Previous studies have shown that the halophilic world is highly diverse, but this diversity is reduced with increasing salt concentrations [46]. In the case of intermediate-salinity environments, several moderately halophilic bacteria have been reported, including *Halomonas*, *Salinivibrio*, *Halobacillus*, *Thalassobacillus*, *Bacillus*, *Salinicoccus*, *Idiomarina*, *Chromohalobacter*, and *Salinicoccus* [7,38,47–49]. In the metagenomic samples from Maras, bacteria from the phylum *Proteobacteria* (38%) were the most abundant, followed by *Actinobacteria* (11.58%), *Firmicutes* (2.68%), *Cyanobacteria* (0.40%), *Bacteroidetes* (0.40%), *Deinococcus-Thermus* (0.26%), and *Verrucomicrobia* (0.26%) (Figure 2).

At the species level in Maras salterns, it was interesting that the most abundant bacterium was *Thiohalorhabdus denitrificans* (11.51%), which is an extremely halophilic species [50], followed by *Thiohalospira halophila* (0.87%) [51]. Both of these species are chemolithoautotrophic sulfur-oxidizing bacteria which use thiosulfate as the electron donor [50,51], and neither has been reported previously in intermediate-salinity settings.

Other halophilic bacteria, such as *Pseudomonas* (2.15%) and *Halomonas* (0.94%), were identified in lower proportions than previously reported [1,44,47]. Even the main bacteria described in hypersaline systems, such as *Salinibacter ruber* [52,53] and *Rhodovibrio salinarum* [1], were found in low abundance (~0.07%, each species) in our study, probably because the altitudes of these sites affect bacterial structures, as we have shown.

In addition, predominant non-halophilic bacteria found included *Lawsonella clevelandensis* (7.06%), *Escherichia coli* (2.08%), *Clostridium difficile* (1%), *Cutibacterium acnes* (0.9%), and *Ralstonia solanacearum* (0.8%). The presence of non-halophilic bacteria in hypersaline environments has been previously described in the Santa Pola saltern (19% NaCl), and some of these organisms have developed adaptation mechanisms, such as a strong GC bias, as has been identified in halophilic organisms as a strategy to avoid UV-induced thymidine dimer formation [44,45,54,55].

The two samples from Acos exhibited similar compositions of microorganisms: *Proteobacteria* corresponded to ~59% of identified sequences, followed by *Bacteroidetes* (11%), *Balneolaeota* (6%), *Firmicutes* (5%), and *Actinobacteria* (2%). Both Acos metagenomes had the same composition as environments of intermediate salinity previously reported, showing a high diversity and abundance of bacteria [7,38]. Interestingly, in Acos salterns members of the *Balneolaeota* phylum were identified, including moderate halophiles (5–10% NaCl) abundant in sediments, saline soils, and marine habitats [55,56] (Figure 2b,c).

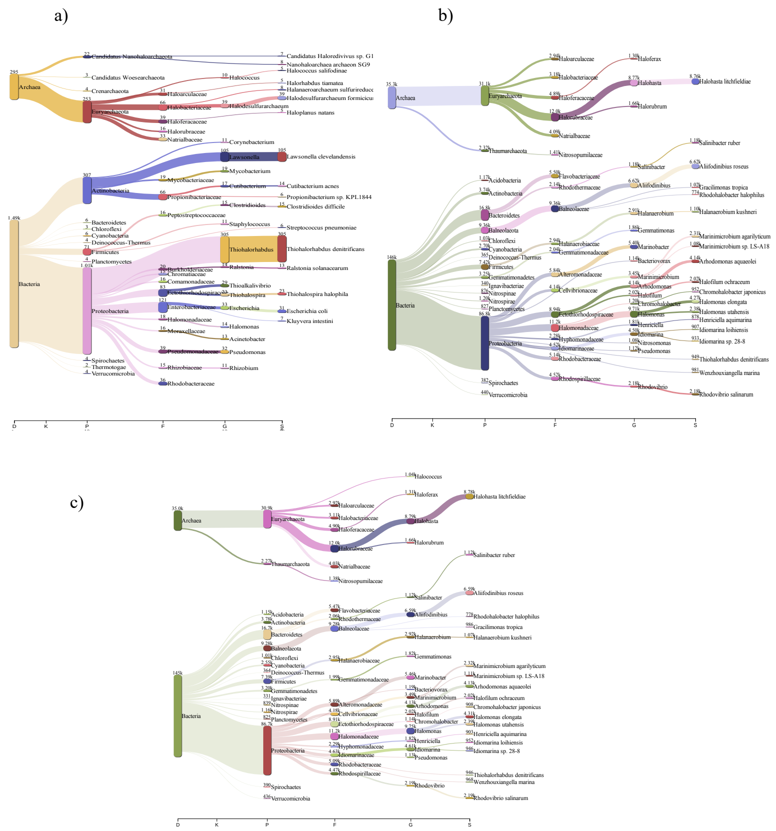


Figure 2. Taxonomic profile in hypersaline metagenomes from Cusco, Perú. (**a**) Maras; (**b**) Acos 1; (**c**) Acos 2. On the *x*-axis are the taxonomic levels: D, domain; P, phylum; C, class; O, order; F, family; G, genus; S, species. Numbers correspond to the assigned contigs.

At the level of genus, *Halomonas* was the most abundant (8.4%), with more than 70 different species identified in Acos; *Halomonas elongata* (2.8%) was the most abundant, followed by *Halomonas utehensis* (1.6%). In this regard, organisms of the *Halomonas* genus are aerobic heterotrophic, halo-alkaliphilic, and sulfur-oxidizing bacteria and are commonly found in intermediate-salinity, high-altitude environments [57,58]; they are also a source for the production of bioplastic polyhydroxyalkanoates [59].

In contrast, at the species level, the most abundant bacteria were *Aliifodinibius roseus* (~5%) within the phylum *Balneolaeota*; this species is considered moderately halophilic (6–10% NaCl for optimal growth). Also abundant were two species, *Halomonas elongata* (2.93%) and *Arhodomonas aquaeolei* (2.84%), an obligately halophilic bacterium with optimal growth at 15% NaCl; both of these species have been shown to degrade phenol [60]. To our knowledge, only a few reports have described these bacteria in a metagenome from an intermediate-salinity environment. *Marinimicrobium agarilyticum*, *Rhodovibrio salinarum* (1.50%), *Salinibacter ruber* (0.80%), and *Idiomarina* sp. (0.64%) were in low abundance. *Idiomarina loihiensis* is a bacterium identified in environments a wide range of temperatures (from 4 °C to 46 °C) and salinities (from 5% to 21%) that presents polyextremophile behavior [61].

In both the Maras and Acos sites, the low abundance of *S. ruber* is understandable, since this bacterium prefers environments with higher salinity.

Therefore, different species of moderately halophilic bacteria were found in Acos, with *Proteobacteria* the most abundant. These results correlate with findings from another high-altitude saltern located in Atacama, Chile, at 2,700 m above sea level, where halophilic bacteria able to grow at intermediate salinity were isolated [62]. In general, the moderately halophilic bacteria are aerobic or facultative anaerobic microorganisms that belong to different genera, as part of a physiologically heterogeneous group of bacteria [47].

In intermediate-salinity salterns, such as the Peruvian hypersaline systems, the abundance of archaeal organisms is low, as found in the Maras samples, where *Euryarchaeota* organisms were found to be highly abundant, followed by "*Candidatus* Nanohaloarchaeota," and "*Candidatus* Woesearchaeota." In both samples from Acos, *Euryarchaeota* organisms were the most abundant, followed by *Thaumarchaeota*, *Crenarchaeota*, and "*Candidatus* Bathyarchaeota." Within the *Euryarchaeota* phylum, the *Halobacterium* family was found to be predominant, similar to findings from other salterns and salty lakes [6,45,63,64].

In Maras, *Halodesulfurarchaeum formicicum* was the most abundant species. *Halodesulfurarchaeum* is a novel anaerobic genus that was discovered in a deep-sea salt-saturated anoxic environment and in sediments from hypersaline lakes [65].

In Acos, the most abundant archaeon was *Halohasta litchfieldiae* (~3.5%), a chemoorganotrophic aerobic that can grow in salt concentrations around 12–28%, presenting adaptation to low temperatures [66–68] as occurs in the area of the Peruvian Andes where minimum temperatures reach between −7 °C and −4.4 °C.

The taxonomic assignment analysis was also carried out with MG-RAST; the abundance of archaea was low, in accordance with the results of Kaiju. From the class *Halobacteria*, 14 different genera were identified, with *Haloarcula* genus the most predominant in the three samples (Figure 3). The presence of this genus is interesting because it has been reported to be involved in recombination processes. This process could be occurring between bacteria and archaea, among their sharing genes, for example, rhodopsin family genes which are common and have different functions such ion pumps, channels, enzymes, photosensory receptors that could favor the adaptation [69,70].

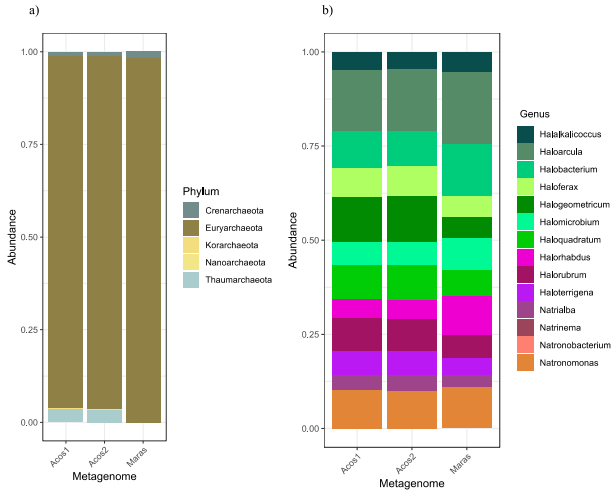


Figure 3. Taxonomic classification of archaea according to MetaGenome Rapid Annotation Subsystems Technology (MG-RAST). (**a**) Composition of the archaeal community at the phylum level, where *Euryarchaeota* prevail. (**b**) Diversity within the class *Halobacteria.* The genus *Haloarcula* prevails in all samples.

3.4. Composition of the Viral Community

Although viruses are sources of genetic variation, as they can modify a genome's plasticity and alter the structure of populations and also biogeochemical cycles, few reports have described the structure of virus communities in hypersaline environments [71–74]. In this work, the taxonomic assignment performed with Kaiju revealed that 0.2% to 1.32% of the reads were associated with viruses (Table 1). This was probably because we did not perform a viral enrichment with our samples; however, it was possible to find viruses, because they would be included within the host cells or in the form of proviruses [75].

Because of the low percentage of detected viruses in the samples, we used Virsorter, which detects the viral signal in metagenomic datasets [18]. From the assembly of reads, we identified the viral contigs according to VirSorter, and they were subsequently classified taxonomically with MEGAN (See Materials and Methods).

The results identified the order *Caudovirales*, specifically, the *Siphoviridae*, *Podoviridae*, and *Myoviridae* families, in the samples; indeed, these families seem to be ubiquitous in marine environments [76].

Interestingly, in Maras eukaryotic viruses such as *Adenovirus* and *Herpesvirus* were identified, probably as a consequence of the composition of eukaryotic organisms in the samples, as also reported for Red Sea brines [77]. Additional double-stranded DNA (dsDNA) viruses associated with eukaryotes were also found in the Acos samples, mainly viruses from the *Phycodnaviridae, Poxviridae*, *Mimiviridae*, and *Pandoravidae* families (Figure 4); all of these are Megavirales, which are nucleocytoplasmic large DNA virus (NCLDVs). NCLDVs infect animals and unicellular eukaryotes [78] found in other hypersaline environments, such as the Salton Sea in the United States and Organic Lake in the Antarctic [79].

Another important group of viruses found in Acos was an unclassified archaeal dsDNA virus (Figure 4); this virus has been reported in high abundance in hypersaline environments, with spindle-shaped morphologies of *Haloarchaea* viruses, but this happens when salt concentration reaches saturation, where in general *Archaea* are predominant [12,72,80].

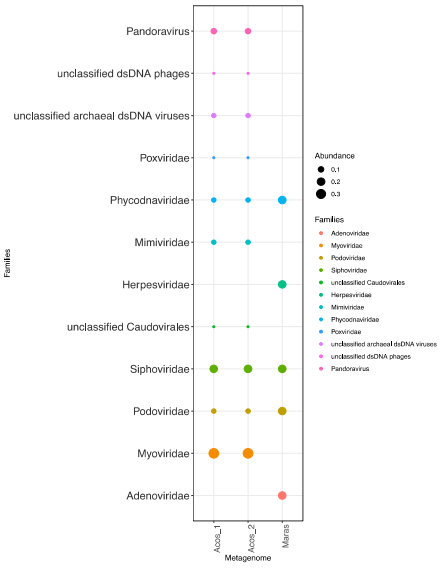


Figure 4. Compositions of the viral communities at the species level in Acos and Maras.

In summary, we identified six virus families associated with eukaryotic cells and five families that infect *Bacteria* and *Archaea*. This last group was the most abundant, according to the microbial composition in this environment.

3.5. Composition of Fungal Communities

The diversity of microorganisms in intermediate-salinity systems is not restricted to prokaryotes. Approximately 2% of the reads corresponded to eukaryotes. According to the Megan classification system, two phyla of fungi were found, *Ascomycota* (with 85%) and *Basidiomycota* (with 10%), as has been reported for other hypersaline environments [81]. At the family level, the most abundant were *Arpergillacea*, followed by *Sordariaceae*, *Sporidiobolaceae*, and *Chaetomiaceae* (Figure 5). *Aspergillus* has been reported to be dominant in salterns of Slovenia, along with *Cladosporium* and *Penicillium* [82]. These filamentous fungi are ubiquitous and have been isolated with high frequency in hypersaline environments [83]. Some species in the *Sordariaceae* family have also been isolated from hypersaline environments. The *Sporidiobolaceae* family is within the *Basidiomycota* phylum, which has been recovered from sea water, glacier ice, and extremophile environments. *Rhodotorula* was recovered from hypersaline ponds in Israel [84]. The *Chaetomiaceae* family was recovered together with 19 inhabiting hyphomycetes fungi in soils from the hypersaline Urmia Lake [85]. Thus, a high diversity of fungi has been found in hypersaline environments, where the most abundant are melanized *Aspergillus*, which is a ubiquitous genus used in biotechnology applications for its production of citric acid and enzymes, and non-melanized *Rhodotorula*, which comprises several species that can be used in bioremediation [86].

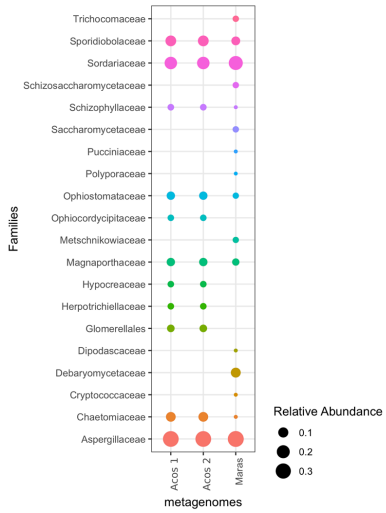


Figure 5. Diversity taxonomy of fungi. The *Aspergillaceae* family was the most abundant in Acos and Maras.

In summary, the results show an important diversity of fungi within the hypersaline environments; however, their functions are still unclear.

A comparison between *Halomonas elongata* strain HEK1 (MAJD01000001.1), *H. elongata* strain MH25661 (QJUB00000000.1), and the two genomes recovered from this work revealed that all strains share a core of 2,984 genes. Indeed, both genomes recovered from Acos share 916 genes, a significant number of genes in comparison to the other genomes. In addition, *Idiomarina loihiensis*-like, reported as a predominant genome in a saline environment, was also assembled. The coverage was 74–78% with a reference genome (*I. loihiensis L2TR* GCA_000008465.1), and the deep coverage was 5-6x for Acos 1 and Acos 2. Both strains share 93.0% of identity according to their ANIs. Finally, the two genomes share 2151 common genes (Supplementary Figure S3).

Interestingly, the reconstructed genomes of *Halomonas elongata* and *Idiomarina loihiensis* show different strategies for maintaining osmotic equilibrium, according to the annotation; de novo synthesis of the ectoine pathway is complete in *H. elongata*. Ectoine is a compatible solute of low molecular weight of aspartate metabolism, which is produced when there are increased K^+-glutamate levels [89]. In contrast, in *Idiomarina loihiensis*, de novo synthesis of ectoine was absent; however, we identified genes encoding ABC transporters such as the ATP-dependent Na+ exporter *natAB*, in addition to other iron transporters, which promote detoxification in hypersaline environments. These findings show different adaptation strategies of bacteria in hypersaline environments.

The annotation of the genes exclusively shared between two *Halomonas* genomes from Acos revealed that most of them were related to nitrogen metabolism, chemical reactions, and pathways involving organic acids. Regarding the genes related to the metabolism of nitrogen, genes encoding a nitrate/nitrite sensor protein, nitrate reductase, and ammonia monooxygenase were found. This is interesting since *Halomonas* use nitrogen as the last acceptor of electrons even in conditions of low oxygen, as is the case in hypersaline environments [90].

In general, these *Proteobacteria* play an important role in the nitrogen cycle, through recycling of nitrogen by assimilation of gaseous nitrogen from the atmosphere and decomposition of organic matter, causing nitrogen to be constantly available [90].

3.6.2. Reconstruction of Viral Genomes

Traditional techniques limited us in obtaining viral genomes, but through metagenomics it was possible to reconstruct these genomes, allowing us to expand knowledge about the influence of viruses in this particular environment. In this regard, the viral contigs identified correspond to bacteriophages, as expected, since bacteria were more abundant in our metagenomes. In the Maras sample, a genome with 97% similarity with the lambda phage of *Enterobacteria* (*Siphoviridae* family) was found. This phage infects *Escherichia coli*, a non-halophilic bacterium that was abundant in this sample (Figure 7a).

In Acos samples, around 100 different contigs with viral signals were identified; because many of these could be fragments of viral sequences, different criteria were used, such as the presence of inverted terminal repeats in the case of circular genomes and similarities in size lengths with a reference genome (no more than 10% of size length) [18].

In the Acos samples, two phages of *Halomonas elongata* were recovered. This finding was somewhat expected, since *H. elongata* is abundant in these intermediate-salinity environments, but this is the first time that bacteriophages have been reported in this bacterium. The two recovered phages have a size length of approximately 28 Kbp, and a comparative analysis with two ΦHAP-1 reference genomes revealed that they have the same pattern of synteny and a protein identity greater than 65%. (Figure 7b).

The phage used for comparison was *Halomonas* phage ΦHAP-1. This is a *Hapunavirus* belonging to the family *Myoviridae* and was isolated from *Halomonas aquamarina*. The GC content in ΦHAP-1 is 59%, which is slightly lower than other phages such as ΦHAP-1, found in Acos with a 64% GC content, and similar to the GC content of the host genome (*H. elongata*) [91].

The ΦHAP-1-type phages from Acos have 40 putative open reading frames (ORFs) with 6 genes fewer than the reference genome. Genes coding for proteins such as the RepA replication protein, the prophage repressor, the prophage antirepressor, and the protelomerase were not identified; the latter is necessary for the maintenance of the linear state of the prophage within the host genome [92].

In addition, inverted repeats were found at positions 28,428–28,452 to 28,455–28,479 with a length size of 25 bp and an identity of 100%, suggesting that the genome is in a circular form, because this kind of inverted repeated sequence is usually found in regions processed by protelomerases and originates from the release of phage with covalently closed ends. All of these findings suggest that the phage could be in their free form.

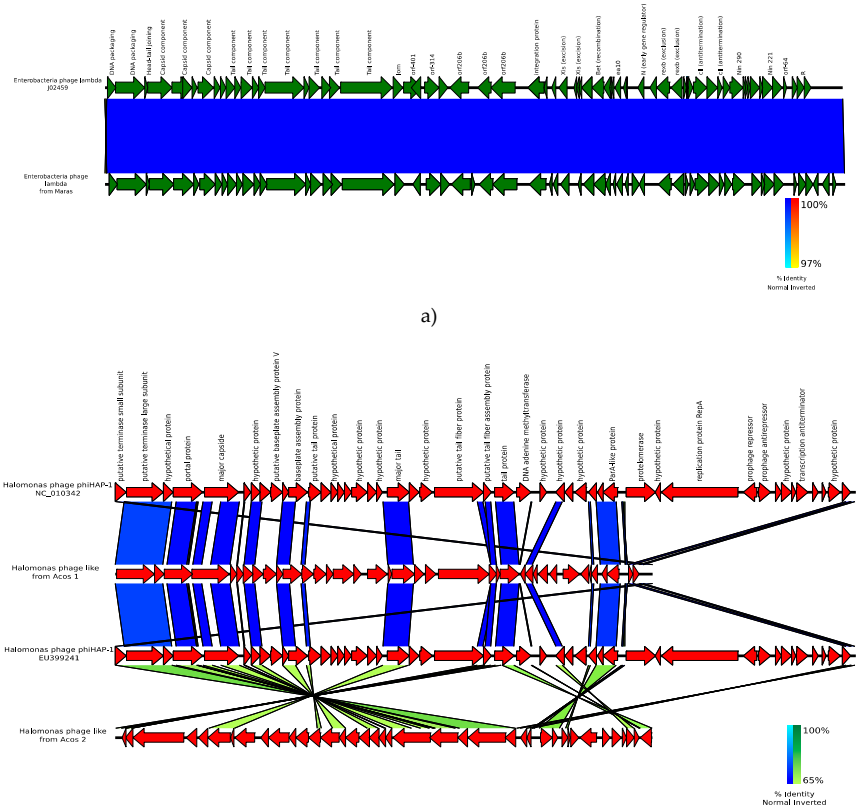


Figure 7. Genomes of novel bacteriophages. (**a**) *Enterobacteria* phage lamda-like from Maras; (**b**) *Halomonas* phage-like (phiHAP-1) from Acos.

The rest of the viral sequences obtained in the metagenomes through BLAST analysis had very poor identity with sequences in the NCBI database, and they were used for taxonomic or functional allocation. Thus, the contigs of >10 kb was clustered using the Viral RefSeq database and vConTACT2; this tool allows classification of viral sequences with protein comparisons. In Figure 8, two examples of viral assignation taxonomy are presented. In Figure 8 a viral sequence with ~27 Kpbs shares identity with proteins from *Cellulophaga* phage, which infects algae typically found in marine environments. In Figure 8 are four viruses with size lengths of about ~11 to 30 Kpbs that shared identities with different enterophages, showing a mosaicism as a reflection of horizontal gene transfer. In total, 27 sequences could have a taxonomic assignment as new viruses with this strategy (Supplementary Figure S4).

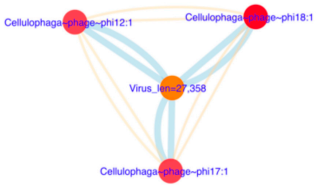


Figure 8. Protein-sharing network of genome of *Cellulophaga* phage. Yellow lines indicate strong similitarity, and blue lines indicate weak similarity. Thus, the virus of length 27,358 bp could be a novel virus.

3.6.3. MAGs

Another strategy to retrieve new genomes from metagenomic sequences with little or no identity with sequences already reported is by binning, in which genomes are assembled without a reference sequence. The binning method has the aim to classify contig sequences in a specific taxon, called metagenome-assembled genomes (MAGs). The binning methods can also describe novel species in these environments. The binning of metagenomic sequences was performed only for Acos, because at least two samples with the same origin are necessary to enrich the data. From this, a total of 42 bins were assembled in annotated draft genomes, and their ribosomal genes were extracted. However, some of this process resulted in a low degree of completeness. Therefore, we performed a phylogenetic analysis that revealed that 31 MAGs were classified within a specific domain and seven of them were closely related to the *Halobacteria* class (Bin 5, Bin 14, Bin 15, Bin 31, Bin 30, Bin 34, and Bin 41) (Figure 9) within the *Euryarchaeota* phylum, which is predominant in hypersaline environments.

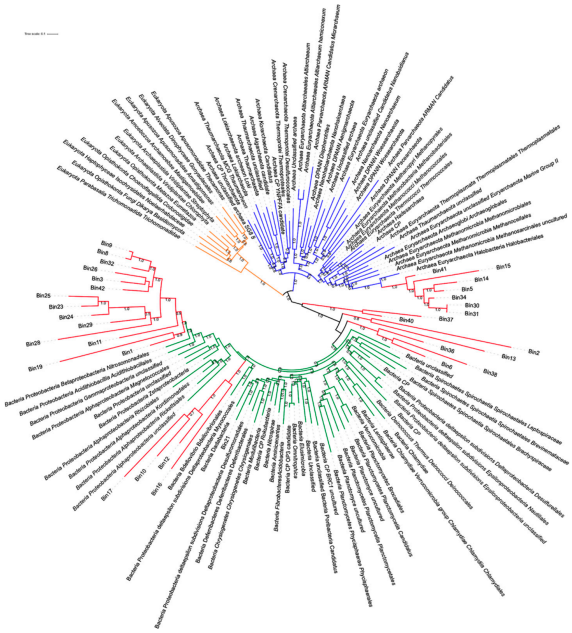


Figure 9. Phylogenetic tree, including the binned sequences associated with different taxa clades. Branches in orange correspond to Eukaryota; branches in blue correspond to Archaea; branches in green correspond to Bacteria; Branches in red correspond to bins. Bootstrap levels are noted.

Four MAGs were closely related with *Alphaproteobacteria* unclassified (Bin 10, Bin 12, Bin 16, Bin 17), and 14 MAGs were closely related to *Gammaproteobacteria.* Indeed, most of the found bacteria corresponded to *Proteobacteria*, in accordance with the bacteria found in our metagenomes (Bin 1, Bin 3, Bin 8, Bin 9, Bin 11, Bin 19, Bin 23, Bin 24, Bin 25, Bin 26, Bin 28, Bin 29, Bin32, Bin 42). Therefore, in this analysis the *Proteobacteria* phylum prevails over the *Euryarchaeota* phylum in Acos samples, indicating that salinity plays an important role in the structure of the community of microorganisms that inhabits this ecosystem.

The strategies for the reconstruction of genomes, such as fragment recruitment and MaxBin, offered different results, since the first strategy is a targeted search and for the second strategy the search starts from scratch in obtaining draft genomes. Another important difference is that the latter assemblies, coming from more than one metagenome, could build chimeric genomes. However, with the two strategies, genomes of *Gammaproteobacteria* similar to *Halomonas* could be reconstructed.

3.7. Functional Community Composition

The strategies that halophilic organisms use to survive in hypersaline environments are diverse and include thickening of the cell wall, increase in pigmentation, production of compatible solutes, solute transport mechanisms, and production of antibiotic proteins to limit the growth of other populations [93]. Therefore, we analyzed the functional composition of microorganisms in intermediate-salinity environments in order to determine how these mechanisms are potentially used by microorganisms in these environments.

Thus, the contigs from hypersaline metagenomes were annotated using SEED subsystems, and these results revealed that 11–13% of coding sequences from Acos and 14% of those from Maras were related to metabolism of carbohydrates (central carbohydrate metabolism, synthesis of monosaccharides and polysaccharides) (Figure 10).

The genes classified into the category of amino acids and derivatives functions were present in ~8% to 12% in Acos and ~11% in Maras. Overall, in the three metagenomes the synthesis of lysine, threonine, methionine, and cysteine were the more abundant categories. This correlated with the fact that in some halophilic bacteria there is a preferential use of codons to encode these amino acids [94]. In this regard, most of these amino acids are hydrophobic, found on the inside of proteins, especially in hypersaline environments, which strengthen the hydrophobic interactions [92].

Other categories overrepresented, with ~6.7% to ~9.77% abundance in samples, were respiration, functions related to donating/accepting electrons, and ATP synthases. All of these participate in the transfer of electrons to obtain energy.

Interestingly, the category related to pigment functions was found in 8% to 10% Acos sequences and 7% of Maras sequences. The class *Halobacteriaceae* is mainly responsible for α-bacterioruberin pigment, a pink-red product in hypersaline environments. In addition, *Salinibacter ruber* is responsible for producing salinixanthin carotenoid, a C-40 acyl glycoside carotenoid that also contributes to the coloration of salterns. This bacteria and these pigments are important in hypersaline environments as they reduce the UV irradiation that damages DNA, which tends to be high in these saline environments [9].

The category of membrane transport was present in ~3% to ~5% abundance; in particular, the membrane proteins in Gram-negative bacteria were more abundant than in Gram-positive bacteria, including the YrbG Na^+/Ca^{2+} cation antiporter, a very important protein in this kind of saline environment. This system has been reported in *Haloarchaea*, which have a wide variety of ion transporters, to have a role in regulating fluctuating salinity levels and avoiding osmotic shock [95]. In other salterns with intermediate salinity, such as Santa Pola (13% NaCl), genes related to this function have been reported to be overrepresented [45].

The function of resistance to antibiotics and toxic compounds was also found to be abundant in these metagenomes, including pathways involved in sulfur heavy metal cycling, cobalt–zinc–cadmium resistance, and also copper homeostasis and resistance to arsenic. Since heavy metals such as arsenic

do not have biological roles, low concentrations are toxic to the cell, and therefore microorganisms have mechanisms for reduction. Many Archaea have different heavy metal transporters [96].

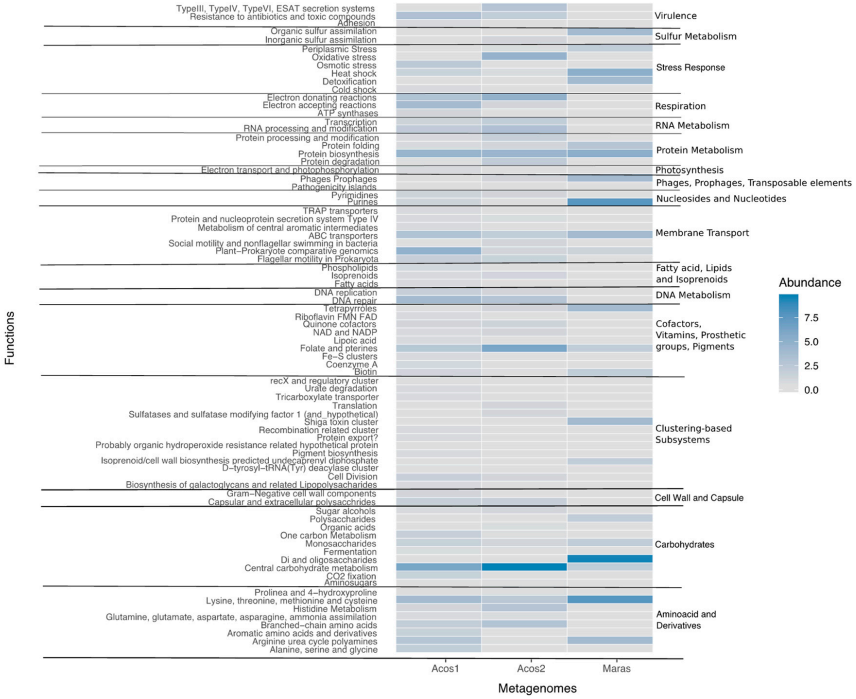


Figure 10. Heatmap of the relative abundance of proteins based on SEED classifications.

In addition, the stress category was more abundant in Maras (~14%) than in Acos (~6%). This category includes predominant functions such as oxidative stress, osmotic stress, heat stress, detoxification, cold shock, and periplasmic stress. These types of functions were prevalent in Maras, and although the concentration of salt was higher in Maras than in Acos, a greater presence of non-halophilic organisms was identified in Maras. It is well known that organisms growing in high concentrations of salt accumulate stress molecules, such as reactive oxygen species, and the organisms must therefore have mechanisms for their detoxification [97].

In the same way, samples from Acos presented abundant oxidative stress functions. Reactive oxygen species in hypersaline environments are common, thus organisms in these environments have detoxification mechanisms. In particular, in microaerophilic and anaerobic metagenomes, oxygen-detoxifying enzymes have been identified, such as superoxide dismutases, catalases, peroxidases, and glutathione peroxidase [98]. In the Acos metagenome, we identified enzymes involved in the response to oxidative stress, such as 5-oxoprolinase, and enzymes responsible for maintaining the reducing environment, such as glutathione reductase, glutathione hydrolase (involved in reduction of glutathione disulfide), and hydroperoxide resistance (responsible for detoxification of organic hydroperoxides).

Since the microorganisms are under oxidative stress, it is common to identify redundant enzymes responsible for DNA repair [99]. However, reactive oxygen species are not the only compounds that modify the genetic material; other agents that produce data in the genetic material include UV light

exposure and desiccation, and so, as expected, the functions of DNA synthesis and DNA repair are the most represented in proteins found in the Acos samples.

According to our analysis with MG-RAST, detoxification enzymes were identified in Archaea, within the classification "housecleaning nucleoside triphosphate pyrophosphatases"; all of these belonged to class Nudix hydrolases, including nucleoside 5-triphosphatase and 5-nucleotidase SurE. In Bacteria, these enzymes were found in high abundance, as was the dimeric dUTPase enzyme. Interestingly, viruses also possess detoxification enzymes of this category, in particular the enzyme deoxyuridine 5′-triphosphate nucleotidohydrolase, that decreases the intracellular concentration of dUTP so that uracil cannot be incorporated into viral progeny DNA.

All of the above enzymes are responsible for the elimination of damaged nucleotides caused by reactive oxygen species. For viruses, the incorporation of damaged nucleotides in nucleic acids is detrimental to replication of viral progeny. In this way, the virus could contribute to the adaptation of the host to its environment.

Regarding DNA repair, we found bacterial systems that contribute to this function, among which were base excision repair (BER), repair of DNA double-strand breaks (DSBs) (RecBCD pathway and RecFOR pathway), nucleotide excision repair (NER), and DNA mismatch repair (MutL-MutS system). However, the mechanisms of nucleotide excision repair (NER) and DNA mismatch repair (MutL-MutS system) were more abundant in Bacteria. The function related to nucleotide excision repair has also been reported to be overrepresented in hypersaline environments [100].

In addition, we identified eight proteins related to DNA DSB repair in the annotations for viral sequences; this is one of the most common damaging events [101]. However, bacteriophages and some NCLDV possess homologous proteins, such as Rad50/SbcC, which is probably involved in the processing of dsDNA ends for processing during recombination [102]. These proteins were also identified in circular genomes of bacteriophages, such as *Vibrio parahaemolyticus* bacteriophage [103], which could indicate that these proteins are also propagated in this type of virus and could have implications in the repair of genetic material in stress environments. Other genes for methyltransferase enzymes, which are ubiquitous in the prokaryotic world and are associated with host protection of DNA damage, were also identified in our viral sequences.

Other functionally important genes found in viral sequences were auxiliary metabolic genes (AMGs) originally from the genome host. The AMGs found were ribonucleotide reductases and *phoH*, among others. The ribonucleotide reductases are associated with lytic rather than temperate viruses, and the *phoH* gene plays a role in the transport of phosphate in conditions of starvation. *Synechococcus* and *Prochlorococcus* (cyanophages) carry AMGs; however, in this study we found these families were in low abundance, as they are predominantly found in marine environments, so it correlates with the abundance of these families reported above [104].

3.8. Metabolic Pathway Involved in Biogeochemical Cycles

In order to evaluate the contribution of different metabolic pathways in the biogeochemical cycles associated the metagenomes, MEBS software was used to analyze the three samples. From this analysis, only two complete pathways of the carbon cycle were identified (Figure 11), while the nitrogen and sulfur cycles in the samples were more highly represented (Figure 11). In the case of nitrogen, the pathways of denitrification and the reduction of nitrate by assimilation were found to be more prevalent, since that nitrite is generally produced under anoxic conditions such as in hypersaline environments [6]. On the other hand, the reduction of dissimilatory nitrate (nitrite-ammonia) involving the proteins encoded by the genes *nirB*, *nirD*, *nrfA*, and *nrfH* is generally more highly expressed in *Proteobacteria*, *Bacteroidetes*, *Euryarchaeota*, and *Verrucomicrobia* [90]. Those were found as complete in our metagenomes, which correlates with the great abundance of *Proteobacteria* in the metagenomes.

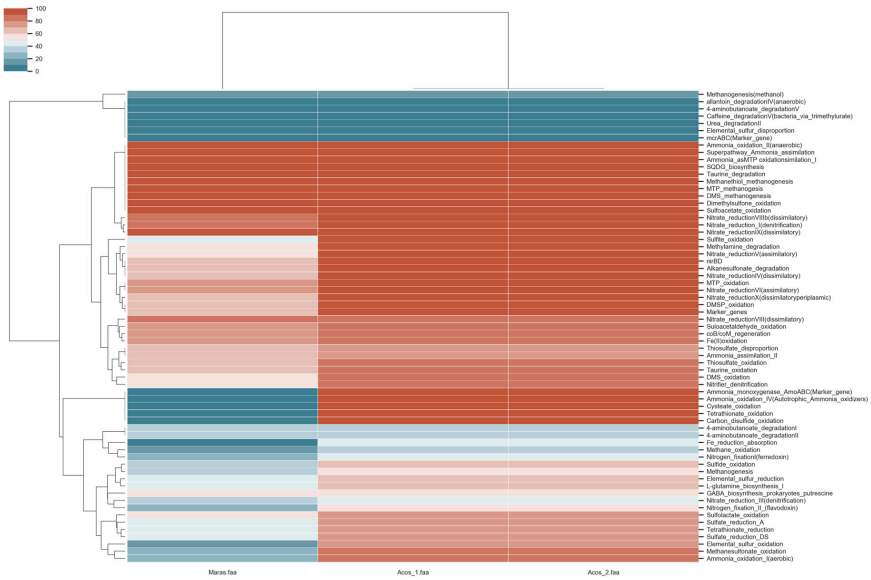


Figure 11. Completeness pathways of biogeochemical cycles. Mainly the pathways of nitrogen and sulfur cycles are complete within the hypersaline samples.

Because oxygen is limited, denitrification (nitrate-nitrite) is another pathway that contributes to the nitrogen cycle. In addition, species in the environment use nitrogen as a source of growth [105,106]. These pathways were also found as complete in the metagenomes of Acos, which indicates the importance of nitrogen in hypersaline environments. In this pathway, the proteins encoded by the genes *narGHIJ*, *napAB*, *nirKS*, *norBC*, and *nosZ* are included; these genes are expressed by *Bacteroidetes*, *Euryarchaeota*, and *Proteobacteria*. In addition, the *narL* gene in the virus compensates for the metabolic pathways of the microorganisms for nitrogen metabolism [90,107]. Finally, in the Maras sample, partially complete denitrification pathways were found (a 40-60% of representation), indicating that microorganisms can contribute to the reduction of nitrate and nitrite for the production of N_2.

Some organisms, such as *Proteobacteria* and *Thaumarchaeota*, are responsible for producing nitrate by nitrification at high salt concentrations [93,108], as well as the route of nitrogen fixation; however, they were partially complete, despite nitrite being an important energy source (Figure 11). In this regard, nitrite is not the only source of energy in this environment, since many Archaea and some bacteria use sulfur compounds as donors or electron acceptors for energy production [109]. In this case, the pathways related to sulfite oxidation, oxidation of sulfur DMS, and oxidation of dimethylsulfoniopropionate (DMSP) were found to be complete. Mainly, DMSP has been reported in abundance, which indicates that DMSP is an important source of carbon and energy [110]. Therefore, Bacteria and Archaea contribute to the oxidation of DMSP as an energy source, at different proportions.

4. Conclusions

In this study, we present a snapshot of microbial and functional diversity of two intermediate hypersaline environments in the Peruvian Andes, based on a metagenomics shotgun approach. The intermediate salinity environments show a great diversity and abundance of bacteria, more so than the archaea in the samples. At the level of phylum, *Proteobacteria* are the most abundant and predominated over other bacteria and archaea. However, the *Balneolaeota* phylum was found only in

Acos in great abundance, but was not diverse. In addition, we reconstructed the draft genomes of *H. elongata* and *I. loihiensis*, which have different mechanisms of adaptation to hypersaline environments, via de novo synthesis of ectoine and *natAB* transporters, respectively. Also, we obtained whole genomes from bacteriophages. Functional analysis indicated that microorganism in hypersaline environments contribute to the biogeochemical cycles involving carbon and nitrogen as the source of energy. We also found genes related to oxidative stress and DNA repair. Interestingly, viruses also had such repair protein genes, which are otherwise exclusive to eukaryotes and bacteria. This study contributes to the current knowledge of intermediate-salinity environments at high altitudes.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/10/11/891/s1, Figure S1: Rarefaction curves based at level species diversity, Figure S2: Dendrogram of all samples. Analysis of beta-diversity was carried out at species level using hclust and Bray-Curtis dissimilarity, Figure S3: Pangenome of draft genome (a) *Halomonas elongata* (b) *Idiomarina loihienis*, Figure S4: Protein-sharing viral network of virus from samples of Acos, Table S1: Indexes of diversity.

Author Contributions: Conceptualization, H.G.C.-S., M.A.Q.-R., and S.D.-R.; Data curation, R.A.B.-G.; Formal analysis, H.G.C.-S. and E.P.-R.; Funding acquisition, M.A.Q.-R. and S.D.-R.; Investigation, H.G.C.-S. and S.D.-R.; Methodology, H.G.C.-S., P.E., P.R., A.L.-T., J.L.S., R.A.B.-G., G.L.-U., E.P.-R., and M.A.Q.-R.; Resources, I.V., E.P.-R., M.A.Q.-R., and S.D.-R.; Software, H.G.C.-S., E.P.-R., and S.D.-R.; Supervision, S.D.-R.; Writing—original draft, H.G.C.-S., R.A.B.-G., S.T.-S., and E.P.-R.; Writing—review and editing, M.A.Q.-R. and S.D.-R.

Funding: This research was partially funded by contract number 227-2015-FONDECYT approved by resolution CU-005-2016_UNSAAC and contract number 23-2018-UNSAAC approved by resolution R-392-2018-UNSAAC to M.A.Q.R.; and by the Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México, PAPIIT IN-201117 to E.P.R. H.G.C.S. is a scholarship recipient of Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT Number 227229) program fellowship.

Acknowledgments: We thank the Unidad de Secuenciación Masiva y Bioinformática of Instituto de Biotecnología-UNAM for giving us access to its computer cluster and the Centro de Investigación en Dinámica Celular for giving us access to its server.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Maturrano, L.; Santos, F.; Rosselló-Mora, R.; Antón, J. Microbial diversity in Maras salterns, a hypersaline environment in the Peruvian Andes. *Appl. Environ. Microbiol.* **2006**, *72*, 3887–3895. [CrossRef] [PubMed]
2. Hernández, L.M. Caracterización de la Microbiota de las Salinas de Maras, un Ambiente Hipersalino de los Andes de Perú. Ph.D. Thesis, Universitat d'Alacant-Universidad de Alicante, Alicante, Spain, 2004.
3. Grant, W.D. Life at low water activity. *Philos. Trans. R. Soc. B Biol. Sci.* **2004**, *359*, 1249–1267. [CrossRef] [PubMed]
4. Demergasso, C.; Casamayor, E.O.; Chong, G.; Galleguillos, P.; Escudero, L.; Pedrós-Alió, C. Distribution of prokaryotic genetic diversity in athalassohaline lakes of the Atacama Desert, Northern Chile. *FEMS Microbiol. Ecol.* **2004**, *48*, 57–69. [CrossRef] [PubMed]
5. Ventosa, A.; de la Haba, R.R.; Sánchez-Porro, C.; Papke, R.T. Microbial diversity of hypersaline environments: A metagenomic approach. *Curr. Opin. Microbiol.* **2015**, *25*, 80–87. [CrossRef] [PubMed]
6. Naghoni, A.; Emtiazi, G.; Amoozegar, M.A.; Cretoiu, M.S.; Stal, L.J.; Etemadifar, Z.; Shahzadeh Fazeli, S.A.; Bolhuis, H. Microbial diversity in the hypersaline Lake Meyghan, Iran. *Sci. Rep.* **2017**, *7*, 11522. [CrossRef] [PubMed]
7. Fernández, A.B.; Vera-Gargallo, B.; Sánchez-Porro, C.; Ghai, R.; Papke, R.T.; Rodriguez-Valera, F.; Ventosa, A. Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front. Microbiol.* **2014**, *5*. [CrossRef] [PubMed]
8. Quillaguamán, J.; Hashim, S.; Bento, F.; Mattiasson, B.; Hatti-Kaul, R. Poly(β-hydroxybutyrate) production by a moderate halophile, Halomonas boliviensis LC1 using starch hydrolysate as substrate. *J. Appl. Microbiol.* **2005**, *99*, 151–157. [CrossRef]
9. Oren, A. Saltern evaporation ponds as model systems for the study of primary production processes under hypersaline conditions. *Aquat. Microb. Ecol.* **2009**, *56*, 193–204. [CrossRef]

10. Ventosa, A.; Fernández, A.B.; León, M.J.; Sánchez-Porro, C.; Rodriguez-Valera, F. The Santa Pola saltern as a model for studying the microbiota of hypersaline environments. *Extrem. Life Extreme Cond.* **2014**, *18*, 811–824. [CrossRef]
11. Haferburg, G.; Gröning, J.A.D.; Schmidt, N.; Kummer, N.-A.; Erquicia, J.C.; Schlömann, M. Microbial diversity of the hypersaline and lithium-rich Salar de Uyuni, Bolivia. *Microbiol. Res.* **2017**, *199*, 19–28. [CrossRef]
12. Emerson, J.B.; Thomas, B.C.; Andrade, K.; Allen, E.E.; Heidelberg, K.B.; Banfield, J.F. Dynamic Viral Populations in Hypersaline Systems as Revealed by Metagenomic Assembly. *Appl. Environ. Microbiol.* **2012**, *78*, 6309–6320.11. [CrossRef] [PubMed]
13. Sarwar, K.; Azam, I.; Tahir, I. Biology and Applications of Halophilic Bacteria and Archea: A Review. *Electron. J. Biol.* **2015**, *11*, 98–103.
14. Bioinformatics, B. *FastQC: A Quality Control. Tool for High. Throughput Sequence Data*; Babraham Institute: Cambridge, UK, 2011.
15. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef] [PubMed]
16. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef] [PubMed]
17. Glass, E.M.; Wilkening, J.; Wilke, A.; Antonopoulos, D.; Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010**, *2010*, Pdb.prot5368. [CrossRef]
18. Roux, S.; Enault, F.; Hurwitz, B.L.; Sullivan, M.B. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **2015**, *3*, e985. [CrossRef]
19. Huson, D.H.; Beier, S.; Flade, I.; Górska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.-J.; Tappu, R. MEGAN Community Edition-Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **2016**, *12*, e1004957. [CrossRef] [PubMed]
20. McMurdie, P.J.; Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **2013**, *8*, e61217. [CrossRef]
21. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. 2011. Available online: http://www.R-project.org (accessed on 30 September 2019).
22. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
23. Bushnell, B. *BBMap Short-Read Aligner, and Other Bioinformatics Tools*; University of California: Berkeley, CA, USA, 2015.
24. Golosova, O.; Henderson, R.; Vaskin, Y.; Gabrielian, A.; Grekhov, G.; Nagarajan, V.; Oler, A.J.; Quiñones, M.; Hurt, D.; Fursov, M.; et al. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. *PeerJ* **2014**, *2*, e644. [CrossRef]
25. Minot, S.S.; Krumm, N.; Greenfield, N.B. One Codex: A Sensitive and Accurate Data Platform for Genomic Microbial Identification. *Bioinformatics* **2015**. [CrossRef]
26. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
27. Zhou, Y.; Liang, Y.; Lynch, K.H.; Dennis, J.J.; Wishart, D.S. PHAST: A Fast Phage Search Tool. *Nucleic Acids Res.* **2011**, *39*, W347–W352. [CrossRef] [PubMed]
28. Bolduc, B.; Jang, H.B.; Doulcier, G.; You, Z.-Q.; Roux, S.; Sullivan, M.B. vConTACT: An iVirus tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*. *PeerJ* **2017**, *5*, e3243. [CrossRef] [PubMed]
29. Wu, Y.-W.; Tang, Y.-H.; Tringe, S.G.; Simmons, B.A.; Singer, S.W. MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2014**, *2*, 26. [CrossRef]
30. Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K.; et al. A new view of the tree of life. *Nat. Microbiol* **2016**, *1*, 16048. [CrossRef]

31. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
32. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* **2009**, *26*, 1641–1650. [CrossRef]
33. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [CrossRef]
34. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]
35. Silva, G.G.Z.; Green, K.T.; Dutilh, B.E.; Edwards, R.A. SUPER-FOCUS: A tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* **2016**, *32*, 354–361. [CrossRef] [PubMed]
36. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
37. De Anda, V.; Zapata-Peñasco, I.; Poot-Hernandez, A.C.; Eguiarte, L.E.; Contreras-Moreira, B.; Souza, V. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: Unraveling the sulfur cycle. *GigaScience* **2017**, *6*. [CrossRef] [PubMed]
38. Fernandez, A.B.; Ghai, R.; Martin-Cuadrado, A.B.; Sanchez-Porro, C.; Rodriguez-Valera, F.; Ventosa, A. Metagenome Sequencing of Prokaryotic Microbiota from Two Hypersaline Ponds of a Marine Saltern in Santa Pola, Spain. *Genome Announc.* **2013**, *1*. [CrossRef]
39. Fernández, A.B.; Ghai, R.; Martin-Cuadrado, A.-B.; Sánchez-Porro, C.; Rodriguez-Valera, F.; Ventosa, A. Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* **2014**, *88*, 623–635. [CrossRef] [PubMed]
40. Wang, J.; Yang, D.; Zhang, Y.; Shen, J.; van der Gast, C.; Hahn, M.W.; Wu, Q. Do Patterns of Bacterial Diversity along Salinity Gradients Differ from Those Observed for Macroorganisms? *PLoS ONE* **2011**, *6*, e27597. [CrossRef] [PubMed]
41. Fierer, N.; Jackson, R.B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 626–631. [CrossRef] [PubMed]
42. Pagaling, E.; Wang, H.; Venables, M.; Wallace, A.; Grant, W.D.; Cowan, D.A.; Jones, B.E.; Ma, Y.; Ventosa, A.; Heaphy, S. Microbial Biogeography of Six Salt Lakes in Inner Mongolia, China, and a Salt Lake in Argentina. *Appl. Environ. Microbiol.* **2009**, *75*, 5750–5760. [CrossRef]
43. Mora-Ruiz, M.D.; Cifuentes, A.; Font-Verdera, F.; Pérez-Fernández, C.; Farias, M.E.; González, B.; Orfila, A.; Rosselló-Móra, B. Biogeographical patterns of bacterial and archaeal communities from distant hypersaline environments. *Syst. Appl. Microbiol.* **2018**, *41*, 139–150. [CrossRef]
44. Ghai, R.; Pašić, L.; Fernández, A.B.; Martin-Cuadrado, A.-B.; Mizuno, C.M.; McMahon, K.D.; Papke, R.T.; Stepanauskas, R.; Rodriguez-Brito, B.; Rohwer, F.; et al. New Abundant Microbial Groups in Aquatic Hypersaline Environments. *Sci. Rep.* **2011**, *1*, 135. [CrossRef]
45. Vera-Gargallo, B.; Ventosa, A. Metagenomic Insights into the Phylogenetic and Metabolic Diversity of the Prokaryotic Community Dwelling in Hypersaline Soils from the Odiel Saltmarshes (SW Spain). *Genes* **2018**, *9*, 152. [CrossRef]
46. Ma, Y.; Galinski, E.A.; Grant, W.D.; Oren, A.; Ventosa, A. Halophiles 2010: Life in Saline Environments. *Appl. Environ. Microbiol.* **2010**, *76*, 6971–6981. [CrossRef] [PubMed]
47. Ventosa, A.; Nieto, J.J.; Oren, A. Biology of Moderately Halophilic Aerobic Bacteria. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 504–544. [PubMed]
48. Arahal, D.R.; García, M.T.; Vargas, C.; Cánovas, D.; Nieto, J.J.; Ventosa, A. Chromohalobacter salexigens sp. nov., a moderately halophilic species that includes Halomonas elongata DSM 3043 and ATCC 33174. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 1457–1462. [CrossRef] [PubMed]
49. Babavalian, H.; Amoozegar, M.A.; Pourbabaee, A.A.; Moghaddam, M.M.; Shakeri, F. Isolation and identification of moderately halophilic bacteria producing hydrolytic enzymes from the largest hypersaline playa in Iran. *Microbiology* **2013**, *82*, 466–474. [CrossRef]
50. Sorokin, D.Y.; Tourova, T.P.; Galinski, E.A.; Muyzer, G.; Kuenen, J.G. Thiohalorhabdus denitrificans gen. nov., sp. nov., an extremely halophilic, sulfur-oxidizing, deep-lineage gammaproteobacterium from hypersaline habitats. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 2890–2897. [CrossRef]
51. Sorokin, D.Y.; Tourova, T.P.; Muyzer, G.; Kuenen, G.J. Thiohalospira halophila gen. nov., sp. nov. and Thiohalospira alkaliphila sp. nov., novel obligately chemolithoautotrophic, halophilic, sulfur-oxidizing gammaproteobacteria from hypersaline habitats. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 1685–1692. [CrossRef]

52. Oren, A. Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* **2013**, *4*. [CrossRef]
53. González-Torres, P.; Gabaldón, T. Genome Variation in the Model Halophilic Bacterium Salinibacter ruber. *Front. Microbiol.* **2018**, *9*. [CrossRef]
54. Kennedy, S.P.; Ng, W.V.; Salzberg, S.L.; Hood, L.; DasSarma, S. Understanding the Adaptation of Halobacterium Species NRC-1 to Its Extreme Environment through Computational Analysis of Its Genome Sequence. *Genome Res.* **2001**, *11*, 1641–1650. [CrossRef]
55. Paul, S.; Bag, S.K.; Das, S.; Harvill, E.T.; Dutta, C. Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol.* **2008**, *9*, R70. [CrossRef]
56. Vera Gargallo, B.; Roy Chowdhury, T.; Brown, J.; Fansler, S.J.; Durán Viseras, A.; Sánchez-Porro Álvarez, C.; Bailey, V.L.; Jansson, J.K.; Ventosa, A. Spatial distribution of prokaryotic communities in hypersaline soils. *Sci. Rep.* **2019**, *9*. [CrossRef] [PubMed]
57. Mamani, J.I.; Pacheco, K.B.; Elorrieta, P.; Romoacca, P.; Castelan, H.; Davila, S.; Sierra, J.L.; Quispe-Ricalde, M.A. Draft Genome Sequence of Halomonas elongata MH25661 Isolated from a Saline Creek in the Andes of Peru. *Microbiol. Resour. Announc.* **2019**, *8*. [CrossRef] [PubMed]
58. Menes, R.J.; Viera, C.E.; Farías, M.E.; Seufferheld, M.J. Halomonas vilamensis sp. nov., isolated from high-altitude Andean lakes. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 1211–1217. [CrossRef] [PubMed]
59. Xiao-Ran, J.; Jin, Y.; Xiangbin, C.; Guo-Qiang, C. Chapter Eleven—Halomonas and Pathway Engineering for Bioplastics Production. In *Methods in Enzymology*; Scrutton, N., Ed.; Academic Press: Cambridge, MA, USA, 2018; pp. 309–328.
60. Adkins, J.P.; Madigan, M.T.; Mandelco, L.; Woese, C.R.; Tanner, R.S. Arhodomonas aquaeolei gen. nov., sp. nov., an aerobic, halophilic bacterium isolated from a subterranean brine. *Int. J. Syst. Bacteriol.* **1993**, *43*, 514–520. [CrossRef] [PubMed]
61. Donachie, S.P.; Hou, S.; Gregory, T.S.; Malahoff, A.; Alam, M. Idiomarina loihiensis sp. nov., a halophilic γ-Proteobacterium from the Lō'ihi submarine volcano, Hawai'i. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 1873–1879. [CrossRef] [PubMed]
62. Prado, B.; Del Moral, A.; Quesada, E.; Ríos, R.; Monteoliva-Sanchez, M.; Campos, V.; Ramos-Cormenzana, A. Numerical Taxonomy of Moderately Halophilic Gram-negative Rods Isolated from the Salar de Atacama, Chile. *Syst. Appl. Microbiol.* **1991**, *14*, 275–281. [CrossRef]
63. Minegishi, H.; Echigo, A.; Nagaoka, S.; Kamekura, M.; Usami, R. Halarchaeum acidiphilum gen. nov., sp. nov., a moderately acidophilic haloarchaeon isolated from commercial solar salt. *Int. J. Syst. Evol. Microbiol.* **2010**, *60*, 2513–2516. [CrossRef]
64. Rensing, C. Adaption to life at high salt concentrations in Archaea, Bacteria and Eukarya. *Saline Syst.* **2005**, *1*, 6. [CrossRef]
65. Sorokin, D.Y.; Messina, E.; Smedile, F.; Roman, P.; Damsté, J.S.S.; Ciordia, S.; Mena, M.C.; Ferrer, M.; Golyshin, P.N.; Kublanov, I.V.; et al. Discovery of anaerobic lithoheterotrophic haloarchaea, ubiquitous in hypersaline habitats. *ISME J.* **2017**, *11*, 1245–1260. [CrossRef]
66. Mou, Y.-Z.; Qiu, X.-X.; Zhao, M.-L.; Cui, H.-L.; Oh, D.; Dyall-Smith, M.L. Halohasta litorea gen. nov. sp. nov., and Halohasta litchfieldiae sp. nov., isolated from the Daliang aquaculture farm, China and from Deep Lake, Antarctica, respectively. *Extremophiles* **2012**, *16*, 895–901. [CrossRef]
67. Oren, A. Taxonomy of halophilic Archaea: Current status and future challenges. *Extremophiles* **2014**, *18*, 825–834. [CrossRef] [PubMed]
68. Williams, T.J.; Liao, Y.; Ye, J.; Kuchel, R.P.; Poljak, A.; Raftery, M.J.; Cavicchioli, R. Cold adaptation of the Antarctic haloarchaea Halohasta litchfieldiae and Halorubrum lacusprofundi. *Environ. Microbiol.* **2017**, *19*, 2210–2227. [CrossRef] [PubMed]
69. Fuchsman, C.A.; Collins, R.E.; Rocap, G.; Brazelton, W.J. Effect of the environment on horizontal gene transfer between bacteria and archaea. *PeerJ* **2017**, *5*, e3865. [CrossRef] [PubMed]
70. Pushkarev, A.; Inoue, K.; Larom, S.; Flores-Uribe, J.; Singh, M.; Konno, M.; Tomida, S.; Ito, S.; Nakamura, R.; Tsunoda, S.P.; et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **2018**, *558*, 595–599. [CrossRef] [PubMed]
71. Santos, F.; Yarza, P.; Parro, V.; Meseguer, I.; Rosselló-Móra, R.; Antón, J. Culture-Independent Approaches for Studying Viruses from Hypersaline Environments. *Appl. Environ. Microbiol.* **2012**, *78*, 1635–1643. [CrossRef]

72. Garcia-Heredia, I.; Martin-Cuadrado, A.-B.; Mojica, F.J.M.; Santos, F.; Mira, A.; Antón, J.; Rodriguez-Valera, F. Reconstructing Viral Genomes from the Environment Using Fosmid Clones: The Case of Haloviruses. *PLoS ONE* **2012**, *7*, e33802. [CrossRef]
73. Ramos-Barbero, M.D.; Martínez, J.M.; Almansa, C.; Rodríguez, N.; Villamor, J.; Gomariz, M.; Escudero, C.; dC Rubin, S.; Antón, J.; Martínez-García, M.; et al. Prokaryotic and viral community structure in the singular chaotropic salt lake Salar de Uyuni. *Environ. Microbiol.* **2019**, *21*, 2029–2042. [CrossRef]
74. Weitz, J.S.; Wilhelm, S.W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* **2012**, *4*. [CrossRef]
75. Roux, S.; Adriaenssens, E.M.; Dutilh, B.E.; Koonin, E.V.; Kropinski, A.M.; Krupovic, M.; Kuhn, J.H.; Lavigne, R.; Brister,, J.R.; Varsani, A.; et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* **2019**, *37*, 29–37. [CrossRef]
76. Castelán-Sánchez, H.G.; Lopéz-Rosas, I.; García-Suastegui, W.A.; Peralta, R.; Dobson, A.D.W.; Batista-García, R.A.; Dávila-Ramos, S. Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. *Mar. Genom.* **2019**, *46*, 16–28. [CrossRef]
77. Antunes, A.; Ngugi, D.K.; Stingl, U. Microbiology of the Red Sea (and other) deep-sea anoxic brine lakes. *Environ. Microbiol. Rep.* **2011**, *3*, 416–433. [CrossRef] [PubMed]
78. Claverie, J.-M.; Abergel, C. Mimiviridae: An Expanding Family of Highly Diverse Large dsDNA Viruses Infecting a Wide Phylogenetic Range of Aquatic Eukaryotes. *Viruses* **2018**, *10*, 506. [CrossRef] [PubMed]
79. Yau, S.; Lauro, F.M.; DeMaere, M.Z.; Brown, M.V.; Thomas, T.; Raftery, M.J.; Andrews-Pfannkoch, C.; Lewis, M.; Hoffman, J.M.; Gibson, J.A.; et al. Virophage control of antarctic algal host-virus dynamics. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6163–6168. [CrossRef] [PubMed]
80. Le Romancer, M.; Gaillard, M.; Geslin, C.; Prieur, D. Viruses in extreme environments. In *Life in Extreme Environments*; Amils, R., Ellis-Evans, C., Hinghofer-Szalkay, H., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2007; pp. 99–113.
81. Buchalo, A.S.; Nevo, E.; Wasser, S.P.; Oren, A.; Molitoris, H.P. Fungal life in the extremely hypersaline water of the Dead Sea: ®first records. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1998**, *265*, 1461–1465. [CrossRef]
82. Gunde-Cimerman, N.; Zalar, P. Extremely Halotolerant and Halophilic Fungi Inhabit Brine in Solar Salterns around the Globe. *Food Technol. Biotechnol.* **2014**, *52*, 170–179.
83. Butinar, L.; Sonjak, S.; Zalar, P.; Plemenitaš, A.; Gunde-Cimerman, N. Melanized halophilic fungi are eukaryotic members of microbial communities in hypersaline waters of solar salterns. *Bot. Mar.* **2005**, *48*. [CrossRef]
84. Lahav, R.; Fareleira, P.; Nejidat, A.; Abeliovich, A. The Identification and Characterization of Osmotolerant Yeast Isolates from Chemical Wastewater Evaporation Ponds. *Microb. Ecol.* **2002**, *43*, 388–396. [CrossRef]
85. Asem, A.; Eimanifar, A.; Wink, M. Update of "Biodiversity of the Hypersaline Urmia Lake National Park (NW Iran)". *Diversity* **2016**, *8*, 6. [CrossRef]
86. Tkavc, R.; Matrosova, V.Y.; Grichenko, O.E.; Gostinčar, C.; Volpe, R.P.; Klimenkova, P.; Gaidamakova, E.K.; Zhou, C.E.; Stewart, B.J.; Lyman, M.G.; et al. Prospects for Fungal Bioremediation of Acidic Radioactive Waste Sites: Characterization and Genome Sequence of Rhodotorula taiwanensis MD1149. *Front. Microbiol.* **2018**, *8*. [CrossRef]
87. Jünemann, S.; Kleinbölting, N.; Jaenicke, S.; Henke, C.; Hassa, J.; Nelkner, J.; Stolze, Y.; Albaum, S.P.; Schlüter, A.; Goesmann, A.; et al. Bioinformatics for NGS-based metagenomics and the application to biogas research. *J. Biotechnol.* **2017**, *261*, 10–23. [CrossRef]
88. McNair, K.; Edwards, R.A. Genome Peek—An online tool for prokaryotic genome and metagenome analysis. *PeerJ* **2015**, *3*, e1025. [CrossRef] [PubMed]
89. Kindzierski, V.; Raschke, S.; Knabe, N.; Siedler, F.; Scheffer, B.; Pflüger-Grau, K.; Pfeiffer, F.; Oesterhelt, D.; Marin-Sanguino, A.; Kunte, H.J. Osmoregulation in the Halophilic Bacterium Halomonas elongata: A Case Study for Integrative Systems Biology. *PLoS ONE* **2017**, *12*, e0168818. [CrossRef] [PubMed]
90. Ren, M.; Zhang, Z.; Wang, X.; Zhou, Z.; Chen, D.; Zeng, H.; Zhao, S.; Chen, L.; Hu, Y.; Zhang, C.; et al. Diversity and Contributions to Nitrogen Cycling and Carbon Fixation of Soil Salinity Shaped Microbial Communities in Tarim Basin. *Front. Microbiol.* **2018**, *9*. [CrossRef] [PubMed]
91. Mobberley, J.M.; Authement, R.N.; Segall, A.M.; Paul, J.H. The Temperate Marine Phage ΦHAP-1 of Halomonas aquamarina Possesses a Linear Plasmid-Like Prophage Genome. *J. Virol.* **2008**, *82*, 6618–6630. [CrossRef] [PubMed]

92. Oren, A. Microbial life at high salt concentrations: Phylogenetic and metabolic diversity. *Saline Syst.* **2008**, *4*, 2. [CrossRef]
93. Rhodes, M.E.; Spear, J.R.; Oren, A.; House, C.H. Differences in lateral gene transfer in hypersaline versus thermal environments. *BMC Evol. Biol.* **2011**, *11*, 199. [CrossRef]
94. Reed, C.J.; Lewis, H.; Trejo, E.; Winston, V.; Evilia, C. Protein Adaptations in Archaeal Extremophiles. *Archaea* **2013**, *2013*. [CrossRef]
95. Becker, E.A.; Seitzer, P.M.; Tritt, A.; Larsen, D.; Krusor, M.; Yao, A.I.; Wu, D.; Madern, D.; Eisen, J.A.; Darling, A.E.; et al. Phylogenetically Driven Sequencing of Extremely Halophilic Archaea Reveals Strategies for Static and Dynamic Osmo-response. *PLOS Genet.* **2014**, *10*, e1004784. [CrossRef]
96. Srivastava, P.; Kowshik, M. Mechanisms of Metal Resistance and Homeostasis in Haloarchaea. *Archaea* **2013**, *2013*, 732864. [CrossRef]
97. Spooner, R.; Yilmaz, Ö. The Role of Reactive-Oxygen-Species in Microbial Persistence and Inflammation. *Int. J. Mol. Sci.* **2011**, *12*, 334–352. [CrossRef]
98. Lipson, D.A.; Haggerty, J.M.; Srinivas, A.; Raab, T.K.; Sathe, S.; Dinsdale, E.A. Metagenomic Insights into Anaerobic Metabolism along an Arctic Peat Soil Profile. *PLoS ONE* **2013**, *8*, e64659. [CrossRef] [PubMed]
99. Jones, D.L.; Baxter, B.K. DNA Repair and Photoprotection: Mechanisms of Overcoming Environmental Ultraviolet Radiation Exposure in Halophilic Archaea. *Front. Microbiol.* **2017**, *8*. [CrossRef] [PubMed]
100. Plominsky, A.M.; Henríquez-Castillo, C.; Delherbe, N.; Podell, S.; Ramirez-Flandes, S.; Ugalde, J.A.; Santibañez, J.F.; van den Engh, G.; Hanselmann, K.; Ulloa, O.; et al. Distinctive Archaeal Composition of an Artisanal Crystallizer Pond and Functional Insights Into Salt-Saturated Hypersaline Environment Adaptation. *Front. Microbiol.* **2018**, *9*. [CrossRef] [PubMed]
101. Kinoshita, E.; van der Linden, E.; Sanchez, H.; Wyman, C. RAD50, an SMC family member with multiple roles in DNA break repair: How does ATP affect function? *Chromosome Res. Int J. Mol. Supramol. Evol. Asp. Chromosome Biol.* **2009**, *17*, 277–288. [CrossRef]
102. de Souza, R.F.; Iyer, L.M.; Aravind, L. Diversity and evolution of chromatin proteins encoded by DNA viruses. *Biochim. Biophys. Acta* **2010**, *1799*, 302–318. [CrossRef]
103. Ramírez-Orozco, M.; Serrano-Pinto, V.; Ochoa-Álvarez, N.A.; Makarov, R.Y.; Martínez-Díaz, S.F. Genome sequence analysis of the Vibrio parahaemolyticus lytic bacteriophage VPMS1. *Arch. Virol.* **2013**, *158*, 2409–2413. [CrossRef]
104. Gao, E.-B.; Huang, Y.; Ning, D. Metabolic Genes within Cyanophage Genomes: Implications for Diversity and Evolution. *Genes* **2016**, *7*, 80. [CrossRef]
105. Cabello, P. Nitrate reduction and the nitrogen cycle in archaea. *Microbiology* **2004**, *150*, 3527–3546. [CrossRef]
106. DasSarma, S.; DasSarma, P. Halophiles. *eLS* **2017**. [CrossRef]
107. He, T.; Li, H.; Zhang, X. Deep-Sea Hydrothermal Vent Viruses Compensate for Microbial Metabolism in Virus-Host Interactions. *MBio* **2017**, *8*. [CrossRef]
108. Andrei, A.-Ş.; Banciu, H.L.; Oren, A. Living with salt: Metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiol. Lett.* **2012**, *330*, 1–9. [CrossRef] [PubMed]
109. Liu, Y.; Beer, L.L.; Whitman, W.B. Sulfur metabolism in archaea reveals novel processes. *Environ. Microbiol.* **2012**, *14*, 2632–2644. [CrossRef] [PubMed]
110. Yau, S.; Lauro, F.M.; Williams, T.J.; DeMaere, M.Z.; Brown, M.V.; Rich, J.; Gibson, J.A.; Cavicchioli, R. Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *ISME J.* **2013**, *7*, 1944–1961. [CrossRef] [PubMed]

ARTICULO 4

microorganisms

MDPI

Article

The Microbial Composition in Circumneutral Thermal Springs from Chignahuapan, Puebla, Mexico Reveals the Presence of Particular Sulfur-Oxidizing Bacterial and Viral Communities

Hugo Gildardo Castelán-Sánchez [1,*], Pablo M. Meza-Rodríguez [1], Erika Carrillo [1], David I. Ríos-Vázquez [2], Arturo Liñan-Torres [1], Ramón Alberto Batista-García [1], Ernesto Pérez-Rueda [3,4], Norma Elena Rojas-Ruíz [2] and Sonia Dávila-Ramos [1,*]

1 Centro de Investigación en Dinámica Celular, Instituto de Investigación en Ciencias Básicas y Aplicadas, Universidad Autónoma del Estado de Morelos, Cuernavaca 62209, Morelos, Mexico; pablo.manuel.mz@gmail.com (P.M.M.-R.); cerika996@gmail.com (E.C.); arturo21lt@gmail.com (A.L.-T.); rabg@uaem.mx (R.A.B.-G.)

2 Centro de Investigaciones en Ciencias Microbiológicas, del Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Puebla 72570, Mexico; david.rios.vazquez@gmail.com (D.I.R.-V.); normaelena_rojas@yahoo.com.mx (N.E.R.-R.)

3 Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Sede Mérida, Universidad Nacional Autónoma de México, Unidad Académica Yucatán, Mérida, Yucatán C.P. 97302, Mexico; ernesto.perez@iimas.unam.mx

4 Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, Providencia, Santiago C.P. 7500000, Chile

* Correspondence: hcastelans@gmail.com (H.G.C.-S.); sonia.davila@uaem.mx (S.D.-R.)

Received: 1 August 2020; Accepted: 17 September 2020; Published: 29 October 2020

check for updates

Abstract: Terrestrial thermal springs are widely distributed globally, and these springs harbor a broad diversity of organisms of biotechnological interest. In Mexico, few studies exploring this kind of environment have been described. In this work, we explore the microbial community in Chignahuapan hot springs, which provides clues to understand these ecosystems' diversity. We assessed the diversity of the microorganism communities in a hot spring environment with a metagenomic shotgun approach. Besides identifying similarities and differences with other ecosystems, we achieved a systematic comparison against 11 metagenomic samples from diverse localities. The Chignahuapan hot springs show a particular prevalence of sulfur-oxidizing bacteria from the genera *Rhodococcus*, *Thermomonas*, *Thiomonas*, *Acinetobacter*, *Sulfurovum*, and *Bacillus*, highlighting those that are different from other recovered bacterial populations in circumneutral hot springs environments around the world. The co-occurrence analysis of the bacteria and viruses in these environments revealed that within the *Rhodococcus*, *Thiomonas*, *Thermonas*, and *Bacillus* genera, the Chignahuapan samples have specific species of bacteria with a particular abundance, such as *Rhodococcus erytropholis*. The viruses in the circumneutral hot springs present bacteriophages within the order Caudovirales (Siphoviridae, Myoviridae, and Podoviridae), but the family of Herelleviridae was the most abundant in Chignahuapan samples. Furthermore, viral auxiliary metabolic genes were identified, many of which contribute mainly to the metabolism of cofactors and vitamins as well as carbohydrate metabolism. Nevertheless, the viruses and bacteria present in the circumneutral environments contribute to the sulfur cycle. This work represents an exhaustive characterization of a community structure in samples collected from hot springs in Mexico and opens opportunities to identify organisms of biotechnological interest.

Keywords: thermophilic bacteria; AMG viral genes; terrestrial thermal spring

1. Introduction

Terrestrial thermal springs are widely distributed throughout the world. They harbor a significant number of microorganisms of biotechnological interest. These ecosystems have been classified in low-temperature (<55 °C) and high-temperature (>55 °C) springs; in terms of pH, the springs are acidic (pH < 4), intermediate (pH ~4), circumneutral or neutral (pH ~7), or alkaline (pH > 7) [1–3]. Additionally, the thermal springs are classified according to their origin in magmatic waters, which are born in volcanic areas and at high temperatures (>50 °C), and telluric waters, which are formed when underground water currents pass along deep hot rocks [4].

The most studied thermophilic environment in the world is Yellowstone National Park (YNP) (pH 2, 75 °C) [1,5], where pioneering studies opened the possibility of exploring the diversity of microorganisms in extreme environments, as well as the genes that encode enzymes with biotechnological applications [3,6–8]. Recently, a growing interest in studying these ecosystems has emerged [1,9–22], and these studies have focused on the sites of high temperatures and acidic or alkaline pH [15,16,19,23–25], and have identified a high diversity of microorganisms [22,26]. Intermediate or circumneutral hot springs also exhibit a high diversity of microorganisms; biodiversity generally decreases with increasing temperature and decreasing pH [21].

Mexico contains a wide diversity of thermal springs, steam vents, geothermally heated soils, boiling mud pools, and geothermal zones [25,27,28]; however, few studies have described the diversity of microbial communities in thermal environments. In particular, in the acidic hot spring "*Los Azufres*" (pH 3.6 and 65 °C) located in the state of Michoacan, *Rhodobacter*, *Acidithiobacillus*, and *Lysobacter* [25], among other bacteria, have been identified; whereas the *Sulfolobales archaeon* has been discovered in "*Los Azufres*" [29]. Finally, the viruses identified correspond to archaeal *Fusellovirus*, archaeal *Rudivirus*, and *Sulfolobales Archaeon AZ1* [30]. In addition, in a hot spring located in the Araro region, Michoacan, different genera of bacteria were found, such as *Bacillus*, *Aeromonas*, and *Pseudomonas* [31,32], whereas in the "*Carrizal*" thermal pool hot spring and in "*Los Baños*" in Veracruz, Mexico, bacteria of the genera *Geobacillus*, *Anoxybacillus*, and *Aeribacillus* have been identified [33].

In this work, we explore the microbial community and functional composition in the thermal spring "*Baños Termales de Chignahuapan*", which is located in the geothermal region of Tulancingo-Acoculco, Sierra Norte, Puebla. This is a mountainous complex, whose origin dates back to the Pleistocene (1.7–0.9 million years ago), where there is a magmatic hot spring with travertine sediment compositions and rocks of dacites and rhyolites [27,28].

We consider that the hot springs in Chignahuapan present a particular combination of physicochemical characteristics, such as high concentrations of calcium, carbonate, and sulfur, making it an excellent spot to determine the microorganisms and viruses that make up that ecosystem, as well as their functional potential. In addition, we achieved a comparative analysis with 11 circumneutral hot springs, to determine differences in composition and diversity in microorganisms and highlight the influence of the environment in the community structure.

2. Materials and Methods

2.1. Physicochemical Characterization

The temperature of thermal water was measured in situ. A 1 liter water sample was analyzed in the laboratory to determine the physicochemical parameters. The pH was determined by the method established in NMX-AA-008-SCFI-2016 [34]; electrical conductivity reported in deci-siemens (dS) was measured on a HANNA conductivity meter; the ion's calcium (Ca^{2+} mg L^{-1}) and magnesium (Mg^{2+} mg L^{-1}) were determined by the EDTA method; sodium (Na^{+} mg L^{-1}) and potassium (K^{+} mg L^{-1}) were evaluated with flamometry; nitrates (NO_3 mg L^{-1}) were determined by the methodology cited in NMX-AA-079-SCFI-2001 [35]. Sulfates ($SO_4{}^{-2}$ mg L^{-1}) were determined by the NMX-AA-074-SCFI-2014 [36]; carbonates ($CO_3{}^{-2}$ mg L^{-1}) were evaluated according to

NMX-AA-029-SCFI-2001 [37], and bicarbonates (HCO_3^{-1} mg L^{-1}) were determined by volumetric methods; chlorides (Cl^{-1} mg L^{-1}) were analyzed according to NMX-AA-073-SCFI-2001 [38].

2.2. Sample Collection and Processing from Chignahuapan Puebla, Mexico

Two samples of water (20 L) were collected from the recreational center "Baños Termales de Chignahuapan" located at coordinates 19°50′30′ N 97°59′41" W, at an altitude of 2136 m above sea level, during April 2019. The samples were obtained with sterilized tools in 1 L containers, from the water emerged in the mountain before the pools were supplied, and transported at room temperature to the laboratory of the Benemérita Universidad Autónoma de Puebla, where they were filtered through 0.22 µm Millipore filters. The DNA was obtained from the filters and isolated using ZymoBIOMICS DNA kits (MoBio, West Carlsbad, CA, USA). The DNA concentration was determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific), and fluorometry was measured using a Qubit 4 fluorometer (Invitrogen). DNA was sequenced using the Illumina NextSeq 500 platform with the Nextera reagent kit V3.0 for a read length 2 × 75 bp at the Instituto de Biotecnología of Universidad Nacional Autónoma de México.

2.3. Taxonomic Annotation of Metagenome

We analyzed two metagenomic samples from Chignahuapan, Puebla, and 11 shotgun metagenomic samples retrieved from the Sequence Read Archive (SRA) database, from thermal environments with physicochemical characteristics of circumneutral hot spring and pH 7 (Table S1). The quality control of sequences and the removal of adapters was performed by using Trimmomatic v.036 [39] with a sliding window of 4 bp, an average quality per base of 30, and a minimum read length of 75 bp. Reads were assembled in contigs using MEGAHIT [40], under default parameters in paired-end mode, and contigs with a minimum length of 1000 bp were considered for further analysis.

Taxonomic assignments were performed with the software Kaiju v1.7.3 [41] against the nonredundant protein database v1.7 sourced from the National Center for Biotechnology Information (NCBI) databases using the maximum exact matches, and 11 as a minimum match length. Finally, the results were displayed with the library Pavian in R [42].

The composition of the prokaryotic communities was evaluated using statistical analyses in R [43]. The nonmetric multidimensional scaling (NMDS) plot was performed in Vegan v2.3-1, with the stress function, to determine the goodness.

The taxonomic profiles at the genus level were used to calculate the diversity indices from all data. Diverse alpha-diversity descriptors were obtained using the Phyloseq function in R [44].

The beta diversity was determined by Bray–Curtis dissimilarity, and the sampling effort was evaluated through the rarefaction curves using the Vegan library implemented in R [45].

The metagenomes were deposited in the Joint Genome Institute (JGI) Integrated Microbial Genomes and Microbiomes database, with accession number: Gs014786.

2.4. Co-Occurrence Network Analysis

The co-occurrence analysis was implemented using the igraph library [46] and bipartite library in R [47], implemented under development Virome Network Analysis (ViNA) [48]. In brief, we computed a table of incidences of the relevant bacteria and viruses at the species level. These tables indicate the presence or absence of each taxon in the metagenome. After that, the network displayed the taxon associations and locations, which was built using the Kamada–Kawai algorithm layout [49].

2.5. Identification and Annotation of Viral Genomes

For virus classification, two approaches were implemented. In the first approach, Virsorter [50] was used to determine the viral contigs, using the viral hallmark genes annotated as "main capsid protein", "portal", "large subunit of the terminase", "tail", and "envelope", among others. The entire contig was considered viral if more than 80% of predicted genes on a contig had a viral signal.

This software finds new viruses with different confidence categories from 1 to 6, with 3 and 6 as the least confidence level.

The categories 1, 2, 4, and 5 were concatenated, and contigs were compared with BLASTn against the nonredundant database (nr), with the following parameters: -num_alignments 20, -num_descriptions 20, e-value 0.0001, -word_size 11. The results were visualized in MEGAN v5.10.6 considering the lowest common ancestor (LCA) method with the following parameters that reduce the rate of false positives and false negatives [51]: minimum support of 2; minimum score of 70; top percent of 10.

In the second approach, the viral contigs were recovered and the auxiliary metabolic genes (AMG) were obtained using the program VIBRANT [52]. This program is a hybrid machine-learning algorithm and similarity comparisons of protein sequences. It annotates the genes supporting metabolism and recovers the metabolic pathways where these genes are involved. We considered a minimum length of 1000 bp, with summary plots on and function virome off. The retrieved contigs were compared with BLASTn against the Viral RefSeq database and analyzed using the software MEGAN v5.10.6 with the same conditions [51].

2.6. Functional Annotation

To predict protein-coding genes in the assembled contigs, we used Prodigal v2.6.3 [53] with the metagenomic mode. The functional annotation was achieved using SUPERFOCUS [54], which contains the SEED database with an E-value of 0.0003 and 60% identity. The results were displayed in a heatmap using the ggplot2 library in R [55]. The metabolic pathways were displayed using MG-RAST server with the database KO [56].

3. Results and Discussion

3.1. Field Sampling and Physicochemical Characterization

The water samples were collected from two thermal springs at Chignahuapan, Puebla, Mexico. The first sample was collected from the thermal spring (Mex_Chig_S1) with a temperature of 49–50 °C, and a pH of 7.02. The second sample was collected from the water that supplies the pool (Mex_Chig_S2). This sample temperature was 45 °C and the pH was 6.66.

The compositions of both samples were compared with respect to the content of different salts, quantified in Table 1. The sulfate was found with a concentration of 25.6 and 30.2 mg L^{-1} in Mex_Chig_S1 and Mex_Chig_S2, respectively; in the site, there was an intense smell of hydrogen sulfide (H_2S), indicating that the sulfate was being reduced to hydrogen sulfide by sulfate-reducing microorganisms. These concentrations were in the range of thermal spring waters. In contrast, when Ca^{2+}, carbonate ($HCO_3{}^{-1}$), and Na^{2+} were quantified, these ions were present in high concentrations, suggesting that they are out of range according to water quality standards in Mexico.

The presence of high levels of carbonates and calcium in both samples collected could be associated with hydrothermal travertine deposits found in hot springs, as these deposits are mainly composed of $CaCO_3$ (calcite) [6,19]. Likewise, there are previous reports of the presence of travertine deposits (calcium carbonate), rhyolite, and dacite in the Chignahuapan springs [28], and the presence of these carbonates could be involved in the modification of the microbial structures of the communities.

Table 1. Physicochemical parameters of the Chignahuapan hot springs.

Chemical Properties	Mex_Chig_S1	Mex_Chig_S2
Temperature in °C	49–50	45
pH	7.02	6.66
Electrical conductivity dS m^{-1}	1.52	1.50
Ca^{2+} mg L^{-1}	203.1	81.6
Mg^{2+} mg L^{-1}	33.2	17.3
Na^{1+} mg L^{-1}	102.0	212.9
K^{1+} mg L^{-1}	14.6	12.8
$NO_3{}^{-1}$ mg L^{-1}	16.9	12.2
$SO_4{}^{-2}$ mg L^{-1}	25.6	30.2
$PO_4{}^{-3}$ mg L^{-1}	4.2	2.5
$CO_3{}^{-2}$ mg L^{-1}	0	0
$HCO_3{}^{-1}$ mg L^{-1}	780.8	634.4
Cl^{-1} mg L^{-1}	196.0	98.9

The structure of the microbial community is also modified by the concentration of Na^+; it is known that decreased concentrations of salts lead to a higher diversity of bacteria, while archaea are abundant in a higher concentration of salts [57,58]. In the thermal environments of Puebla, a concentration outside the range for consumption according to the Mexican regulations was also found; thus, the high salt concentrations found in both samples will determine the diversity and structure of microbiomes in thermal springs.

3.2. Microbial Community Composition of Chignahuapan Metagenomes

The diversity and abundance of microorganisms in the circumneutral thermal spring samples from Chignahuapan, Puebla, were determined by shotgun metagenomic sequencing. From these metagenomes obtained from two locations, the diversity of the microbiome was determined. To this end, the assembled contigs were classified with Kaiju (Table 2), and the results at the domain level showed that Bacteria represent between 88.4 and 91.8% of the total microorganisms, followed by Archaea (1.3–1.8%), Viruses (0.8–0.9%), and Eukarya (0.7–0.8%) (Figure S1). The microbial composition in the metagenomes was in concordance with that found in similar environments with a neutral pH, moderate temperature (50 °C), and similar chemical composition, such as the Jordanian hot springs [17,21] (Figure 1).

Table 2. Total of contigs classified by Kaiju of Metagenome from Chignahuapan.

Sample	No. Total of Contigs	Total Contigs		% Contigs Classified within Domain			
		Classified	Unclassified	Bacteria	Archaea	Eukarya	Virus
Mex_Chig_S1	8474	8082 (95.5%)	392 (4.51%)	91.8%	1.8%	0.802%	1.08%
Mex_Chig_S2	8361	7645 (91.4%)	716 (8.56%)	88.4%	1.3%	0.7299%	0.957%

In the sample Mex_Chig_S1, bacteria from the phylum *Actinobacteria* (64.21%) were the most abundant, followed by Proteobacteria (36.09%). In contrast, in the sample Mex_Chig_S2, Proteobacteria was found as the most abundant (76.81%), followed by *Actinobacteria* (20.14%) (Figure 1). This result is consistent with previous descriptions in circumneutral water, sulfur water springs, and volcanic terrain, respectively, where Proteobacteria and Actinobacteria are predominant in sulfur water springs and volcanic terrain, respectively [17,48].

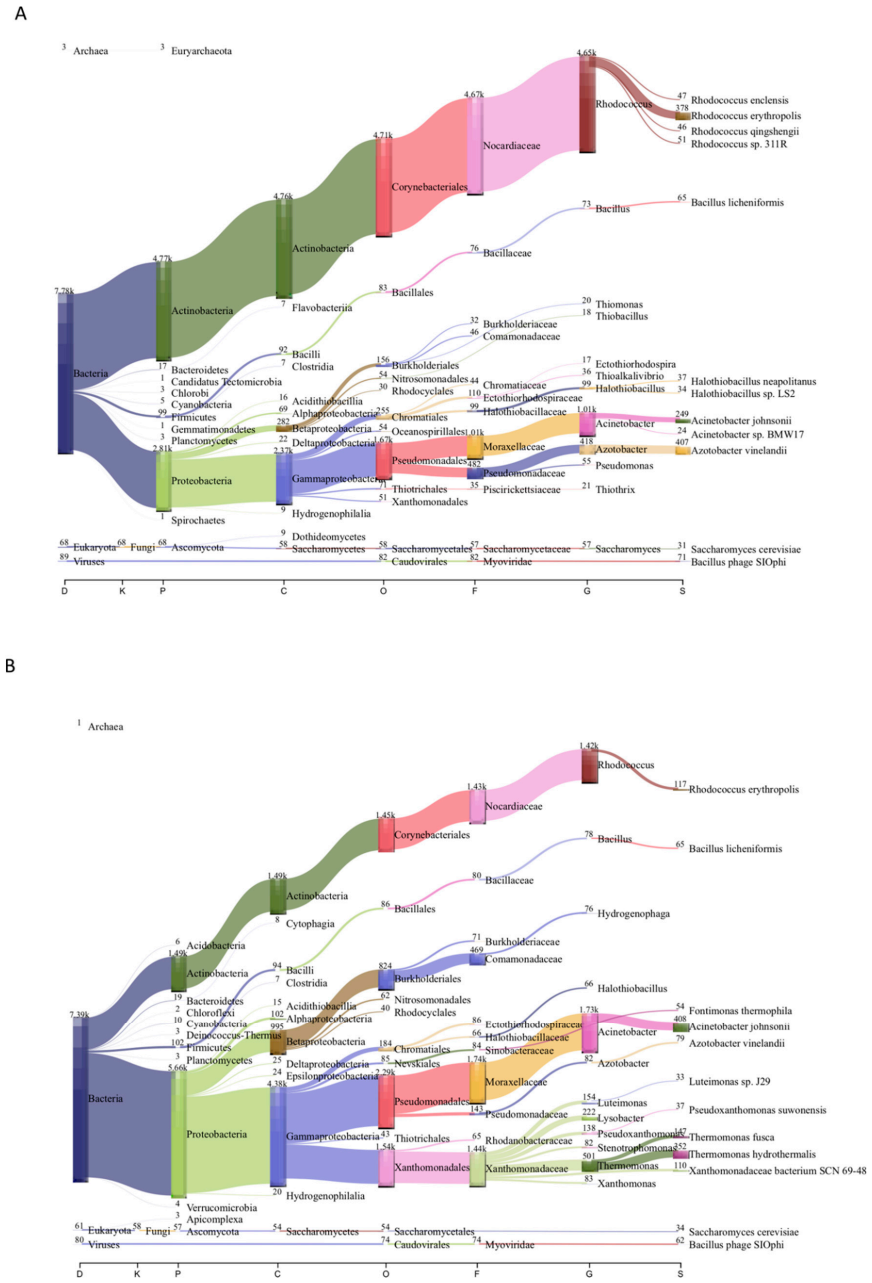


Figure 1. Taxonomic profile in circumneutral thermal spring metagenomes from Chignahuapan, Mexico. (**A**) Mex_Chig_S1. (**B**) Mex_Chig_S2. On the x-axis are the taxonomic levels: D, domain; P, phylum; C, class; O, order; F, family; G, genus; S, species. The numbers correspond to the assigned contigs.

In Mex_Chig_S1, the most abundant genera were *Rhodococcus* (59%), followed by *Acinetobacter* (13%), *Azotobacter* (5%), *Halothiobacillus* (1.2%), and *Bacillus* (0.9%). At the species level, *Rhodococcus erythropolis* was identified as the most abundant (Figure 1). It is worth mentioning that *R. erythropolis*, originally isolated from crude oil [59], is a biologically important bacterium because it possesses selective desulfurization activity and the capacity to degrade alkanes (C_8 to C_{20} n-alkanes) and methyl benzenes such as toluene [59,60].

Another interesting genus was *Halothiobacillus*, which is an obligate chemolithoautotroph and sulfur oxidizer. In particular, *Halothiobacillus neapolitanus* was found in the samples that encode a complete SOX complex, involved in sulfur oxidation [61].

At the species level, *Bacillus licheniformis* (0.8%) was also found—an interesting bacterium which was also isolated from Jordanian hot springs [17]. *Bacillus licheniformis* is widely distributed in thermal springs, and it is considered for commercial use as it has been used in the production of enzymes, antibiotics, and detergents, but some species of *Bacillus* are involved in carbon metabolism [62,63].

In Mex_Chig_S2, *Acinetobacter* (23%) was the most abundant genus, followed by *Rhodococcus* (19%), *Thermomonas* (6.8%), *Lysobacter* (3%), *Luteimonas* (2%), *Pseudoxanthomonas* (1.8%), and *Xanthomonas* (1.1%).

At the species level, the most abundant were *Thermomonas hydrothermalis* (4.1%), *Thermomonas fusca* (1.7%), *Rhodococcus erythropolis* (1.3%), *Xanthomonadaceae bacterium* SCN 69-48 (1.3%), and in less than 1% of species such as *Fontimonas thermophila*, *Sulfurovum* sp. enrichment culture clone C5, *Thiomonas bhubaneswarensis*, *Thiomonas intermedia*, and *Sulfurovum* sp. AS07-7, were found; these bacteria are important because they are involved in the sulfur metabolism and the capability to oxidize sulfur [64] (Figure 1, Figure S2).

Overall, these bacteria are moderate thermophilic organisms isolated from hydrothermal springs with diverse enzymatic activities characterized as amylases, cellulases, and lectinases [65]. *T. bhubaneswarensis* and *T. intermedia* are widely distributed in hot springs from India, which are rich in arsenic and contain low levels of organic matter; these bacteria are sulfur and thiosulfate-oxidizing [66].

The sulfur-oxidizing bacteria were highly abundant in both samples, correlating with the concentrations of sulfate in water. These bacteria can oxidize sulfur compounds (thiosulfate, tetrathionate, sulfide, and polysulfide) to produce energy, which has been previously reported in high-temperature sulfidic hot springs [14].

In general, these bacteria belong to Actinobacteria, Gammaproteobacteria, Betaproteobacteria, and Epsilonproteobacteria phyla, chemoheterotrophs or chemolithoautotrophs in the microbial community, with the ability to use electrons from inorganic compounds as an energy source. Overall, many of them are sulfur-oxidizing bacteria.

In the circumneutral thermal spring samples from Chignahuapan, Puebla, archaeal organism abundance was low, similar to findings in other circumneutral hot springs [17,50]. This archaeal composition is similar to those found in Malaysia's circumneutral hot spring, where a low proportion of archaea was found [67,68]. The most abundant classes identified in our samples correspond to Halobacteria within the Euryarchaeota phylum. Within the few species found were *Halolamina sediminis* and *Candidatus altiarchaeum* sp. Both are interesting because the first is associated with hypersaline aquatic environments, which may be possible due to the high salt concentrations present in Chignahuapan, and the second is involved with carbon fixation and plays an essential role in biogeochemical cycles [69,70]. These results contrast with previous reports indicating that Crenarchaeota is most abundant in terrestrial thermophilic environments in acid hot springs with high temperatures [63,65].

3.3. Comparative Analysis and Ecological Indices

To determine whether the samples of circumneutral hot springs from Chignahuapan, Puebla, share similarities with other hot springs samples, we compared the microbial community with 11 circumneutral metagenomes from the SRA database. These were obtained from different sites in the world, selected based on physicochemical characteristics similar to those of the sample's Mex_Chig_S1 and Mex_Chig_S2 (Table S1). The aquatic thermal terrestrial metagenomes were selected with a nearly neutral pH.

The comparison showed that the Puebla hot springs metagenomes did not have a similar bacterial composition to the other metagenomes. To analyze this, we carried out a nonmetric multidimensional scaling (nMDS) analysis. The nMDS analysis showed that the Chignahuapan samples were grouped together, while the other metagenomes were grouped according to their geographical area, with the stress of 0.03; the lower the stress value, the better the goodness of fit (Figure 2A).

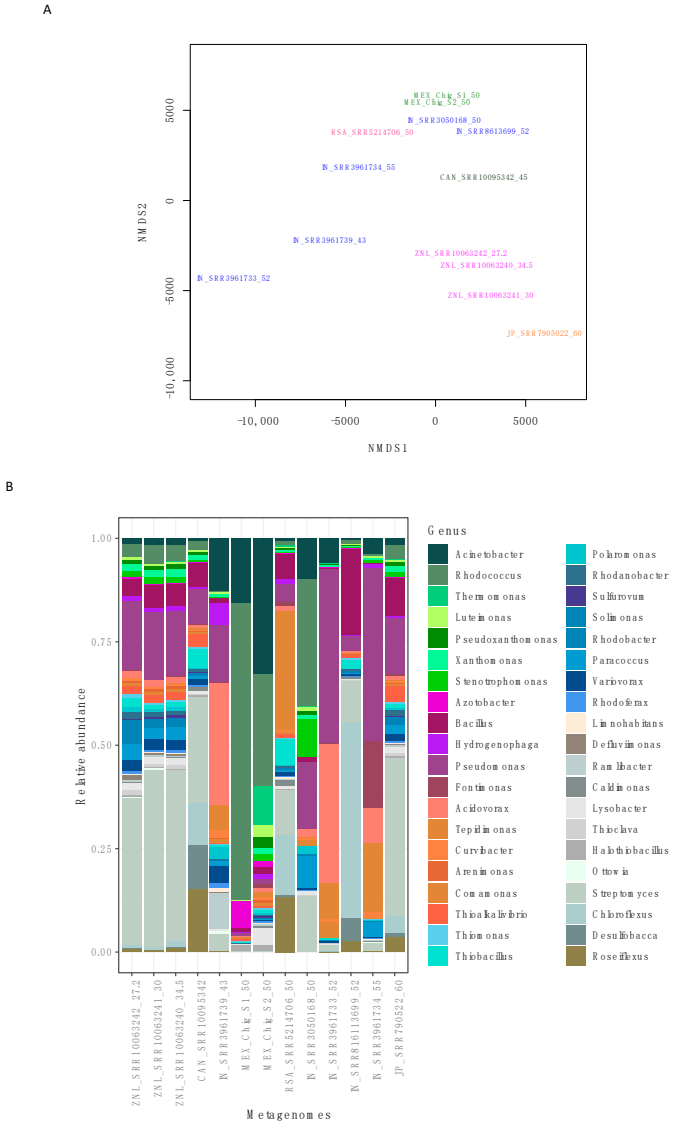


Figure 2. Comparison of hot spring metagenomes at the genera level. (**A**) nMDS analysis, using a Bray–Curtis distance between samples. On the x-axis is dimension 1 and on the y-axis is dimension 2. (**B**) Relative abundances of 30 bacterial genera of metagenomes.

Likewise, differences were found in the composition at the genus level in the bacterial communities. The microbial community structure was particular in the Chignahuapan samples, mainly of the *Rhodococcus* genus, which was found in great abundance, followed by genera such as *Acinetobacter*, *Thermomonas*, and *Azotobacter* (Figure 2B).

This particular community of microorganisms is possibly associated with the physicochemical characteristics, including abiotic factors such as Ca, SO_4, HCO_3 ion levels, that have previously been reported to have an essential role in the microbial composition and have been associated with some bacteria, such as *Thermomonas hydrotermalis*, *Bacillus licheniformis*, *Bacillus subtilis*, and *Anoxybacillus kamchatkensis* [6]. The composition of the bacteria and minerals reported is similar to those of Chignahuapan; therefore, the water's physicochemical factors very likely influence the structure of the microbial community.

Concerning the other metagenomes, the IN_SRR3050168_50 and IN_SRR8613699_52 from India were clustered near the Mexican samples, and the others, IN_SRR3961733_52, IN_SRR3961734_55, and IN_SRR3961739_43 samples were separated from all others (Figure 2A). In this regard, the genera *Acidovorax*, *Microbacterium*, *Pseudomonas*, and *Caulobacter* were associated with India's location. A similar finding was observed with the samples from New Zealand, ZNL_SRR10063242_27.2, ZNL_SRR10063241_30, and ZNL_SRR10063240_34.5, where *Streptomyces*, *Phenylobacterium*, and *Asticcacaulis* were clustered. The hot springs of Canada (CAN_SRR10095342_45) and Japan (JP_SRR7905022_60) were clustered together, whereas the genera, *Streptomyces*, *Roseiflexus*, *Desulfobacca*, and *Chloroflexus*, were clustered (Figure 2B and Figure S3).

The results show differences in the recovered genera and that they are grouped by geographic area. However, this grouping can also be driven by the composition of ions, minerals, and elements present in the water. In previous studies, measurements of different elements, compounds, and ions were taken in the water samples. For example, in India's selected samples, which have high concentrations of Co, La, Fe, Hg, and Si, the predominant bacteria were *Pseudomonas stutzeri* and *Acidovorax* sp.; these findings are in accordance with the taxonomic assignment made by us [9]. Furthermore, samples from another study from the same country were nearly clustered with them. However, in this other site, different dissolved solids were measured: there were high concentrations of phosphorus and sulfur (Figure S3), and the genera *Microbacterium*, *Propionibacterium*, *Caulobacter*, and *Rhodococcus* predominated, among others [71].

Whereas the New Zealand samples had different values in the parameters evaluated, one of the samples had a high concentration of methane (ZNL_SRR10063242_27.2), and ammonia (ZNL_SRR10063240_34.5) and iron were present in all the samples. The genera that predominated in these metagenomes were different than those found in the other metagenomes [72].

On the other hand, in the Japanese metagenome (JP_SRR7905022_60), high levels of iron and dissolved oxygen have been reported, and these findings correlate with the more abundant bacteria genus *Chloroflexus*, which has photosynthetic activity [73].

However, the differences observed in the microbial community are associated with the temperature presented by thermal environments. For example, a study of the hot springs in Canada and New Zealand showed that some phyla had trends that changed with temperature, where Cyanobacteria, Acidobacteria, Verrucomicrobia, and Planctomycetes were absent at high temperatures, while other phyla did not show changes [74]. In our cluster analysis, we can also observe that there is certain proximity of the metagenomes that share a similar temperature (Figure 2).

The multivariate approaches performed revealed that diversity patterns changed in each geographical location, where specific genera predominated in each of the metagenomes, and possibly this predominance was a consequence of the different physicochemical compositions and temperatures present in the water [26].

These differences can also be clearly distinguished in the relative abundance analysis. Proteobacteria were predominant, correlating with other moderate temperature circumneutral springs, such as the Jordanian Main springs, where about 50% corresponds to these phyla [17]. However, Puebla's microbial community structure was mainly compared to other terrestrial hot springs (Figure S3). Overall,

these results indicate that the microorganism communities change depending on the geographical area and physicochemical composition (Figure S3).

The circumneutral hot springs were evaluated through the ecological indices or diversity indices, and species accumulation curves were compared among all the metagenomes. The rarefaction curve associated with the Chignahuapan metagenomes showed a low number of species, probably due to a low yield obtained from the sequencing run; therefore, these samples did not have asymptotic behavior (Figure 3A). When the alpha and beta biodiversity indices were analyzed, a variation was shown along with the α-index.

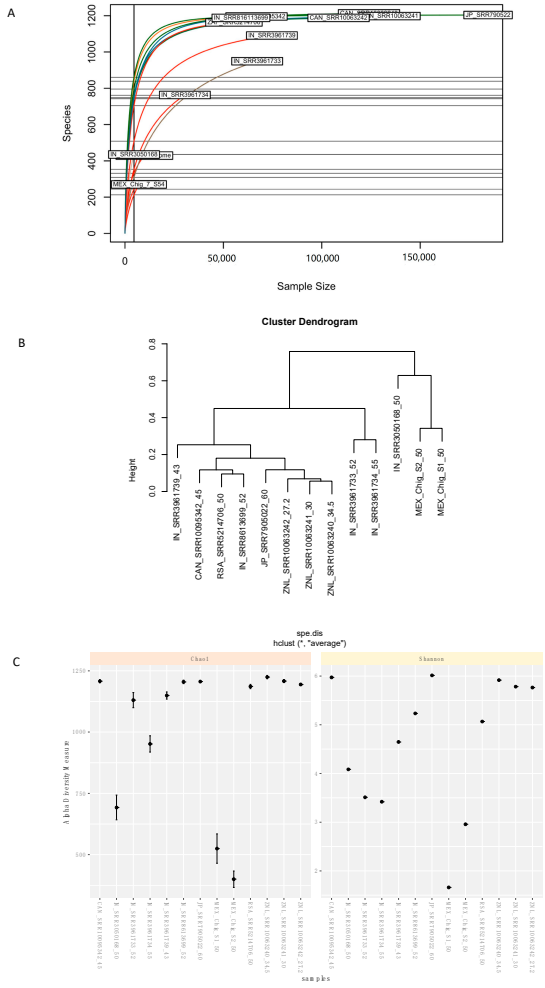


Figure 3. Analysis of diversity and enrichment of species. (**A**) Species-accumulation curves at the genus level. Rarefaction curves of the 13 hot spring metagenomes; the curves from Chignahuapan do not have asymptotic behavior compared to the others. (**B**) Box plots of alpha diversity. (Chao1 and Shannon). (**C**) Beta diversity dendrogram, showing dissimilarity between the metagenomes, on the x-axis is the metagenomes, and on the y-axis is alpha diversity.

Alpha-diversity values were lower for the Chignahuapan samples compared to the other metagenomes. However, it has been reported that diversity decreases with increasing temperature; these studies showed that the alpha-index is higher in those places with a neutral pH and a temperature of 50 °C [74]. Therefore, it was expected that the samples from Mexico had high diversity indices compared to other samples. In the case of the samples analyzed, the sample from Japan had the highest alpha index, compared to the other metagenomes. Therefore, it can be suggested that while diversity is modified by physicochemical changes, it is also modified by experimental yields and conditions.

The low values obtained for the β-diversity suggest that Chignahuapan metagenomes are less diverse than the other metagenomes (Figure 3B). In this regard, the clustering using β-diversity showed that the hot spring metagenomes were clustered primarily based on geographic locations and temperature. The Chignahuapan thermal springs clustered independently, indicating that they have a different microorganism community than the other metagenomes (Figure 3C). Interestingly, in this analysis, the sample from India was grouped with the samples from Mexico. The Indian sample also had low diversity indices indicating less diversity.

In summary, comparative analysis shows that Chignahuapan metagenomes are less diverse than other hydrothermal environments. However, they have a unique microorganism composition, because they do not group with the other samples with similar temperatures. Therefore, the chemical composition of water could determine the microbial structure of Chignahuapan hot springs.

3.4. Co-Occurrence Network Analysis

We performed a co-occurrence network analysis to evaluate possible genera interactions between the 13 metagenomes. Each of them evaluated the association between species and metagenomes; the genus *Rhodococcus* (Actinobacteria) was considered because it is the most abundant bacterial genus in Chignahuapan samples. Another of the abundant phyla was Proteobacteria. However, two genera of sulfur-oxidizing Beta-proteobacterium were considered in this analysis (*Thiomonas* and *Thiobacillus*) and *Bacillus* (Firmicutes). Although these genera were not the most abundant, they are sulfur-oxidizing bacteria, and so it is interesting to observe their interaction; the main network metrics were evaluated, which were connectance, nestedness, modularity index, and weighted closeness.

The *Rhodococcus* network showed a connectivity value of 0.54, weighted nestedness of 0.73, and modularity index of 0.39. The higher weighted closeness in the network was for *Rhodococcus erythropolis* (node A), with a value of 0.32. This result suggests that *R. erythropolis* is present in all circumneutral host springs; however, the value of weighted nestedness was closer for the Ching_S1_Mex and Ching_S2_Mex metagenomes, indicating the main predominance of this bacterium in metagenomes of Chignahuapan in comparison with other metagenomes. These results also confirm that *R. erythropolis* is the most abundant species (Figure 4A).

The *Thiomonas* cluster showed the following network metric values: connectance 0.66; modularity index 0.10; weighted nestedness 0.65 (node B); *Thiomonas intermedia* and *Thiomonas* sp. FB-6 had the highest value-weighted closeness with 0.13 and 0.19, respectively. It was interesting that it was closer to most of the nodes in the graph or present in all terrestrial hot springs. In contrast, *Thiomonas* sp. CB6 (node H), *Thiomonas* sp. ACO7 (node I), and *Thiomonas* sp. B1 (node J) had the lowest value for weighted closeness, with 0.0067. Overall, in the Mexico metagenomes, *Thiomonas* species were less shared with the other metagenomes, indicating the microbial composition is particular (Figure 4D).

The *Acidithiobacillus* cluster showed a connectivity of 0.58, modularity index of 0.06, and weighted nestedness of 0.71. The weighted closeness lowest values were obtained for *Thiobacillus* sp. 0-1251 (node K) at 0.0014, *Thiobacillus* sp. SCN 64-35 (node I) at 0.002, and *T. denitrificans* ATCC 25259 (node J) at 0.007. Higher values of weighted closeness were determined for *Thiobacillus* sp. 65-29 (node A) at 0.43 and *T. denitrificans* (node B) at 0.20. These results indicate that all hot springs share *Acidithiobacillus* and highlight a particular species from this environment. In general, the most abundant species were *Thiobacillus* sp. 65-29 (node A), *T. denitrificans* (node B), *Thiobacillus* sp. 63-78 (node D), *Thiobacillus* sp. GWE1_62_9 (node E), mainly involved in the sulfur oxidation systems (Figure 4C).

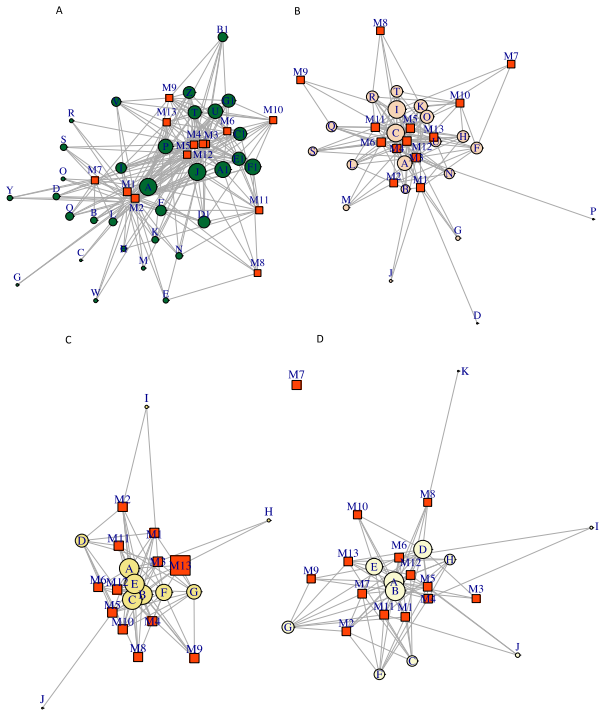


Figure 4. Co-occurrence network analysis by genus. The boxes represent the analyzed metagenomes and circle the species. (M1) Ching_S2_Mex, (M2) Ching_S1_Mex, (M3) ZNL_SRR10063240_34.5, (M4) ZNL_SRR10063241_30, (M5) ZNL_SRR10063242_27.2, (M6) CAN_SRR10095342_45, (M7) IN_SRR3050168_50, (M8) IN_SRR3961733_50, (M9) IN_SRR3961734_55, (M10) IN_SRR3961739_43, (M11) RSA_SRR5214706_50, (M12) JP_SRR7905022_60 (M13) IN_SRR816113699_52. (**A**) ***Rhodococcus*** **genus**. (A) *R. erythropolis*, (B) *R. erythropolis* SK121, (C) *R. erythropolis* DN1, (D) *R. erythropolis* PR4, (E) *R. erythropolis* CCM2595, (F) *R. qingshengii*, (G) *R. qingshengii* BKS 20-40, (H) *Rhodococcus* sp. 008, (I) *R. enclensis*, (J) *R. fascians*, (K) *Rhodococcus* sp. P27, (L) *Rhodococcus* sp. 311R, (M) *Rhodococcus* sp. ARP2, (N) *Rhodococcus* sp. ADH, (O) *Rhodococcus* sp. 164Chir2E, (P) *R. hoagii*, (Q) *Rhodococcus* sp. YH3-3, (R) *Rhodococcus* sp. BH4, (S) *Rhodococcus* sp. 66b, (T) *R. rhodochrous*, (U) *Rhodococcus* sp. RD6.2, (V) *Rhodococcus* sp. YL-1, (W) *Rhodococcus* sp. EPR-134, (Y) *Rhodococcus* sp. 1139, (Z) *Rhodococcus ruber*, (A1) *R. opacus*, (B1) *R. opacus* PD630, (C1) *Rhodococcus rhodnii*, (D1) *Rhodococcus* sp. AD45, (E1) *R. jostii*, (F1) *R. tukisamuensis*, (G1) *R. yunnanensis*. (**B**) ***Thiobacillus*** **genus**. (A) *Thiobacillus* sp. 65-29, (B) *T. denitrificans*, (C) *Thiobacillus* sp. 65-1059, (D) *Thiobacillus* sp. 63-78, (E) *Thiobacillus* sp. GWE1_62_9, (F) *T. thioparus*, (G) *Thiobacillus* sp. SCN 62-729, (H) *Thiobacillus* sp. SCN 63-374, (I) *Thiobacillus* sp. SCN 64-35, (J) *T. denitrificans* ATCC 25259, (K) *Thiobacillus* sp. 0-1251, (L) *Thiobacillus* sp. 65-1402, (M) *Thiobacillus* sp. SCN 64-317, (N) *Thiobacillus* sp. SCN 63-57, (O) *Thiobacillus* sp. SCN 63-1177, (P) uncultured Thiobacillus sp., (Q) *Thiobacillus* sp. 65-69. (**C**) ***Thiomonas*** **genus**. (A) *Thiomonas bhubaneswarensis*, (B) *Thiomonas intermedia*, (C) *Thiomonas* sp. FB-Cd, (D) *Thiomonas* sp. SCN 64-16, (E) *Thiomonas* sp. FB-6, (F) *Thiomonas* sp. CB2, (G) *Thiomonas* sp. CB3, (H) *Thiomonas* sp. CB6, (I) *Thiomonas* sp. ACO7, (J) *Thiomonas* sp. B1. (**D**) ***Bacillus*** **genus**. (A) *B. subtilis*, (B) *B. licheniformis*, (C) *B. cereus (D) B. cereus* R309803, (E) *B. mycoides*, (F) *B. wiedmannii*, (G) *Bacillus* sp. OxB-1, (H) *B. megaterium*, (I) *Bacillus* sp. F56, (J) *B. testis*, (K) *B. amyloliquefaciens group*, (M) *B. amyloliquefaciens*, (N) *B. salsus*, (O) *B. thuringiensis*, (P) *B. thuringiensis serovar israelensis* ATCC 35646, (Q) *B. weihenstephanensis*, (R) *B. anthracis*, (S) *B. gaemokensis*, (T) *B. manliponensis*, (U) *B. cytotoxicus*.

The *Bacillus* cluster showed the following network metric values: connectance, 0.51; modularity index, 0.471; weighted nestedness, 0.47. Furthermore, the value-weighted closeness of most higher values was 0.13 for to *Bacillus cytotoxicus* (node U), *Bacillus subtilis* (node A) with 0.09, and *Bacillus cereus* (node C) with 0.09; these showed higher centrality or connections within the network in all metagenomes. The Chignahuapan samples were closer to *Bacillus licheniformis* (node B) at 0.06, whereas the lower weighted closeness was *B. cereus* R309803 (node D) at 0.0009, *Bacillus* sp. OxB-1 (node G) at 0.003, *B. amyloliquefaciens* (node M) at 0.006, *B. testis* (node J) at 0.001, and *Bacillus thuringiensis* serovar *israelensis* ATCC 35646 (node P) at 0.0009, which indicates that these species of bacteria are poorly connected in the network and are unique within the metagenomes Chig_S2_Mex (M1) and ZNL_SRR10063240 (M3). Interestingly, *Bacillus cereus* R309803 is a unique species in the Chignahuapan metagenome (Figure 4B). The value of modularity suggests that the network of *Bacillus* has a modular structure in this cluster. Modularity with values above 0.44 indicates that the networks are more connected.

3.5. Functional Metagenomics Analysis

A functional analysis was performed with all metagenomes to determine whether the functional activity was similar in all the metagenomes; the amino acid sequences were annotated using SEED subsystems. The SEED subsystem is a classification system that organizes the coding sequences for functional categories into a hierarchy with 5 levels of resolution; in level 1, the families of proteins that share function. The results shown in Figure 5 correspond to the level 2 families.

The metabolism of carbohydrates (~13–22%) (central carbohydrate metabolism, CO_2 fixation, and fermentation) are the hot springs' main functional processes. From these, the metabolic pathways identified in high abundance correspond to the Calvin–Benson cycle, CO_2 uptake carboxysome, Tricarboxylic acid (TCA) cycle, and oxidative phosphorylation pathways involved in CO_2 fixation. Carbon fixation is a process where inorganic carbon (in the form of CO_2) is transformed into organic compounds and is an essential process for the production of anabolism precursors. Four metabolic pathways in bacteria have the capacity of Carbon fixation: Calvin–Benson cycle, the reverse TCA cycle, the Wood–Ljungdahl pathway, and the 3-hydroxypropionate (3-HP) bicycle [75,76]. Two carbon fixation pathways (Calvin–Benson cycle and TCA cycle) were identified in the metagenomes, whereas only the gene encoding the ribulose bisphosphate carboxylase (RuBisCO) was found in the samples. The results suggest that the Calvin–Benson cycle is the primary metabolic pathway in the Chignahuapan terrestrial hot springs water ecosystem (Figure S4).

Carbonate was found in high concentrations in the hot water spring of Chignahuapan; the high abundance of HCO_3 could correlate with the presence of pathways involved in the fixation of CO_2, and the carbonic anhydrase, involved in the interconversion of CO_2 to carbonate [77]. Additionally, the CO_2 is dissolved in water and can form different compounds such as carbon dioxide, carbonic acid, bicarbonate, and carbonate, and sulfur-oxidizing bacteria can fix it. In this context, *Acidithiobacillus ferrooxidans* was found in metagenomes and uses CO_2 for growth [78]. Furthermore, other thermophilic bacteria have been reported, including *Bacillus schlegelii* and *P. thermocarboxydovorans* growing with CO_2 as a unique carbon and energy source [79,80].

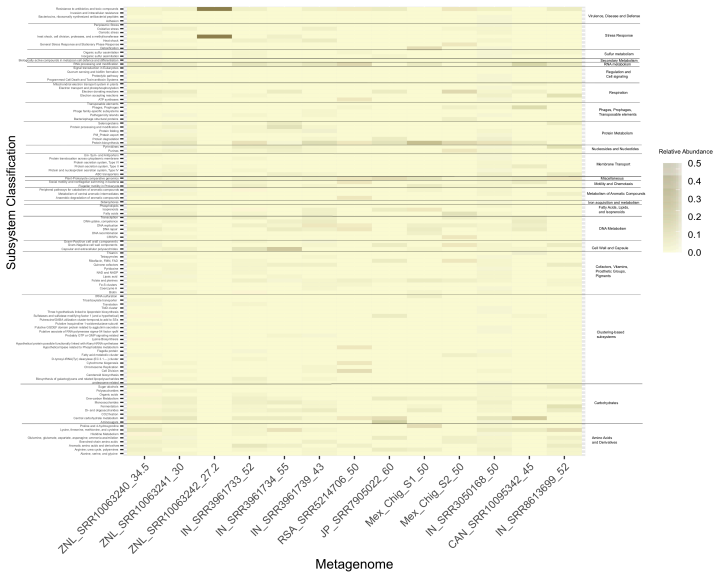


Figure 5. Functional annotation of circumneutral hot springs using the SEED subsystem database. On the x-axis are the different circumneutral metagenomes; on the y-axis is the classification into subsystems.

At the first level within the SEED category, the amino acids and derivatives, cofactors, vitamins, prosthetic groups, and pigments were the most abundant categories (~2–14%), followed by lysine, threonine, methionine, and cysteine biosynthesis. Most of these amino acids involved in protein synthesis and as cofactors in many metabolic reactions; however, methionine and cysteine biosynthesis are involved in the sulfur metabolism for biosynthesis of sulfur-containing biomolecules, such as the cysteine sulfinate-dependent pathways to produce sulfite or turine [81].

This metabolic category has also been reported in abundance in the hot springs of the Araró region, located in the Trans-Mexican Volcanic Belt [32], suggesting a central role in the biosynthesis of diverse compounds, for assimilatory sulfate reduction, SOX pathway, and for obtaining energy in thermal water environments (Figure 5).

Sulfur metabolism has been reported in microbial mats, hydrothermal vents, and YNP hot springs, where microorganisms called sulfur oxidizers live, as well as many chemolithotrophic Proteobacteria [77,82]. Overall, in the metagenomes analyzed, the proportion of sulfur metabolism was present in the metagenomes (0.9–1.9%).

Chignahuapan hot springs contain the complete SOX pathway. This pathway is involved in the oxidation of sulfide (S^{2-}) and thiosulfate ($S_2O_3^{2-}$) to sulfate (SO_4^{2-}) (Figure S5). This pathway has been reported present in alpha and epsilon-proteobacteria [18,59,83], identified in the two metagenomes of Puebla, suggesting that *epsilon-proteobacteria* are contributing to sulfide and thiosulfate oxidation, as an energy source. It has been reported that some bacteria can remove inorganic sulfur from oil, and it has been investigated that the enzymes to carry out this process are within the sox pathway; this was characterized in *Rhodococcus* sp. strain IGTS8 [84], the most abundant genera identified in the hot springs. The bacteria *H. neapolitanus* and *Acidithiobacillus caldus* have sulfur-oxidizing enzyme systems involved in the SOX pathway, and both bacteria were present in Chignahuapan.

In summary, sulfur oxidation is carried out by bacteria within the phylum Proteobacteria and Actinobacteria in Chignahuapan thermal springs. Where the assimilatory sulfate pathway involved in the reduction in sulfate (SO_4^{2-}) to sulfide (S^{2-}) was complete (Figure S5) [85,86]. These results contrast with bacteria involved in sulfur metabolism such as Deltaproteobacteria and Firmicutes found in Yellowstone National Park [87,88].

Nitrogen fixation is associated with carbon fixation in the microbial mat communities, and the nitrogen fixation occurred in many environments [89]. Nitrogen-fixing enzymes could be expected to be present in metagenomes. However, in general, the nitrogen fixation pathway enzymes found to be abundant in the samples of Chignahuapan were the enzymes involved with the pathway assimilatory nitrate reduction nitrate to ammonia (*nasA* and *nirB* genes) and ammonia-lyases also involved in ammonia production. As mentioned above, this pathway is associated with nitrogen fixation, and carbonic anhydrase was found here, which is associated with converting carbon dioxide to carbonates (Figure S6).

The stress oxidative and heat shock category was highly abundant in Chignahuapan of the reactive oxygen species (ROS) and can cause irreversible damage to cells, and indifferent thermophilic bacteria, and it has been reported to present a superoxide dismutase; this enzyme catalyzes the dismutation of the superoxide (O_2^{1-}) into oxygen (O_2). In the case of the samples, the bacteria that presented the putative enzyme were *Thioalkalivibrio* spp., and *Acinetobacter* spp., while *Rhodococcus opacus* B4 had NrdH, which mediates resistance to oxidative stresses. Previously, bacteria with superoxide dismutases had been reported, such as *Aquifex pyrophilus*, *Hydrogenobacter thermophilus*, *Thermus thermophilus*, *Propionibacterium shermanii*, and *Rhodothermus* sp. among others. This system is essential to avoid damage to extreme environments [90]. This abundant category of functions also relates to other thermal environments, such as the mats from Araró Mexico [32].

Similarly, as there are mechanisms to prevent cell damage, there are DNA repair mechanisms; previously in hypersaline waters, we observed that microorganisms have different enzymes involved in DNA repair [58]. In the case of thermophilic microorganisms, there are efficient mechanisms to prevent DNA repair and proteins and the lipid membrane from preventing these damages: for

example, the composition of fatty acids changes with increasing temperature [90]. The thermophilic microorganisms of Chignahuapan also have repair systems; however, they are abundantly presented with the enzyme exonuclease SbcC, which is involved in DNA repair when alkylation is damaged.

Overall, the annotation allowed us to predict the functional potential of the thermophilic community of circumneutral metagenomes. Whereas mainly fixation of carbon is the crucial pathway and amino acids and derivatives, many of them contribute to sulfur metabolism and fixation of carbon. Likewise, the pathway to oxidation and reduction in metabolites of sulfur was completed. Many of the thermophilic microorganisms have a mechanism to prevent oxidative stress damage and repair DNA damage.

3.6. Viral Community Composition

For the viral community in a circumneutral terrestrial hot spring, the analysis shows that overall, nine viral families were retrieved, and many of them infect bacteria, such as Siphoviridae, Myoviridae, Podoviridae, Corticoviridae, and Herelleviridae. These viruses also infect invertebrates Baculoviridae, eukaryotic algae Phycodnaviridae, and Protozoan Mimiviridae.

Clustering analysis revealed that most abundant families are Myoviridae and Siphoviridae, which are ubiquitous in all metagenomes. These samples also showed that the metagenomes of Mexico can be grouped with samples from India and Canada with similar temperatures (Figure 6).

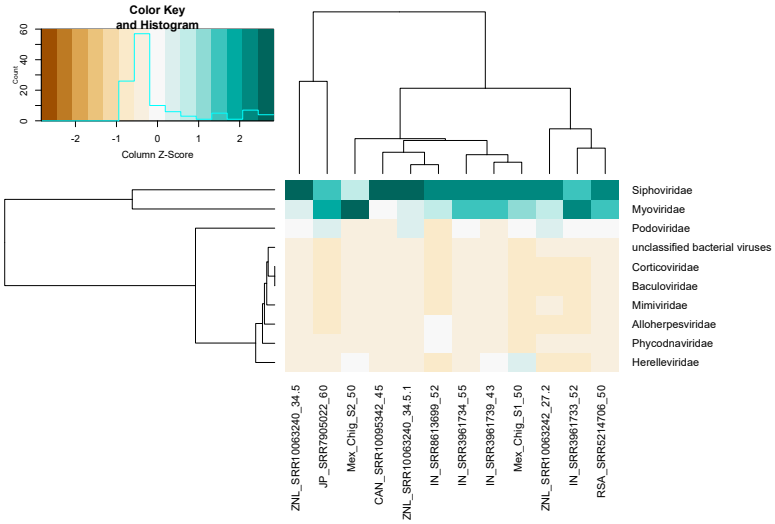


Figure 6. Heat map abundance of viral families retrieved from Vibrant. Myoviridae and Siphoviridae families were ubiquitous in metagenomes.

These results contrast with previous reports in thermal or geothermal terrestrial environments; it has been determined that the main families of viruses present are Fuselloviridae, Bicaudaviridae, Turriviridae, Ampullaviridae, Guttaviridae, Lipothrixviridae, Rudiviridae, and Globuloviridae [3,91]. Most of these virus families infect Archaea that live in thermal environments with a temperature above 80 °C, and these viruses have been called Archeoviruses [92,93].

However, it was expected that the viruses recovered in these metagenomes mainly infect bacteria, since the temperature and pH are also moderate in the analyzed metagenomes. Thus, the dominance of viruses that infect prokaryotes found are moderate thermophilic bacteria. Nevertheless, in the

hot springs of Chignahuapan, besides Siphoviridae and Myoviridae, the Herelleviriadae family was abundant.

The Herelleviriadae family was recently described: the phylogenetic evidence performed at the genome and proteome level, and by a single gene (capsid, tail protein, DnaB395 like-helicase), showed that these viruses are polyphyletic since they are grouped into five different clades or subgroups. This family contains linear viral genomes, with a length of 125–170 kbp, that infect the genus *Bacillus* [94].

The abundance of bacteriophages of the Herelleviriadae family correlates with the taxonomic assignment results, where the genus *Bacillus* was present in both metagenomes, and according to the co-occurrence analysis, the main associated species of the metagenomes of Mexico were mainly *B. subtilis* and *B. licheniformis*.

VIRSorter and VIBRANT programs were used in order to retrieve viral contigs. This software failed to recover the complete genome from samples of Chignahuapan, only partial sequences, probably because of the low performance in the metagenome sequencing. However, according to the taxonomic classification from viral contig retrieval, the most abundant viruses were *Acidithiobacillus* phage AcaML1, *Bacillus* phage SIOphi, *Bacillus* virus Bobb, *Acinetobacter* virus R3177, and *Bacillus* phage Shbh1 in the samples from Mexico, correlating with what was found in bacteria. According to the taxonomic assignment, 200 viruses were recovered with Vibrant and 125 viruses with Virsorter in all samples. In the Japanese metagenome, there was a greater abundance of viruses (Figure S7).

Overall, the viral communities in moderate thermophilic environments that infected moderate to thermophilic bacteria demonstrated distinct viral community structures among the circumneutral thermal springs, compared with acid or hyperthermophile hot springs such as Yellowstone [92]. These results indicate that the circumneutral thermal springs harbor viral communities with phage double-stranded DNA that infect mainly bacteria, followed by viruses that infect invertebrates or eukaryotic algae.

Through an analysis of occurrence carried out with the recovered viruses from Vibrant, the results from the analysis revealed there are few connections between virus species in the metagenomes, each having many particular species. These results correlate with the classifications of bacteria, where each one observed that each environment has a genus of particular bacteria, and that terrestrial hot springs have a common *Acidithiobacillus phage* AcaML1, which infects *Acidithiobacillus caldus* (Figure S8).

Acidithiobacillus caldus has been reported in the various thermal environments. It is a moderately thermoacidophilic bacteria that contributes to the carbon and sulfur cycles, as it obtains energy from the oxidation of elemental sulfur for carbon dioxide fixation, and it is ubiquitous in sulfide mineral environments [95].

3.7. Auxiliary Metabolic Genes and Whole Viral Genomes

We evaluated the presence of AMGs in the viral contigs recovered from all the metagenomes, but in the case of the Chignahuapan metagenomes, no AMGs were obtained. The viral genes recovered in these correspond mainly to structural parts of the virion. Thus, this comparison was carried out only with those viral contigs that contained AGMs.

The AMGs have been related to an increase in the fitness and altering or complementing of their host's metabolism, facilitating adaptation under adverse conditions [96–98], and they have been related to photosynthesis, carbon fixation [99], and sulfur and nitrogen biogeochemical cycles [100].

The most abundant AMGs were classified within the functional categories metabolism of cofactors and vitamins (MCV), carbohydrates metabolism (CM), amino acids metabolism (AAM), metabolism of terpenoids, and polyketides (Figure 7).

In the MCV category, the virus contributes to the folate biosynthesis pathway (KEGG entries identified: K00287, K01495, K01737, K06920, K09457, K10026). Tetrahydrofolate is a cofactor present in all bacteria, and it is essential to the growth; synthesis of formylmethionyl tRNAfMet, carried out by the enzyme dihydrofolate reductase (DHFR); and in thermophilic bacteria, has been identified to use modified folates [101], and within the folate biosynthesis pathway, the AMG *7-cyano-7-deazaguanine*

synthase was also found, a hypermodified 7-deazaguanosine, which has been previously reported in viruses; it is proposed that phages have taken the 7-deazaguanine from bacteria to evade the restriction-modification system (RM system) [101,102]. Interestingly, viruses found in thermophilic environments have these AGMs.

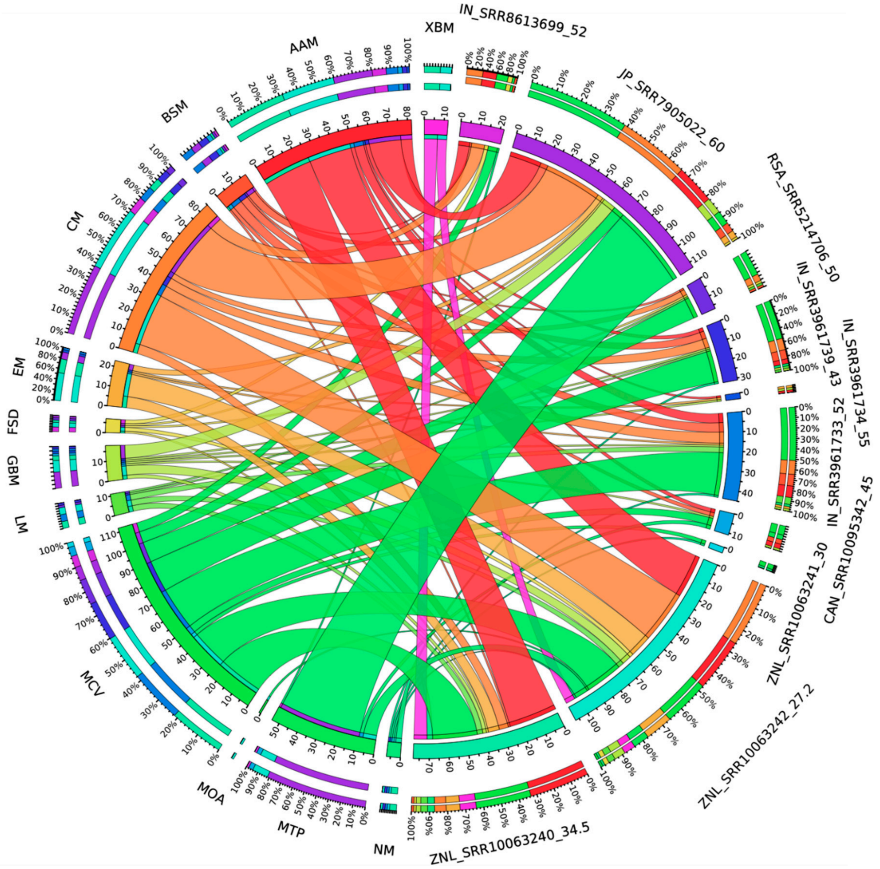


Figure 7. Auxiliary metabolic genes recovered from Vibrant. Representation of the metagenomes that contain the viral AMG. AAM: Amino acid metabolism; GBM: Glycan biosynthesis and metabolism; MCV: Metabolism of cofactors and vitamins; MTP: Metabolism of terpenoids and polyketides; BSM: Biosynthesis of other secondary metabolites; NM: Nucleotide metabolism; CM: Carbohydrate metabolism; LP: Lipid metabolism; EM: Energy metabolism; FSD: Folding, sorting and degradation; XBM: Xenobiotics biodegradation and metabolism; MOA: Metabolism of other amino acids.

Additionally, other metabolic pathways were found within this category MCV such as Thiamine metabolism (K03153, K04487), biotin metabolism (K00059, K09458), pantothenate and coenzyme (CoA) biosynthesis (K13038), and retinol metabolism (K11153), and within nicotinate and nicotinamide metabolism (K00858, K01916, K03462, K13522) viruses have a gene that could modify or complement these metabolic pathways.

In the category carbohydrate metabolism (CM), AMGs were involved in the pentose phosphate pathway (K00616, K01053), galactose metabolism (K01784), ascorbate and alternate metabolism

(K00012), propanoate metabolism (K00822), and glyoxylate and dicarboxylate metabolism. For example, in the case of the phosphate pathway transaldolase (*talC* gene), it catalyzes glyceraldehyde-3-phosphate, sedoheptulose-7-phosphate into fructose-6-phosphate and erythrose-4-phosphate. This class of transaldolases has been reported in cyanomyoviruses [97].

The categories least abundant but not less interesting, are those for obtaining energy and degradation of hydrocarbons. Some viruses have been reported in oil reserves, which could have interesting functions with the biotechnological applications [103].

The energy category was found in a low proportion; the pathways involved were sulfur metabolism and methane metabolism. One of the most important enzymes that we found was *cysH* phosphoadenosine phosphosulfate reductase, which is involved in the reduction in 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to sulfite, an intermediate step of the pathway assimilatory sulfate reduction. Additionally, *sufS* which encodes cysteine desulfurase was found: this enzyme mobilizes the sulfur from L-cysteine.

In previous work, it has been reported that the viruses inside hydrothermal vents carry out AMG related to dissimilatory sulfite; specifically, the enzyme reverse dissimilatory sulfite reductases encoded by *rdsrA* and *rdsrC* in charge of oxidizing elemental sulfur, and some oxidizing sulfur bacteria lack the Sox system and use *rdsrA* for oxidizing elemental sulfur [104]. Nevertheless, here, the AMGs found were *cysH* and *sufS* involved in sulfur metabolism. Therefore, viruses and bacteria present in circumneutral environments contribute to the sulfur cycle.

Finally, there are some bacteria that are capable of carrying out the degradation of hydrocarbons; interestingly, we found viral AMGs in low abundance involved in the degradation of xenobiotics: for example, in the Benzoate degradation pathway (K01055), *pcaD* genes encoding to 3-oxoadipate enol-lactonase. Toluene degradation and fluorobenzoate degradation pathway (K01061) were found through the hydrolase carboxymethylenebutenolidase. This represents a biotechnologically important finding since these sequences could be obtained for use in the industry. These results reveal that viruses contribute to carbohydrate metabolism, sulfur cycles, and the degradation of aromatic compounds.

4. Conclusions

The composition of microorganisms in Chignahuapan is driven by chemical composition and geographic location. Since the microbial community's structure was particular where the bacteria of the genera *Rhodococcus*, *Acinetobacter*, *Thermomonas*, *Tepidimonas*, and *Azotobacter* predominated, in comparison to other circumneutral environments, these bacteria are sulfur oxidizers, which is consistent with the functional analysis where the sulfur reduction and oxidation pathways were complete. The functional analysis also predicted that the Calvin–Benson cycle metabolic pathways are the main pathways to contribute to carbon fixation. Furthermore, the microorganisms present in circumneutral environments have mechanisms that prevent cellular and DNA damage. Therefore, the microbial community structure in particular in each location is driven by physicochemical properties, but many metabolic pathways were common in circumneutral terrestrial hot springs, which contribute to carbon fixation. In the hot springs' viral community, prokaryotic viruses predominate overall, but in Chignahuapan, the Herelleviridae family was mainly abundant. The analysis of the auxiliary genes revealed that the viruses also contribute to metabolic pathways and the sulfur cycle. This study shows the information on the microbial and viral diversity in Mexican hot springs and compares these microbial communities with other metagenomic samples, thus providing an opportunity to understand the role of the viral AMGs and the structure of the viruses in the adaptation process to circumneutral hot springs. Additionally, this report serves as a reference to viromes in extremophile environments from Mexico.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2607/8/11/1677/s1, Figure S1: Taxonomic profiling and community structure comparison at the domain level. Our profiling showed a dominance of Bacteria, at around 88.4 to 91.8%, followed by Archaea, viruses, and eukaryotes; Figure S2: Stack plots of the relative abundances of Archaea and Bacterial at the level of phylum. (A) Archaea domain. The Euryarchaeota phylum was dominant in the samples. (B) Bacteria domain. The Proteobacteria phylum was dominant in the samples; Figure S3: Biplot showing an association of genera based on location. The samples are grouped based on location; Figure S4: Pathway carbon metabolism from Chignahuapan. Colors indicate enzymes present in metagenomes; Figure S5: Pathway sulfur metabolism from Chignahuapan the green color indicates the enzymes present in the metagenome; Figure S6: Pathway nitrogen metabolism from Chignahuapan. Colors indicate enzymes present in metagenomes; Figure S7: Species of virus identification (A) Vibrant and (B) Virsorter; Figure S8: Virus network analysis. Metagenomes (M1) ZNL_SRR10063242_27.2, (M2) JP_SRR7905022_60, (M3) IN_SRR3961734_55, (M4) Mex_Chig_S1, (M5) ZNL_SRR10063241_30, (M6) ZNL_SRR10063240_34.5, (M7) IN_SRR3961739_43, (M8) RSA_SRR5214706_50, (M9) IN_SRR3961733_52, (M10), IN_SRR8613699_52, (M11) CAN_SRR10095342_45, (M11) Mex_Chig_S2. Virus species (1) Bacillus virus Bobb, (2) Bacillus virus Bcp1, (3) Nitunavirus, (4) Siminovitchvirus, (5) Enterobacteria phage vB_KleM-RaK2, (6) Bixzunavirus, (7) Acidovorax virus ACP17, (8)Sinorhizobium virus M12, (9) Pectobacterium phage CBB, (10) Campylobacter virus Los1, (11) Campylobacter phage PC5, (12) Enterobacter phage Arya, (13) Pseudomonas phage PPpW-3, (14) Erwinia phage vB_EamM_Parshik, (15) Acinetobacter virus ME3, (16) Acinetobacter virus LZ35, (17) Burkholderia phage KS5, (18) Burkholderia phage vB_BceM_AP3, (19) Ralstonia phage RSY1, (20) Stenotrophomonas phage Smp131, (21) Pseudoalteromonas phage C5a, (22) Agrobacterium virus Atuph07, (23) Svunavirus, (24) Enterobacter virus PG7, (25) unclassified Tequatrovirus, (26) Yersinia virus PST, (27) Acinetobacter phage Acj9, (28) Morganella phage vB_MmoM_MP1, (29) Acidithiobacillus phage AcaML1, (30) Acinetobacter phage Ab105-1phi, (31) Alteromonadaceae phage B23, (32) Aurantimonas phage AmM-1, (33) Bacillus phage AR9, (34) Bordetella phage vB_BbrM_PHB04, (35) Bradyrhizobium phage BDU-MI-1, (36) Caulobacter phage Cr30, (37) Cyanophage S-RIM12, (38) Cyanophage S-RIM14, (39) Cyanophage S-RIM32, (40) Cyanophage S-RIM44, (41) Deep-sea thermophilic phage D6E, (42) Faecalibacterium phage FP_Mushu, (43) Lake Baikal phage Baikal-20-5m-C28, (44) Ochrobactrum phage POA1180, (45) Prochlorococcus phage P-HM1, (46) Prochlorococcus phage P-SSM2, (47) Prochlorococcus phage P-SSM7, (48) Prochlorococcus phage P-TIM68, (49) Pseudomonas phage Lu11, (50) Pseudomonas phage PaBG, (51) Rhizobium phage vB_RleM_PPF1, (52) Salicola phage SCTP-2, (53) Shewanella phage SFCi1, (54) Shewanella sp. phage 1/40, (55) Shigella phage SfIV, (56) Stenotrophomonas phage IME-SM1, (57) Stenotrophomonas phage vB_SmaS-DLP_6, (58) Synechococcus phage ACG-2014f, (59) Synechococcus phage ACG-2014g, (60) Thermus phage TMA, (61) Xanthomonas phage XacN1, (62) Yersinia phage phiR1-37, (63) Edwardsiella virus MSW3, (64) Enterobacteria phage J8-65, (65) Pseudomonas phage vB_PaeP_PAO1_Ab05, (66) Ralstonia phage RS-PII-1, (67) Ralstonia phage RsoP1EGY, (68) Rhizobium phage RHEph01, (69) Erwinia phage vB_EamP-S2, (70) Escherichia virus Pollock, (71) Erwinia virus Frozen, (72) Pseudomonas virus KPP25, (73) Escherichia phage APC_JM3.2, (74) Bordetella virus BPP1, (75) Enterobacter phage Tyrion, (76) Aeromonas phage phiARM81mr, (77) Agrobacterium phage Atu_ph08, (78) Burkholderia phage vB_BmuP_KL4, (79) Cellulophaga phage phi46:3, (80) Cellulophaga phage phi14:2, (81) Delftia phage RG-2014, (82) Pseudoalteromonas phage HP1, (83) Pseudomonas phage AF, (84) Pseudomonas phage TC6, (85) Pseudomonas phage ZC08, (86) Puniceispirillum phage HMO-2011, (87) Ralstonia phage DU_RP_II, (88) Ralstonia phage RSK1, (89) Sinorhizobium phage PBC5, (90) Sinorhizobium phage phiM5, (91) Xanthomonas citri phage CP2, (92) Xylella phage Xfas53, (93) Burkholderia virus AH2, (94) Streptomyces phage Maneekul, (95) Mycobacterium virus Vincenzo, (96) Mycobacterium virus Godines, (97) Mycobacterium phage 40BC, (98) Pseudomonas phage JBD18, (99) Gordonia virus Bowser, (100) Gordonia virus Britbrat, (101) Pseudomonas phage MP42, (102) Vibrio virus pVp1, (103) Rhodobacter virus RcCronus, (104) Stenotrophomonas virus DLP5, (105) Pseudomonas phage phi1, (106) Doucettevirus, (107) Mycobacterium virus Pukovnik, (108) Mycobacterium virus Timshel, (109) Mycobacterium phage HINdeR, (110) Escherichia phage ST2, (111) Microbacterium virus Koji, (112) unclassified Laroyevirus, (113) Vibrio virus MAR10, (114) Marvinvirus, (115) Dinoroseobacter virus D5C, (116) Mycobacterium virus Panchino, (117) Pseudomonas virus PaMx25, (118) Pseudomonas phage JG012, (119) Gordonia virus Zirinka, (120) Gordonia phage BatStarr, (121) Pseudomonas phage AAT-1l, (122) Xanthomonas phage Xoo-sp2, (123) unclassified Pbi1virus, (124) Escherichia phage YDC107_1, (125) Streptomyces virus Jay2Jay, (126) Gordonia virus OneUp, (127) Burkholderia phage Bcep176, (128) Burkholderia phage KS9, (129) unclassified Timquatrovirus, (130) Arthrobacter phage vB_ArS-ArV2, (131) Azospirillum phage Cd, (132) Bacillus phage vB_BhaS-171, (133) Bifidobacterium phage Bbif-1, (134) Caulobacter phage CcrColossus, (135) Caulobacter phage Sansa, (136) Cellulophaga phage phi19:1, (137) Corynebacterium phage LGCM-V4, (138) Croceibacter phage P2559Y, (139) Erysipelothrix phage phi1605, (140) Geobacillus virus E2, (141) Geobacillus virus E3, (142) Gordonia phage Confidence, (143) Gordonia phage GMA1, (144) Gordonia phage GMA2, (145) Gordonia phage McGonagall, (146) Halomonas phage QHHSV-1, (147) Klebsiella phage 5 LV-2017, (148) Lactobacillus phage PLE3, (149) Lactococcus phage PLg-TB25, (150) Microbacterium phage Paschalis, (151) Pseudomonas phage JBD25, (153) Pseudomonas phage JBD44, (154) Pseudomonas phage JBD68, (155) Pseudomonas phage phiPSA1, (156) Pseudomonas phage PS-1, (157) Pseudomonas phage YMC11/02/R656, (158) Psychrobacter phage Psymv2, (159) Ralstonia phage RS138, (160) Rhizobium phage 16-3, (161) Rhizobium phage vB_RleS_L338C, (161) Rhodobacter phage RcapMu, (162) Rhodococcus phage Jace, (163) Rhodovulum phage vB_RhkS_P1, (164) Sinorhizobium phage phi2LM21, (165) Sinorhizobium phage phi3LM21, (166) Sinorhizobium phage phiLM21, (167) Stenotrophomonas phage S1, (168) Streptococcus phage phiZJ20091101-1, (169) Streptomyces phage Chymera, (170) Streptomyces phage Ibantik, (171) Streptomyces phage mu1/6, (172) Synechococcus phage S-CBS3, (173) Synechococcus virus S-ESS1, (174) Thiobacimonas phage vB_ThpS-P1, (175) Paenibacillus

phage Dragolir, (176) Paracoccus virus Shpa, (177) Acinetobacter virus R3177, (178) Acinetobacter phage Ab105-2phi, (179) Acinetobacter phage vB_AbaS_TRS1, (180) Gordonia virus Vivi2, (181) Roseobacter virus RDJL1, (182) Pseudomonas virus LKO4, (183) Pseudomonas virus MP1412, (184) Pseudomonas virus PAE1, (185) Bordetella phage FP1, (186) Pseudomonas phage AN14, (187) Pseudomonas phage vB_PaeS_S218, (188) Marinomonas phage YY, (189) Cyprinid herpesvirus 2, (190) Ictalurid herpesvirus 1, (191) Megavirus chiliensis, (192) Moumouvirus, (193) Tupanvirus soda lake, (194) Paramecium bursaria Chlorella virus 1, (195) Ostreococcus tauri virus 2, (196) Yellowstone lake phycodnavirus 1, (197) Yellowstone lake phycodnavirus 2, (198) Campylobacter phage A18a, (199) Synechococcus phage S; Table S1: Metagenome data considered for the analysis.

Author Contributions: Design of the work, H.G.C.-S., S.D.-R., and E.P.-R. Collection and the metagenomics, D.I.R.-V., H.G.C.-S., and N.E.R.-R. Processing of samples, DNA extraction, and preparation, A.L.-T., D.I.R.-V., and H.G.C.S. Data analysis with bioinformatic tools, H.G.C.-S., P.M.M.-R., E.C., and A.L.-T. Interpretation of the experimental data, H.G.C.-S., S.D.-R., E.P.-R., N.E.R.-R., and R.A.B.-G. Resources, S.D.-R., E.P.-R., and N.E.R.-R. Writing—original draft and development of the manuscript, H.G.C.-S. All authors contributed to the preparation of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: H.G.C.-S. is a doctoral student from Programa de Doctorado en Ciencias Universidad Autónoma del Estado de Morelos, and received a fellowship number 227229 from Consejo Nacional de Ciencia y Tecnología (CONACYT). E.P.-R. was supported by the Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN-209620).

Acknowledgments: We thank the Instituto de Biotecnología—UNAM for giving us access to its computer cluster. The authors gratefully acknowledge Arturo Flores for the facilities granted to carry out the sampling in "*Baños Termales de Chignahuapan*".

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Inskeep, W.P.; Jay, Z.J.; Tringe, S.G.; Herrgard, M.; Rusch, D.B. The YNP Metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Front. Microbiol.* **2013**, *4*, 67. [CrossRef] [PubMed]
2. Mardanov, A.V.; Gumerov, V.M.; Beletsky, A.V.; Ravin, N.V. Microbial diversity in acidic thermal pools in the Uzon Caldera, Kamchatka. *Antonie Van Leeuwenhoek* **2018**, *111*, 35–43. [CrossRef] [PubMed]
3. Strazzulli, A.; Fusco, S.; Cobucci-Ponzano, B.; Moracci, M.; Contursi, P. Metagenomics of microbial and viral life in terrestrial geothermal environments. *Rev. Environ. Sci. Biotechnol.* **2017**, *16*, 425–454. [CrossRef]
4. Arsanova, G.I. The origin of thermal waters in volcanic areas. *J. Volcanol. Seismol.* **2014**, *8*, 361–374. [CrossRef]
5. Inskeep, W.P.; Jay, Z.J.; Herrgard, M.J.; Kozubal, M.A.; Rusch, D.B.; Tringe, S.G.; Macur, R.E.; deM. Jennings, R.; Boyd, E.S.; Spear, J.R.; et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front. Microbiol.* **2013**, *4*, 95. [CrossRef] [PubMed]
6. Valeriani, F.; Crognale, S.; Protano, C.; Gianfranceschi, G.; Orsini, M.; Vitali, M.; Spica, V.R. Metagenomic analysis of bacterial community in a travertine depositing hot spring. *New Microbiol.* **2018**, *41*, 126–135.
7. López-López, O.; Cerdán, M.E.; González-Siso, M.I. Hot spring metagenomics. *Life* **2013**, *3*, 308–320. [CrossRef]
8. Marsh, C.L.; Larsen, D.H. CHARACTERIZATION of some thermophilic bacteria from the hot springs of yellowstone national park. *J. Bacteriol.* **1953**, *65*, 193–197. [CrossRef]
9. Saxena, R.; Dhakan, D.B.; Mittal, P.; Waiker, P.; Chowdhury, A.; Ghatak, A.; Sharma, V.K. Metagenomic analysis of hot springs in central india reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.* **2017**, *7*, 2123. [CrossRef]
10. Mangrola, A.; Dudhagara, P.; Koringa, P.; Joshi, C.G.; Parmar, M.; Patel, R. Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genom. Data* **2015**, *4*, 153–155. [CrossRef]
11. Sahoo, R.K.; Subudhi, E.; Kumar, M. Investigation of bacterial diversity of hot springs of Odisha, India. *Genom. Data* **2015**, *6*, 188–190. [CrossRef] [PubMed]

12. Sangwan, N.; Lambert, C.; Sharma, A.; Gupta, V.; Khurana, P.; Khurana, J.P.; Sockett, R.E.; Gilbert, J.A.; Lal, R. Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator-prey genotypes: Metagenomic recovery of predator prey genotypes. *Environ. Microbiol. Rep.* **2015**, *7*, 812–823. [CrossRef] [PubMed]
13. Amin, A.; Ahmed, I.; Salam, N.; Kim, B.-Y.; Singh, D.; Zhi, X.-Y.; Xiao, M.; Li, W.-J. Diversity and distribution of thermophilic bacteria in hot springs of pakistan. *Microb. Ecol.* **2017**, *74*, 116–127. [CrossRef] [PubMed]
14. Panda, A.K.; Bisht, S.S.; De Mandal, S.; Kumar, N.S. Bacterial and archeal community composition in hot springs from Indo-Burma region, North-East India. *AMB Express* **2016**, *6*, 111. [CrossRef] [PubMed]
15. Jiménez, D.J.; Andreote, F.D.; Chaves, D.; Montaña, J.S.; Osorio-Forero, C.; Junca, H.; Zambrano, M.M.; Baena, S. Structural and functional insights from the metagenome of an acidic hot spring microbial Planktonic Community in the Colombian Andes. *PLoS ONE* **2012**, *7*, e52069. [CrossRef] [PubMed]
16. Paul, S.; Cortez, Y.; Vera, N.; Villena, G.K.; Gutiérrez-Correa, M. Metagenomic analysis of microbial community of an Amazonian geothermal spring in Peru. *Genom. Data* **2016**, *9*, 63–66. [CrossRef]
17. Hussein, E.I.; Jacob, J.H.; Shakhatreh, M.A.K.; Al-razaq, M.A.A.; Juhmani, A.F.; Cornelison, C.T. Exploring the microbial diversity in Jordanian hot springs by comparative metagenomic analysis. *MicrobiologyOpen* **2017**, *6*, e00521. [CrossRef]
18. Kubo, K.; Knittel, K.; Amann, R.; Fukui, M.; Matsuura, K. Sulfur-metabolizing bacterial populations in microbial mats of the Nakabusa hot spring, Japan. *Syst. Appl. Microbiol.* **2011**, *34*, 293–302. [CrossRef]
19. Nishiyama, E.; Higashi, K.; Mori, H.; Suda, K.; Nakamura, H.; Omori, S.; Maruyama, S.; Hongoh, Y.; Kurokawa, K. The relationship between Microbial Community Structures and Environmental Parameters Revealed by Metagenomic Analysis of hot spring water in the Kirishima Area, Japan. *Front. Bioeng. Biotechnol.* **2018**, *6*, 202. [CrossRef]
20. Tobler, D.J.; Benning, L.G. Bacterial diversity in five Icelandic geothermal waters: Temperature and sinter growth rate effects. *Extremophiles* **2011**, *15*, 473. [CrossRef]
21. Chan, C.S.; Chan, K.-G.; Ee, R.; Hong, K.-W.; Urbieta, M.S.; Donati, E.R.; Shamsir, M.S.; Goh, K.M. Effects of physiochemical factors on prokaryotic biodiversity in malaysian circumneutral hot springs. *Front. Microbiol.* **2017**, *8*, 1252. [CrossRef] [PubMed]
22. Menzel, P.; Gudbergsdóttir, S.R.; Rike, A.G.; Lin, L.; Zhang, Q.; Contursi, P.; Moracci, M.; Kristjansson, J.K.; Bolduc, B.; Gavrilov, S.; et al. Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* **2015**, *70*, 411–424. [CrossRef] [PubMed]
23. Inskeep, W.P.; Rusch, D.B.; Jay, Z.J.; Herrgard, M.J.; Kozubal, M.A.; Richardson, T.H.; Macur, R.E.; Hamamura, N.; deM. Jennings, R.; Fouke, B.W.; et al. Metagenomes from High-Temperature Chemotrophic Systems Reveal Geochemical Controls on Microbial Community Structure and Function. *PLoS ONE* **2010**, *5*, e9773. [CrossRef] [PubMed]
24. López-López, O.; Knapik, K.; Cerdán, M.-E.; González-Siso, M.-I. Metagenomics of an alkaline hot spring in Galicia (Spain): Microbial diversity analysis and screening for novel lipolytic enzymes. *Front. Microbiol.* **2015**, *6*, 1291. [CrossRef]
25. Brito, E.M.S.; Villegas-Negrete, N.; Sotelo-González, I.A.; Caretta, C.A.; Goñi-Urriza, M.; Gassie, C.; Hakil, F.; Colin, Y.; Duran, R.; Gutiérrez-Corona, F.; et al. Microbial diversity in Los Azufres geothermal field (Michoacán, Mexico) and isolation of representative sulfate and sulfur reducers. *Extremophiles* **2014**, *18*, 385–398. [CrossRef]
26. Massello, F.L.; Chan, C.S.; Chan, K.-G.; Goh, K.M.; Donati, E.; Urbieta, M.S. Meta-Analysis of microbial communities in hot springs: Recurrent taxa and complex shaping factors beyond PH and temperature. *Microorganisms* **2020**, *8*, 906. [CrossRef]
27. Sánchez-Córdova, M.M.; Canet, C.; Rodríguez-Díaz, A.; González-Partida, E.; Linares-López, C. Water-rock interactions in the Acoculco geothermal system, eastern Mexico: Insights from paragenesis and elemental mass-balance. *Geochemistry* **2020**, *80*, 125527. [CrossRef]
28. López-Hernández, A.; García-Estrada, G.; Aguirre-Díaz, G.; González-Partida, E.; Palma-Guzmán, H.; Quijano-León, J.L. Hydrothermal activity in the Tulancingo–Acoculco Caldera Complex, central Mexico: Exploratory studies. *Geothermics* **2009**, *38*, 279–293. [CrossRef]
29. Servín-Garcidueñas, L.E.; Martínez-Romero, E. Draft genome sequence of the Sulfolobales Archaeon AZ1, obtained through metagenomic analysis of a Mexican Hot Spring. *Genome Announc.* **2014**, *2*, e00164-14. [CrossRef]

50. Roux, S.; Enault, F.; Hurwitz, B.L.; Sullivan, M.B. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* **2015**, *3*, e985. [CrossRef] [PubMed]
51. Huson, D.H.; Auch, A.F.; Qi, J.; Schuster, S.C. MEGAN analysis of metagenomic data. *Genome Res.* **2007**, *17*, 377–386. [CrossRef]
52. VIBRANT: Automated Recovery, Annotation and Curation of Microbial Viruses, and Evaluation of Virome Function from Genomic Sequences|bioRxiv. Available online: https://www.biorxiv.org/content/10.1101/855387v1 (accessed on 10 June 2020).
53. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]
54. Silva, G.G.Z.; Green, K.T.; Dutilh, B.E.; Edwards, R.A. SUPER-FOCUS: A tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* **2016**, *32*, 354–361. [CrossRef]
55. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis, Use R!* 2nd ed.; Springer: Cham, Switzerland, 2016; ISBN 978-3-319-24277-4.
56. Keegan, K.P.; Glass, E.M.; Meyer, F. MG-RAST, a metagenomics service for analysis of microbial community structure and function. In *Microbial Environmental Genomics (MEG)*; Martin, F., Uroz, S., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2016; pp. 207–233. ISBN 978-1-4939-3369-3.
57. Wang, J.; Yang, D.; Zhang, Y.; Shen, J.; van der Gast, C.; Hahn, M.W.; Wu, Q. Do Patterns of bacterial diversity along salinity gradients differ from those observed for macroorganisms? *PLoS ONE* **2011**, *6*, e27597. [CrossRef]
58. Castelán-Sánchez, H.G.; Elorrieta, P.; Romoacca, P.; Liñan-Torres, A.; Sierra, J.L.; Vera, I.; Batista-García, R.A.; Tenorio-Salgado, S.; Lizama-Uc, G.; Pérez-Rueda, E.; et al. Intermediate-Salinity systems at high altitudes in the peruvian andes unveil a high diversity and abundance of bacteria and viruses. *Genes* **2019**, *10*, 891. [CrossRef]
59. Ohhata, N.; Yoshida, N.; Egami, H.; Katsuragi, T.; Tani, Y.; Takagi, H. An extremely oligotrophic bacterium, rhodococcus erythropolis N9T-4, isolated from crude oil. *J. Bacteriol.* **2007**, *189*, 6824–6831. [CrossRef]
60. Castorena, G.; Suárez, C.; Valdez, I.; Amador, G.; Fernández, L.; Le Borgne, S. Sulfur-selective desulfurization of dibenzothiophene and diesel oil by newly isolated Rhodococcus sp. strains. *FEMS Microbiol. Lett.* **2002**, *215*, 157–161. [CrossRef]
61. Veith, A.; Botelho, H.M.; Kindinger, F.; Gomes, C.M.; Kletzin, A. The sulfur oxygenase reductase from the mesophilic bacterium halothiobacillus neapolitanus is a highly active thermozyme. *J. Bacteriol.* **2012**, *194*, 677–685. [CrossRef]
62. Mohammad, B.T.; Al Daghistani, H.I.; Jaouani, A.; Abdel-Latif, S.; Kennes, C. Isolation and Characterization of Thermophilic Bacteria from Jordanian Hot Springs: Bacillus Licheniformis and Thermomonas Hydrothermalis Isolates as Potential Producers of Thermostable Enzymes. Available online: https://www.hindawi.com/journals/ijmicro/2017/6943952/ (accessed on 16 June 2020).
63. Rey, M.W.; Ramaiya, P.; Nelson, B.A.; Brody-Karpin, S.D.; Zaretsky, E.J.; Tang, M.; de Leon, A.L.; Xiang, H.; Gusti, V.; Clausen, I.G.; et al. Complete genome sequence of the industrial bacterium Bacillus licheniformis and comparisons with closely related Bacillus species. *Genome Biol.* **2004**, *5*, 1–12. [CrossRef] [PubMed]
64. Inagaki, F.; Takai, K.; Nealson, K.H.; Horikoshi, K. Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the ε-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 1477–1482. [CrossRef]
65. Vieille, C.; Zeikus, G.J. Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1–43. [CrossRef] [PubMed]
66. Panda, S.K.; Jyoti, V.; Bhadra, B.; Nayak, K.C.; Shivaji, S.; Rainey, F.A.; Das, S.K. Thiomonas bhubaneswarensis sp. nov., an obligately mixotrophic, moderately thermophilic, thiosulfate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2171–2175. [CrossRef] [PubMed]
67. Chan, C.S.; Chan, K.-G.; Tay, Y.-L.; Chua, Y.-H.; Goh, K.M. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* **2015**, *6*, 177. [CrossRef]
68. Kvist, T.; Mengewein, A.; Manzei, S.; Ahring, B.K.; Westermann, P. Diversity of thermophilic and non-thermophilic crenarchaeota at 80 Â°C. *FEMS Microbiol. Lett.* **2005**, *244*, 61–68. [CrossRef]

69. Probst, A.J.; Weinmaier, T.; Raymann, K.; Perras, A.; Emerson, J.B.; Rattei, T.; Wanner, G.; Klingl, A.; Berg, I.A.; Yoshinaga, M.; et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat. Commun.* **2014**, *5*, 5497. [CrossRef]
70. Baker, B.J.; De Anda, V.; Seitz, K.W.; Dombrowski, N.; Santoro, A.E.; Lloyd, K.G. Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* **2020**, *5*, 887–900. [CrossRef]
71. Mehetre, G.T.; Paranjpe, A.S.; Dastager, S.G.; Dharne, M.S. Complete metagenome sequencing based bacterial diversity and functional insights from basaltic hot spring of Unkeshwar, Maharashtra, India. *Genom. Data* **2016**, *7*, 140–143. [CrossRef]
72. Stewart, L.C.; Stucker, V.K.; Stott, M.B.; de Ronde, C.E.J. Marine-influenced microbial communities inhabit terrestrial hot springs on a remote island volcano. *Extremophiles* **2018**, *22*, 687–698. [CrossRef] [PubMed]
73. Ward, L.M.; Idei, A.; Nakagawa, M.; Ueno, Y.; Fischer, W.W.; McGlynn, S.E. Geochemical and metagenomic characterization of jinata onsen, a proterozoic-analog hot spring, reveals novel microbial diversity including iron-tolerant phototrophs and thermophilic lithotrophs. *Microbes Environ.* **2019**, *34*, 278–292. [CrossRef] [PubMed]
74. Sharp, C.E.; Brady, A.L.; Sharp, G.H.; Grasby, S.E.; Stott, M.B.; Dunfield, P.F. Humboldt's spa: Microbial diversity is controlled by temperature in geothermal environments. *ISME J.* **2014**, *8*, 1166–1174. [CrossRef] [PubMed]
75. Thiel, V.; Hügler, M.; Ward, D.M.; Bryant, D.A. The dark side of the mushroom spring microbial mat: Life in the Shadow of Chlorophototrophs. II. Metabolic Functions of abundant community members predicted from metagenomic analyses. *Front. Microbiol.* **2017**, *8*, 943. [CrossRef] [PubMed]
76. Nunoura, T.; Chikaraishi, Y.; Izaki, R.; Suwa, T.; Sato, T.; Harada, T.; Mori, K.; Kato, Y.; Miyazaki, M.; Shimamura, S.; et al. A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* **2018**, *359*, 559–563. [CrossRef]
77. Gai, C.S.; Lu, J.; Brigham, C.J.; Bernardi, A.C.; Sinskey, A.J. Insights into bacterial CO_2 metabolism revealed by the characterization of four carbonic anhydrases in Ralstonia eutropha H16. *AMB Express* **2014**, *4*, 2. [CrossRef]
78. Bryan, C.G.; Davis-Belmar, C.S.; van Wyk, N.; Fraser, M.K.; Dew, D.; Rautenbach, G.F.; Harrison, S.T.L. The effect of CO_2 availability on the growth, iron oxidation and CO_2-fixation rates of pure cultures of Leptospirillum ferriphilum and Acidithiobacillus ferrooxidans. *Biotechnol. Bioeng.* **2012**, *109*, 1693–1703. [CrossRef]
79. Schenk, A.; Aragno, M. Bacillus schlegelii, a New species of thermophilic, facultatively chemolithoautotrophic bacterium oxidizing molecular hydrogen. *J. Gen. Microbiol.* **1979**, *115*, 333–341. [CrossRef]
80. Hamana, K.; Matsuzaki, S. Polyamines of carbon monoxide-utilizing bacteria, Pseudomonas thermocarboxydovorans and Pseudomonas carboxydohydrogena. *FEMS Microbiol. Lett.* **1990**, *70*, 353–356. [CrossRef]
81. Stipanuk, M.H.; Ueki, I. Dealing with methionine/homocysteine sulfur: Cysteine metabolism to taurine and inorganic sulfur. *J. Inherit. Metab. Dis.* **2011**, *34*, 17–32. [CrossRef]
82. Ruby, E.G.; Jannasch, H.W. Chemolithotrophic sulfur-oxidizing bacteria from the galapagos rift hydrothermal ventst. *Appl. Environ. Microbiol.* **1981**, *42*, 8. [CrossRef] [PubMed]
83. D'Auria, G.; Artacho, A.; Rojas, R.A.; Bautista, J.S.; Méndez, R.; Gamboa, M.T.; Gamboa, J.R.; Gómez-Cruz, R. Metagenomics of bacterial diversity in villa luz caves with sulfur water springs. *Genes* **2018**, *9*, 55. [CrossRef] [PubMed]
84. Denome, S.A.; Oldfield, C.; Nash, L.J.; Young, K.D. Characterization of the desulfurization genes from Rhodococcus sp. strain IGTS8. *J. Bacteriol.* **1994**, *176*, 6707–6716. [CrossRef] [PubMed]
85. Tian, H.; Gao, P.; Chen, Z.; Li, Y.; Li, Y.; Wang, Y.; Zhou, J.; Li, G.; Ma, T. Compositions and abundances of sulfate-reducing and sulfur-oxidizing microorganisms in water-flooded petroleum reservoirs with different temperatures in China. *Front. Microbiol.* **2017**, *8*, 143. [CrossRef]
86. Yamamoto, M.; Takai, K. Sulfur metabolisms in epsilon- and gamma-proteobacteria in deep-sea hydrothermal fields. *Front. Microbiol.* **2011**, *2*, 192. [CrossRef]
87. Fishbain, S.; Dillon, J.G.; Gough, H.L.; Stahl, D.A. Linkage of high rates of sulfate reduction in Yellowstone Hot Springs to unique sequence types in the dissimilatory sulfate respiration pathway. *Appl. Environ. Microbiol.* **2003**, *69*, 3663–3667. [CrossRef] [PubMed]

88. Dillon, J.G.; Fishbain, S.; Miller, S.R.; Bebout, B.M.; Habicht, K.S.; Webb, S.M.; Stahl, D.A. High rates of sulfate reduction in a Low-Sulfate Hot Spring Microbial Mat are driven by a low level of diversity of sulfate-respiring microorganisms. *Appl. Environ. Microbiol.* **2007**, *73*, 5218–5226. [CrossRef]
89. Nishihara, A.; Haruta, S.; McGlynn, S.E.; Thiel, V.; Matsuura, K. Nitrogen fixation in thermophilic chemosynthetic microbial communities depending on hydrogen, sulfate, and carbon dioxide. *Microbes Environ.* **2018**, *33*, 10–18. [CrossRef]
90. Ranawat, P.; Rawat, S. Stress response physiology of thermophiles. *Arch. Microbiol.* **2017**, *199*, 391–414. [CrossRef]
91. Dávila-Ramos, S.; Castelán-Sánchez, H.G.; Martínez-Ávila, L.; Sánchez-Carbente, M.d.R.; Peralta, R.; Hernández-Mendoza, A.; Dobson, A.D.W.; Gonzalez, R.A.; Pastor, N.; Batista-García, R.A. A Review on viral metagenomics in extreme environments. *Front. Microbiol.* **2019**, *10*, 2403. [CrossRef]
92. Snyder, J.C.; Stedman, K.; Rice, G.; Wiedenheft, B.; Spuhler, J.; Young, M.J. Viruses of hyperthermophilic Archaea. *Res. Microbiol.* **2003**, *154*, 474–482. [CrossRef]
93. Pina, M.; Bize, A.; Forterre, P.; Prangishvili, D. The archeoviruses. *FEMS Microbiol. Rev.* **2011**, *35*, 1035–1054. [CrossRef] [PubMed]
94. Barylski, J.; Kropinski, A.M.; Alikhan, N.-F.; Adriaenssens, E.M. ICTV report consortium ICTV virus taxonomy profile: Herelleviridae. *J. Gen. Virol.* **2020**, *101*, 362–363. [CrossRef] [PubMed]
95. Mangold, S.; Valdés, J.; Holmes, D.; Dopson, M. Sulfur metabolism in the extreme acidophile acidithiobacillus caldus. *Front. Microbiol.* **2011**, *2*, 17. [CrossRef] [PubMed]
96. Castelán-Sánchez, H.G.; Lopéz-Rosas, I.; García-Suastegui, W.A.; Peralta, R.; Dobson, A.D.W.; Batista-García, R.A.; Dávila-Ramos, S. Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. *Mar. Genom.* **2019**, *46*, 16–28. [CrossRef]
97. Crummett, L.T.; Puxty, R.J.; Weihe, C.; Marston, M.F.; Martiny, J.B.H. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **2016**, *499*, 219–229. [CrossRef] [PubMed]
98. He, T.; Li, H.; Zhang, X. Deep-Sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. *mBio* **2017**, *8*, 00893-17. [CrossRef]
99. Puxty, R.J.; Evans, D.J.; Millard, A.D.; Scanlan, D.J. Energy limitation of cyanophage development: Implications for marine carbon cycling. *ISME J.* **2018**, *12*, 1273–1286. [CrossRef]
100. Roux, S.; Brum, J.R.; Dutilh, B.E.; Sunagawa, S.; Duhaime, M.B.; Loy, A.; Poulos, B.T.; Solonenko, N.; Lara, E.; Poulain, J.; et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **2016**, *537*, 689–693. [CrossRef]
101. White, R.H. Distribution of folates and modified folates in extremely thermophilic bacteria. *J. Bacteriol.* **1991**, *173*, 1987–1991. [CrossRef]
102. Hutinet, G.; Kot, W.; Cui, L.; Hillebrand, R.; Balamkundu, S.; Gnanakalai, S.; Neelakandan, R.; Carstens, A.B.; Fa Lui, C.; Tremblay, D.; et al. 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nat. Commun.* **2019**, *10*, 5442. [CrossRef] [PubMed]
103. Pannekens, M.; Kroll, L.; Müller, H.; Mbow, F.T.; Meckenstock, R.U. Oil reservoirs, an exceptional habitat for microorganisms. *New Biotechnol.* **2019**, *49*, 1–9. [CrossRef] [PubMed]
104. Anantharaman, K.; Duhaime, M.B.; Breier, J.A.; Wendt, K.A.; Toner, B.M.; Dick, G.J. Sulfur oxidation genes in diverse deep-sea viruses. *Science* **2014**, *344*, 757–760. [CrossRef] [PubMed]