



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

**Análisis de la diversidad viral presente en muestras de agua y sedimento del
Golfo de México.**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Maestra en Ciencias

PRESENTA:

Raquel Neri Barranco

TUTOR PRINCIPAL

Dr. Lorenzo Segovia Forcella,
Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR

Dra. Blanca Taboada Ramírez,
Instituto de Biotecnología, UNAM

Dr. Ismael Hernández Lucas,
Instituto de Biotecnología, UNAM

Ciudad de México, Junio, 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Few of us may ever live on the sea or under it, but all of us are making increasing use of it either as a source of food and other materials, or as a dump. As our demands upon the ocean increase, so does our need to understand the ocean as an ecosystem. Basic to the understanding of any ecosystem is knowledge of its food web, through which energy and materials flow.

Pomeroy, 1974

Agradecimientos

Agradezco a la UNAM y al posgrado en Ciencias Bioquímicas, por haberme permitido seguir siendo estudiante de esta casa máxima de estudios. Asimismo, gracias a la beca de posgrado del CONACyT (CVU 927513) por haberme apoyado económicamente estos cuatro semestres, sin ese apoyo este trabajo no hubiera sido posible. Gracias al Instituto de Biotecnología (IBt) de la UNAM por haberme dejado ser parte de su comunidad y por brindarme todos los recursos que necesitaba para realizar este proyecto.

Gracias al Dr. Lorenzo Segovia por haberme permitido hacer esta tesis de maestría bajo su dirección y por haberme brindado tanta libertad para realizarla. Lorenzo, en estos casi cinco años trabajando contigo aprendí muchas cosas, pero sobre todo me llevo tu actitud pragmática ante la vida y la apertura mental que te caracterizan. Gracias a lxs miembrxs de mi comité tutorial, Dra. Blanca Taboada Ramírez y Dr. Ismael Hernández Lucas, por haberme brindado valiosas opiniones y recomendaciones para mi proyecto. Gracias Dra. Blanca, por su tiempo y disponibilidad en todo momento para apoyarme, asesorarme y ayudarme con todos los percances bioinformáticos a los que me fui enfrentando, sin su participación, este proyecto no hubiera sido posible. Gracias al nodo FOS, al cluster Teopanzolco y al M. en C. Jerome Veryelen por haberme permitido realizar todos los análisis computacionales que necesité. Gracias a lxs miembrxs de mi jurado: Dr. Adrián Ochoa Leyva, Dra. Ernestina Godoy Lozano, Dra. Liliana Pardo López, Dr. Pavel Isa y Dra. Sonia Dávila Ramos, por su tiempo y dedicación para leer mi tesis, sus aportaciones fueron muy valiosas y sin duda, construyeron una mejor versión de mi tesis. También agradezco a la M. en C. Blanca Ramos Cerrillo por su apoyo en la elaboración de este proyecto. Gracias a Antonio Bolaños Guillén, por su apoyo extraordinario y eficiente en los trámites administrativos que requerí a lo largo de la maestría.

Gracias a todos los miembrxs, ex miembrxs del laboratorio 12, compañerxs y amigxs de la maestría y del IBt que hicieron de esta una experiencia más amena. Gracias a mi familia Rose, Ani, Nachi y Papá, por haberme apoyado en todo momento en cualquier decisión que tomara, sin duda, soy muy afortunada de existir en esta familia. Gracias Miguel por todo lo que aprendimos juntos estos casi tres años de posgrado y porque la maestría hubiera sido muy diferente sin tu compañía. Gracias Alf, Mishel, JuanJo, Raúl y Stella por todos esos malos y buenos momentos juntos.

Esta tesis no ganó ningún bono de titulación y probablemente no ganará ningún premio a la mejor tesis o mención honorífica, pero ninguno de esos reconocimientos podrá reflejar lo mucho que me costó aprender a realizar este proyecto, darle sentido biológico y llegar hasta este punto. Por siempre estaré agradecida conmigo misma por no haber tirado la toalla y haber aguantado hasta el final.

CONTENIDO

1. RESUMEN	6
2. ABSTRACT	7
3. INTRODUCCIÓN	8
3.1 Importancia de los virus en ecosistemas marinos	8
3.2 Metagenómica	15
3.3 Golfo de México	17
3.4 Virus en zonas con presencia de hidrocarburos	18
4. JUSTIFICACIÓN	21
5. OBJETIVOS	21
6. MÉTODO	21
6.1 Muestras	22
6.1.1 Muestras originales del CIGoM	22
6.1.2 Muestras de agua tomadas durante el derrame del DWH (Mason <i>et al.</i> , 2012)	23
6.1.3 Muestras de sedimento tomadas durante el derrame del DWH (Mason <i>et al.</i> , 2014)	23
6.1.4 Muestras de emanaciones naturales de hidrocarburos en el noreste del GoM (Dong <i>et al.</i> , 2019)	25
6.1.5 Muestras de emanaciones naturales de hidrocarburos en el noroeste del GoM (Zhao <i>et al.</i> , 2020)	25
6.2 Procesamiento bioinformático	26
6.2.1 Calidad de las secuencias	27
6.2.2 Caracterización taxonómica total	30
6.2.3 Caracterización taxonómica viral	30
6.2.4 Asignación taxonómica	33
6.2.5 Análisis de diversidad y metagenómica comparativa	33
7 RESULTADOS Y DISCUSIÓN	35
7.2 Control de calidad	35
7.3 Ensamblajes y predicción de ORF's	36
7.4 Composición taxonómica total	38
7.5 Composición taxonómica viral	42

7.6	Análisis de diversidad y ecología comparativa	50
7.6.1	Riqueza, abundancia y alfa diversidad de especies virales	51
7.6.2	Análisis de diversidad según el tipo de sustrato	59
7.6.3	Beta diversidad	62
7.6.4	Análisis de diversidad relacionado con la presencia o ausencia de hidrocarburos	65
8	CONCLUSIONES	73
	Perspectivas	74
	Referencias	74
	ANEXO	82

1. RESUMEN

Los virus son las entidades biológicas más abundantes y más ampliamente distribuidas en todos los ecosistemas. En los últimos años se ha ampliado el estudio de virus en ambientes marinos y se ha visto que su papel ecológico es muy importante. Al lizar a sus hospederos provocan la liberación de nutrientes y los hacen disponibles para otras formas de vida, así contribuyen al bienestar integral de estos ecosistemas. Sin embargo, la gran limitante en el estudio de los virus es que la mayoría no son cultivables en laboratorio y la metagenómica representa una alternativa a esta problemática. El Golfo de México (GoM) es un área que naturalmente emana hidrocarburos y que también ha sido perturbada por derrames petroleros, por lo cual resulta un ecosistema marino interesante para el estudio de los microorganismos que lo habitan y viven bajo estas condiciones. No obstante, existen muy pocos estudios sobre virus en el GoM. En este trabajo, se utilizaron muestras metagenómicas del GoM previamente secuenciadas, provenientes de agua y de sedimento, a partir de las cuales se extrajo la diversidad viral. En términos de diversidad, las muestras se agrupan según el tipo de sustrato del que provengan (agua o sedimento), siendo las muestras de sedimento más diversas que las de agua. Asimismo, se encontró que para los estudios donde se tenían muestras con presencia de hidrocarburos y muestras control (ausencia de hidrocarburos) había diferencias significativas en diversidad viral entre los dos grupos de muestras. Dentro de las muestras con hidrocarburos hay enriquecimiento de especies virales que previamente han sido reportadas en zonas con hidrocarburos o provenientes de ambientes extremos. Lo anterior, sugiere que las comunidades virales al igual que las de procariontes, son un reflejo de las condiciones de un ecosistema dado, en este caso, ecosistemas perturbados con hidrocarburos.

2. ABSTRACT

Viruses are the most abundant and ubiquitous biological entity across all ecosystems. In recent years the study of marine viruses has set off and it has made it clear the relevance of their ecological role. By lysing their hosts and releasing organic compounds back into the food chain, making them available for other lifeforms, viruses are a big player in the wellbeing and health of these ecosystems. Nevertheless, the biggest constrain in viral research is that most are not culturable in laboratory conditions because their hosts are also unculturable. Metagenomics helps to solve this problem. The Gulf of Mexico (GoM), is a geographic area where hydrocarbons seeps occur naturally and because of that it has also been subject to multiple human-caused oil spills. These conditions make the GoM an interesting marine ecosystem for the study of the microorganisms that thrive under those conditions. But there are very few viral analyses in the GoM. In this study, viral diversity was obtained from water and sediment samples across the GoM. Samples clustered according to their origin (water or sediment), with sediment samples being more diverse than water samples. In addition, it was found that for the studies where there were samples with hydrocarbons and control samples without hydrocarbons, there were significant differences in terms of viral diversity related to the presence or absence of hydrocarbons. Such that in the samples with hydrocarbons, there is an enrichment of viral species that were already seen in areas with hydrocarbons or from extreme environments. So, viral communities as well as prokaryotes, are a reflection of the environmental conditions in a given ecosystem, in this case, ecosystems with high levels of hydrocarbons.

3. INTRODUCCIÓN

3.1 Importancia de los virus en ecosistemas marinos

Durante la década de los 80's se creía que la concentración de virus en aguas ambientales no contaminadas era baja, por lo cual se consideraba que no eran importantes en términos ecológicos. Sin embargo, unos años después con nuevas técnicas de conteo de partículas virales por medio de microscopía electrónica, se descubrió que en realidad la abundancia de virus de muestras de aguas ambientales era mayor que la concentración de bacterias (Bergh *et al.*, 1989). Desde entonces, se intuyó que el papel ecológico de los virus en aguas ambientales era más importante de lo que se creía y se desarrolló una nueva área de investigación en biología, la virología marina (Breitbart, 2012).

Hoy en día se sabe que los virus son la entidad biológica más abundante de los océanos y también representan el reservorio de la mayor parte de la diversidad genética de los ecosistemas marinos (Suttle, 2007). En promedio hay 10^7 partículas virales por mililitro en la superficie oceánica, es decir, un orden de magnitud más abundante que los procariontes (Wommack & Colwell, 2000; Breitbart, 2012). Sin embargo, debido al tamaño pequeño de los virus, en términos de biomasa son el segundo grupo más grande en los océanos, superados por los procariontes (Fig. 1) (Suttle, 2005).

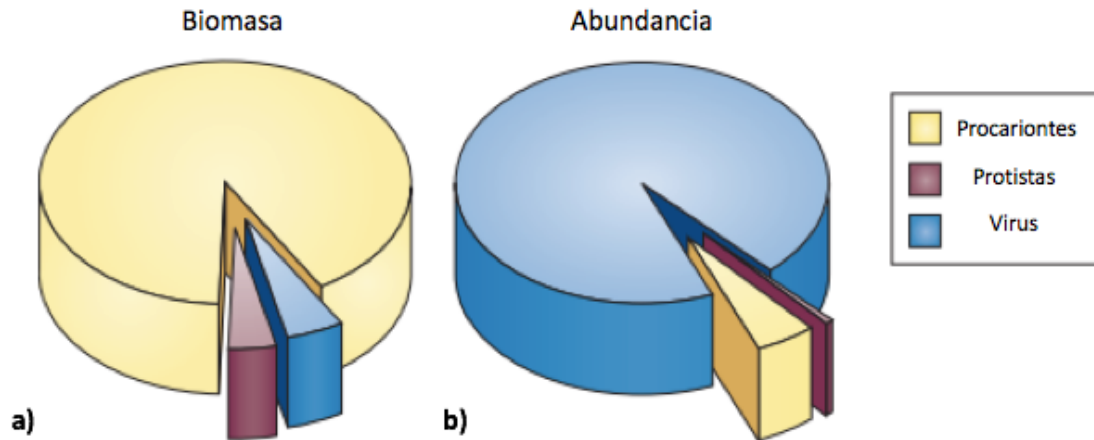


Figura 1. Representación gráfica de la proporción en términos de biomasa y abundancia de los tres grupos biológicos principales en los océanos: procariontes, protistas y virus. En cuanto a biomasa, el grupo más representado son los procariontes (a), sin embargo, en abundancia, los virus representan la mayor parte de los tres grupos microbiológicos (b) (modificada de Suttle, 2007).

La mayor parte de los virus marinos son bacteriófagos (virus que infectan bacterias y arqueas), también conocidos como fagos, lo cual tiene sentido debido a que las bacterias son el segundo grupo más abundante en los océanos después de los virus (Angly, 2004; Weinbauer, 2004; Breitbart, 2012). Los virus que más se reportan en estudios ambientales pertenecen a fagos del orden *Caudovirales*, distribuidos principalmente en tres familias de fagos de DNA de doble cadena (dsDNA) con cola: *Myoviridae*, *Siphoviridae* y *Podoviridae* (Fig. 2) (Anderson *et al*, 2011; Corinaldesi *et al*, 2012; Taboada *et al*, 2018).

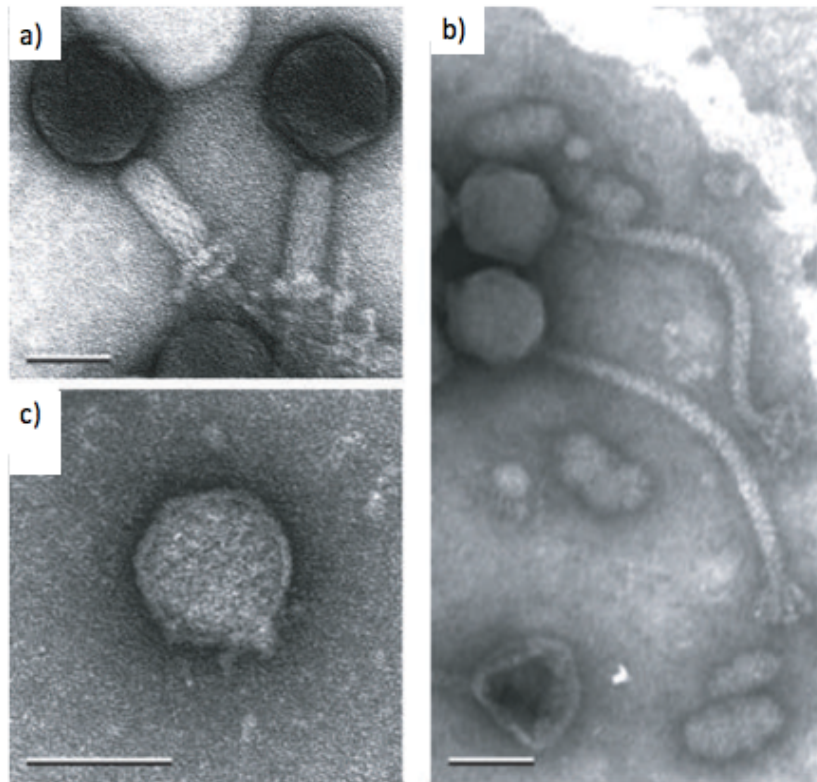


Figura 2. Las tres familias de virus más encontradas en muestras ambientales. Estas son familias de fagos pertenecientes al orden de los caudovirales: a) *Myovirus*, b) *Siphovirus* y c) *Podovirus*. Los tres son fagos de DNA de doble cadena y dos de ellos (*Myovirus* y *Siphovirus*) tienen la estructura de cola (modificada de Suttle, 2005).

Se ha visto que los fagos marinos, al igual que en ecosistemas terrestres, pueden seguir dos tipos de ciclos de vida: el ciclo lítico y el ciclo lisogénico, los cuales funcionan de la siguiente manera. Existen partículas virales libres en el ambiente que encuentran a un hospedero y lo infectan, se anclan e inyectan su material genómico y de aquí se puede seguir una de las dos vías posibles. En el ciclo lítico, el fago se replica utilizando la maquinaria del hospedero y produce muchas partículas virales nuevas que provocarán la lisis, liberándolas en el ambiente. Mientras que, en el ciclo lisogénico, el genoma del virus se inserta dentro del genoma del hospedero (profago) o existe como plásmido dentro del hospedero y permanece ahí replicándose junto con el hospedero hasta que, por algún factor, ambiental por ejemplo, el virus se induce, es decir, sale del ciclo lisogénico y entra

al ciclo lítico (Fig. 3) (Ofir & Zorek, 2018). En ambientes marinos, se ha encontrado que ambas estrategias de vida son comunes (Suttle, 2007).

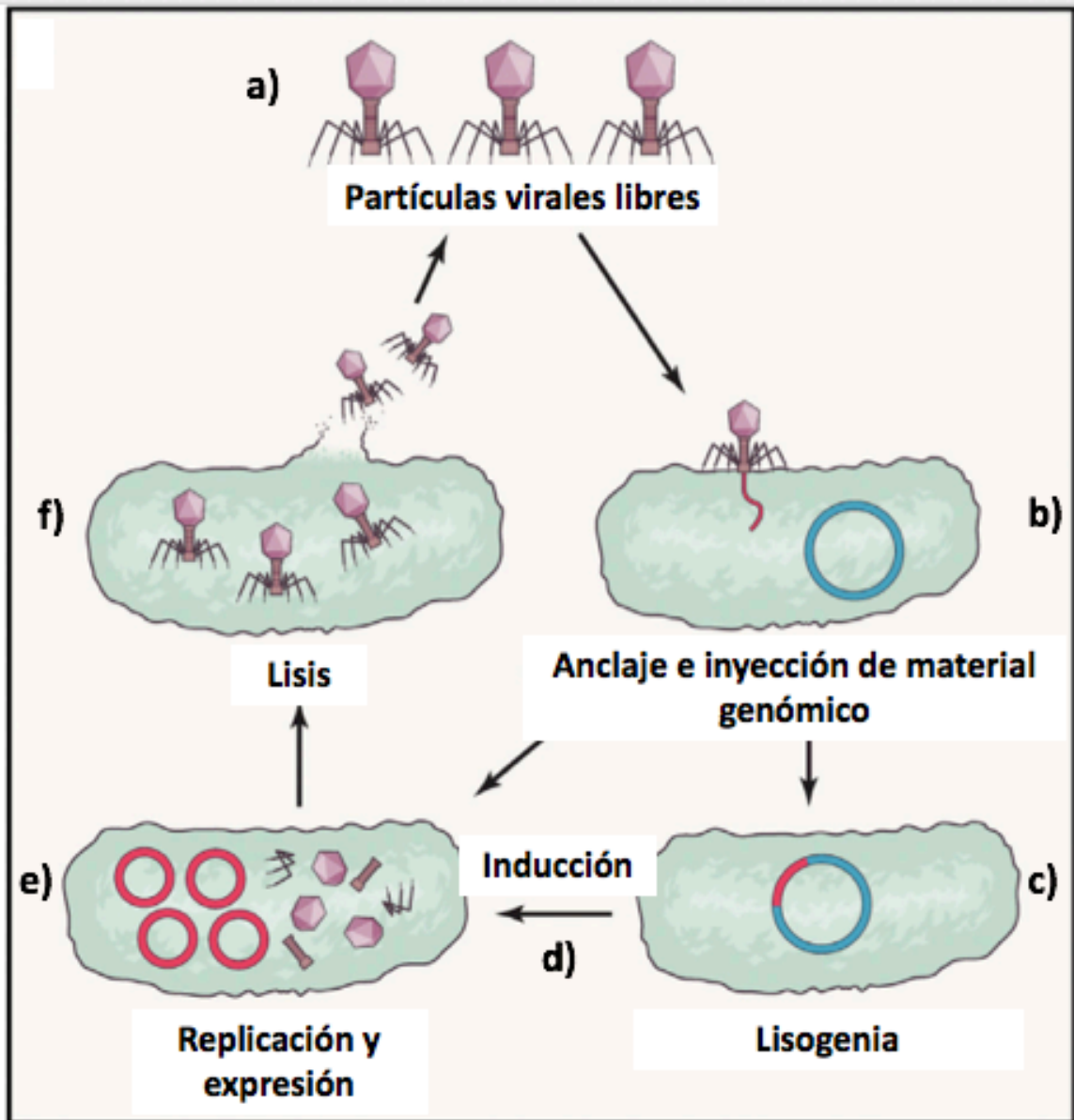


Figura 3. Pasos de la infección viral mediante el ciclo lítico y lisogénico. Al inicio existen partículas virales libres (a), estas encuentran un hospedero, se anclan e inyectan su material genómico (b). A partir de este paso el ciclo puede seguir dos vías: en el ciclo lítico, el virus se replica (e) hasta provocar la lisis del hospedero, liberando muchas partículas nuevas (f). En el ciclo lisogénico, el genoma del virus se inserta al genoma del hospedero (c) hasta que por algún factor se induce (d) la entrada al ciclo lítico y sigue los pasos previamente descritos (modificada de Ofir & Zorek, 2018).

Además de los fagos, la mayor parte de los virus marinos tienen como hospederos a otros microorganismos, tales como arqueas o microeucariontes, y debido a la gran abundancia que los virus tienen en los océanos, ejercen un papel importante en la diversidad y en la dinámica poblacional de las comunidades de sus hospederos (Weitz & Wilhelm, 2012).

En cuanto a la dinámica poblacional, es a través de la lisis viral de hospederos específicos que se crean nichos disponibles en los que otras especies menos competitivas pueden crecer. Este fenómeno se basa en el modelo de Lotka-Volterra y ha sido descrito con el modelo ecológico "*Kill the winner*" (KtW), para el caso específico de fagos y bacterias. Bajo este modelo, las comunidades bacterianas están formadas por dos grupos principales que se encuentran compitiendo por los mismos recursos: especialistas en competencia (estrategas-r) y especialistas en defensa (estrategas-k). En ausencia de depredadores, los estrategas-r dominarían, acaparando todos los recursos. Sin embargo, en la dinámica de KtW, los estrategas-r, son depredados por fagos, controlando sus poblaciones y permitiendo que los estrategas-k puedan recibir recursos y aumentar su crecimiento (Breitbart, 2012). También al lisar hospederos microeucariontes que se alimentan de bacterias, liberan la presión de depredación permitiendo que otros subgrupos de bacterias florezcan (Suttle, 2007).

Por otro lado, los virus influyen en la diversidad de sus hospederos al transducir genes metabólicos o morfológicos entre hospederos. Por ejemplo, por medio de ciclos lisogénicos, los fagos pueden introducir parte de su material genómico dentro de genomas bacterianos, transfiriendo genes de importancia ecológica como inmunidad a la invasión por otros fagos o genes fotosintéticos (Weinbauer & Rassoulzadegan, 2004; Suttle, 2007).

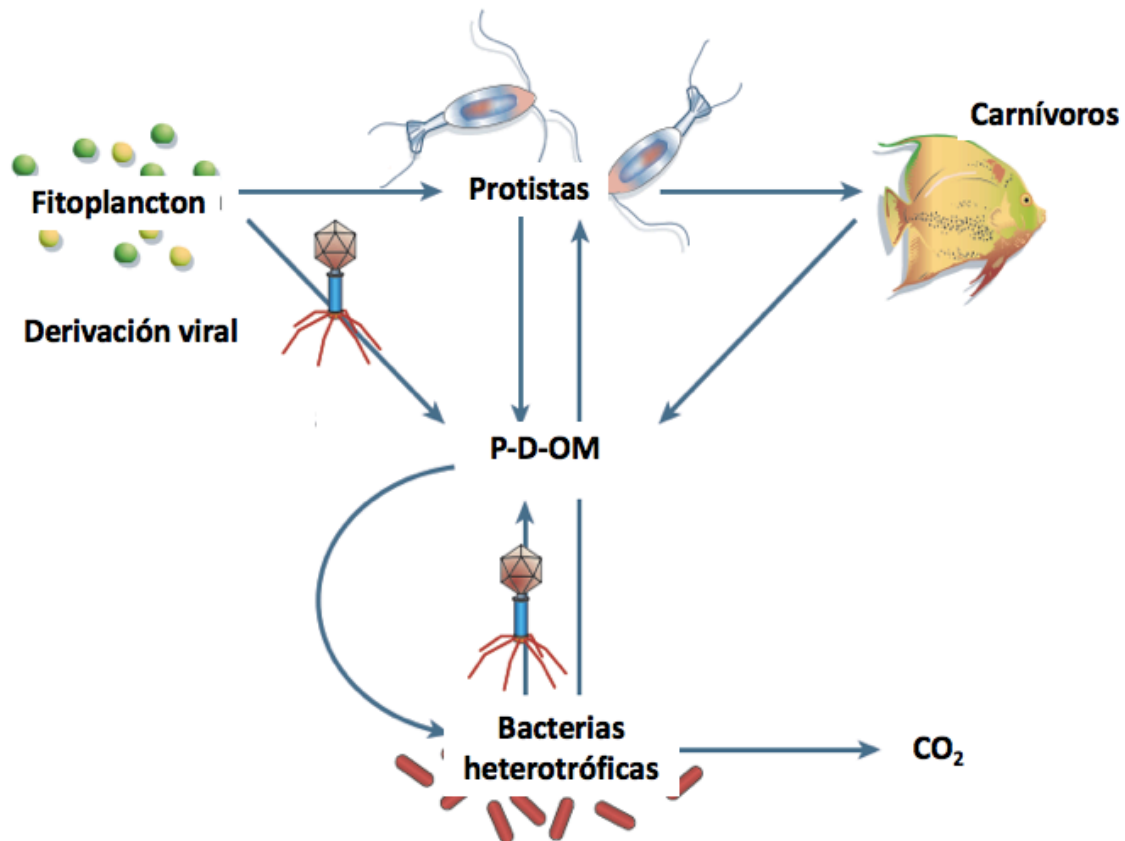


Figura 4. Diagrama del funcionamiento de la derivación viral. Por medio de la lisis viral, los virus catalizan el movimiento de nutrientes desde sus hospederos (bacterias heterotróficas y fitoplancton) hacia los reservorios de DOM (*Dissolved Organic Matter*, DOM, tamaño entre 0.22-0.7 μm) y POM (*Particulate Organic Matter*, POM, tamaño entre 0.053-2 mm) en el océano. Esto evita que los nutrientes lleguen directamente a niveles tróficos más altos y sean aprovechados por otras comunidades de bacterias (modificada de Suttle, 2005).

Mediante la lisis viral masiva, los virus también tienen un papel de suma importancia en los ciclos biogeoquímicos globales (Fig. 4). La lisis viral media la transferencia de carbono orgánico y otros nutrientes provenientes de heterótrofos y fotoautótrofos al reservorio de materia orgánica disuelta (*Dissolved Organic Matter*, DOM, tamaño entre 0.22-0.7 μm) y de materia orgánica particulada (*Particulate Organic Matter*, POM, tamaño entre 0.053-2 mm) que existe en el océano. Este movimiento de materia es conocido como la derivación viral (*viral shunt*); se le conoce de esa forma debido a que la lisis viral evita que el carbono y los demás nutrientes del fitoplancton

y de bacterias lisadas llegue directamente a niveles tróficos más altos (peces, otros protistas) y en cambio, quede biodisponible para el consumo de otras bacterias heterotróficas, estimulando su crecimiento. Se estima que al día se liberan cerca de 10^9 toneladas de carbono por medio de la derivación viral, por lo cual este proceso funciona como un catalizador del reciclaje de microorganismos, estimulando el ciclo de nutrientes y carbono (Wilhelm & Suttle, 1999; Suttle, 2005; Suttle, 2007; Breitbart, 2012; Weitz & Wilhelm, 2012; Coutinho et al., 2017).

Aunque en menor abundancia, los virus marinos que infectan eucariontes también son importantes en el océano, infectando desde algas y camarones, hasta ballenas. Dentro de los virus que infectan protistas, se encuentran familias de virus gigantes con genomas de más de 10^6 de pares de bases (bp), como *Mimiviridae*, *Marseilleviridae* o *Pithoviridae* (Fig. 5) (Aherfi et al., 2016). De algunos, se conoce mucho, ya que infectan especies de interés comercial, tal como el virus del síndrome de mancha blanca (*white spot syndrome virus – WSSV*) que afecta al camarón peneido. Sin embargo, en general, se sabe poco de los virus marinos que infectan eucariontes que no son de interés comercial (Suttle, 2007).

En general, una de las mayores causas del poco conocimiento acerca de virus que infectan microorganismos se debe a que para poder estudiarlos es necesario poder cultivar al hospedero. Sin embargo, para muchos de ellos no se conoce el hospedero y además la mayoría de esos hospederos aún son incultivables en condiciones de laboratorio, por ejemplo, en el caso de los fagos, se estima que menos del 1% de las bacterias acuáticas que existen son cultivables en placas (Staley & Konopka, 1985). No obstante, este problema ha sido en gran parte resuelto gracias a la metagenómica, al no requerir del cultivo para poder estudiarlos.

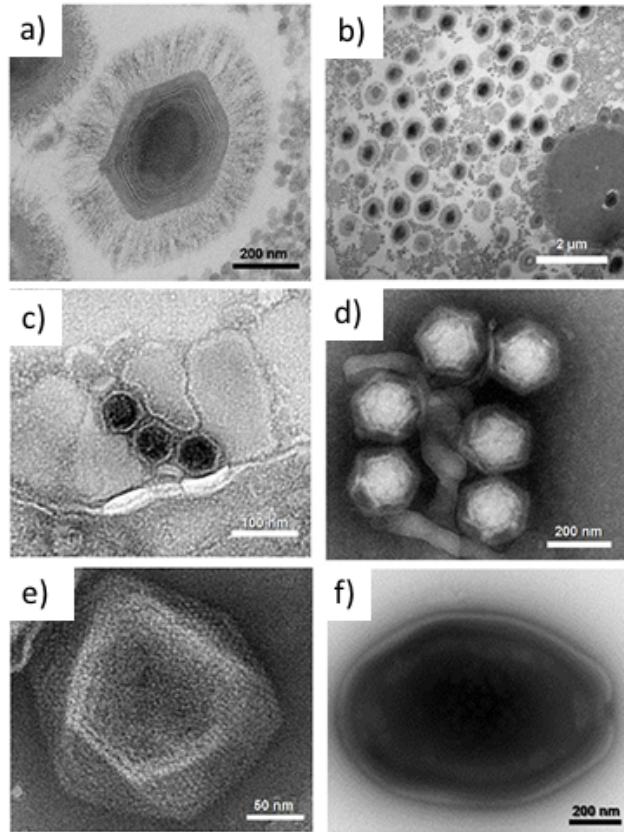


Figura 5. Micrografías electrónicas de virus gigantes. a) *Mimivirus*; b) fábrica viral de *Mimivirus* en *Acanthamoeba polyphaga*; c) *Marseillevirus*; d) Virofagos, i.e, virus que infectan otros virus; e) *Faustovirus*; f) *Pandoravirus* sp. (modificada de Aherfi *et al.*, 2016).

3.2 Metagenómica

Actualmente los análisis metagenómicos y las estrategias de secuenciación masiva (*Next generation sequencing*) han permitido cambiar la forma en la que la diversidad viral y microbiana es analizada. La metagenómica es el análisis genómico de microorganismos a través de extracción directa de DNA a partir de un conjunto de microorganismos aislados directamente de una muestra ambiental -sin necesidad de cultivarlos- y su posterior secuenciación y análisis de datos por medio de técnicas bioinformáticas (Fig. 6) (Handelsman, 2004).

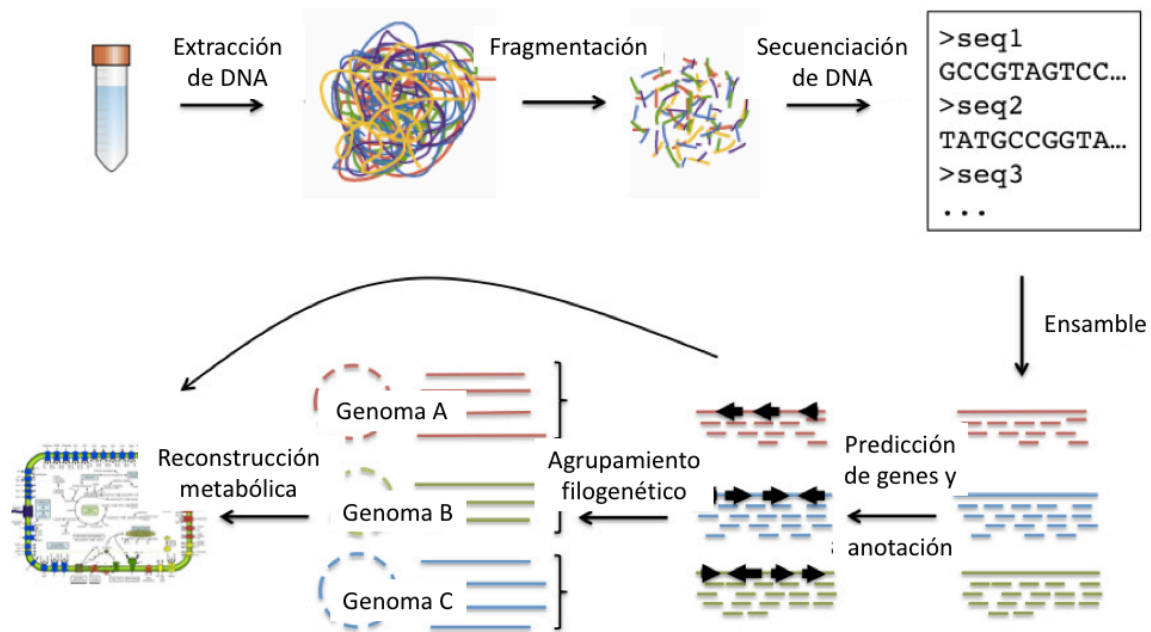


Figura 6. Descripción esquemática de los pasos involucrados en los estudios de metagenómica. En general, primero se lleva a cabo la extracción de DNA (o de RNA) de alguna muestra ambiental, después, ese material genómico se fragmenta y se secuencia. Los pasos posteriores a la secuenciación dependerán de lo que se quiera estudiar. En este ejemplo, las secuencias obtenidas se ensamblan, se hace predicción de genes y anotación de los mismos, luego se agrupan filogenéticamente para crear grupos de secuencias que potencialmente pertenecen al mismo genoma, para así poder hacer una reconstrucción metabólica de esa muestra ambiental (modificada de ENVGEN, s.f.).

De esta forma, se evita el problema del cultivo de microorganismos en condiciones de laboratorio y es posible obtener información sobre la composición microbiológica que existe en un ambiente en particular en un determinado momento. Es por lo anterior, que los estudios metagenómicos han sido ampliamente utilizados para identificar comunidades virales a partir de muestras marinas o de ambientes acuáticos. Ejemplos de lo anterior son los estudios hechos en Cuatro Ciénegas, Coahuila, México (Taboada *et al.*, 2018), en ventilas hidrotermales del fondo oceánico (Anderson *et al.*, 2011), con muestras de regiones oceánicas diferentes (Angly *et al.*, 2006), a partir de sedimento de mar profundo (Corinaldesi *et al.*, 2017) o con muestras alrededor del mundo (Gregory *et al.*, 2019). Todos estos estudios

hechos en diferentes ambientes han demostrado la diversidad y abundancia de virus en la mayoría de los hábitats.

3.3 Golfo de México

El Golfo de México (GoM) es un mar semicerrado y es uno de los reservorios más grandes de petróleo en el mundo. Estos reservorios se originaron en el jurásico tardío y el cretácico temprano, cuando debido a la formación geológica del GoM, tapetes de fitoplancton ricos en lípidos quedaron enterrados dentro de la Tierra, dando lugar a rocas ricas en gas y petróleo (Guzman-Vega & Mello, 1999). Por tal razón, el GoM es un ecosistema que constantemente está expuesto a un flujo de hidrocarburos continuo, ya sea a través de emanaciones naturales o debido a causas antropogénicas, como derrames ocasionados por el manejo de pozos petroleros (Macdonald *et al.*, 2015).

Las condiciones anteriores hacen del GoM un nicho particular, con comunidades microbianas que viven y florecen en condiciones de exposición continua a hidrocarburos. Sin embargo, en términos de diversidad microbiológica, las aguas mexicanas del GoM han sido poco estudiadas en comparación de la costa noreste, perteneciente a Estados Unidos, muchos de estos estudios desarrollados a partir del derrame de petróleo del *Deepwater Horizon* (DWH) (Raggi *et al.*, 2020). El derrame del DWH ocurrió en el 2010 en la costa de Luisiana, E.U.A., causado por la explosión de una plataforma de perforación a cargo de *British Petroleum*. Alrededor de 4.9 millones de barriles de petróleo se liberaron en el GoM y ha sido el derrame accidental en aguas marinas más grande de la historia de Estados Unidos (Liu *et al.*, 2011; Zukunft, 2010).

Sin embargo, hoy en día, no se sabe con certeza la magnitud de los daños y efectos causados por dicho derrame en las aguas mexicanas del GoM, debido a la falta de estudios previos que caracterizaran las condiciones en las que se encontraban antes del derrame. Es por lo que se formó el Consorcio de Investigación de Golfo

de México en el 2015 (CIGoM), el cual surgió como una iniciativa académica y pública, por parte de científicos de diferentes instituciones nacionales y personal de Petróleos Mexicanos (PEMEX), para poder entender y conocer mejor las regiones del GoM pertenecientes a México y así tener una línea base caracterizada, en caso de que otro derrame de tal magnitud volviera a suceder (CIGoM, s.f.).

Uno de los objetivos del CIGoM fue estudiar por medio de metagenómica, la diversidad taxonómica y funcional de las comunidades bacterianas de muestras de agua y sedimento de dos regiones del GoM, la Costa de Campeche y la zona de Perdido en la costa de Tamaulipas, (Godoy-Lozano *et al.*, 2018; Raggi *et al.*, 2020). Sin embargo, a pesar del papel tan importante que los virus tienen en los ecosistemas marinos, aún no han sido estudiados en estas muestras en específico. Aún menos se ha estudiado el papel de los virus bajo el contexto de muestras provenientes de lugares con aporte de hidrocarburos altos, ya sea por vías naturales o antropogénicas como lo es el GoM.

Además, el hecho de contar con muestras de sedimento marino enriquece este estudio, ya que se ha reportado que la riqueza viral parece ser mucho mayor en los sedimentos que en la columna de agua adyacente, por lo cual se ha sugerido que los ecosistemas bénticos pueden representar una de los más grandes reservorios de diversidad viral en el planeta (Corinaldesi *et al.*, 2017). Asimismo, la ecología viral en ambientes sedimentarios ha sido poco estudiada, a pesar de que los sedimentos de suelo marino cubren casi dos tercios de la superficie de la Tierra y funcionan como una interfaz altamente vital y dinámica en los ciclos biogeoquímicos de los océanos (Gage & Tyler, 1991).

3.4 Virus en zonas con presencia de hidrocarburos

Dentro de las investigaciones que se han enfocado en estudiar virus en zonas con presencia de hidrocarburos, ya sea por causas naturales o antropogénicas, se ha encontrado lo siguiente. En Head *et al.*, 2006, se encontró que la mayoría de los

estudios de biorremediación se centran en estudiar a los microorganismos que directamente degradan los contaminantes, sin embargo, estos son parte de una red ecológica sin la cual la degradación no podría ocurrir y es aquí donde la lisis viral juega un papel importante, ya que promueven el reciclaje de nutrientes que son indispensables para incrementar la degradación, tales como el nitrógeno y el fósforo (Fig. 7).

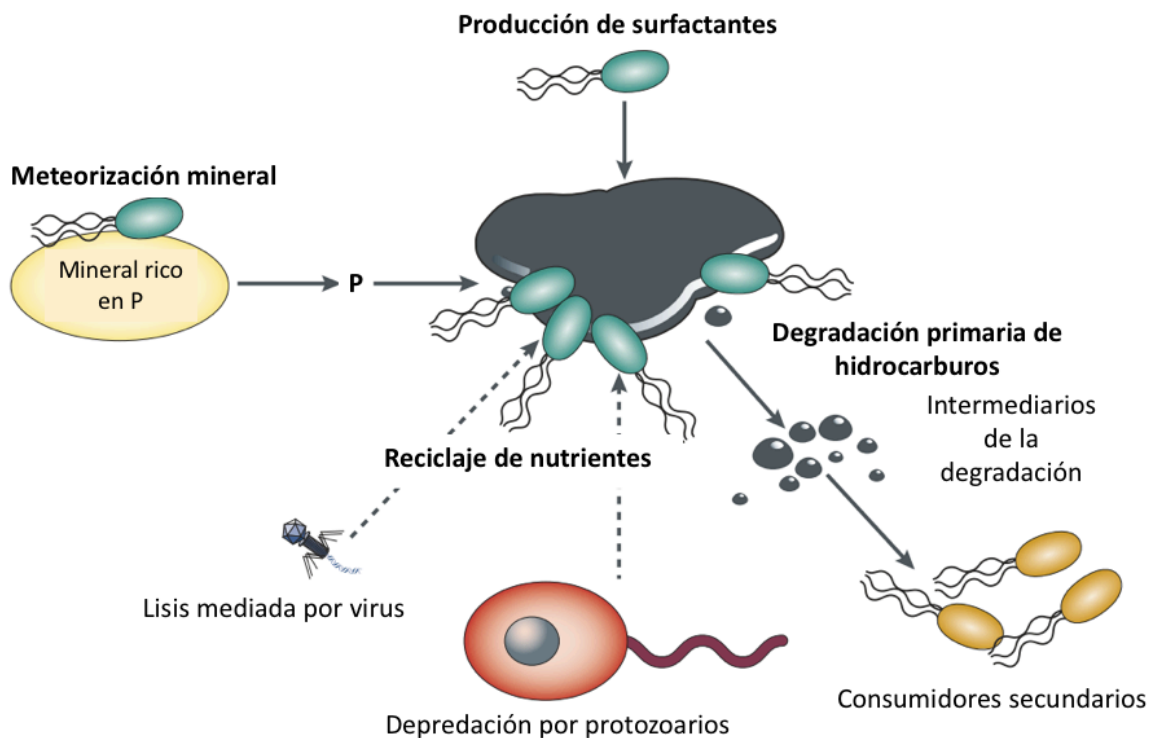


Figura 7. Red ecológica microbiana de degradación de hidrocarburos. En la degradación de petróleo, existen más componentes biológicos que solo los microorganismos que directamente degradan el petróleo. En verde están esquematizadas las bacterias consumidoras primarias del petróleo, estas pueden producir surfactantes que emulsifican el petróleo en agua y también generan compuestos intermediarios que pueden ser ingeridos por consumidores secundarios. Por otro lado, la lisis mediada por virus y la de depredación por protozoarios de las bacterias involucradas, permite el reciclaje de nutrientes y así como la meteorización mineral llevada a cabo por otras bacterias, aseguran el abasto de elementos necesarios para la degradación del petróleo, tales como fósforo y nitrógeno (modificada de Head *et al.*, 2006).

En ese mismo contexto, se probó de forma experimental que la presencia de fagos aceleraba la conversión del carbono orgánico total (COT) a dióxido de carbono, en un biorreactor de biorremediación de aguas contaminadas provenientes de procesos de extracción de petróleo celular (Rosenberg *et al.*, 2010). El 90% del COT se mineralizó en dióxido de carbono, lo cual era mucho más de lo esperado, ya que las bacterias generalmente convierten del 40-60% del carbono disponible en carbono celular (Rosenberg *et al.*, 2010). Sin embargo, la presencia de virus y sus interacciones con otros microorganismos en reservorios de petróleo naturales, aún no ha sido estudiada en gran detalle, hasta ahora ningún virus ha sido aislado de un reservorio de petróleo (Pannekens *et al.*, 2019). No obstante, los estudios a nivel metagenómico, han permitido revelar la presencia de virus en este tipo de ambientes. Como lo sucedido en el derrame del DWH, donde se encontró que los genes de fagos involucrados en replicación eran significativamente más abundantes en las muestras contaminadas que en las no contaminadas (Lu *et al.*, 2012) o en Molnár *et al.*, 2020, donde reconstruyó el genoma de un virus que infecta a arqueas del género *Methanosarcina*, característico de zonas contaminadas por hidrocarburos.

A pesar de la importancia que los virus tienen en la degradación de hidrocarburos y en todos los ecosistemas en general, aún no existe un estudio dónde se caracterizen los virus del GoM, región naturalmente abundante en hidrocarburos. Es por eso que en este estudio se pretende caracterizar varias regiones del GoM en cuanto a diversidad viral, para definir una línea base que servirá para estudios posteriores en materia viral del GoM. Con el fin de ampliar la caracterización viral del GoM, además de las muestras obtenidas por el CIGoM, se añadieron al análisis otros estudios de metagenómica externos al CIGoM, hechos en distintas zonas del GoM. Los cuales son: un estudio con muestras de agua y sedimento tomadas durante el derrame del DWH (Mason *et al.*, 2012; Mason *et al.*, 2014), y otros dos de sedimentos en sitios con salidas naturales de hidrocarburos (Dong *et al.*, 2019; Zhao *et al.*, 2020). Es importante recalcar que en ninguno de estos proyectos se estudió la parte viral, este es el primer reporte respecto a eso.

4. JUSTIFICACIÓN

Los virus tienen un papel de suma importancia en la regulación de los ecosistemas y de las poblaciones de sus hospederos, además se ha demostrado que también son de gran relevancia en los procesos de degradación de hidrocarburos. Por lo anterior, es necesario considerarlos en los análisis de diversidad biológica de metagenomas ambientales de zonas con concentraciones altas de hidrocarburos, tal como el Golfo de México, para tener un panorama completo de las interacciones biológicas que están sucediendo en ese ecosistema.

5. OBJETIVOS

- **General**

Analizar la diversidad viral de muestras de agua y sedimento en diferentes zonas del Golfo de México.

- **Particulares**

- Determinar la abundancia viral de las muestras metagenómicas seleccionadas
- Comparar las muestras en términos de alfa y beta diversidad viral
- Analizar los patrones de diversidad de las muestras en relación con la presencia o ausencia de hidrocarburos

6. MÉTODO

La evaluación de la diversidad microbiana, a través de análisis metagenómicos involucra cuatro pasos principales: i) toma de muestras, ii) extracción y purificación de DNA, iii) secuenciación y iv) análisis bioinformáticos. Las muestras ya habían sido procesadas (paso i, ii y iii) previamente. Por lo anterior, el objetivo de este

trabajo fue realizar análisis bioinformáticos enfocados a identificar virus de estas muestras. Es importante recalcar que originalmente todas estas muestras fueron procesadas experimentalmente, con el fin de estudiar a las comunidades de procariontes que ahí habitaban, por lo cual ninguna había sido previamente estudiada en términos de diversidad viral.

6.1 Muestras

Este trabajo forma parte del CIGoM, por lo cual se tenía acceso directo a las muestras obtenidas por dicho consorcio (9 muestras). El resto de las muestras (30 muestras) se seleccionaron haciendo una búsqueda en línea de artículos existentes hasta septiembre del 2020, que tuvieran las siguientes características: 1) estudios de metagenómica en el GoM, 2) secuenciación de metagenoma completo realizadas con la plataforma de Illumina, 3) disponibilidad de secuencias, ya sea, en línea o contactando a los autores y 4) metadatos disponibles. A continuación se presentan descripciones breves de cada grupo de muestras y en la Tabla 1, se encuentran algunos de los metadatos disponibles para todas las muestras. Para mayores detalles sobre el procesamiento experimental (muestreo, extracción de DNA y secuenciación) de estas muestras, acudir a la Tabla A1 en el Anexo.

6.1.1 Muestras originales del CIGoM

En el contexto del CIGoM, con el objetivo de estudiar mejor las aguas mexicanas del GoM después del derrame del DWH, se tomaron muestras de agua y sedimento con transectos en profundidades de 100 a 3000 metros. Las muestras se tomaron en septiembre de 2015 y marzo del 2016, en la región de Perdido y la región de la Costa de Campeche. En total, nueve muestras tuvieron rendimiento alto de DNA, tres de agua: D18_MAX, A04_MIN y A04_AAIW (MAX: máxima fluorescencia, MIN: zona de mínimo oxígeno y AAIW: agua intermedia antártica) y seis de sedimento: D18_SED1-3, D18_SED, A04_SED y E03_SED y se usaron para llevar a cabo secuenciación de metagenoma completo (*whole metagenome shotgun*) (Raggi *et*

al., 2020). Las muestras D18_MAX, D18_SED1-3 y D18_SED provienen de la región de la Costa de Campeche, mientras que las muestras A04_MIN, A04_AAIW y E03_SED son de la región de Perdido. De estas muestras no se tiene información sobre si provenían de sitios con presencia o ausencia de hidrocarburos, ya que no se hizo una cuantificación al respecto.

6.1.2 Muestras de agua tomadas durante el derrame del DWH (Mason *et al.*, 2012)

Durante el derrame del DWH, se generó un penacho de hidrocarburos en el mar localizado entre 1000-1300 metros de profundidad, que causó un cambio en la composición de las comunidades microbianas. Es por eso que en este estudio, se trabajó con muestras duplicadas de agua tomadas entre el 27 y 31 de mayo del 2010: dos muestras del penacho, una a 1.5 km de distancia (DWH_W1-DWH_W2), una a 11 km de distancia (DWH_W3-DWH_W4) y una zona no contaminada a 40 km de distancia (DWH_W5-DWH_W6).

6.1.3 Muestras de sedimento tomadas durante el derrame del DWH (Mason *et al.*, 2014)

También bajo el contexto del derrame del DWH, en este estudio se tomaron núcleos de sedimento del 16 de septiembre al 20 de octubre del 2010, cinco meses después del derrame. Las muestras se obtuvieron a una profundidad promedio de aproximadamente 1500 m debajo de la superficie del mar, tomando la primera rebanada (0-1 cm) de cada núcleo de sedimento. En total se tomaron 14 muestras (DWH_S1-14), todas presentaban hidrocarburos en concentraciones diferentes. Las muestras provenían de distancias diferentes respecto al penacho de hidrocarburos: las más cercanas en una distancia de 0.7 a 2.7 km (DWH_S1-8), otras de 10.1 a 24.8 km de distancia (DWH_S9-11) y las más lejanas de 33.9 a 59.5 km de distancia (DWH_S12-14).

Tabla 1. Descripción de las muestras utilizadas en este estudio, organizadas por el tipo de muestra (agua o sedimento) y por estudio del cual provienen las muestras.

Nombre	Estudio	Tipo	Profundidad (m)	Tipo de muestra*
D18_MAX	CIGoM	Agua	50-75	NA
A04_MIN	CIGoM	Agua	300-400	NA
A04_AAIW	CIGoM	Agua	1000	NA
DWH_W1	Mason <i>et al.</i> , 2012	Agua	1207	Hidrocarburos
DWH_W2	Mason <i>et al.</i> , 2012	Agua	1207	Hidrocarburos
DWH_W3	Mason <i>et al.</i> , 2012	Agua	1179	Hidrocarburos
DWH_W4	Mason <i>et al.</i> , 2012	Agua	1179	Hidrocarburos
DWH_W5	Mason <i>et al.</i> , 2012	Agua	1020	Control
DWH_W6	Mason <i>et al.</i> , 2012	Agua	1020	Control
D18_SED1	CIGoM	Sedimento	1281	NA
D18_SED2	CIGoM	Sedimento	1281	NA
D18_SED3	CIGoM	Sedimento	1281	NA
D18_SED	CIGoM	Sedimento	1295	NA
A04_SED	CIGoM	Sedimento	2967	NA
E03_SED	CIGoM	Sedimento	3010	NA
ZHAO_S1	Zhao <i>et al.</i> , 2020	Sedimento	1400	Control
ZHAO_S2	Zhao <i>et al.</i> , 2020	Sedimento	1400	Control
ZHAO_S3	Zhao <i>et al.</i> , 2020	Sedimento	1400	Hidrocarburos
ZHAO_S4	Zhao <i>et al.</i> , 2020	Sedimento	1400	Control
ZHAO_S5	Zhao <i>et al.</i> , 2020	Sedimento	1400	Hidrocarburos
ZHAO_S6	Zhao <i>et al.</i> , 2020	Sedimento	1400	Hidrocarburos
ZHAO_S7	Zhao <i>et al.</i> , 2020	Sedimento	1400	Hidrocarburos
DWH_S1	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S2	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S3	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S4	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S5	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S6	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S7	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S8	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S9	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S10	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S11	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S12	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S13	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DWH_S14	Mason <i>et al.</i> , 2014	Sedimento	1500	Hidrocarburos
DONG_S1	Dong <i>et al.</i> , 2019	Sedimento	2800	Hidrocarburos
DONG_S3	Dong <i>et al.</i> , 2019	Sedimento	3000	Hidrocarburos
DONG_S2	Dong <i>et al.</i> , 2019	Sedimento	3200	Hidrocarburos

*La categoría "Tipo de muestra", se refiere a sí en el estudio de procedencia de las muestras, se menciona si provienen de sitios con presencia de hidrocarburos (confirmado por mediciones, por fotografías submarinas o por derrames petroleros recientes) o si eran muestras control.

6.1.4 Muestras de emanaciones naturales de hidrocarburos en el noreste del GoM (Dong *et al.*, 2019)

En este estudio se investigó el potencial metabólico de comunidades de bacterias y arqueas provenientes de tres muestras de sedimento de filtraciones naturales de petróleo en el fondo marino (aproximadamente tres km de profundidad) en el este del GoM. Las tres muestras (DONG_S1, DONG_S2 y DONG_S3) tenían hidrocarburos y mostraron que estas comunidades estaban potencialmente sostenidas a través de la metabolización anaeróbica de hidrocarburos, acetato e hidrógeno.

6.1.5 Muestras de emanaciones naturales de hidrocarburos en el noroeste del GoM (Zhao *et al.*, 2020)

En este estudio se enfocaron en estudiar las comunidades microbianas de sedimentos profundos que están bajo requerimientos de energía extremadamente bajos, pero que pueden ser estimuladas por fuentes de carbono externas, tales como las salidas naturales de petróleo en el fondo marino. Son siete muestras (ZHAO_S1, ZHAO_S3, ZHAO_S4, ZHAO_S5, ZHAO_S6 y ZHAO_S7), que van de los 10-15 cm de cada núcleo de sedimento y se ubican en el noroeste del Golfo de México. De estas muestras, tres de ellas (ZHAO_S1, ZHAO_S2 y ZHAO_S4), son como controles, ya que se tomaron en zonas donde no había hidrocarburos. Con base en las muestras de referencia se vio que las salidas naturales de hidrocarburos pueden cambiar en gran forma la estructura de la comunidad microbiana al estimular la fijación de nitrógeno.

Así se tienen, 39 muestras, nueve del estudio original del CIGoM y 30 de muestras descargadas en línea, distribuidas en distintas zonas del GoM (Fig. 8).

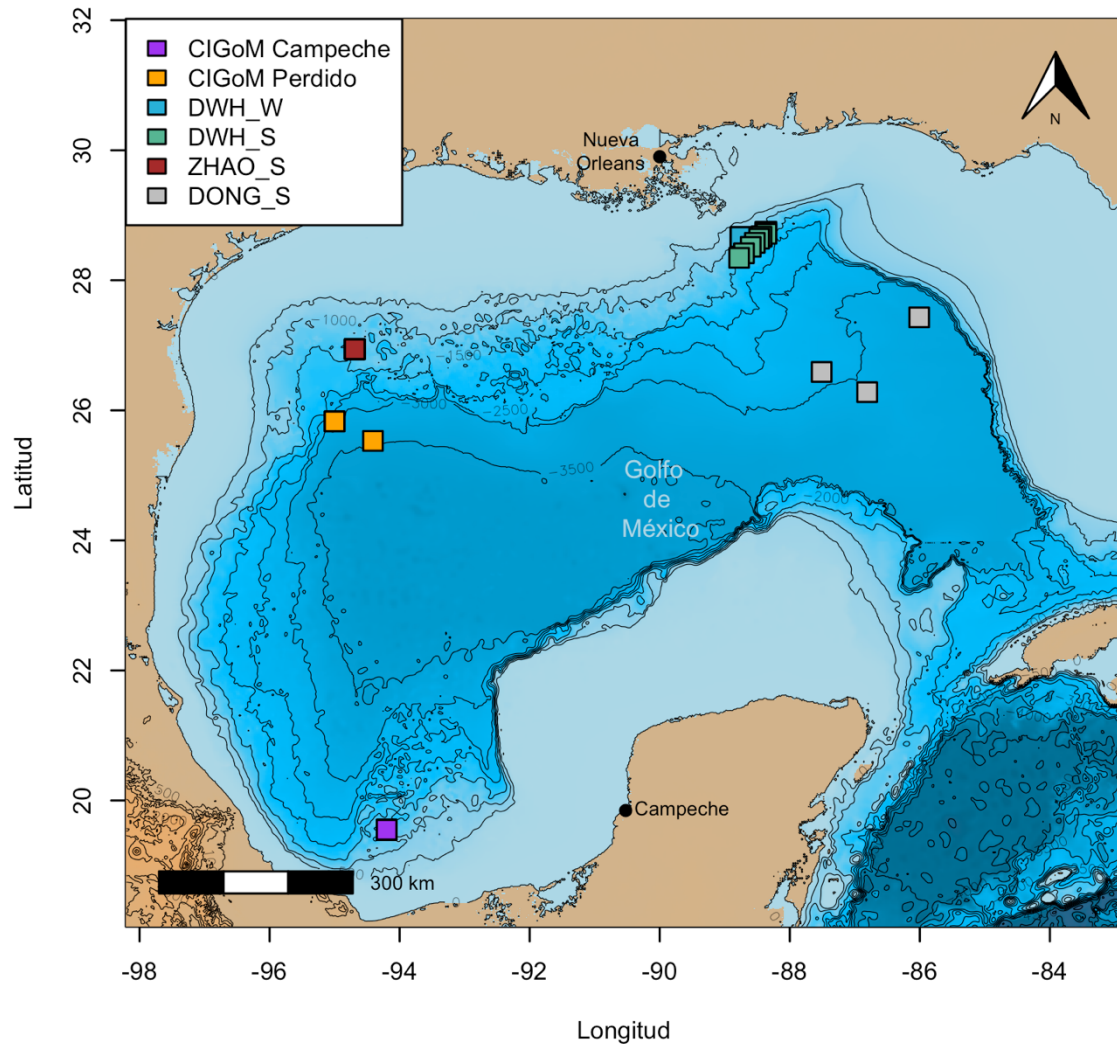


Figura 8. Distribución geográfica de las muestras de estudio en el GoM. Las muestras del CIGoM se encuentran localizadas en dos regiones del GoM pertenecientes a México, zona de Perdido (CIGoM Perdido) y la zona de Campeche (CIGoM Campeche). Las muestras de Mason *et al.*, 2012 y Mason *et al.*, 2014, están en el noreste del GoM (DWH_W y DWH_S), en las costas de Mississippi y Luisiana, E.U.A. Las muestras de Zhao *et al.*, 2020, en el noroeste del GoM en las costas de Texas (ZHAO_S) y las de Dong *et al.*, 2019 en el noreste del GoM (DONG_S).

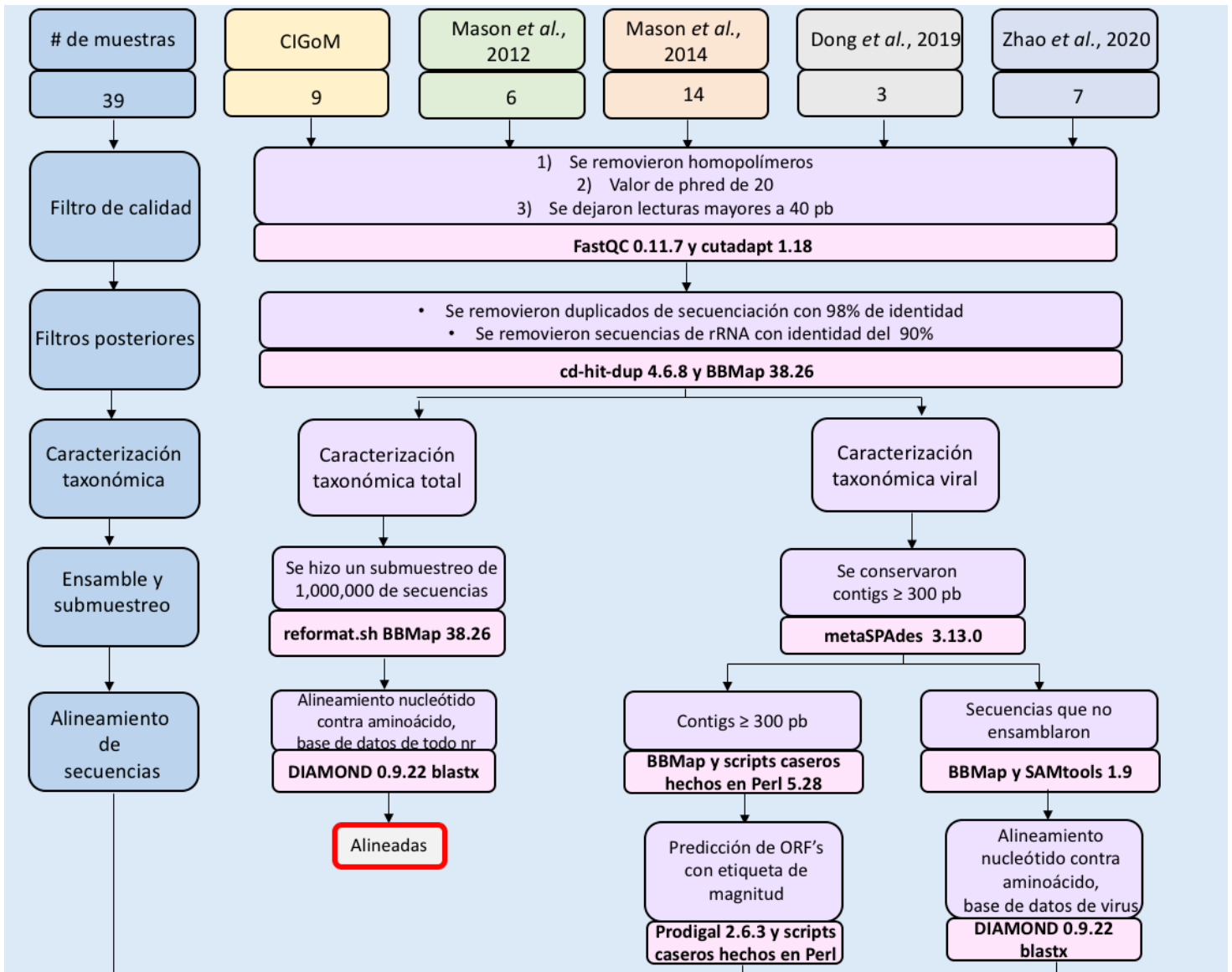
6.2 Procesamiento bioinformático

El procesamiento de las lecturas obtenidas por medio de la secuenciación se llevó a cabo basándose en el flujo de trabajo descrito en el artículo de Taboada *et al.*,

2018 con modificaciones que se adecuan a las muestras de este estudio en particular y el análisis computacional se hizo usando el cluster del Instituto de Biotecnología de la UNAM. El diagrama del flujo de trabajo bioinformático se muestra en la Fig. 9, todos los comandos utilizados para correr los programas se encuentran en la Tabla A2 en el Anexo.

6.2.1 Calidad de las secuencias

El primer paso fue realizar el análisis de la calidad de las secuencias, por medio del programa FastQC (versión 0.11.7; Andrews, 2010). Con base en lo obtenido, se realizó el control de calidad de las muestras.



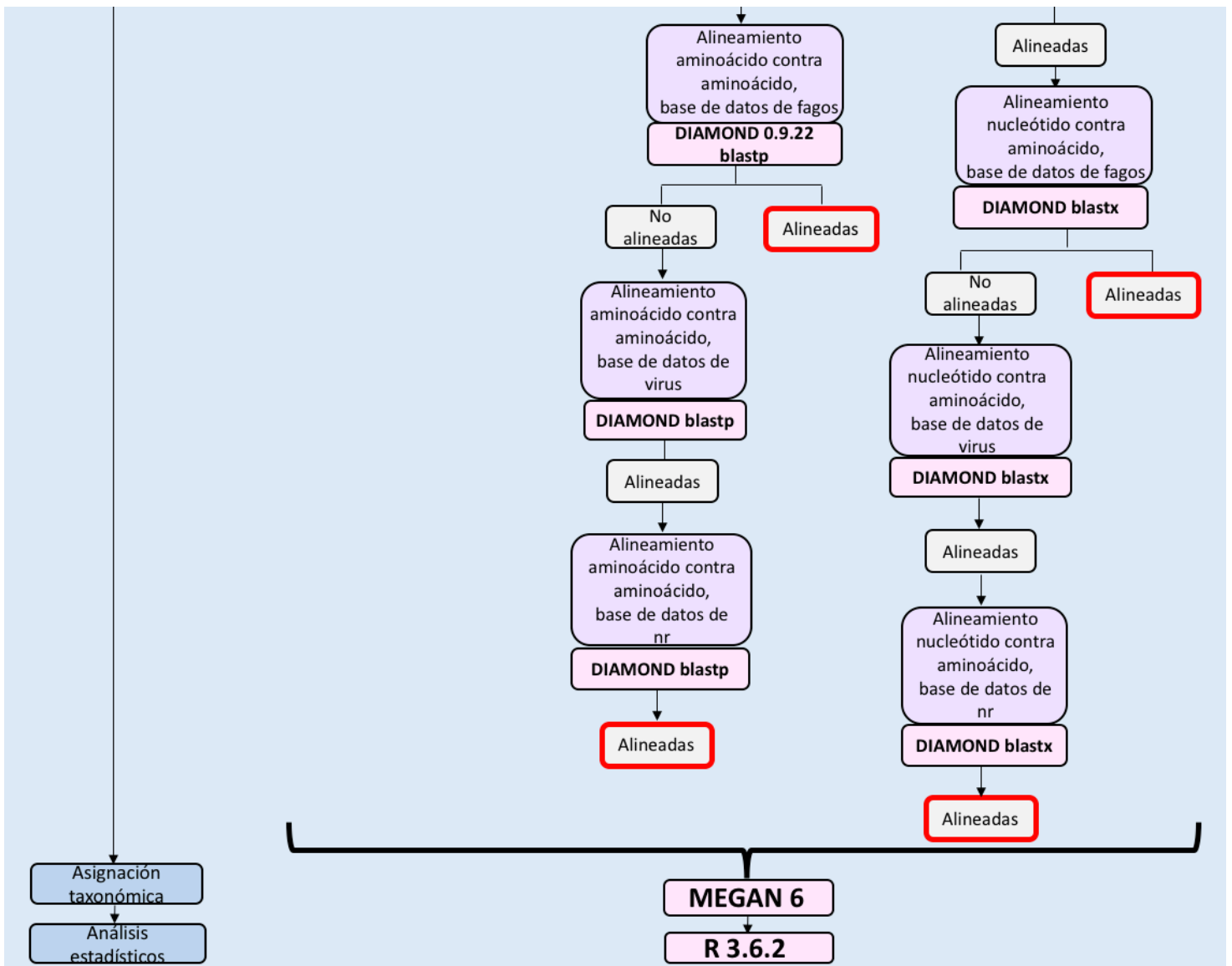


Figura 9. Flujo de trabajo bioinformático. En general para los cinco grupos de muestras, se realizaron los pasos de: filtrado por calidad, remoción de duplicados, remoción de secuencias de rRNA, ensamble, mapeo de lecturas, alineamiento contra bases de datos, asignación taxonómica, análisis de diversidad y metagenómica comparativa. Los recuadros con margen rojo, con la leyenda “Alineadas” fueron los archivos a partir de los cuales se realizaron los pasos de asignación taxonómica y análisis de diversidad.

Se utilizó el programa cutadapt (versión 1.18; Martin, 2011) para limpiar las secuencias según lo requerido. Las secuencias se mantuvieron como documentos pareados y si una de las secuencias del par no cumplía con las características de calidad dadas, el par completo se eliminaba.

Se fijó el nivel de calidad de las bases de las secuencias a un valor de Phred de 20, se quitaron secuencias sobrerrepresentadas que en general consistían en homopolímeros de nucleótidos (p.eg., GGGGGGG, AAAAAA), así como las bases en los extremos 5' y 3' en las que no se leyó ningún nucleótido (N), se eliminaron kmeros sobrerrepresentados (secuencias cortas de 5 nucleótidos, que superan la frecuencia esperada de aparición) y se eliminaron las secuencias de longitud menor a 40 pares de bases (pb).

Después de haber aplicado los filtros anteriores, se volvió a correr el programa FastQC para ver si la calidad había mejorado, si no mejoraba se volvían a aplicar los filtros de calidad necesarios. Una vez que las secuencias cumplían con los estándares de calidad, se utilizó el programa cd-hit-dup, dentro de la paquetería de CD-HIT (versión 4.6.8; Li & Godzik, 2006) (-e 2, hasta dos *mismatches* entre las secuencias, valor estandarizado en Taboada *et al.*, 2018 basándose en el porcentaje de error de los secuenciadores de Illumina (Glenn, 2011)) para eliminar secuencias duplicadas originadas por PCR y que resultan como lecturas de dos o más copias de exactamente el mismo fragmento de DNA (Taboada *et al.*, 2018).

El siguiente paso consistió en remover las secuencias que correspondieran a RNA ribosomal (rRNA) con el programa BBMap (versión 38.26; Bushnell, 2017) con un porcentaje de identidad del 90%, mapeando contra la base de datos de Silva (<https://www.arb-silva.de/>) de rRNA de la subunidad grande (SSU Ref NR 99, eliminando redundancia al 99%) y de la subunidad pequeña (LSU Parc). La remoción de secuencias de rRNA es un paso estandarizado en el flujo de trabajo de Taboada *et al.*, 2018, el estudio en el que se basó la metodología del trabajo aquí presente y las secuencias de rRNA se eliminan por dos razones principales: 1) debido a que estas son secuencias universales en bacterias estas pueden dar lugar a falsos positivos y 2) remover información innecesaria de las muestras hace el procesamiento más ágil y elimina ruido en la caracterización taxonómica que podría ser interpretado como parte de las secuencias desconocidas de las muestras (Taboada *et al.*, 2018). Aún cuando en este trabajo el objetivo no era trabajar con

bacterias, estas secuencias estarían listas para ser utilizadas en el futuro por alguien más interesado en el estudio de ellas.

Las secuencias restantes después de la aplicación de los filtros anteriores, fueron consideradas como secuencias válidas y se usaron para los análisis posteriores.

6.2.2 Caracterización taxonómica total

Con el objetivo de tener un panorama general de la composición taxonómica de las muestras y del porcentaje de secuencias conocidas y desconocidas, y al mismo tiempo ahorrar tiempo y esfuerzo de cómputo se hizo un submuestreo de un millón de secuencias con el programa `reformat.sh` (`samplereadstarget=1000000`) de BBMap (versión 38.26; Bushnell, 2017). Estas secuencias se alinearon con DIAMOND `blastx` (versión 0.9.22; Buchfink et al, 2015) contra toda la base de datos de proteínas del NCBI (`nr`) y a partir de los archivos de salida se hizo la asignación taxonómica.

6.2.3 Caracterización taxonómica viral

6.2.3.1 Ensamble

Las secuencias válidas se ensamblaron con el programa `metaSPAdes` (versión 3.13.0; Nurk *et al*, 2017), las muestras con secuencias de 75-121 pb (D18_MAX, A04_MIN, A04_AAIW, E03_SED, DWH_W1-6 y DWH_S1-14) se ensamblaron con los kmeros default (`-k 21,33,55`), las secuencias de 150 pb (D18_SED, A04_SED y DONG_S1-3) se ensamblaron con los kmeros (`-k 21,33,55,77`) y las secuencias de 250 pb (ZHAO_S1-7) con los kmeros (`-k 21,33,55,77,121`), según lo recomendado en el manual del programa. El análisis de las estadísticas de los ensamblajes se hizo con QUAST (versión 5.0.2; Gurevich *et al.*, 2013).

Se seleccionaron los contigs mayores o iguales a 300 pb ya que es el tamaño de contig estandarizado en Taboada *et al.*, 2018 (flujo de trabajo bioinformático en el

cual se basó el estudio aquí presente) para caracterización taxonómica viral y se debe a dos razones principales: 1) para cubrir toda la gama posible de genes virales ya que los contigs de 300 pb se traducen en proteínas de 100 aminoácidos, lo cual es el tamaño de los genes más pequeños de fagos, llamados genes accesorios (Iida *et al.*, 1998) y 2) en términos de estadística a nivel de blast, con 100 aminoácidos ya se puede determinar un buen valor de e-value de acuerdo a la base de datos actual (Taboada *et al.*, 2018). El hecho de que no se haya trabajado con los contigs menores a 300 pb se debe a que esto incrementaría de forma substancial el esfuerzo y tiempo de computo en los análisis posteriores y aún cuando existan proteínas virales menores a los 100 aminoácidos estas también podrían ser identificadas por homología con los contigs mayores o iguales a 300 pb (Taboada *et al.*, 2018).

Sobre estos contigs se mapearon las secuencias originales de cada muestra, con el programa BBMap (minid=0.90) y a partir del mapeo se generaron archivos con las secuencias no mapeadas, i.e, las que no habían sido ensambladas.

6.2.3.2 Secuencias ensambladas

6.2.3.2.1 Predicción de marcos de lectura abierta (ORF'S)

A partir de los contigs se predijeron ORF's, es decir, secuencias codificantes, con el programa Prodigal (versión 2.6.3; Hyatt *et al.*, 2010) (-p meta). A estos ORF's se les mapearon las secuencias originales de cada muestra, con el programa BBMap (minid=0.90). A partir del mapeo se generaron archivos con las secuencias mapeadas, i.e, las que habían sido identificadas como parte de los ORF's. Con estos archivos, por medio de scripts caseros hechos en Perl 5.28 por la Dra. Blanca Taboada, se hizo un conteo de cuántas secuencias mapeaban contra cada ORF, es decir, de cuántas secuencias estaba formado cada ORF. A partir de ese conteo, se añadió una etiqueta a los encabezados de cada ORF llamado "magnitud=", de forma que los ORF's pudieran ser comparables con los secuencias crudas no ensambladas en pasos posteriores.

6.2.3.2.2 Alineamiento de ORF's

Para los alineamientos a nivel viral se utilizaron dos bases de datos: una base de datos de proteínas del NCBI (nr) de virus y una base de datos de proteínas del NCBI (nr) de solo fagos (ambas bases curadas por M.C. Jerome Jean Veryelen jerome.verleyen@ibt.unam.mx y disponibles en el cluster Teopanzolco del Instituto de Biotecnología). El objetivo de trabajar con estas dos bases de datos se debe a dos factores: 1) tener las muestras listas en caso de que en estudios posteriores se busque trabajar solo con fagos y 2) el alineamiento contra la base de datos de nr de virus después pasa por un proceso de eliminación de falsos positivos, en el cual se alinean contra toda la base de nr, sin embargo, este paso no puede ser llevado a cabo con las secuencias de fagos ya que algunos de ellos están en forma de profagos y serían clasificados como secuencias de bacterias.

Los ORF's se alinearon contra la base de datos de proteínas del NCBI (nr) de fagos (disponible en el cluster Teopanzolco del Instituto de Biotecnología), a nivel de aminoácido contra aminoácido con el programa usando la opción blastp. Los ORF's que alinearon se mantuvieron y a partir de estos, por medio de scripts caseros hechos en Bash 4.2.46, se generaron archivos con todos los ORF's que no habían alineado. Los ORF's que no alinearon, se mapearon contra la base de datos de nr de virus (disponible en el cluster Teopanzolco del Instituto de Biotecnología) con los mismos parámetros. Asimismo, los ORF's que alinearon contra virus, se volvieron a alinear contra toda la base de datos de nr (-e 0.00001) con el fin de eliminar falsos positivos, los ORF's que correspondían a virus se conservaron. Los ORF's que alinearon contra la base de datos de fagos, no se volvieron a alinear contra toda la base de nr porque esto eliminaría las secuencias de profagos.

6.2.3.1 Secuencias no ensambladas

La mayor proporción de secuencias en las muestras correspondían a secuencias no ensambladas. Para facilitar el análisis (en cuanto términos de uso de memoria

de DIAMOND), estas primero se alinearon contra la base de datos nr de virus, a nivel de nucleótido contra aminoácido con el programa DIAMOND usando la opción blastx, con el fin de crear un subgrupo con cuál fuera más fácil trabajar. Las secuencias que alinearon, se volvieron a alinear contra fagos y las que no alinearon contra virus. Las secuencias que alinearon contra fagos se conservaron y las que alinearon contra virus se volvieron a alinear contra toda la nr (-e 0.00001), las secuencias que correspondían a virus se conservaron.

Así, cuatro archivos de salida de DIAMOND por muestra pasaron a la etapa de asignación taxonómica: ORF's y secuencias crudas que alinearon contra fagos y ORF's y secuencias crudas que alinearon contra virus.

6.2.4 Asignación taxonómica

La asignación taxonómica se hizo con MEGAN Ultimate Edition (versión 6.19.9; Huson *et al.*, 2007), usando el programa daa-meganizer dentro de su paquetería. El parámetro de minscore=50 para las muestras con lecturas de 75-121 pb y minscore=100 para las lecturas y ORF's mayores o iguales a 150 pb y el parámetro de maxExpected=0.001 para los alineamientos contra fagos y virus, y maxExpected=0.00001 para los alineamientos contra toda la nr. Asimismo se uso la opción de readAssignmentMode=readMagnitude, la cual leía la etiqueta de magnitud de los ORF's, es decir que la asignación taxonómica se hizo a nivel de secuencia.

Después se usaron los programas compute-comparison, extract-biome y ExtractReads de MEGAN, para extraer matrices de conteo de secuencias de taxones virales presentes en cada muestra a nivel de familia, género y especie.

6.2.5 Análisis de diversidad y metagenómica comparativa

Los análisis de diversidad, de metagenómica comparativa y las figuras originales de este estudio se hicieron con el lenguaje de programación R 3.6.2 (R Core Team,

2019) y con diversas paqueterías del mismo. Las matrices de abundancia de cada muestra generadas en MEGAN se exportaron a R, en donde por medio de las paqueterías dplyr 1.0.1 (Wickham *et al.*, 2020) y reshape 0.8.8 (Wickham, 2007), se hicieron matrices que conjuntaban a todas las muestras a nivel de familia, género y especie, respectivamente. A partir de la matriz de abundancia a nivel de familia se generaron gráficas de barras apiladas con la paquetería ggplot2 3.3.2 (Wickham, 2016) para visualizar la abundancia taxonómica de las muestras. Debido a las diferencias en profundidad de secuenciación entre las muestras, las matrices se normalizaron con el método cumNormMat de la paquetería metagenomeSeq 1.28.2, en el cual los valores de normalización se calculan a partir de la suma acumulada de los conteos hasta un umbral o cuantil específico basado en cada conjunto de datos en particular (Paulson *et al.*, 2013a). Las pruebas de diversidad (índice de Simpson, alfa diversidad y beta diversidad), así como los métodos de ordenación (PCoA) se hicieron con la paquetería vegan 2.5-6 (Oksanen *et al.*, 2019). La clusterización de las muestras se hizo con la función hclust(method="average") del paquete base de R, con la matriz de disimilitud construida con la distancia del coeficiente de correlación de Pearson, basado en el paquete pvclust (Suzuki *et al.*, 2019). Las gráficas de diversidad se construyeron con las opciones de graficación estándares de R y con ggplot2. Se usaron scripts escritos en R por la Dra. Ernestina Godoy Lozano, que utilizan la paquetería metagenomeSeq (Paulson *et al.*, 2013b) para obtener los géneros y especies virales diferencialmente abundantes en las muestras con la función fitFeatureModel y el valor de P ajustado al número de muestras. La paquetería venn 1.9 (Dusa, 2020), se utilizó para hacer los diagramas de Venn. El mapa de la Fig. 8, se construyó con las siguientes paqueterías: marmap 1.0.4 (Pante & Simon-Bouhet, 2013), raster 3.3-13 (Hijmans, 2020), prettymapr 0.2.2 (Dunnington, 2017) y mapdata 2.3.0 (Becker & Wilks, 2018).

7 RESULTADOS Y DISCUSIÓN

7.2 Control de calidad

El número total de secuencias obtenidas para cada muestra está entre los 20 y 300 millones, siendo las muestras de DWH_S las de mayor número de secuencias. En promedio el porcentaje de secuencias válidas fue de 87% (DE=14), lo que quiere decir que en general todas las muestras tenían buena calidad. A excepción de la muestra A04_MIL, en la cual se mantuvo el 27% de sus secuencias originales (Fig. 10 y Tabla A3, Anexo).

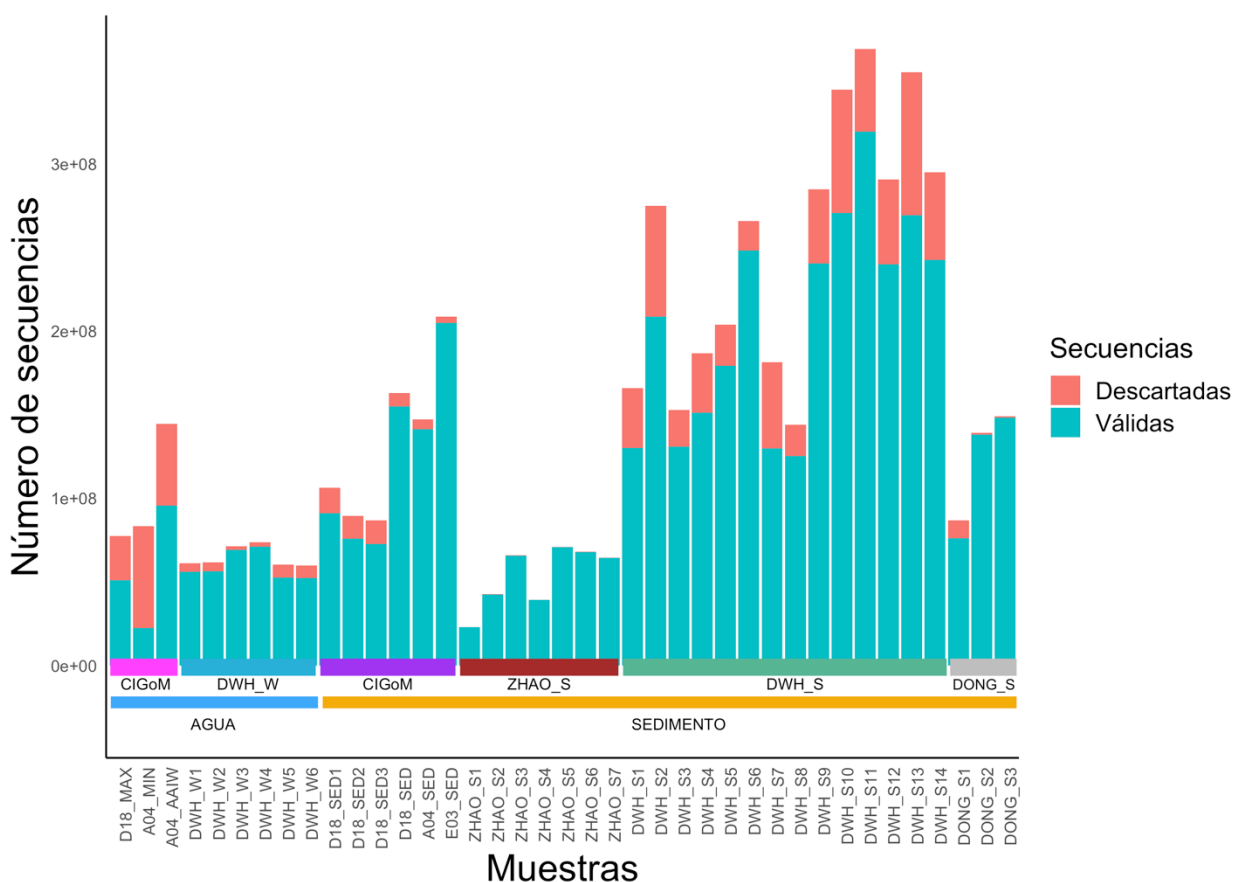


Figura 10. Porcentaje de secuencias válidas por muestra, es decir, secuencias que pasaron los filtros de calidad. Las muestras están organizadas en el eje x, según el el sustrato y el estudio del que provienen.

7.3 Ensamblajes y predicción de ORF's

En cuanto a los ensamblajes, las muestras a partir de las cuales se obtuvieron mayor número de contigs mayores o iguales a 300 pb fueron las del estudio de DONG_S y las que menor número de contigs presentan son las muestras de agua (Fig. 11 y Tabla A4, Anexo).

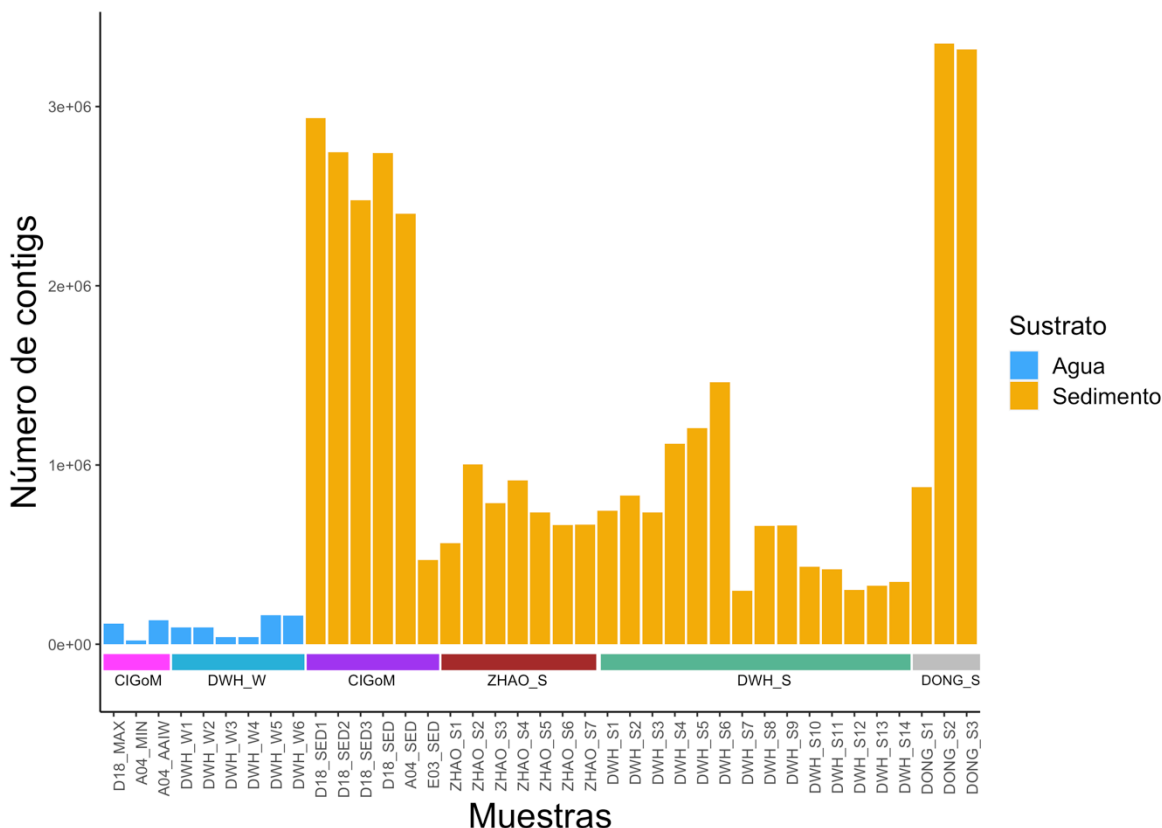


Figura 11. Número de contigs obtenidos en cada muestra. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Asimismo, la muestra con mayor número de ORF's predichos fue A04_SED, lo cual coincide con que haya sido una de las muestras con mayor número de contigs obtenidos y en general las muestras de sedimento, al igual que con el número de contigs, tuvieron los valores más altos de ORF's encontrados (Fig. 12 y Tabla A5, Anexo).

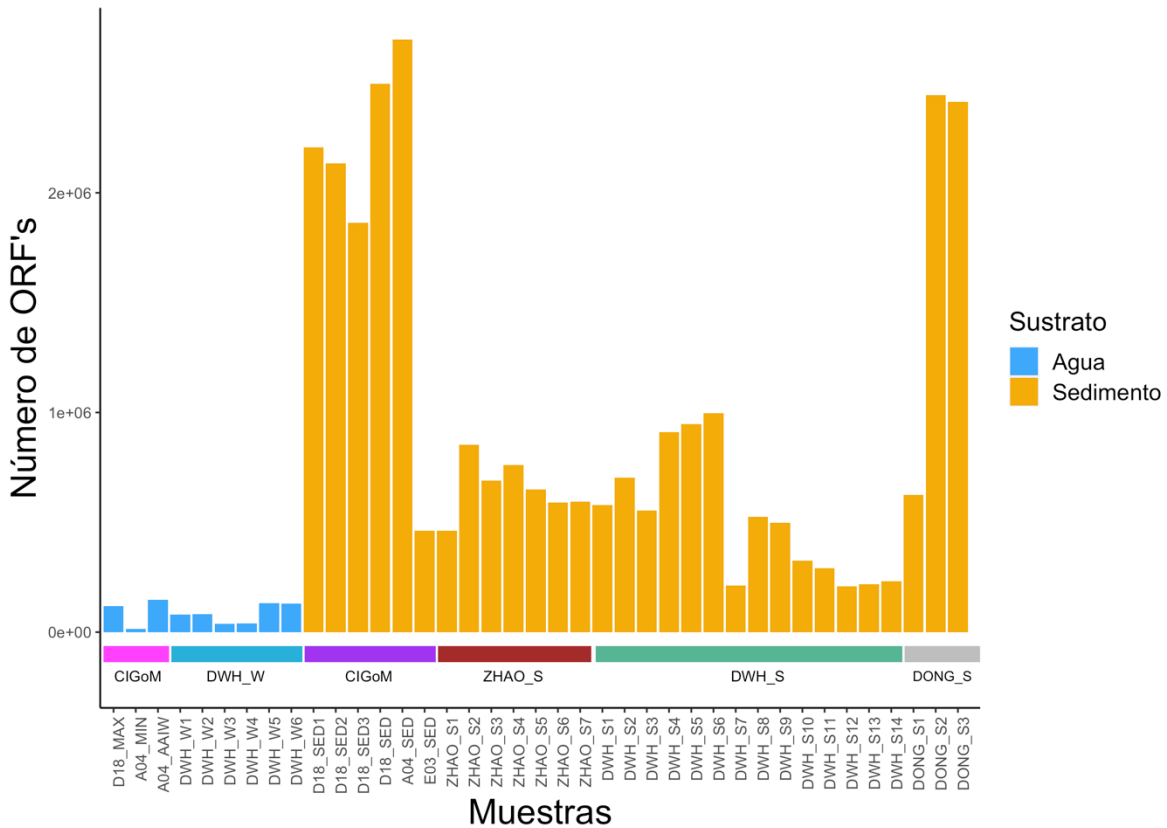


Figura 12. Número de ORF's predichos por muestra. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Sin embargo, a nivel de ORF's encontrados por kilobase (kb) (número de ORF's divididos por la longitud total de pares de bases ensambladas por muestra), se observa que todas las muestras tienen la misma tasa, excepto en A04_SED, cual supera los dos ORF's/kb (Fig. 13).

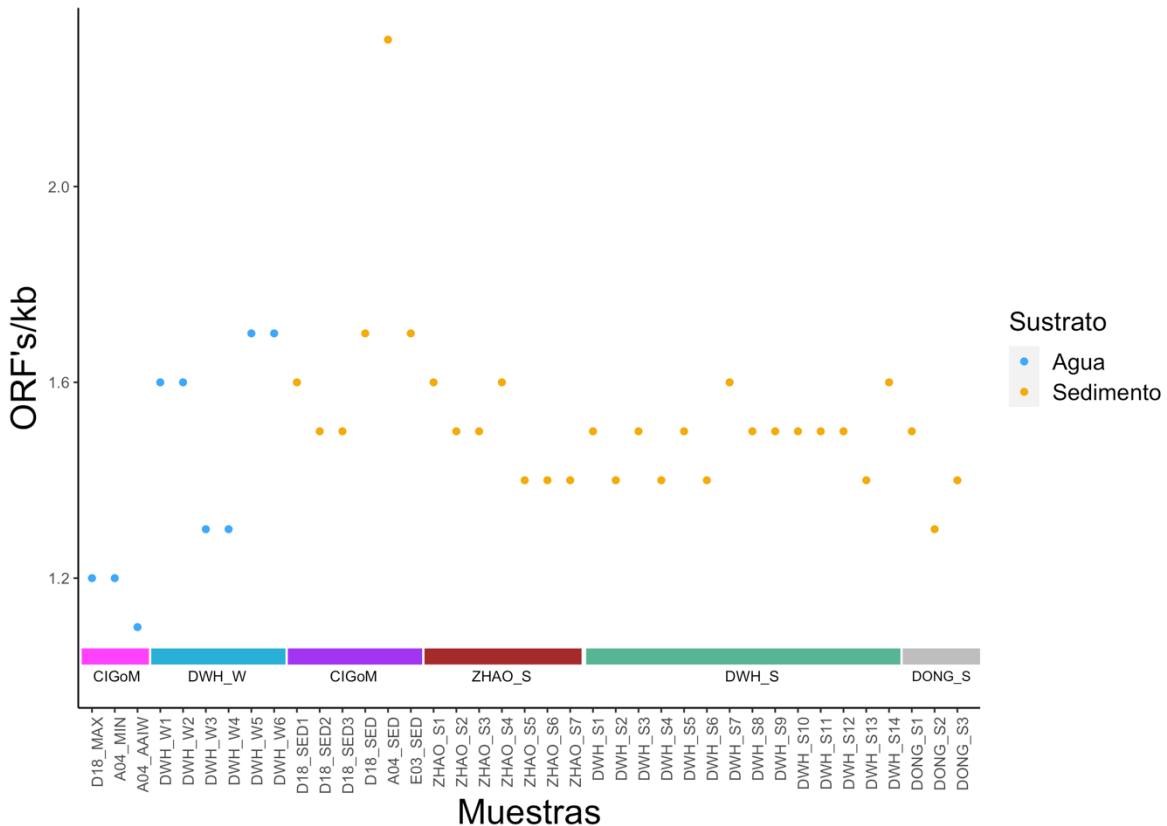


Figura 13. Número de ORF's encontrados por kb ensamblada de cada muestra. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

7.4 Composición taxonómica total

En un millón de secuencias submuestreadas de todas las muestras, el porcentaje de secuencias anotadas como conocidas, va del 4% al 98% del total (Fig. 14). Las muestras del DWH de agua (DWH_W) tuvieron los mayores porcentajes de secuencias conocidas (promedio de 89%, DE=13), seguida de las muestras del noroeste del GoM, ZHAO_S (promedio 42%, DE=7), las muestras del DWH de sedimento (promedio 34%, DE=14), las del noreste del GoM, DONG_S, (promedio 12%, DE=2) y por último las del CIGoM (promedio 4%, DE=2).

De las muestras del CIGoM, la muestra que tuvo mayor porcentaje de identificación fue D18_SED1 (5.3%), seguida de D18_SED3 (5.1%) y la que tuvo el menor porcentaje fue A04_MIN (1.9%).

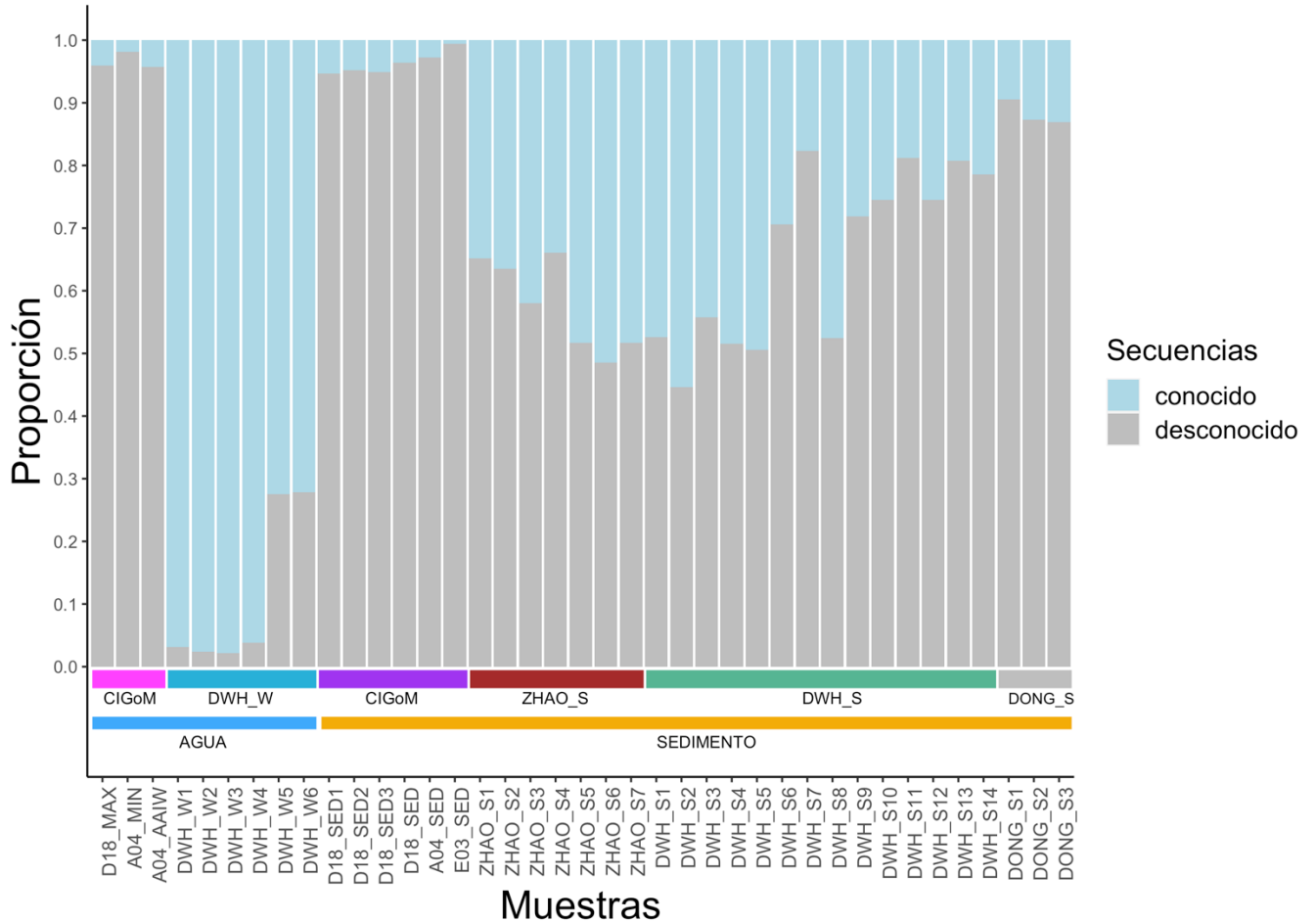


Figura 14. Proporción de secuencias conocidas y desconocidas. Las secuencias conocidas son aquellas que tuvieron un alineamiento con secuencias en bases de datos, mientras que las desconocidas no han sido reportados todavía. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Respecto a las secuencias del CIGoM, esto coincide con lo reportado en otros estudios, en dónde la mayor parte de las secuencias obtenidas son identificadas como desconocidas (Anderson *et al.*, 2011; Corinaldesi *et al.*, 2012; Gregory *et al.*, 2019; Taboada *et al.*, 2018, Yoshida *et al.*, 2013). Sin embargo, respecto al mayor porcentaje de secuencias identificadas como conocidas en el resto de las muestras, esto podría deberse a qué tanto haya sido estudiada determinada área, como por ejemplo, en el estudio de Corinaldesi *et al.*, 2012, se muestra que mientras en regiones como el noreste del océano Atlántico el porcentaje de secuencias conocidas es del 3.9%, en el mar Mediterraneo por el contrario, las secuencias

conocidas representan el 54%. El mar Mediterraneo ha sido ampliamente estudiado (Coutinho *et al.*, 2019; Haro-Moreno *et al.*, 2018; Tully *et al.*, 2017, entre otros) mientras que el único estudio de esa zona específica del océano Atlántico es el de Corinaldesi *et al.*, 2012.

Las aguas mexicanas del GoM han sido poco estudiadas y los sedimentos han sido aún menos estudiados (CIGoM, s.f.; Godoy-Lozano *et al.*, 2018; Raggi *et al.*, 2020), en comparación de las áreas estadounidenses del GoM (Dong *et al.*, 2019; Kotska *et al.*, 2011; Liu *et al.*, 2012; Mason *et al.*, 2012; Mason *et al.*, 2014; Zhao *et al.*, 2020, entre otros), lo que podría explicar porque la mayoría de las muestras de CIGoM tienen más del 90% de sus secuencias como desconocidas.

Dentro del porcentaje de secuencias anotadas como conocidas, la muestra uno de las muestras de sedimento del DWH (DWH_S1) tuvo el mayor porcentaje de secuencias virales (1.2%) (Fig. 15). En general, las secuencias conocidas corresponden a bacterias, seguido por secuencias de arqueas y a su vez estas en algunos casos seguidas por secuencias virales y en otros por secuencias correspondientes a eucariontes. Este fenómeno también se observa en otros estudios de metagenómica (Anderson *et al.*, 2011; Corinaldesi *et al.*, 2012; Gregory *et al.*, 2019), en donde la mayoría de las secuencias conocidas son identificadas como bacterias y dependiendo del caso, es el grupo taxonómico que sigue en abundancia. Sin embargo, esto no quiere decir que en los ecosistemas marinos, las bacterias sean más abundantes que los virus. En realidad, este fenómeno se debe principalmente a dos factores: 1) a lo largo de la historia, las bacterias han sido más estudiadas que los virus, lo cual se ve reflejado en un mayor número de secuencias bacterianas en las bases de datos (RefSeq Growth Statistics, 2021) y 2) el tamaño promedio de una bacteria es mayor que el promedio de un virus, por lo cual las bacterias representan mayor biomasa (Suttle, 2007) y por lo tanto mayor cantidad de DNA.

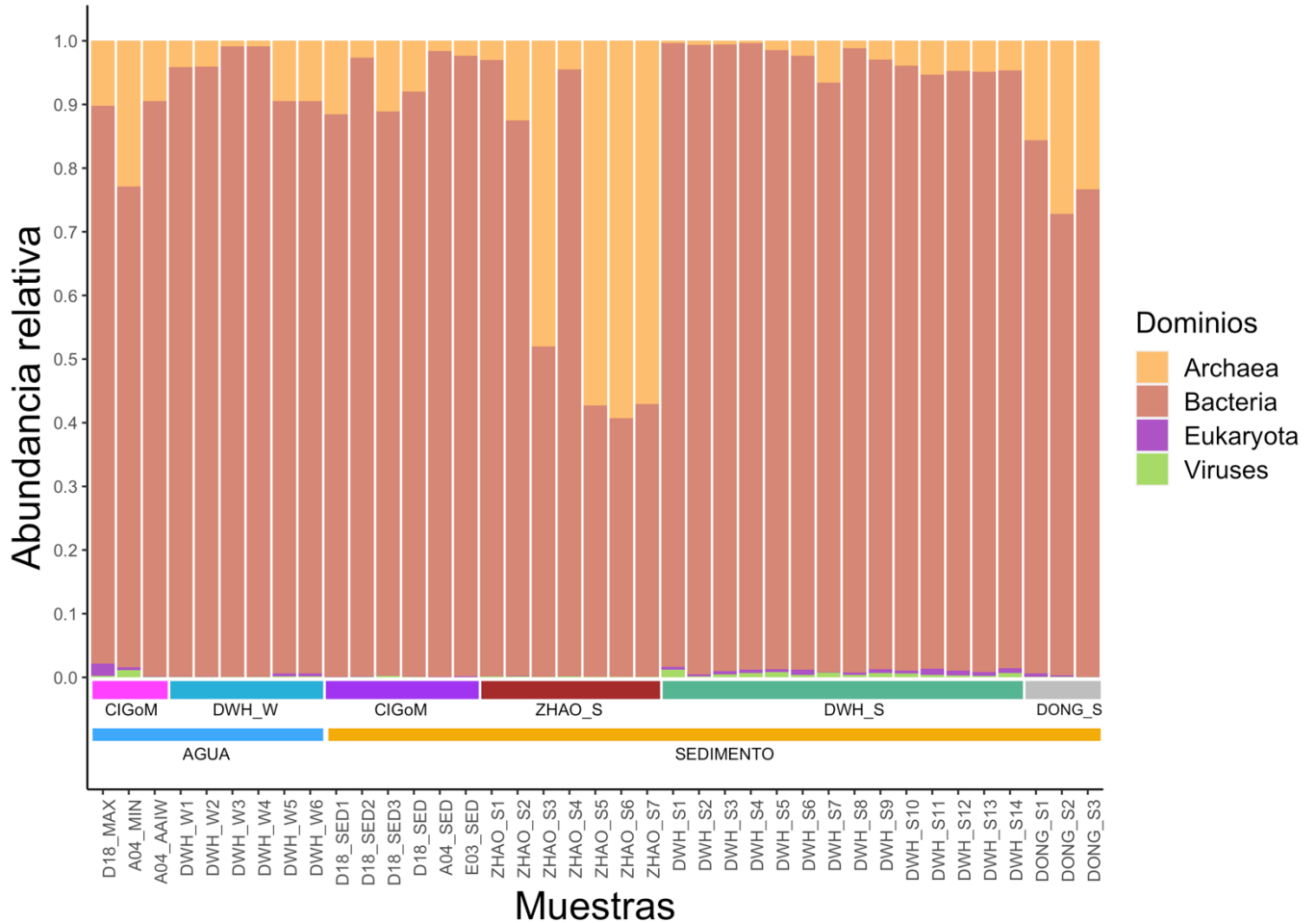


Figura 15. Abundancia relativa de las secuencias conocidas asignadas según el dominio al que pertenecen (Bacteria, Archaea, Eukaryota y Virus). En todas las muestras, la mayor parte de las muestras corresponde a bacterias, seguido de arqueas y en algunos casos virus, y siempre en menor abundancia las secuencias correspondientes a eucariontes.

Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Es importante destacar que, la proporción de secuencias virales encontradas en una determinada muestra, suele ser mayor cuando el procesamiento de la misma está enfocado hacia aislar virus. Es decir, se utiliza un tamaño de filtro determinado (0.2 μm y estudiando el filtrado) y la extracción del material genómico se da mediante tratamientos especiales (Yoshida *et al.*, 2013). Las muestras del CIGoM, se obtuvieron con el objetivo de analizar secuencias de bacterias, lo cual podría ser

una de las explicaciones de por qué el porcentaje de secuencias conocidas identificadas como virales es bajo (del 0% al 1.2%).

Otro punto importante, es que la proporción de eucariontes en la muestra D18_MAX (1.9%) es mayor que en el resto de las otras muestras (promedio 0.3, DE=0.3) y esto muy probablemente se debe a que esta es la muestra de menor profundidad (50-75 m). Se sabe que los protistas (grupo del que provienen la mayoría de las secuencias reconocidas como eucariontes en este tipo de ambiente) abundan en aguas superficiales y conforme aumenta la profundidad disminuye su biomasa (Caron *et al.*, 1995).

7.5 Composición taxonómica viral

Con el objetivo de analizar si el esfuerzo de muestreo en la secuenciación había sido suficiente, se obtuvo una curva de rarefacción a nivel de especies virales (Fig. 16). Las curvas de rarefacción de todas las muestras logran alcanzar una meseta, lo que indica que el esfuerzo de muestreo fue suficiente para todas las muestras. Sin embargo, se observa que para el mismo esfuerzo de muestreo (tamaño de muestreo alrededor de 200,000 secuencias), la muestra en la que se encuentra el menor número de especies es A04_MIN. Debido que en comparación con el resto de las muestras, A04_MIN se comporta diferente en cuanto al número de especies encontradas, es probable que las diferencias se deban a causas a priori, es decir, causas relacionadas con el procesamiento de la muestra, más que a diferencias inherentes a la muestra.

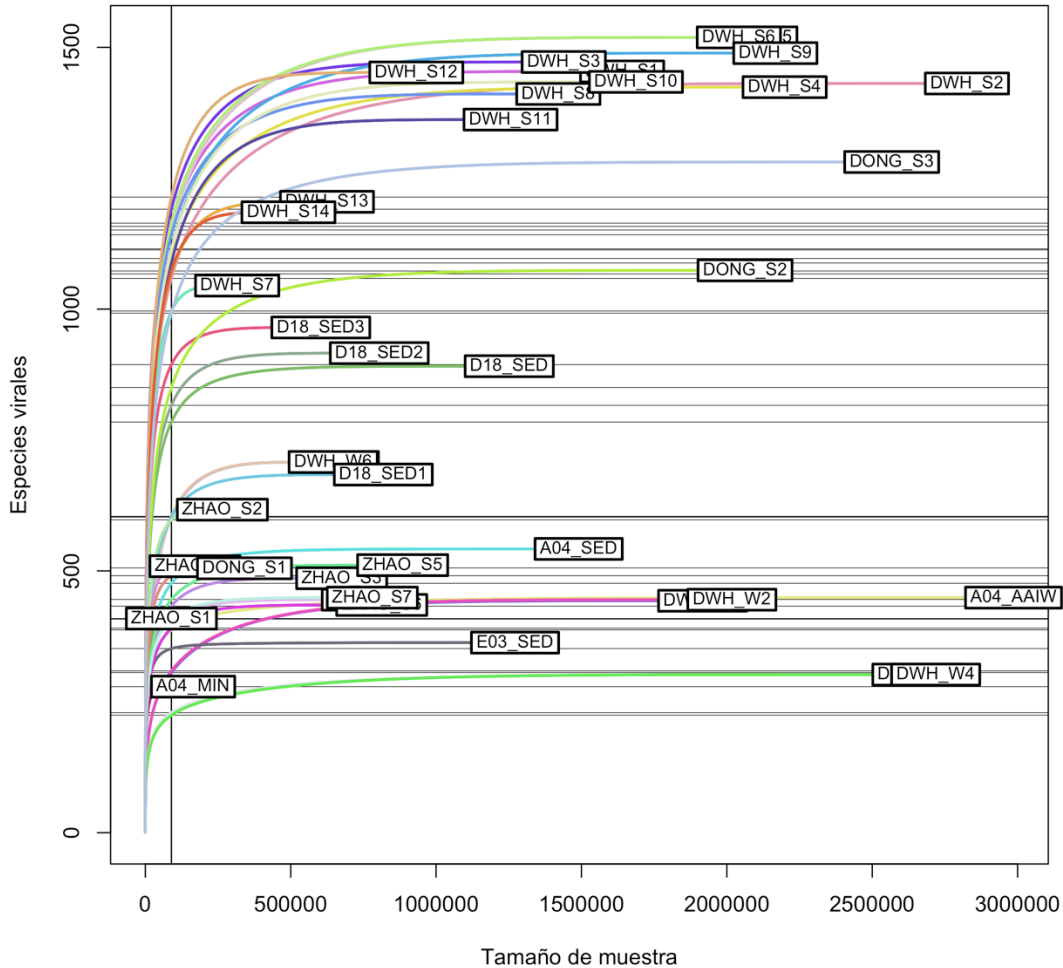


Figura 16. Curva de rarefacción del total de las muestras. Se observa que a pesar de que todas las muestras alcanzan el esfuerzo máximo de muestreo, el número de especies encontradas es diferencial entre las muestras.

La identificación viral se hizo a nivel de familia, género y especie, en la Tabla 2 se muestra la asignación a nivel de familia. De las 31 familias identificadas, 17 pertenecen a familias de fagos y a su vez son las que acumulan la mayoría de las secuencias identificadas. En general se sabe que en los océanos hay mayor cantidad de fagos que de virus que infectan eucariontes, lo cual toma sentido si se toma en cuenta que las bacterias representan en términos de abundancia, el mayor hospedero disponible en el mar (Angly, 2004; Cochlan *et al.*, 1993; Weinbauer, 2004; Breitbart, 2012)

Tabla 2. Listado del total de familias de virus encontradas en las muestras. Se muestra la abundancia relativa promedio de cada familia respecto al resto de las familias en el conjunto de muestras.

Familias virales	Tipo de genoma	Hospedero	Abundancia promedio (%)	Desviación estándar
<i>Myoviridae</i>	dsDNA	Bacterias y arqueas	50.9	10.9
<i>Siphoviridae</i>	dsDNA	Bacterias y arqueas	34.4	4.8
<i>Podoviridae</i>	dsDNA	Bacterias y arqueas	12.2	12.9
<i>Autographiviridae</i>	dsDNA	Bacterias	0.7	0.5
<i>Demereciviridae</i>	dsDNA	Bacterias	0.4	0.7
<i>Microviridae</i>	ssDNA	Bacterias	0.4	1.0
<i>Inoviridae</i>	dsDNA	Bacterias	0.3	0.2
<i>Chaseviridae</i>	dsDNA	Bacterias	0.1	0.2
<i>Ackermannviridae</i>	dsDNA	Bacterias	0.1	0.1
<i>Corticoviridae</i>	dsDNA	Bacterias	<0.1	0.1
<i>Autolykiviridae</i>	dsDNA	Bacterias	<0.1	<0.1
<i>Herelleviridae</i>	dsDNA	Bacterias	<0.1	<0.1
<i>Drexlerviridae</i>	dsDNA	Bacterias	<0.1	<0.1
<i>Tectiviridae</i>	dsDNA	Bacterias	<0.1	<0.1
<i>Plasmaviridae</i>	dsDNA	Bacterias	<0.1	<0.1
<i>Spiraviridae</i>	ssDNA	Arqueas	<0.1	<0.1
<i>Halspiviridae</i>	dsDNA	Arqueas	<0.1	<0.1
<i>Pithoviridae</i>	dsDNA	Amebas	0.2	0.8
<i>Mimiviridae</i>	dsDNA	Protozoarios	0.1	0.3
<i>Marseilleviridae</i>	dsDNA	Protozoarios	0.1	0.1
<i>Phycodnaviridae</i>	dsDNA	Algas	0.1	0.3
<i>Lavidaviridae</i>	dsDNA	Virus	<0.1	0.1
<i>Malacoherpesviridae</i>	dsDNA	Invertebrados	<0.1	0.1
<i>Asfarviridae</i>	dsDNA	Vertebrados	<0.1	<0.1
<i>Iridoviridae</i>	dsDNA	Vertebrados/Invertebrados	<0.1	0.1
<i>Circoviridae</i>	ssDNA	Vertebrados	<0.1	<0.1
<i>Cruciviridae</i>	ssDNA	Eucariontes	<0.1	<0.1
<i>Herpesviridae</i>	dsDNA	Vertebrados	<0.1	<0.1
<i>Parvoviridae</i>	ssDNA	Vertebrados/Invertebrados	<0.1	<0.1
<i>Bacilladnaviridae</i>	ssDNA	Algas	<0.1	<0.1
<i>Retroviridae</i>	ssRNA	Vertebrados	0.0	0.0

La mayoría tienen genoma de dsDNA aunque está el caso de *Retroviridae*, familia con RNA de cadena sencilla (ssRNA), así como las familias *Microviridae*, *Spiraviridae*, *Circoviridae*, *Cruciviridae*, *Parvoviridae* y *Bacilladnaviridae* con genomas de DNA de cadena sencilla (ssDNA). Debido a que en todas las muestras se extrajo dsDNA, el hecho de que se encuentren este tipo de familias con ssDNA y ssRNA, principalmente se debe a los alineamientos de secuencias, en donde las secuencias identificadas como parte de esas familias muy probablemente tienen

porcentajes de homología que las hacen muy parecidas pero que no necesariamente significa que son esas familias.

Las familias más abundantes son: *Myoviridae*, *Podoviridae* y *Siphoviridae* que pertenecen al orden *Caudovirales*. Lo mismo se ha encontrado en estudios previos (Breitbart *et al.*, 2004; Corinaldesi *et al.*, 2012; Gregory *et al.*, 2019; Taboada *et al.*, 2018), en donde estas mismas tres familias son las más abundantes. Se ha visto que en general este es el consenso, que la mayoría de los virus marinos tienen dsDNA y pertenecen a las familias *Myoviridae*, *Podoviridae* y *Siphoviridae* (Weinbauer, 2004; Wommack & Colwell, 2000). Sin embargo, esto también se ve afectado porque la mayoría de los fagos descritos pertenecen al orden *Caudovirales* (Ackermann, 2009) y además tienen dsDNA el cual resulta más fácil extraer en estudios de metagenómica que no están enfocados en virus, lo cual ha hecho que los virus de ssDNA, de RNA y que infectan arqueas estén muy poco estudiados (Breitbart, 2012).

Dentro de las familias que infectan eucariontes, las que tienen los valores más altos de abundancia promedio tienen como hospederos a protistas (*Pithoviridae*, *Mimiviridae*, *Marseilleviridae*, *Phycodnaviridae* y *Bacilladnaviridae*), mientras que una de ellas (*Lavidaviridae*) infecta virus gigantes y el resto (*Malacoherpesviridae*, *Asfarviridae*, *Iridoviridae*, *Circoviridae*, *Cruciviridae*, *Herpesviridae* y *Retroviridae*), son infectivas para vertebrados e invertebrados (Tabla 2). Las familias *Mimiviridae*, *Marseilleviridae*, *Iridoviridae* y *Phycodnaviridae*, previamente han sido reportadas en diferentes estudios de metagenómica, dentro de las familias de virus que infectan eucariontes, más abundantes en mares y océanos (Castelán-Sánchez *et al.*, 2019; Corinaldesi *et al.*, 2017; Flaviani *et al.*, 2017; Gregory *et al.*, 2019; Nigro *et al.*, 2017; Taboada *et al.*, 2018; Yang *et al.*, 2019).

La familia de virus que infecta eucariontes más abundante fue *Pithoviridae*, esta es la familia a la que pertenecen las partículas virales más grandes encontradas hasta ahora (Legendre *et al.*, 2014) y también se ha encontrado abundante en otros

argumento anterior, ya que debido a que en este estudio se trabajó con muestras que fueron procesadas para el estudio de bacterias, se podría creer que los virus encontrados serían solo aquellos integrados a los genomas bacterianos. Sin embargo, al ver la alta proporción de virus líticos se prueba lo contrario.

En algunos casos, como en las muestras DWH_S1-14, se observa que la proporción de *Podoviridae* en relación con *Siphoviridae* y *Myoviridae*, es mayor al resto de las muestras, en las cuales en general, la proporción de *Podoviridae* es mucho menor que *Myoviridae* y *Siphoviridae* (Fig. 17). En general, se observa que dependiendo de factores como el sitio de muestreo, el tipo de ambiente y la temporada del año en la que se tomó la muestra, entre otros, esta proporción varía (Garin-Fernandez *et al.*, 2018; Hwang *et al.*, 2017). Entonces existen varios motivos que podrían explicar por qué las muestras DWH_S1-14 tienen mayor proporción de *Podoviridae* y entre ellos están que las muestras provienen de estudios y sitios diferentes y que no fueron tomadas en la misma estación del año.

También en la Fig. 17, se observa que aparecen nuevas familias que no estaban presentes en las muestras del CIGoM, tales como, las familias de fagos: *Microviridae*, *Plasmaviridae*, *Spiraviridae*, *Halspiviridae* y las familias de virus que infectan eucariontes: *Malacoherpesviridae*, *Circoviridae*, *Cruciviridae*, *Herpesviridae*, *Parvoviridae*, *Bacilladnaviridae* y *Retroviridae*.

Asimismo, con el objetivo de hacer más fácil la visualización de las familias de fagos menos abundantes, se hizo una gráfica con todas las familias de fagos excepto las familias *Myoviridae*, *Podoviridae* y *Siphoviridae*, ya que estas familias representaban en promedio el 97% (DE=2.3) de cada muestra (Fig. 18). Así, se observa que al omitir a las familias de fagos más abundantes (*Myoviridae*, *Podoviridae* y *Siphoviridae*), los patrones de abundancia de las demás familias de fagos son diferentes a lo largo de las muestras.

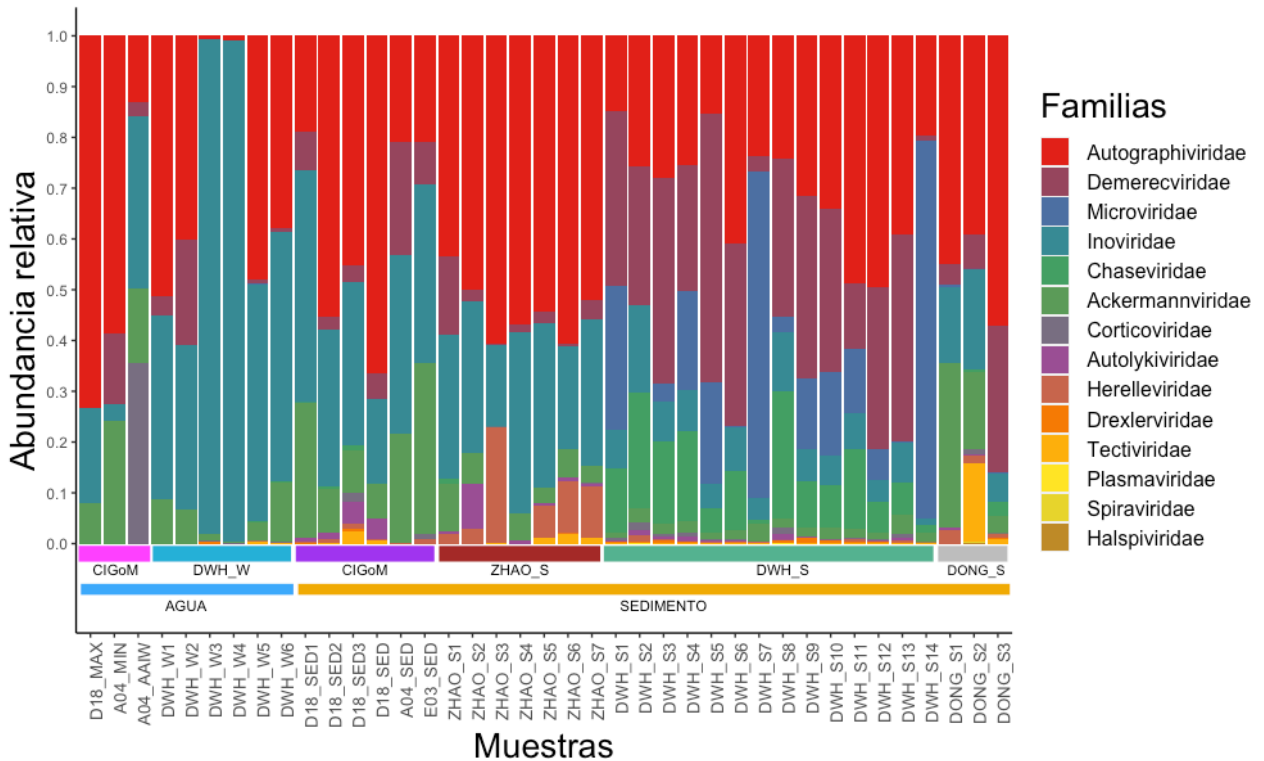


Figura 18. Abundancia relativa de la asignación taxonómica viral a nivel de familias de fagos, excepto *Myoviridae*, *Siphoviridae* y *Podoviridae*. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Asimismo, en la Fig. 18, se observa que omitiendo a las familias *Myoviridae*, *Siphoviridae* y *Podoviridae* la siguiente familia de fagos más abundante es *Autographiviridae*, con una abundancia promedio del 42% (DE= 20). Asimismo, la familia *Demereciviridae*, tiene mayor proporción en las muestras de DWH_S1-14 con una abundancia promedio del 29% (DE=14), en comparación con el resto de las muestras, donde esta familia tiene una abundancia promedio del 6% (DE=7). Lo mismo sucede con la familia *Microviridae*, la cual tiene una abundancia promedio de 19% (DE=23) en las muestras DWH_S1-14, patrón que no se observa en el resto de las muestras (abundancia promedio del 0.04%, DE=0.14) y que en particular en las muestras del CIGoM, esta familia está completamente ausente. De alguna forma, se repite el patrón observado en la Fig. 17, donde las muestras DWH_S1-14 tienen una proporción de familias de fagos diferentes al resto de las muestras. Lo

anterior, se podría deber a que las muestras DWH_S1-14 provienen de un ambiente sumamente alterado, ya que se tomaron durante el derrame del DWH.

Por otro lado, debido a que la mayoría de las secuencias fueron identificadas como fagos (abundancia promedio del 99%, DE=1.2), se realizó una gráfica donde solo se mostraran las familias de virus que infectan eucariontes (Fig. 19). De las 14 familias de virus que infectan eucariontes encontradas, seis corresponden a virus gigantes (*Pithoviridae*, *Mimiviridae*, *Marseilleviridae*, *Phycodnaviridae*, *Asfarviridae* e *Iridoviridae*). Según el grupo de muestras de las que se trate, cambia el patrón de abundancia de familias de virus que infectan eucariontes.

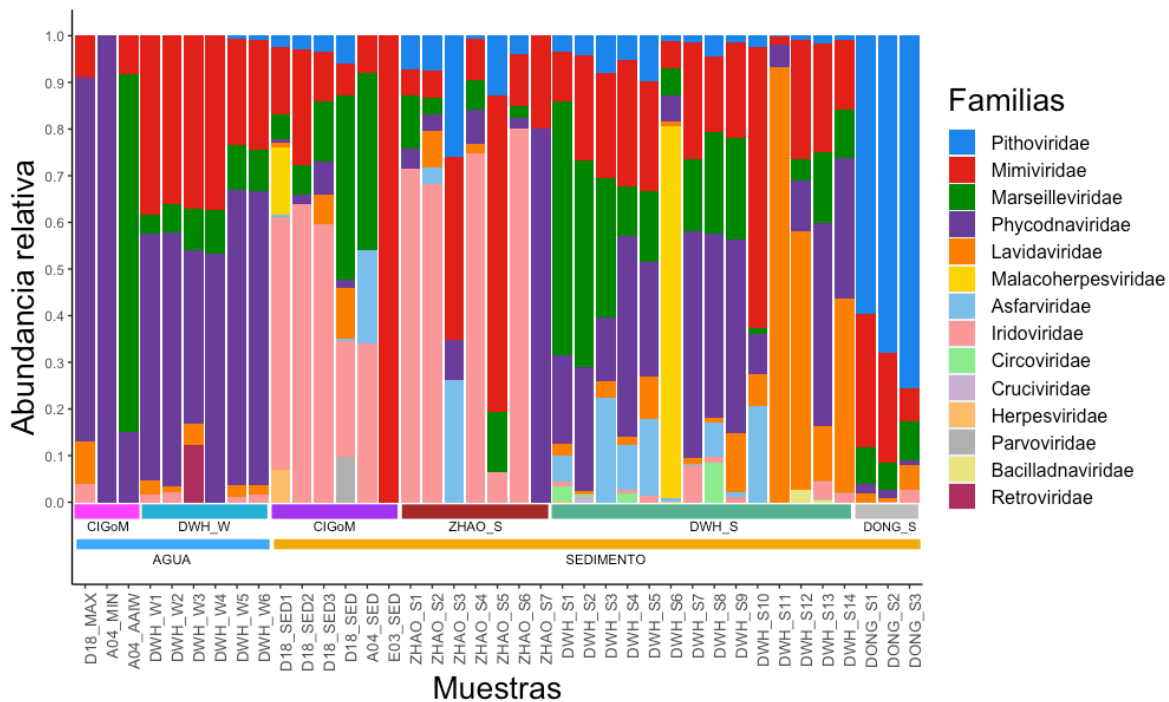


Figura 19. Abundancia relativa de la asignación taxonómica viral a nivel de familias de virus que infectan eucariontes. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

En esta misma figura (Fig. 19), las muestras someras de agua (D18_MAX y A04_MIN), la familia *Phycodnaviridae* es más abundante (abundancia promedio del 89%, DE=16) que en el resto (abundancia promedio del 26%, DE=23), lo cual tiene

sentido ya que esta familia infecta algas y la mayor proporción de algas se encuentra en el agua en la zona fótica o epipelágica (hasta 200 m de profundidad) (Wilson *et al.*, 2009). Aunque también se observa una gran proporción de *Phycodnaviridae*, en muestras más profundas, como A04_AAIW y DWH_W1-6 (1000 m) con una abundancia promedio de 49% (DE=17) y hasta en sedimento en DWH_S1-14 (1500 m) con 26% de abundancia (DE=16). Lo anterior, se ha observado en otros estudios (Gong *et al.*, 2018; Mizuno *et al.*, 2016) y podría explicarse por el flujo continuo de partículas de la superficie que se hunden en lo profundo, por la estabilidad del genoma de *Phycodnaviridae* (dsDNA) o por el hundimiento de los floraciones (*blooms*) de algas senescentes al mar profundo liberando grandes cantidades de partículas virales (Gong *et al.*, 2018; Mizuno *et al.*, 2016; Weinbauer, 2014).

Por otro lado, las muestras de ZHAO_S1-7 tienen un patrón similar a las D18_SED y mientras que en las figuras pasadas, era clara la diferencia del patrón de abundancia para las muestras DWH_S1-14, en esta figura en particular resulta más difuso encontrar una tendencia clara.

Las gráficas anteriores, son una representación visual de los patrones de diversidad viral que existen en las muestras. Es por lo anterior, que se realizaron análisis de diversidad más detallados, los cuales se muestran en la siguiente sección.

7.6 Análisis de diversidad y ecología comparativa

Los análisis de diversidad se hicieron a nivel de especie. Anteriormente se ha reportado que la riqueza viral es mayor en los sedimentos marinos que en la columna de agua adyacente (Breitbart *et al.*, 2004) y con el objetivo de probar ese supuesto, las muestras se separaron según el tipo de sustrato del que provenían: agua o sedimento.

7.6.1 Riqueza, abundancia y alfa diversidad de especies virales

Como se menciona en la metodología, antes de llevar a cabo todos los análisis siguientes, las tablas de abundancia de todas las muestras se normalizaron para contrarrestar las diferencias en profundidad de secuenciación. Dentro de los parámetros básicos de diversidad ecológica, se calcularon la riqueza, abundancia, índice de Shannon e índice de diversidad de Simpson (Tabla A7, Anexo), a partir de los cuales se crearon gráficas separando las muestras según el tipo de sustrato del que provenían. Para poder hacerlo comparable, se obtuvo el porcentaje que representaban el total de secuencias identificadas como virales sobre el total de secuencias válidas por muestra. Así, se observa que la muestra con mayor abundancia de especies virales es A04_AAIW seguida de ZHAO_S4, mientras que la muestra con menor abundancia es ZHAO_S5 (Fig. 20).

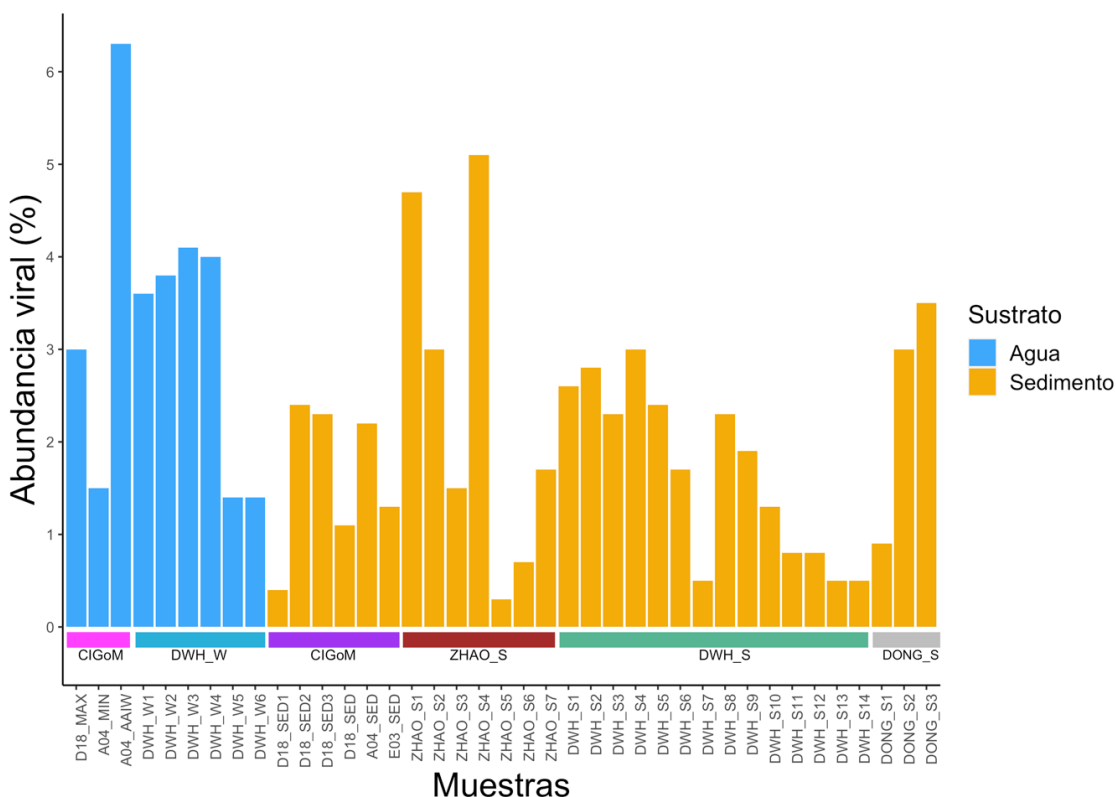


Figura 20. Abundancia de especies virales encontradas por muestra. La abundancia viral se representa como un porcentaje respecto al total de secuencias válidas encontradas por muestra. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Asimismo, se probó si las abundancias entre los dos grupos de agua y sedimento eran significativamente diferente con la prueba de Mann-Whitney-Wilcoxon (Fig. 21). Así se encontró que el agua es significativamente mayor en abundancia de especies virales que el sedimento. Como se observa, la muestra con mayor porcentaje de especies virales es A04_AAIW, donde el 6.3% de sus secuencias válidas son virales.

Aunque depende de cada sitio de muestreo y también del tipo de ambiente acuático del que se trate, en general se ha visto que la abundancia viral varía según la abundancia de procariontes y la concentración de clorofila *a* (Clasen *et al.*, 2008). Lo anterior, se traduce en que en los océanos la abundancia viral, decrece entre más alejado se esté de la costa y entre más profundo se esté en la columna de agua (Suttle, 2007). Así se podría explicar el por qué hay mayor abundancia en las muestras de agua, las cuales tienen profundidades de los 50 a los 1200 m, mientras que las muestras de sedimento rondan entre los 1300 y 3000 m de profundidad. Sin embargo, en la literatura se ha reportado que la abundancia viral en sedimento es mayor que la abundancia viral en la columna de agua adyacente (de 10 a 1000 veces mayor) (Fischer *et al.*, 2003). En otros reportes, se ha visto que los sedimentos de mar profundo y altamente pobres en nutrientes son los que tienen los menores conteos virales (Danovaro *et al.*, 2001). Por lo cual es difícil concluir una tendencia clara sobre la abundancia viral en agua y en sedimento, ya que cada estudio tiene diferencias metodológicas y además pareciera ser que la abundancia viral depende más del ambiente del que se trate que del tipo de sustrato del que provenga (Danovaro *et al.*, 2008).

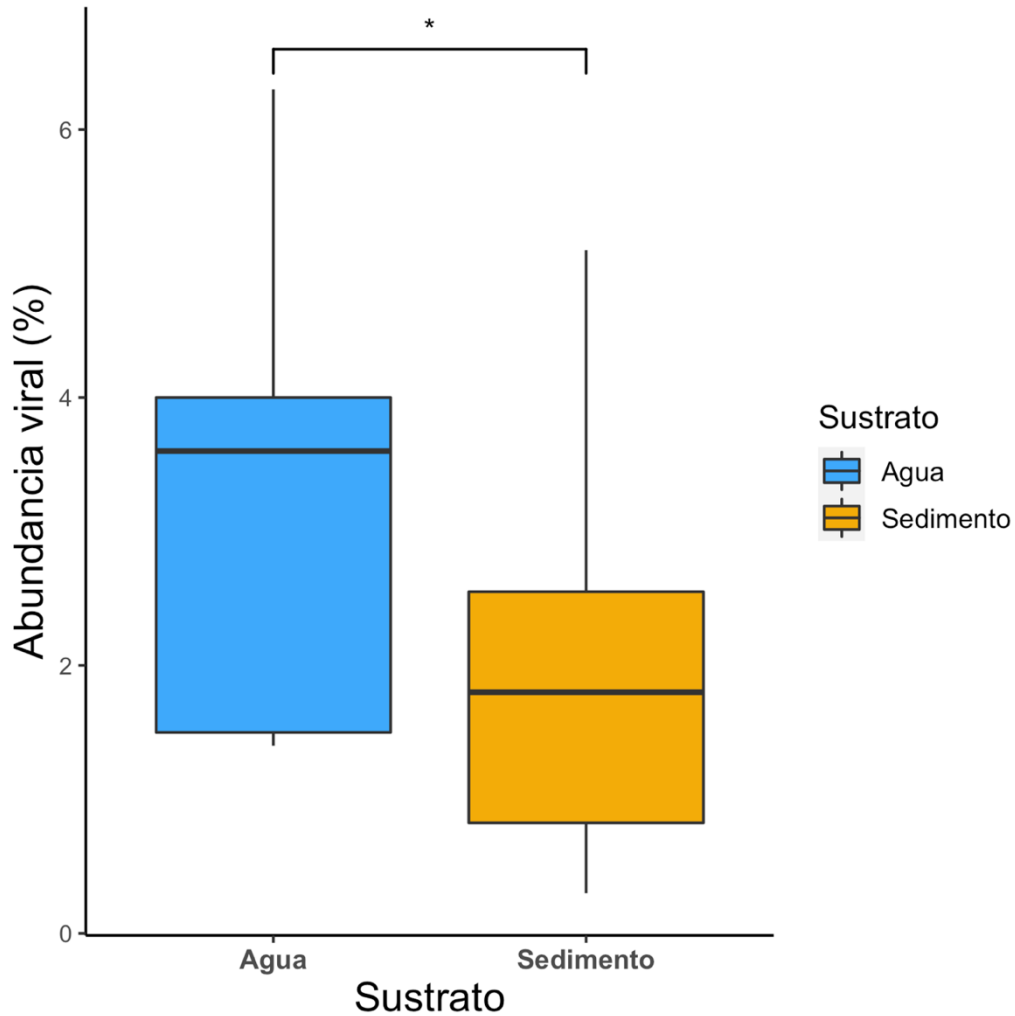


Figura 21. Diferencias en valores abundancia entre las muestras de agua y las de sedimento por medio de la prueba de Mann-Whitney-Wilcoxon. Las muestras de agua son significativamente mayores a las de agua en cuanto valores de riqueza de especies virales. El asterisco (*) denota un valor de $P \leq 0.05$.

La riqueza, por otro lado, se calcula obteniendo el número de especies virales diferentes que hay por muestra y los valores están normalizados según el tamaño de la muestra (Fig. 22). En este caso las muestras con mayor riqueza viral son las de conjunto de DWH_S1-14 y las de menor riqueza son las muestras de agua. Por lo cual se probó si estas diferencias eran estadísticamente significativas con una prueba de Mann-Whitney-Wilcoxon, en donde se encontró que la riqueza viral entre

las muestras de agua y sedimento es significativamente diferente ($P < 0.001$). Como se muestra en la Fig. 23, la riqueza de las muestras de sedimento es mayor a la de las muestras de agua, corroborando lo dicho en Breitbart *et al.*, 2004.

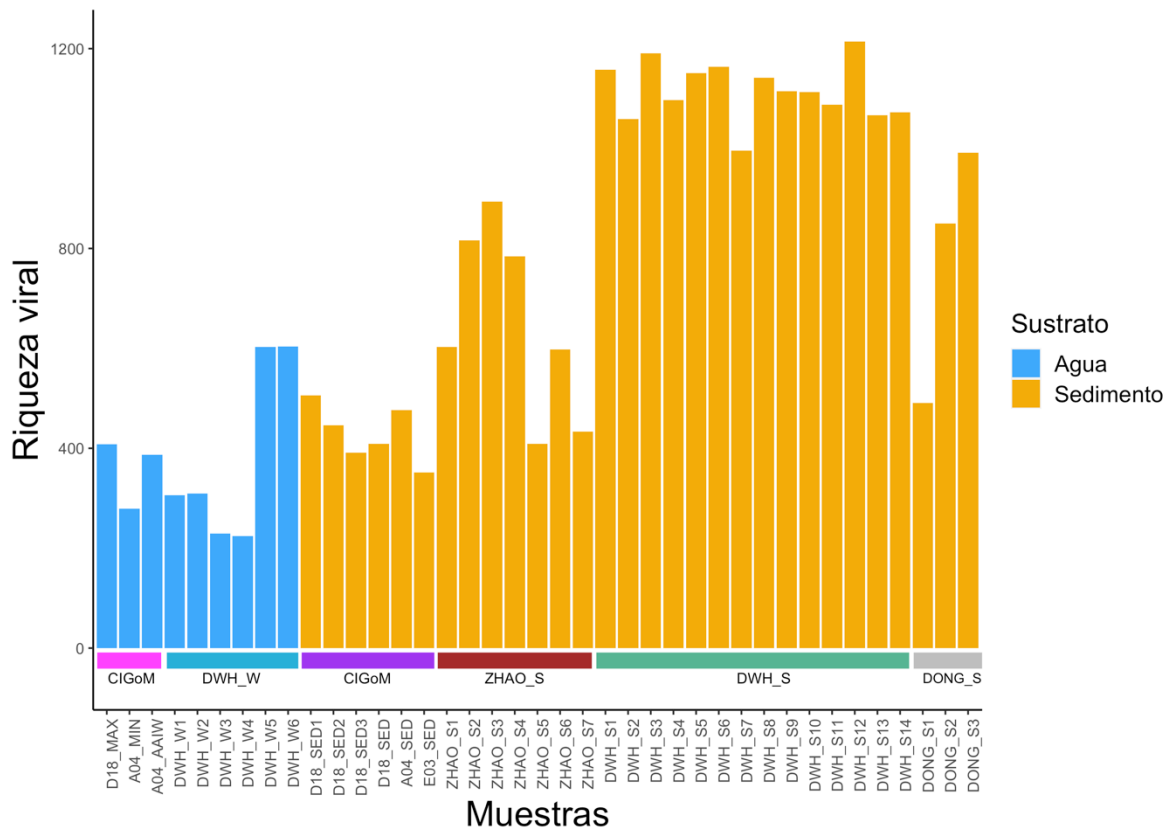


Figura 22. Riqueza de especies virales encontrada por muestra. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

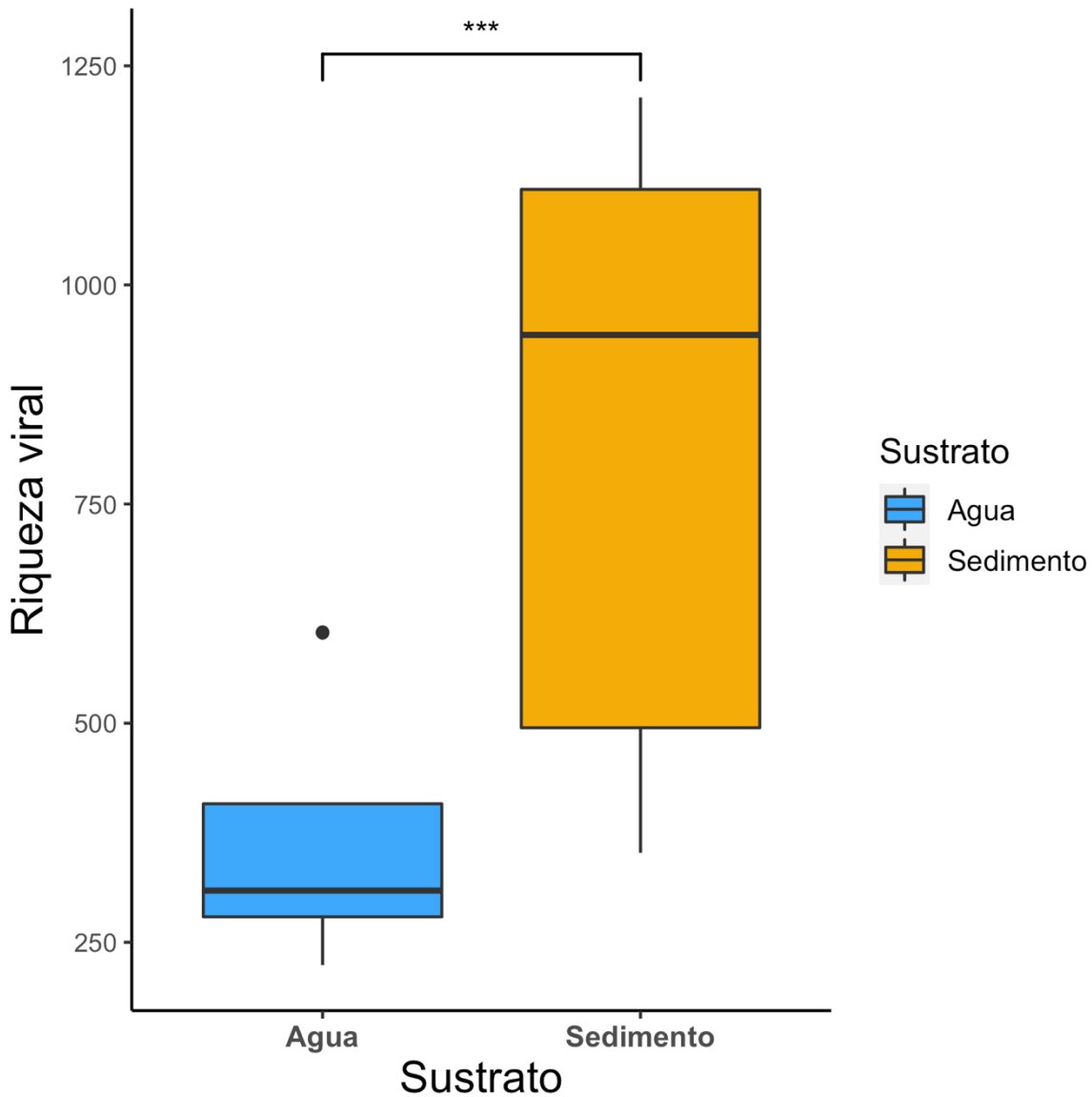


Figura 23. Diferencias en valores de riqueza entre las muestras de agua y las de sedimento por medio de la prueba de Mann-Whitney-Wilcoxon. Las muestras de agua son significativamente mayores a las de agua en cuanto valores de riqueza de especies virales. Los asteriscos (***) denota un valor de $P \leq 0.001$.

El índice de Simpson, indica qué tan homogénea es una muestra, es decir, que todas las especies presentes tengan aproximadamente la misma abundancia. Entre más homogénea una muestra (valores cercano a uno) es más diversa, lo cual a su vez se ve reflejado en el índice de Shannon. Todas las muestras presentan valor cercanos a uno en el índice de Simpson, o sea, son homogéneas (Tabla A7, Anexo).

El índice de Shannon, es uno de los posibles indicadores de alfa diversidad, toma en cuenta la abundancia, la homogeneidad y la riqueza de las especies presentes e indica qué tan diversa es una muestra, entre más el valor sea más alto, la muestra será más diversa. Las muestras de DWH_S son las que presentan los valores alfa diversidad más altos. De manera general, se observa que las muestras de sedimento tienen mayor índice de Shannon que las de agua y esto se ve representado en la Fig. 24.



Figura 24. Valores de alfa diversidad de especies virales denotados por el índice de Shannon de cada muestra. Las muestras están organizadas en el eje x, según el sustrato y el estudio del que provienen.

Las muestras de agua y de sedimento forman dos grupos significativamente diferentes ($P < 0.001$) según su índice de Shannon, es decir, las muestras de sedimento son más diversas que las de agua (Fig. 25).

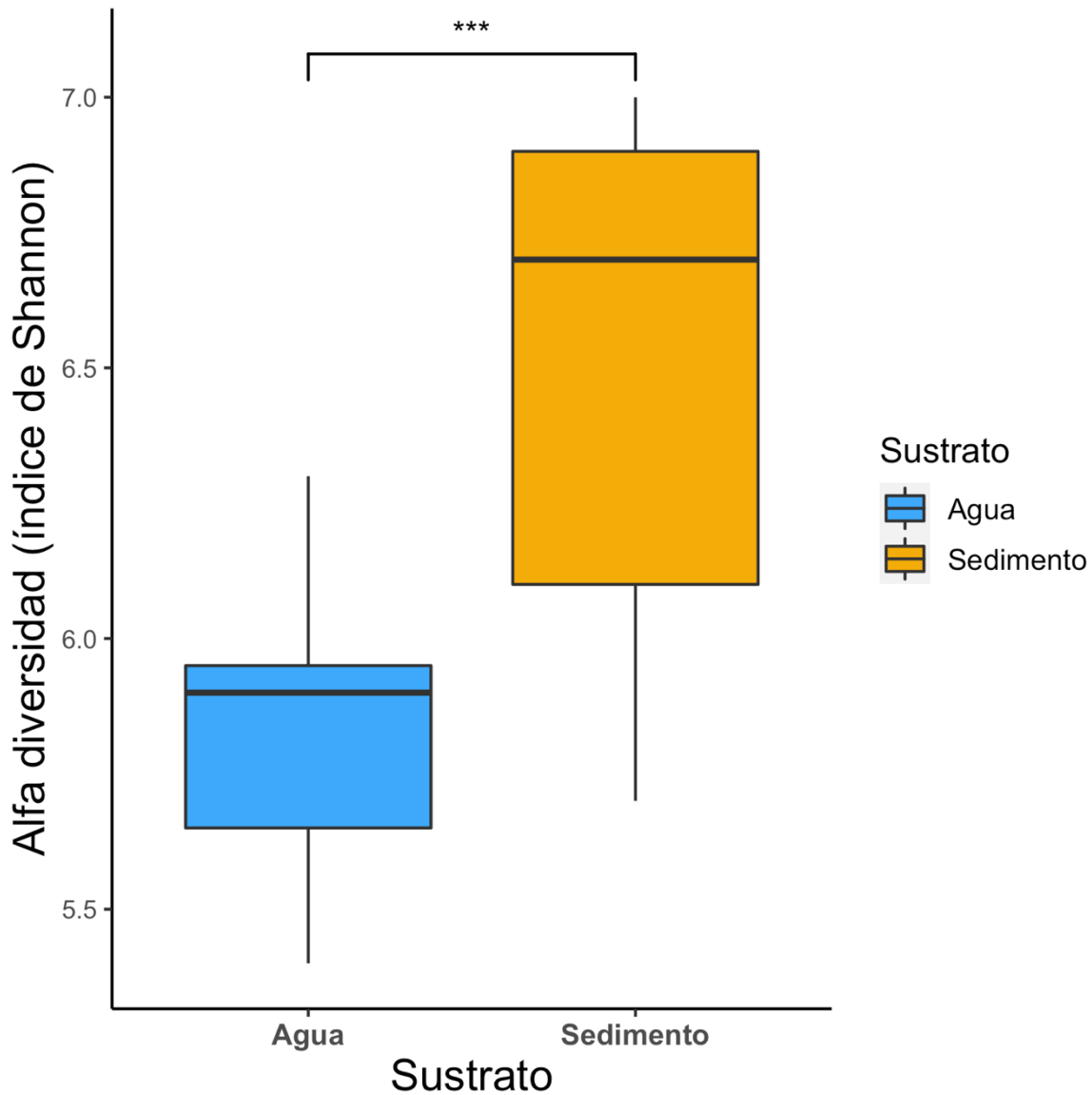


Figura 25. Diferencias de valores de alfa diversidad denotado por el índice de Shannon, entre las muestras de agua y las de sedimento por medio de la prueba de Mann-Whitney-Wilcoxon. Los valores de alfa diversidad (índice de Shannon) de las muestras de sedimento son significativamente mayores a los de las muestras de agua, indicando que las muestras de sedimento son más diversas que las de agua. Los asteriscos (***) denotan un valor de $P \leq 0.001$.

En el mismo estudio de Breitbart *et al.*, 2004, también se reporta que los índices de Shannon son mayores en sedimento que en la columna de agua adyacente. En donde en sedimento los valores van de 9 a 14, mientras que en agua están en el

rango de 7.4-8.1. En otros estudios de metagenómica viral hechos en agua: en la cuenca de agua dulce de Cuatro Ciénegas, México (Taboada *et al.*, 2018), los valores de alfa diversidad están entre 5.9 y 9.1, en un análisis global con muestras de diferentes océanos del mundo (Gregory *et al.*, 2019) los valores rondan entre 5.5 y 7.5, en un muestreo de ventilas hidrotermales en el noreste del Océano Pacífico (Anderson *et al.*, 2011) el rango es de 6.03-8.25 y en otro estudio con muestras de cuatro regiones oceánicas (Angly *et al.*, 2006) los valores están entre 6.05 y 10.8. Los valores de alfa diversidad viral para las muestras presentadas en este estudio están en el rango de 5.4-6.3 y 5.9-7, para las muestras de agua y sedimento respectivamente, por lo cual están dentro de los valores encontrados en los estudios previos alrededor del mundo. Aunque la alfa diversidad de las muestras de sedimento se queda por debajo de los valores de la única referencia de alfa diversidad de metagenómica viral en sedimento (Breitbart *et al.*, 2004).

En cuanto a los valores de alfa diversidad según el grupo de muestra, para las muestras provenientes de agua no se encontraron diferencias significativas en los índices de Shannon de las muestras del CIGoM y las de DWH_W. Mientras que para las muestras de sedimento, se encontró que las muestras de DWH_S son significativamente ($P \leq 0.05$ y $P \leq 0.001$) más diversas que el resto de los otros grupos de muestras (Fig. 26). Lo anterior, podría ser un reflejo de la naturaleza misma de las muestras de DWH_S, ya que fueron tomadas durante un derrame petrolero y como se vio en las figuras anteriores, su diversidad taxonómica viral es diferente al resto de las muestras.

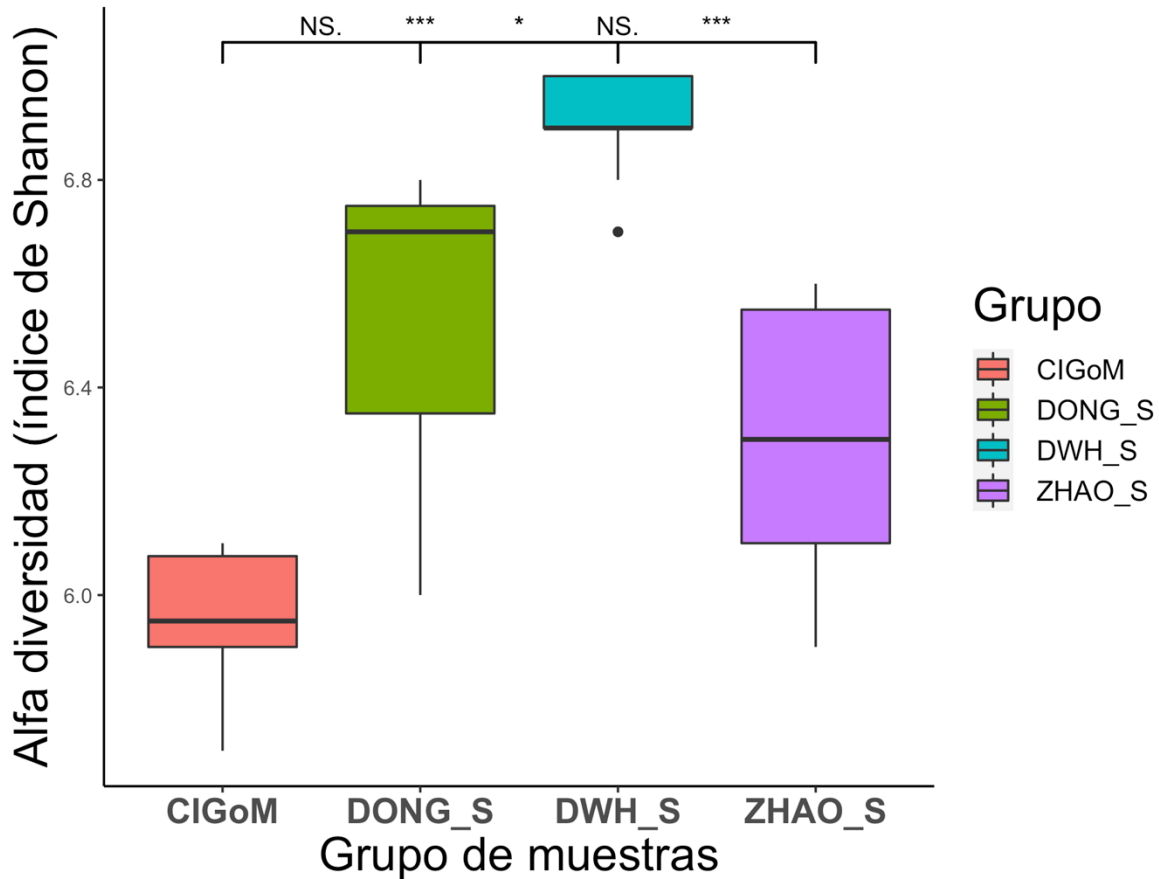


Figura 26. Alfa diversidad representada por el índice de Shannon de los diferentes grupos de muestras provenientes de sedimento. Las muestras de DWH_S tienen una alfa diversidad significativamente mayor al resto de las muestras. Los tres asteriscos (***) denotan un valor de $P \leq 0.001$, un asterisco (*) un valor de $P \leq 0.05$, NS denota una diferencia no significativa.

7.6.2 Análisis de diversidad según el tipo de sustrato

Con base en los resultados anteriores, donde se confirma que las muestras tienen características de diversidad viral diferentes dependiendo del tipo de sustrato del que provienen, se hicieron los siguientes análisis con el objetivo de caracterizar mejor los taxones virales a nivel de género, particulares de cada ambiente.

En la Fig. 27, se muestra la gráfica que representa el análisis hecho para identificar los géneros virales diferencialmente abundantes entre las muestras de agua y

sedimento. Este análisis se representa con la tasa de cambio, es decir, cuántas veces es más abundante un género en una muestra que en la otra. Se encontraron 72 géneros diferencialmente abundantes ($P < 0.05$, valor de P ajustado al número de muestras), sin embargo, con fines ilustrativos, en la Fig. 27 se muestran los géneros encontrados como diferencialmente abundantes en un tamaño de muestra de 50 géneros.

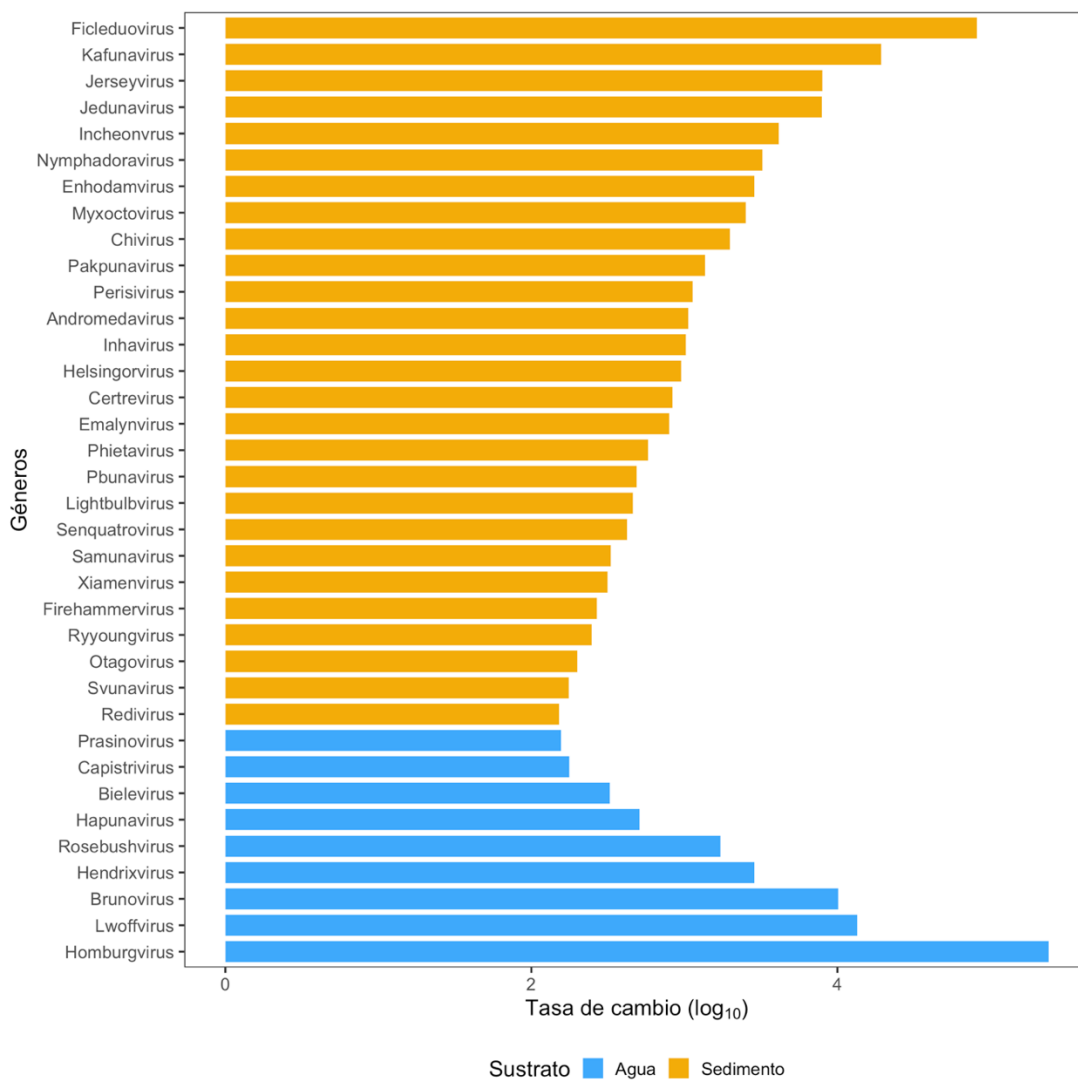


Figura 27. Géneros identificados como diferencialmente abundantes entre las muestras provenientes de agua (azul) y sedimento (amarillo). Las barras representan la abundancia medida en términos de la tasa de cambio, es decir, cuántas veces es más abundante un género en un grupo respecto al otro grupo.

Como se observa hay más géneros diferencialmente más abundantes en sedimento que en agua, entre ellos el género con mayor valor de tasa de cambio es *Ficleduovirus*. Este pertenece a la familia de fagos *Myoviridae*, tienen como hospedero a bacterias del género *Flavobacterium* (Taxonomy Browser, 2021) de la familia *Flavobacteriaceae*, la familia con mayor número de especies acuáticas del filo Bacteroidetes (McBride, 2014). A nivel de especie, se observa lo mismo ya que la especie diferencialmente más abundante en sedimento es *Flavobacterium sp. phage 1/32*, el cual fue caracterizado de agua del mar Báltico (Luhtanen *et al.*, 2014).

En el caso de las muestras de agua, el género con mayor tasa de cambio es *Homburgvirus*, de la familia de fagos *Siphoviridae* y tienen como hospedero a bacterias del género *Listeria* (Taxonomy Browser, 2021), pertenecientes al filo Firmicutes. Dentro de las especies diferencialmente abundantes en agua están *Micromonas pusilla virus 12T* y *Synechococcus phage S-RSM2*, las cuales tienen como hospederos a algas.

Bajo el mismo contexto se buscó saber cuántos géneros están presentes, ausentes o se comparten entre los dos grupos de muestras. Es por eso que se hizo un diagrama de Venn (Fig. 28). Así se observa que el conjunto de agua se encuentra dentro del conjunto de sedimento, es decir, el sedimento contiene todos los géneros presentes en el agua y además tiene 205 géneros que sólo están presentes en sedimento. Lo anterior, concuerda con los valores de riqueza obtenidos previamente, en donde las muestras de sedimento tienen mayor riqueza que las de agua.

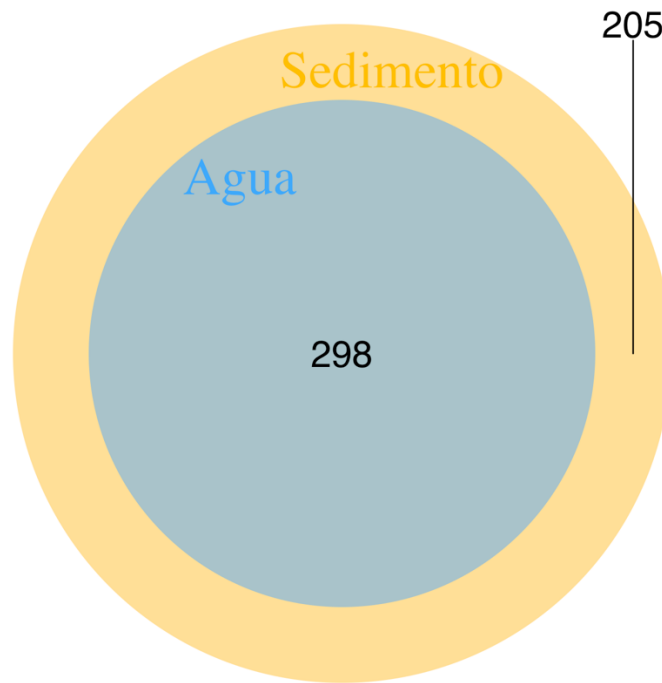


Figura 28. Diagrama de Venn de la comparación de todos los grupos virales a nivel de género presentes en sedimento y en agua. Las muestras de sedimento cuentan con 205 géneros virales que no comparten con las muestras de agua, mientras que todos los taxones virales presentes en agua, también están presentes en sedimento.

7.6.3 Beta diversidad

La beta diversidad permite comparar comunidades diferentes, en términos de diversidad biológica. Con el objetivo de obtener un primer acercamiento sobre las similitudes y diferencias entre las muestras, se hizo una agrupación jerárquica no supervisada a partir de la matriz de distancia a nivel de especie de beta diversidad obtenida usando el coeficiente de correlación de Pearson, ya que este índice reflejaba de la forma más adecuada el origen de las muestras. Esta agrupación se representó con un dendograma circular (Fig. 29).



Figura 29. Dendrograma circular en representación de la agrupación jerárquica de las muestras. Las muestras forman tres grupos principales: A04_MIN como grupo externo (azul), las muestras de agua (azul) y las muestras de sedimento (amarillo).

Como se observa en la Fig. 29, la muestra A04_MIN pareciera formar un grupo externo (*outlier*). Esto podría deberse a que dentro de las muestras del CIGoM, esta fue la muestra con el menor número de secuencias identificadas como conocidas y también obtuvo el menor valor de especies virales encontradas por esfuerzo de muestreo (Fig. A1, Anexo).

En cuanto a las muestras de agua, estas forman un solo grupo que se divide en dos ramas principales, una de ellas contiene a las muestras DWH_W1-4 y la otra a las muestras A04_AAIW y D18_MAX (CIGoM) junto con DWH_W5-6. Mientras que las muestras de sedimento están en dos grupos principales, uno con todas las muestras DWH_S1-14, ZHAO_S2-4 y DONG_S2-3 y otro grupo con DONG_S1, ZHAO_S1,5-7 y todas las muestras de sedimento del CIGoM. Las muestras A04_SED y E03_SED se agrupan con ZHAO_S1, mientras que las demás muestras de sedimento del CIGoM (D18_SED), se agrupan con otras muestras de ZHAO_S5-7 y DONG_S1.

Bajo el mismo contexto, se realizó un Análisis de Coordenadas Principales (PCoA, por sus siglas en inglés), para observar de manera gráfica el agrupamiento de las muestras entre sí (Fig. 30). Se probó agrupar a las muestras por zona, por región del Golfo de México y por profundidad, sin embargo, el único agrupamiento que fue significativo ($P=0.001$) según el análisis de Permanova (Análisis de varianza multivariado permutacional) fue el que se hizo basándose en el tipo de sustrato de las muestras, es decir si las muestras eran de agua o de sedimento.

En las muestras de sedimento se observa que se forman dos grupos principales, uno con todas las muestras de DWH_S y otro con el resto de las muestras de sedimento. Lo que podría indicar que las muestras de DWH_S, tienen patrones de diversidad diferentes al resto de las muestras pero que a su vez comparten algunos grupos virales de forma que aún pueden ser agrupados dentro de las muestras de sedimento. Por otro lado, en las muestras de agua no se observa un comportamiento claro y más bien están dispersas entre sí.

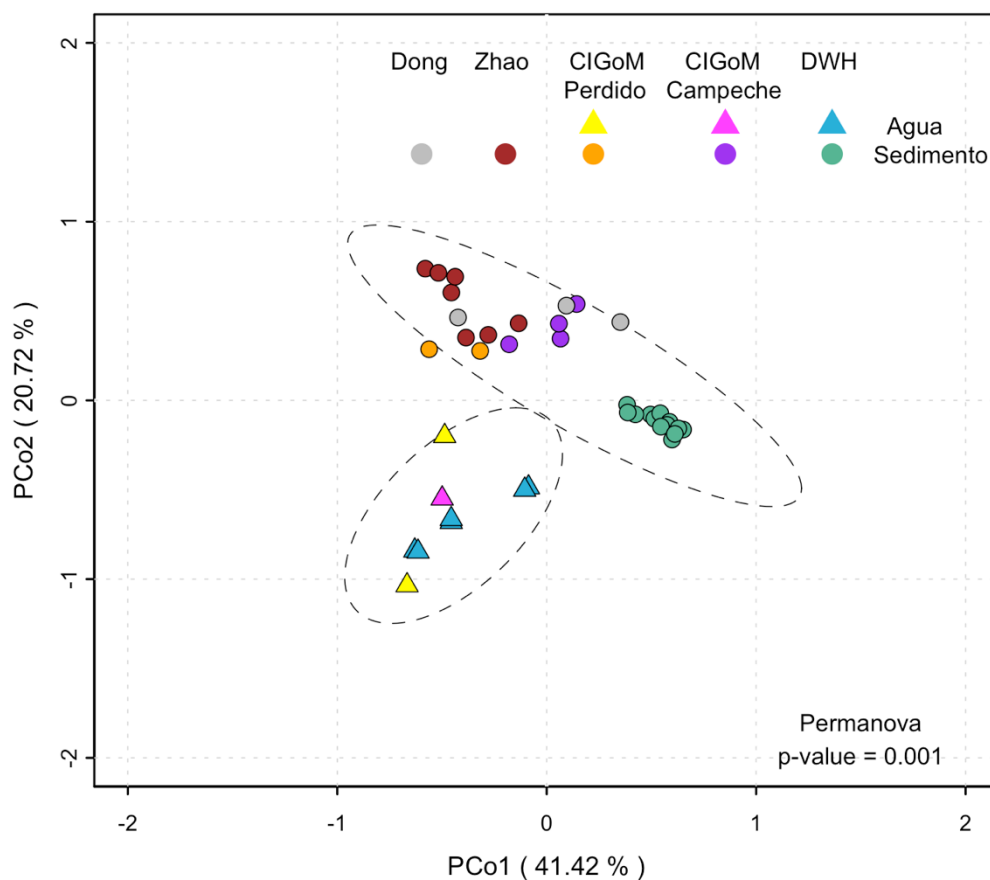


Figura 30. PCoA de la beta-diversidad medida con la distancia del coeficiente de correlación de Pearson estimada de la matriz de abundancia a nivel de especie. El color de cada muestra se dio según el estudio (ZHAO_S, DONG_S, DWH_S, DWH_W y CIGoM) y el sustrato del que provenían (agua o sedimento). Las muestras del CIGoM se dividieron en las zonas de Perdido y Campeche. Las elipses indican un intervalo de confianza igual a 0.95 y representan cada grupo de muestras (agua o sedimento), el valor de Permanova ($P=0.001$) indica las diferencias significativas entre los dos grupos (agua y sedimento).

7.6.4 Análisis de diversidad relacionado con la presencia o ausencia de hidrocarburos

Una de las preguntas que se buscaba responder era si las muestras con presencia de hidrocarburos (confirmada por mediciones o fotografías submarinas) diferían en términos de diversidad en comparación con las muestras control. Los únicos

estudios de los que se tenían muestras con hidrocarburos y muestras control eran el de ZHAO_S y DWH_W.

Entonces se realizó un análisis de correlación en estas muestras, entre la variable de presencia o ausencia de hidrocarburos y el valor de PCo1 (obtenido previamente, Fig. 30) con el fin de encontrar si la variabilidad de las muestras estaba relacionada con que tuvieran o no hidrocarburos. Así se encontró que efectivamente existe una correlación significativa para las muestras de DWH_W ($P < 0.01$ y $R^2 = 0.90$) y para las muestras de ZHAO_S ($P < 0.05$ y $R^2 = 0.67$).

Es por lo anterior que para los casos de DWH_W y ZHAO_S se hicieron los mismos análisis que para las comparaciones entre agua y sedimento, con el objetivo de tener un acercamiento de las diferencias entre muestras provenientes de zonas con presencia o ausencia de hidrocarburos de agua (DWH_W) y sedimento (ZHAO_S) respectivamente.

7.6.4.1 El caso de las muestras de agua del derrame del *Deepwater Horizon* (DWH_W)

Se encontraron 109 géneros diferencialmente abundantes ($P < 0.05$, valor de P ajustado al número de muestras) entre las muestras de DWH_W con presencia (DWH_W_hidrocarburos) y con ausencia de hidrocarburos (DWH_W_control), pero para fines ilustrativos en la Fig. 31 se muestran los géneros encontrados como diferencialmente abundantes en un tamaño de muestra de 50 géneros.

En las muestras de DWH_W_hidrocarburos, el género que tiene el mayor valor de tasa de cambio es *Homburgvirus*, perteneciente a la familia de fagos *Podoviridae*, al igual que en la comparación de géneros diferencialmente abundantes entre agua y sedimento. Asimismo, se hizo el análisis a nivel de especie y la especie que es diferencialmente más abundante en DWH_W_hidrocarburos, es *Bacillus phage vB_BcoS-136*, la cual es una especie sin clasificación dentro de la familia

Siphoviridae (Taxonomy Browser, 2021). Este fago fue originalmente aislado de un lago haloalcalino en Kenia (Akhawale *et al.*, 2019) y no se ha reportado en otros estudios de zonas con hidrocarburos.

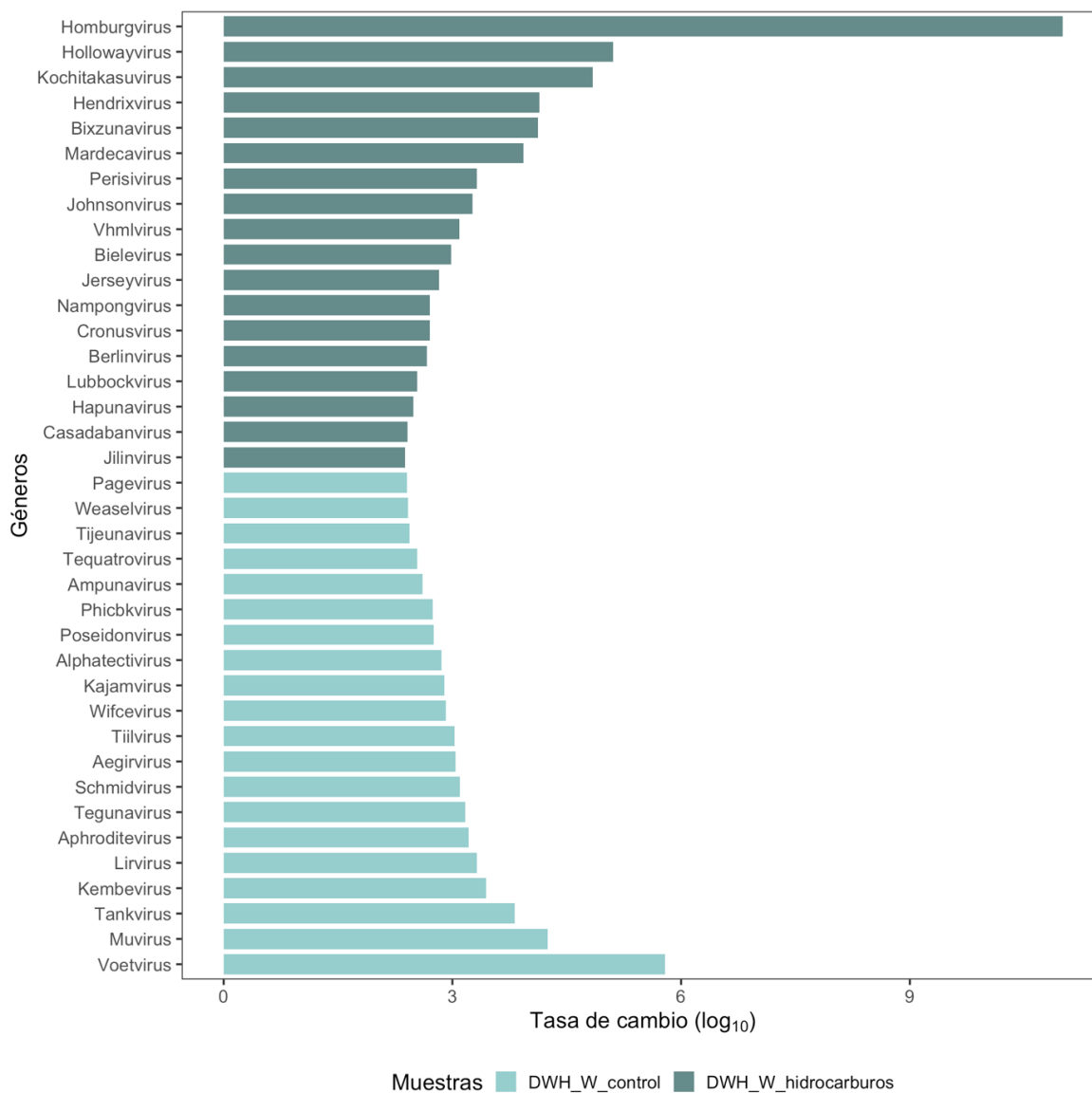


Figura 31. Géneros identificados como diferencialmente abundantes entre las muestras de DWH_W provenientes de zonas con presencia (DWH_W_hidrocarburos) y ausencia (DWH_W_control) de hidrocarburos. Las barras representan la abundancia medida en términos de la tasa de cambio, es decir, cuántas veces es más abundante un género en un grupo respecto al otro grupo.

Mientras que en las muestras de DWH_W_control, el género que es diferencialmente más abundante es Voetvirus, perteneciente a las familias de fagos *Autographiviridae* del orden Caudovirales (Taxonomy Browser, 2021). El género *Voetvirus*, tiene una sola especie, *Synechococcus virus Syn5*, la cual tiene como hospedero a cianobacterias del género *Synechococcus*, ampliamente distribuidas en la capa iluminada superior de los océanos (Mann, 2005). Lo anterior, podría indicar que las muestras de DWH_W_control, tenían mayor cantidad de cianobacterias que las contaminadas con hidrocarburos.

Para saber si habían especies de virus que estuvieran relacionadas con hidrocarburos, se hizo una búsqueda de géneros de bacterias relacionadas con degradación de hidrocarburos (Godoy-Lozano *et al.*, 2018; Raggi *et al.*, 2020; Ramírez *et al.*, 2020) y se vio si dentro de las especies virales encontradas como diferencialmente abundantes había fagos que infectaran a esos géneros de bacterias.

Así se encontró que había fagos diferencialmente más abundantes que infectan bacterias de géneros relacionados con degradación de hidrocarburos, tales como: *Acinetobacter*, *Alteromonas*, *Bacillus*, *Clostridium*, *Cronobacter*, *Erythrobacter*, *Halomonas*, *Marinobacter*, *Pseudomonas*, *Roseobacter*, *Shewanella*, *Synechococcus* y *Vibrio*.

El único fago relacionado con degradación de hidrocarburos, que es diferencialmente más abundante en DWH_W_hidrocarburos que en DWH_W_control, fue *Halomonas phage QHHSV-1* de la familia *Siphoviridae* (Taxonomy Browser, 2021). Este fago fue descubierto en una mina de sal en China e infecta de modo lítico a la bacteria *Halomonas ventosae* QH52-2, la cual ha sido aislada de ambientes de mar profundo en consorcios de degradación de PAH's (Fu *et al.*, 2017; Wang *et al.*, 2008). Además, el género *Halomonas*, pertenece al orden Oceanospirillales, el cual era abundante en estas muestras (Mason *et al.*, 2012).

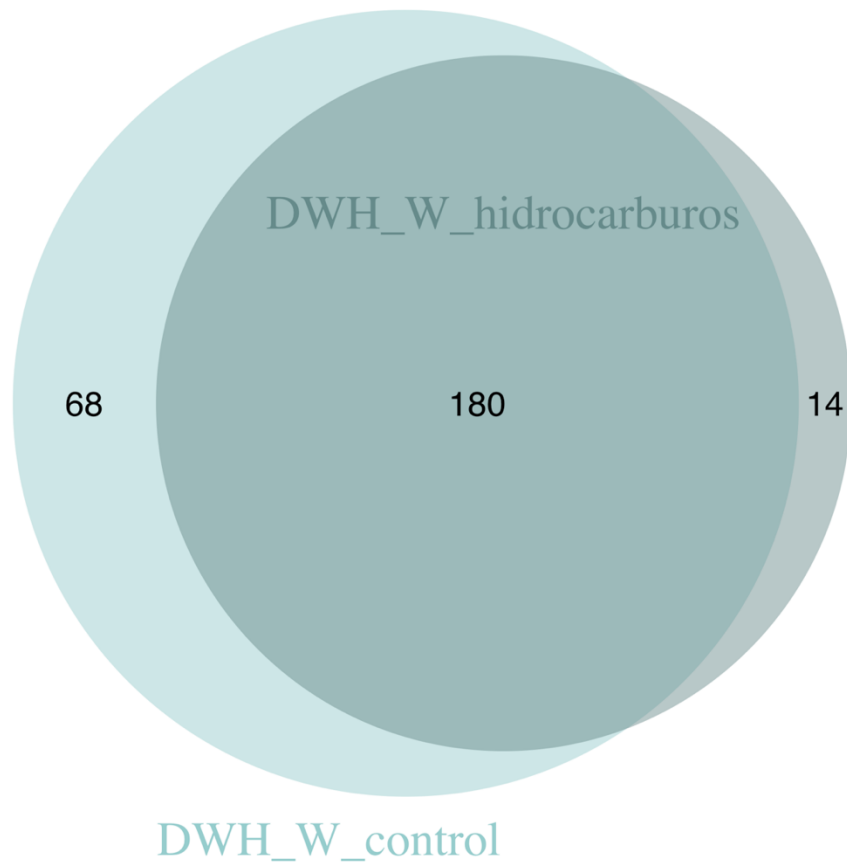


Figura 32. Diagrama de Venn de la comparación de todos los grupos virales a nivel de género presentes en las muestras de DWH_W_control y DWH_H_hidrocarburos. Las muestras de DWH_W_control cuentan con 68 géneros virales que no están en las muestras de DWH_H_hidrocarburos y a su vez, estas últimas tienen 14 géneros virales que no comparten con las muestras control.

También se hizo un diagrama de Venn (Fig. 32) con estas muestras para representar el número de géneros que se comparten o son propios de cada grupo de muestras. Los dos grupos de muestras comparten 181 géneros y en el caso de las muestras DWH_W_control, estas tienen 68 géneros que no están en las muestras con hidrocarburos y a su vez, estas últimas tienen 14 géneros que no se encuentran en las muestras control. El hecho de que las muestras DWH_W_control tengan más géneros propios, es un reflejo de que estas muestras tienen mayor

riqueza que las muestras DWH_W_hidrocarburos, esto probablemente debido a la contaminación por hidrocarburos.

7.6.4.2 El caso de las muestras de sedimento del noroeste del GoM (ZHAO_S)

En el caso del análisis de géneros diferencialmente abundantes para las muestras de ZHAO_S, se encontraron 47 géneros diferencialmente abundantes ($P < 0.05$, valor de P ajustado al número de muestras) de los cuales se representan en su totalidad en la Fig. 33.

Así, en las muestras ZHAO_S_hidrocarburos el género con la mayor tasa de cambio es *Nymphadoravirus*, de la familia de fagos *Siphoviridae* que tiene como hospedero a bacterias del género *Gordonia* (Taxonomy Browser, 2021). Mientras que para las muestras ZHAO_S_control el género con mayor tasa de cambio es *Samunavirus*, también de la familia *Siphoviridae* y que tiene como hospedero a bacterias del género *Pseudomonas* (Taxonomy Browser, 2021).

A nivel de especie, en ZHAO_S_hidrocarburos, la especie diferencialmente más abundante es *ANMV-1 virus*, un virus que infecta arqueas (Taxonomy Browser, 2021) y fue encontrado por primera vez en sedimentos marinos con filtraciones de metano en California (Paul *et al.*, 2015). Lo que a su vez, coincide con lo reportado en Zhao *et al.*, 2020, en donde encuentran que en las muestras con hidrocarburos predominan las arqueas. Mientras que en ZHAO_S_control, es la especie *Iridovirus LCIVAC01*, de la familia de virus gigantes *Iridoviridae* (Taxonomy Browser, 2021). Este virus fue encontrado en sedimento de mar profundo en ventilas hidrotermales del Castillo de Loki (Bäckström *et al.*, 2019).

Se encontraron fagos que infectan bacterias de géneros relacionados con degradación de hidrocarburos, tales como: *Acinetobacter*, *Alteromonas*, *Bacillus*, *Clostridium*, *Cronobacter*, *Erythrobacter*, *Marinobacter*, *Pseudomonas*,

Roseobacter, *Ruegeria*, *Shewanella*, *Synechococcus* y *Vibrio*. Los únicos fagos que infectan bacterias relacionadas con degradación de hidrocarburos y que son diferencialmente más abundantes en ZHAO_S_hidrocarburos que en ZHAO_S_control, fueron fagos que infectan a bacterias del género *Clostridium* y al género *Ruegeria*.

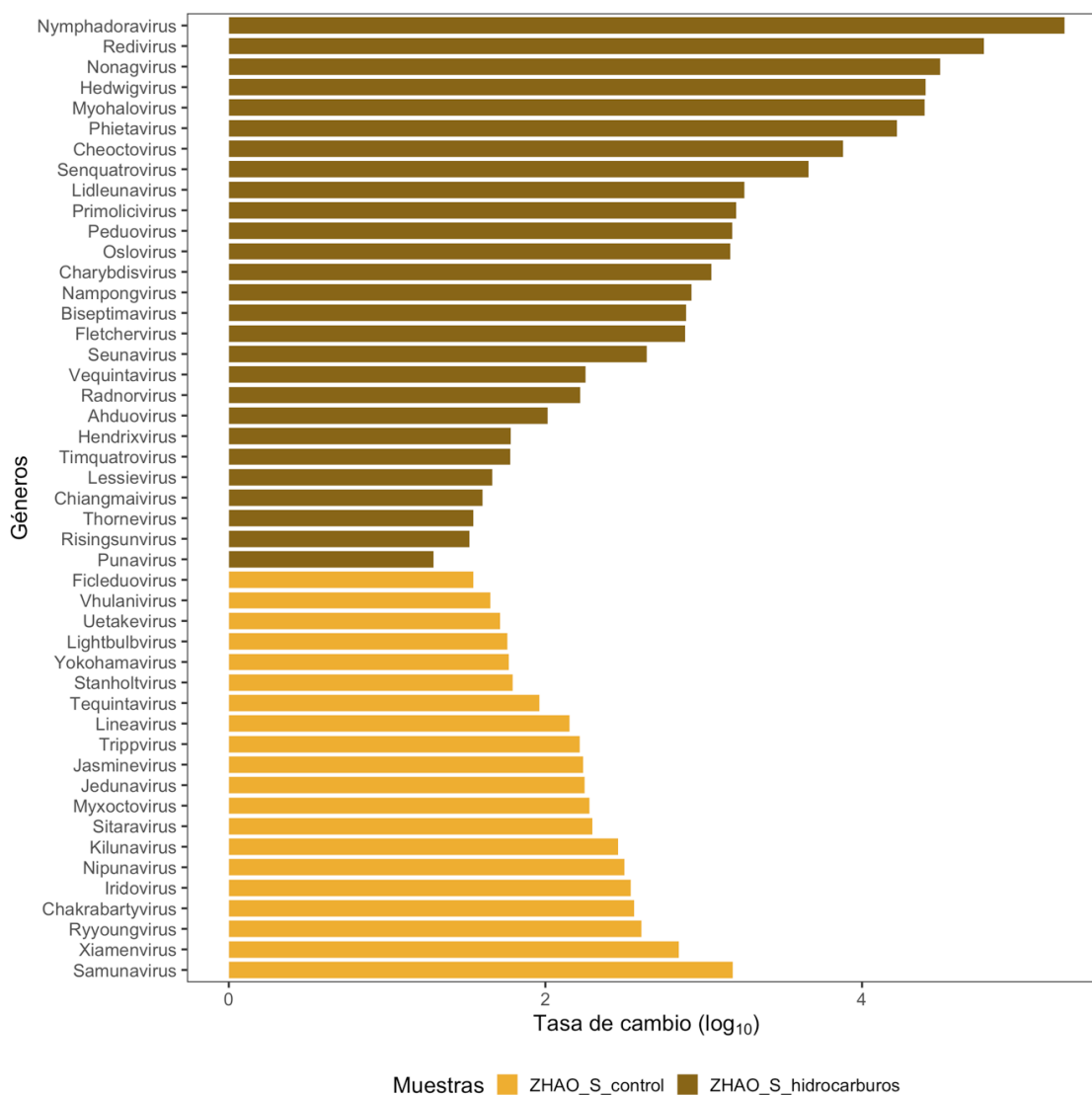


Figura 33. Géneros identificados como diferencialmente abundantes entre las muestras de ZHAO_S provenientes de zonas con presencia (ZHAO_S_hidrocarburos) y ausencia (ZHAO_S_hidrocarburos) de hidrocarburos. Las barras representan la abundancia medida en términos de la tasa de cambio, es decir, cuántas veces es más abundante un género en un grupo respecto al otro grupo.

En específico, se encontró que *Ruegeria phage vB_RpoP-V13*, era diferencialmente más abundante en las muestras con hidrocarburos, pertenece a la familia *Podoviridae* e infecta a la bacteria *Ruegeria pomeroyi* DSS-3. Esta bacteria es una alfa-proteobacteria del clado roseobacter, aislada de agua de mar, que es capaz de degradar compuestos orgánicos de azufre como el DMSP (dimetilsulfoniopropionato) (González et al., 1999; Roseobase, 2021).

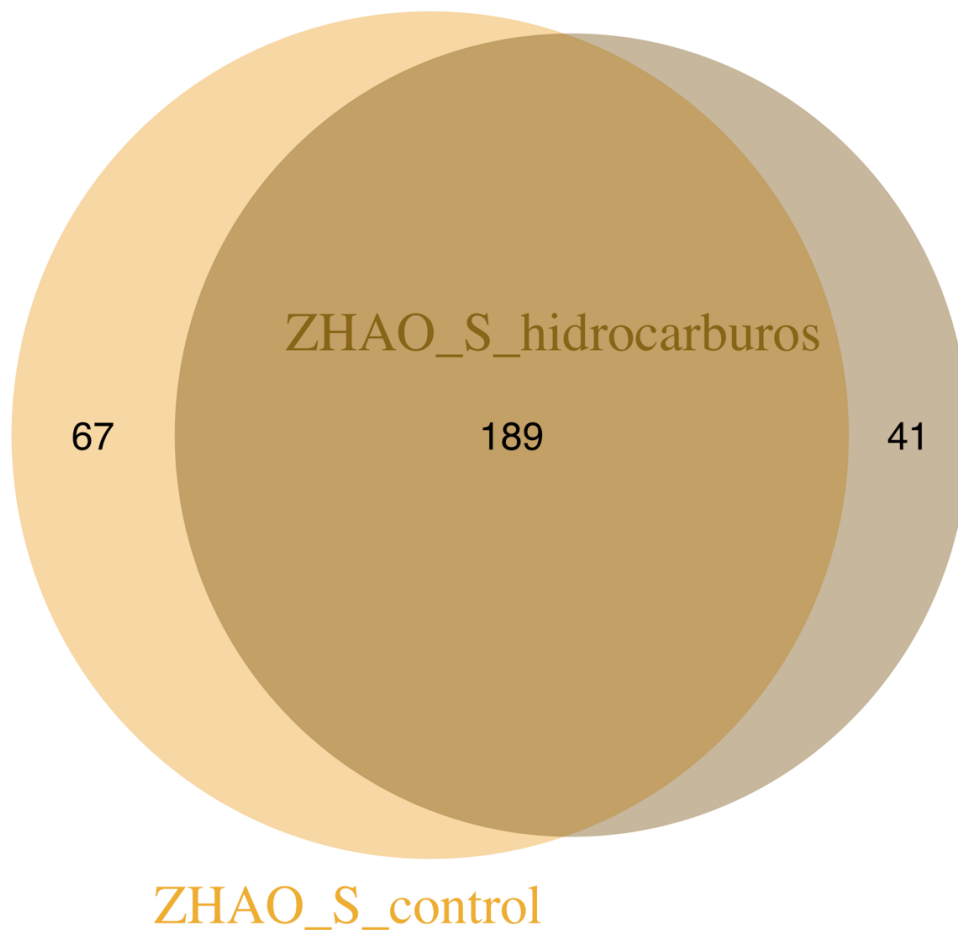


Figura 34. Diagrama de Venn de la comparación de todos los grupos virales a nivel de género presentes en las muestras de ZHAO_S_control y ZHAO_S_hidrocarburos. Las muestras de ZHAO_S_control cuentan con 67 géneros virales presentes que no están en las muestras de ZHAO_S_hidrocarburos y a su vez, estas últimas tienen 41 géneros virales que no comparten con las muestras control.

En el diagrama de Venn (Fig. 34) de estas muestras, se observa que entre los grupos ZHAO_S_control y ZHAO_S_hidrocarburos se comparten 189 géneros. Las muestras control cuentan con 67 géneros que no están presentes en las muestras con hidrocarburos y estas últimas tienen 41 géneros que no están presentes en las muestras control. Lo cual coincide con lo obtenido en términos de riqueza y alfa diversidad, donde las muestras ZHAO_S_control tienen, en general, valores más altos que las muestras ZHAO_S_hidrocarburos.

8 CONCLUSIONES

En este estudio, se seleccionaron diversas muestras de agua y sedimento del Golfo de México y por métodos de bioinformática se analizó la diversidad viral de las mismas. Así se encontró que las muestras formaban grupos separados dependiendo del sustrato del que provenían (agua o sedimento), siendo las muestras de sedimento más diversas y con mayor riqueza viral que las muestras de agua.

Dentro de la diversidad viral se encontró que en general en todas las muestras, las familias de fagos más abundantes en el GoM son *Myoviridae*, *Siphoviridae* y *Podoviridae*. Asimismo, se encontraron familias de virus que infectan eucariontes, en su mayoría pertenecientes a familias de virus gigantes como *Pithoviridae*, *Mimiviridae* y *Phycodnaviridae*.

También se vio que dentro de los estudios donde se tienen muestras control y muestras con hidrocarburos, tal como lo son las muestras de ZHAO_S y DWH_W, existe correlación significativa entre la variabilidad de las muestras y la presencia o ausencia de hidrocarburos. Así, se encontró que existen especies virales que son diferencialmente más abundantes en las muestras con hidrocarburos que en las muestras control y que además son virus relacionados con zonas con hidrocarburos o con ambientes extremos, como ambientes halofílicos o de sedimento profundo. Al

mismo tiempo, parece ser que en estos dos casos, la riqueza y alfa diversidad viral es menor en las muestras con hidrocarburos que en las muestras control. Lo anterior indica que las comunidades virales, de fagos en específico, al igual que las de procariontes, son un reflejo de las condiciones de un ecosistema dado, en este caso, ecosistemas perturbados con hidrocarburos.

Este trabajo es el primero estudio de metagenómica viral del GoM, por lo cual servirá como comparación y referencia para investigaciones futuras en el tema y como catalizador del estudio de la diversidad viral del GoM.

Perspectivas

Se debe hacer un análisis que relacione directamente a las poblaciones de fagos con las de bacterias en las muestras del GoM y así por medio de análisis de abundancias y de diversidad genética viral, poder inferir qué tipo de ciclo (lítico o lisogénico) predomina. También es importante reconstruir genomas virales nuevos a partir de las secuencias no identificadas y ligarlos con posibles hospederos, con el objetivo de reducir la proporción de secuencias desconocidas en las muestras. Además, se deben realizar estudios longitudinales con muestreos a través del tiempo en un solo sitio para ver cómo cambian las poblaciones virales en el GoM, comparando zonas con y sin hidrocarburos. Lo que a su vez lleva al punto de ampliar las muestras a sitios alrededor del mundo que tengan hidrocarburos, para poder obtener un vistazo de la tendencia en general de las poblaciones virales que existen en estas zonas.

Referencias

1. **Ackermann H.W.** (2009). Phage classification and characterization. *Methods in Molecular Biology*, **501**:127-140.
2. **Aherfi S., Colson P., La Scola B. & Raoult D.** (2016). Giant Viruses of Amoebas: An Update. *Frontiers in Microbiology*, **7**:349:1-14.
3. **Akhwale J.K., Rohde M., Rohde C., Bunk B., Spröer C., Boga H.I., Klenk H.P. & Wittmann J.** (2019). Isolation, characterization and analysis of

- bacteriophages from the haloalkaline lake Elmenteita, Kenya. *PLOS ONE*, **14:4**:1-19.
4. **Anderson R.E., Brazelton W.J. & Baross J.A.** (2011). Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *Federation of European Microbiological Societies*, **77**: 120-133.
 5. **Angly F.E., Felts B., Breitbart M., Salamon P., Edwards R.A., Carlson C., Chan A.M., Haynes M., Kelley S., Liu H., Mahaffy J.M., Mueller J.E., Nulton J., Olson R., Parsons R., Rayhawk S., Suttle C.A. & Rohwer F.** (2006). The marine viromes of four oceanic regions. *PLOS Biology*, **4**:2121-2131.
 6. **Andrews S.** (2010). FastQC: a quality control tool for high throughput sequence data. Disponible en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 7. **Bäckström D., Yutin N., Jørgensen S.L., Dharamshi J., Homa F., Zaremba-Niedwiedzka K., Spang A., Wolf Y.I., Koonin E.V. & Ettema T.J.G.** (2019). Virus Genomes from Deep Sea Sediments Expand the Ocean Megavirome and Support Independent Origins of Viral Gigantism. *American Society for Microbiology*, **10:2**:e02497-18.
 8. **Becker R.A. & Wilks A.R.** (2018). mapData: Extra Map Databases. R package version 2.3.0. Disponible en: <https://CRAN.R-project.org/package=mapdata>
 9. **Bergh Ø., Børsheim K.Y., Bratbak G. & Heldal M.** (1989). High abundance of viruses found in aquatic environments. *Nature*, **340**:467-68.
 10. **Breitbart M., Felts B., Kelley S., Mahaffy J.M., Nulton J., Salamon P. & Rohwer F.** (2004). Diversity and population structure of a near-shore marine-sediment viral community. *The Royal Society*, **271**:565-574.
 11. **Breitbart M.** (2012). Marine Viruses: Truth or Dare. *Annual reviews of Marine Science*, **4**:425-448.
 12. **Buchfink B., Xie C. & Huson D.H.** (2015). The DIAMOND protein aligner. Disponible en: <https://github.com/bbuchfink/diamond>
 13. **Bushnell B., Rood J. & Singer E.** (2017). BBtools. Disponible en: <http://www.sourceforge.net/projects/bbmap/>
 14. **Castelán-Sánchez H.G., López-Rosas I., García-Suastegui W.A., Peralta R., Dobson A.D.W., Batista-García R.A. & Dávila-Ramos S.** (2019). Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. *Marine Genomics*, **46**:16-28.
 15. **CIGoM** (2021, 5 de marzo). <https://cigom.org/>
 16. **Clasen J.L., Bridgen S.M., Payet J.P. & Suttle C.A.** (2008). Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshwater Biology*, **53**:1090-110.
 17. **Cochlan W.P., Wikner J., Steward G.F., Smith D.C. & Azam F.** (1993). Spatial distribution of viruses, bacteria and chlorophyll *a* in neritic, oceanic and estuarine environments. *Marine Ecology Progress Series*, **97**:77-87.
 18. **Corinaldesi C., Tangherlini M. & Dell'Anno A.** (2017). From virus isolation to metagenome generation for investigating viral diversity in deep-sea sediments. *Scientific Reports*, **7**:8355-8367.

19. **Coutinho F.H., Silveira C.B., Gregoracci G.B., Thompson C.C., Edwards R.A., Brussaard C.P.D., Dutilh B.E. & Thompson F.L.** (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications*, **8**:15955-15967.
20. **Coutinho F.H., Rosselli R. & Rodriguez-Valera F.** (2019). Trends of Microdiversity Reveal Depth-Dependent Evolutionary Strategies of Viruses in the Mediterranean. *American Society for Microbiology*, **4**:6:e00554-19.
21. **Danovaro R., Manini E. & Dell'Anno A.** (2001). Higher Abundance of Bacteria than of Viruses in Deep Mediterranean Sediments. *Applied and Environmental Microbiology*, **68**:3:1468-1472.
22. **Danovaro R., Corinaldesi C., Filippini M., Fischer U.R., Gessner M.O., Jacquet S., Magagnini M. & Velimirov B.** (2008). Viriobenthos in freshwater and marine sediments: a review. *Freshwater Biology*, **53**:1186-1213.
23. **Dong X., Greening C., Rattray J.E., Chakraborty A., Chuvochina M., Mayumi D., Dolfig J., Li C., Brooks J.M., Bernard B.B., Groves R.A., Lewis I.A. & Hubert C.R.J.** (2019). Metabolic potential of uncultured bacteria and archaea associated with petroleum seepage in deep-sea sediments. *Nature Communications*, **10**:1816:1-12.
24. **Dunnington D.** (2017). prettymapr: Scale Bar, North Arrow, and Pretty Margins in R. R package version 0.2.2 Disponible en: <https://CRAN.R-project.org/package=prettymapr>
25. **Dusa A.** (2020). venn-package: Draw Venn Diagrams. *R package version 1.9*. Disponible en: <https://cran.r-project.org/web/packages/venn/venn.pdf>
26. **ENVGEN** (2021, 4 de marzo). SciLifeLab. <http://envgen.github.io/metagenomics>
27. **Fischer U.R., Wieltsching C., Kirschner A.K.T. & Velimirov B.** (2003) Does Virus-Induced Lysis Contribute Significantly to Bacterial Mortality in the Oxygenated Sediment Layer of Shallow Oxbow Lakes?. *Applied and Environmental Microbiology*, **69**:9:5281-5289.
28. **Flaviani F., Schroeder D.C., Balestreri C., Schroeder J.L., Moore K., Paszkiewicz K., Pfaff M.C. & Rybicki E.P.** (2017). A Pelagic Microbiome (Virus to Protists from a Small Cup of Seawater. *Viruses*, **9**(3):47:1-19.
29. **Fu C., Zhao Q., Li Z., Wang Y., Zhang S., Lai Y., Xiao W. & Cui X.** (2017). Complete genome sequence of *Halomonas ventosae* virulent halovirus QHHSV-1. *Archives of Virology*, **162**:3215-3219.
30. **Gage J.D. & Tyler P.A.** (1991). Deep-sea biology: A natural history of organisms at the deep-sea floor. *Cambridge University Press*, 504 p.
31. **Glenn T.C.** (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**:759-769.
32. **Gregory A.C., Zayed A.A., Conceição-Neto N., Temperton B., Bolduc B., Alberti A., Ardyna M., Arkhipova K., Carmichael M., Cruaud C., Dimier C., Dominguez-Huerta G., Ferland J., Kandels S., Liu Y., Marec C., Pesant S., Picheral M., Pisarev S.,... Sullivan M.B.** (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, **177**:1-15.
33. **Guenther S., Huwyler D., Richard S. & Loessner M.J.** (2009). Virulent Bacteriophage for Efficient Biocontrol of *Listeria monocytogenes* in Ready-To-Eat Foods. *Applied and Environmental Microbiology*, **75**:1:93-100.

34. **Gurevich A., Saveliev V., Vyahhi N. & Tesler G.** (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics*, **29:8**:1072-1075.
35. **Guzman-Vega M.A & Mello M.R.** (1999). Origin of Oil in the Sureste Basin, Mexico. *American Association of Petroleum Geologists Bulletin*, **83:7**:1068-1094.
36. **Godoy-Lozano E.E., Escobar-Zepeda A., Raggi L., Merino E., Gutierrez-Rios R.M., Juarez K., Segovia L., Licea-Navarro A.F., Gracia A., Sanchez-Flores A. & Pardo-Lopez L.** (2018). Bacterial diversity and the geochemical landscape in the Southwestern Gulf of Mexico. *Frontiers in Microbiology*, **9:2528**:1-15.
37. **Gong Z., Liang Y., Wang M., Jiang Y., Yang Q., Xia J., Zhou X., You S., Gao C., Wang J., He J., Shao H. & McMinn A.** (2018). Viral Diversity and Its Relationship With Environmental Factors at the Surface and Deep Sea of Prydz Bay, Antarctica. *Frontiers in Microbiology*, **9:2981**:1-17.
38. **González J.M., Kiene R.P. & Moran M.A.** (1999). Transformation of Sulfur Compounds by an Abundant Lineage of Marine Bacteria in the α -Subclass of the Class *Proteobacteria*. *Applied and Environmental Microbiology*, **65:9**:3810-3819.
39. **Haro-Moreno J.M., López-Pérez M., R. de la Torre J., Picazo A., Camacho A. & Rodriguez-Valera F.** (2018). Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome*, **6**:128.
40. **Handelsman J.** (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, **68**:669-685.
41. **Head I. M., Jones D.M. & Røling W.F.M.** (2006). Marine microorganisms make a meal of oil. *Nature Reviews*, **4**:173-182.
42. **Hijmans R.J.** (2020). raster: Geographic Data Analysis and Modeling. R package version 3.3-13. Disponible en: <https://CRAN.R-project.org/package=raster>
43. **Huson D., Auch A. F., Qi J. & Schuster S. C.** (2007). MEGAN analysis of metagenomic data. Disponible en: <https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/megan6/>
44. **Hyatt D., Chen G., LoCascio P.F., Land M.L., Larimer F.W. & Hauser L.J.** (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. Disponible en: <https://github.com/hyattpd/Prodigal>
45. **Kornienko M., Kupstov N., Gorodnichev R., Bespiatykh D., Guliaev A., Letarova M., Kulikov E., Veselovsky V., Malakhova M., Letarov A., Ilina E. & Shitikov E.** (2020). Contribution of *Podoviridae* and *Myoviridae* bacteriophages to the effectiveness of anti-staphylococcal therapeutic cocktails. *Scientific reports*, **10:18612**:1-11.
46. **Kostka J. E., Prakash O., Overholt W. A., Green S. J., Freyer G., Canion A., Delgardio J., Norton N., Hazen T.C. & Huettel M.** (2011). Hydrocarbon-degrading bacteria and the bacterial community response in gulf of mexico

- beach sands impacted by the deepwater horizon oil spill. *Applied Environmental Microbiology*, **77**:7962-7974.
47. **Iida S., Hiestand-Nauer R., Sandmeier H., Lehnherr H. & Arber W.** (1998). Accessory Genes in the *darA* Operon of Bacteriophage P1 Affect Antirestriction Function, Generalized Transduction, Head Morphogenesis, and Host Cell Lysis. *Virology*, **251**:49-58.
 48. **Legendre M., Bartoli J., Shmakova L., Jeudy S., Labadie K., Adrait A., Lescot M., Poirot O., Bertaux L., Bruley C., Couté Y., Rivkina E., Abergel C. & Claverie J.** (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences*, **111**:11:4274-4279.
 49. **Li W & Godzik A.** (2006). Cd-hit: a fast program for clustering and comparing large sets of proteinL or nucleotide sequences. Disponible en: <http://weizhongli-lab.org/cd-hit/>
 50. **Liu Y., Weisberg R.H., Hu C. & Zheng L.** (2011). Tracking the Deepwater Horizon Oil Spill: A Modeling Perspective. *Eos Transactions American Geophysical Union*, **92**:6:45-52.
 51. **Liu Z., Liu J., Zhu Q., & Wu W.** (2012). The weathering of oil after the Deepwater Horizon Oil spill: insights from the chemical composition of the oil from the sea surface, salt marshes and sediments. *Environmental Research Letters*, **7**:1-14.
 52. **Lu Z., Deng Y., Nostrand J.D.V., He Z., Voordeckers J., Zhou A., Lee Y-J., Mason O.U., Dubinsky E.A., Chavarria K.L., Tom L.M., Fortney J.L., Lamendella R., Jansson J.K., D'haeseleer P., Hazen T.C. & Zhou J.** (2012). Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *The International Society for Microbial Ecology Journal*, **6**:451-460.
 53. **Luhtanen A.M., Eronen-Rasimus E., Kaartokallio H., Rintala J.M., Autio R. & Roine E.** (2014). Isolation and characterization of phage-host systems from the Baltic Sea ice. *Extremophiles*, **18**:1:121-130.
 54. **MacDonald I. R., Garcia-Pineda O., Beet A., Daneshgar Asl S., Feng L., Graettinger G., French-McCay D., Holmes J., Hu C., Huffer F., Leifer I., Muller-Karger F., Solow A., Silva M. & Swayze G.** (2015). Natural and unnatural oil slicks in the Gulf of Mexico. *Journal of Geophysical Research: Oceans*, **120**:8364-8380.
 55. **Mann N.H.** (2005). The Third Age of Phage. *PLoS Biology*, **3**:5:0753-0755.
 56. **Martin M.** (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. Disponible en: <http://journal.emblnet.org/index.php/embnetjournal/article/view/200/458>
 57. **Mason O.U., Hazen T.C., Borglin S., Chain P.S.G., Dubinsky E.A., Fortney J.L., Han J., Holman H.N, Hultman J., Lamendella R., Mackelprang R., Malfatti S., Tom L.M., Tringe S.G., Woyke T., Zhou J., Rubin E.M. & Jansson J.K.** (2012). Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *The International Society for Microbial Ecology Journal*, **6**:1715-1727.
 58. **Mason O.U., Scott N.M., Gonzalez A., Robbins-Pianka A., Bælum J., Kimbrel J., Bouskill N.J., Prestat E., Borglin S., Joyner D.C., Fortney**

- J.L., Jurelevicius D., Stringfellow W.T., Alvarez-Cohen L., Hazen T.C., Knight R., Gilbert J.A. & Jansson J.K. (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *The International Society for Microbial Ecology Journal*, **8**:1464-1475.
59. **McBride M.J.** (2014). The family Flavobacteriaceae. En el libro **Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F.** The Prokaryotes (pp. 643-676). Berlin/Heidelberg, Germany, Springer.
60. **Molnár J., Maygar B., Schneider G., Laczi K., Valappil S.K., Kovács Á.L., Nagy I.K., Rákhely G. & Kovács T.** (2020). Identification of a novel archaea virus, detected in hydrocarbon polluted Hungarian and Canadian samples. *PLoS ONE*, **15**:1-10.
61. **Nigro O.D., Jungbluth S.P., Lin H., Hsieh C., Miranda J.A., Schvarcz C.R., Rappé M.S. & Steward G.F.** (2017). Viruses in the Oceanic Basement. *American Society for Microbiology*, **8**:2:e02129-16.
62. **Nurk S., Meleshko D., Korobeynikov A. & Pevzner P.A.** (2017). metaSPAdes: a new versatile metagenomic assembler. Disponible en: <https://github.com/ablab/spades>
64. **Ofir G. & Sorek R.** (2018). Contemporary Phage Biology; From Classics Models to New Insights. *Cell*, **172**:1260-1270.
65. **Oksanen J., Guillaume-Blanchet F., Friendly M., Kindt R., Legendre P., McGlenn D., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E., Wagner H.** (2019). vegan: Community Ecology Package. R package version 2.5-6. Disponible en: <https://CRAN.R-project.org/package=vegan>
66. **Pante E., Simon-Bouhet B.** (2013). marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLoS ONE*, **8**:9:e73051.
67. **Paul B.G., Bagby S.C., Czornyj E., Arambula D., Handa S., Sczyrba A., Ghosh P., Miller J.F. & Valentine D.L.** (2015). Targeted diversity generation by intraterrestrial archaea and archaeal viruses. *Nature communications*, **6**:6585:1-8.
68. **Paulson J.N., Stine O.C., Bravo H.C. & Pop M.** (2013a). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, **10**:12:1200-1202.
69. **Paulson J.N., Olson N.D., Braccia D.J., Wagner J., Talukder H., Pop M. & Bravo H.C.** (2013b). metagenomeSeq: Statistical analysis for sparse high-throughput sequencing. Bioconductor package. Disponible en: <http://www.cbcb.umd.edu/software/metagenomeSeq>
70. **Pannekens M., Kroll L., Müller H., Mbow F.T. & Meckenstock R.U.** (2019). Oil reservoirs, an exceptional habitat for microorganisms. *New BIOTECHNOLOGY*, **49**:1-9.
71. **R Core Team.** (2019). R: a language and environment for statistical computing. *R Foundations for Statistical Computing*, Vienna, Austria. Disponible en: <https://www.R-project.org/>
72. **Raggi L., García-Guevara F., Godoy-Lozano E.E., Martínez-Santana A., Escobar-Zepeda A., Gutierrez-Ríos R.M., Loza A., Merino E., Sanchez-Flores A., Licea-Navarro A., Pardo-López L., Segovia L. & Juárez K.**

- (2020). Metagenomic Profiling and Microbial Metabolic Potential of Perdido Fold Belt (NW) and Campeche Knolls (SE) in the Gulf of Mexico. *Frontiers in Microbiology*, **11**:1825:1-18.
73. **Ramírez D., Vega-Alvarado L., Taboada B., Estradas-Romero A., Soto L., Juárez K.** (2020). Bacterial diversity in surface sediments from the continental shelf and slope of the North West gulf of Mexico and the presence of hydrocarbon degrading bacteria. *Marine Pollution Bulletin*, **150**:110590.
 74. **RefSeq Growth Statistics.** (2021, 15 de Febrero). NCBI. <https://www.ncbi.nlm.nih.gov/refseq/statistics/>
 75. **Rosenberg E., Bittan-Banin G., Sharon G., Shon A., Hershko G., Levy Itzik & Ron E.Z.** (2010). The phage-driven microbial loop in petroleum bioremediation. *Microbial biotechnology*, **3**:4:467-472.
 76. **Roseobase** (2021, 3 de marzo). University of Georgia, Marine Sciences. <http://www.roseobase.org/Species/dss3.html>
 77. **Roy B., Ackermann H.W., Pandian S., Picard G. & Goulet J.** (1993). Biological Inactivation of Adhering *Listeria monocytogenes* by Listeriaphages and a Quaternary Ammonium Compound. *Applied and Environmental Microbiology*, **59**:2914-2917.
 78. **Staley, J.T., & Konopka, A.** (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology*, **39**:321–346.
 79. **Suttle C.** (2005). Viruses in the sea. *Nature*, **437**:356-361.
 80. **Suttle C.** (2007). Marine viruses-major players in the global ecosystem. *Nature Reviews Microbiology*, **5**:801-811.
 81. **Suzuki R., Terada Y. & Shimodaira H.** (2019). pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 2.2-0. Disponible en: <https://CRAN.R-project.org/package=pvclust>
 82. **Taboada B., Isa P., Gutierrez-Escolano A.L., del Ángel R.M., Ludert J.E., Vázquez N., Tapia-Palacios M.A., Chávez P., Garrido E., Espinosa A.C., Eguiarte L.E., López S., Souza V. & Arias C.F.** (2018). The geographic structure of viruses in the Cuatro Ciénegas basin, a unique oasis in northern Mexico, reveals a highly diverse population on a small geographic scale. *Applied and Environmental Microbiology*, **11**:1-25.
 83. **Taxonomy Browser** (2021, 21 de febrero). NCBI. <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2560131>
 84. **Tully B.J., Sachdeva R., Graham E. D. & Heidelberg J. F.** (2017). 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. *PeerJ*, **5**:e3558.
 85. **ViralZone.** (2019). Swiss Institute of Bioinformatics. https://viralzone.expasy.org/505?outline=all_by_species
 86. **Wang B., Lai Q., Cui Z., Tan T. & Shao Z.** (2008). A pyrene-degrading consortium from deep-sea sediment of the West Pacific and its ke member *Cycloclasticus* sp. P1. *Environmental Microbiology*, **10**:8:1948-1963.
 87. **Weinbauer M.G.** (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, **28**:127-181.
 88. **Weinbauer M.G. & Rassoulzadegan F.** (2004). Are viruses driving microbial diversification and diversity. *Environmental Microbiology*, **6**:1:1-11.

89. **Weitz J.S. & Wilhelm S.W.** (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports*, **4**:1-17.
90. **Wickham H.** (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, **21**:12.
91. **Wickham H.** (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
92. **Wickham H., François R., Henry L. & Müller K.** (2020). dplyr: A Grammar of Data Manipulation. *R package version 1.0.1*. Disponible en: <https://CRAN.R-project.org/package=dplyr>
93. **Wilhelm S.W. & Suttle C.A.** (1999). Viruses and Nutrient Cycles in the Sea. *Bioscience*, **49**:781-88.
94. **Wilson W.H., Etten J.L. & Allen M.J.** (2009). The *Phycodnaviridae*: The Story of How Tiny Giants Rule the World. *Current Topics in Microbiology and Immunology*, **328**:1-42.
95. **Wommack K. E. & Colwell R.** (2000). Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews*, **64**:1:69-114.
96. **Yang Q., Gao C., Jiang Y., Wang M., Zhou X., Shao H., Gong Z. & McMinn A.** (2019). Metagenomic Characterization of the Viral Community of the South Scotia Ridge. *Viruses*, **11(2)**:95:1-19.
97. **Yoshida M., Takaki Y., Eitoku M., Nunoura T. & Takai K.** (2013). Metagenomic analysis of viral communities in (Hado)pelagic sediments. *PLOS ONE*, **8**:1-14.
98. **Zhao R., Summers Z.M., Christman G.D., Yoshimura K.M. & Biddle J.F.** (2020). Metagenomics views of microbial dynamics influenced by hydrocarbon seepage in sediments of the Gulf of Mexico. *Scientific Reports*, **10**:5772:1-13.
99. **Zukunft P.F.** (2010). Summary Report for Sub-Sea and Sub-Surface Oil and Dispersant Detection: Sampling and Monitoring. *Operational Science Advisory Team (OSAT)*.

ANEXO

Tabla A1. Metodología experimental utilizada para el procesamiento y secuenciación de las muestras en cada estudio.

Muestras	Referencia	Muestreo	Extracción de DNA	Secuenciación
CIGoM	(Raggi <i>et al.</i> , 2020)	Las muestras se tomaron en marzo del 2016. Las muestras de agua se colectaron usando una roseta de muestreo con botellas Niskin de 20 litros y cada muestra de agua se concentró a aproximadamente tres litros con un sistema tangencial de filtración. De este volumen, 100 mL se almacenaron a 4 °C para filtración con filtro Sterivex; otros 100 mL se pasaron a través de una membrana de policarbonato de 0.22 µm (Merck Milipore). Inmediatamente los filtros se almacenaron en N ₂ líquido. El sedimento se extrajo con descorazonador de caja y de este se tomaron los primeros 10 cm de muestra. Todas las muestras se mantuvieron a -20 °C hasta que fueran procesadas.	En el caso de las muestras de agua, la extracción de DNA se hizo a partir de los filtros Sterivex o Milipore usando el kit de aislamiento Power Water DNA (MO BIO-QIAGEN), mientras que para las muestras de sedimento se utilizaron 0.5 g de sedimento superficial y se usó el kit de aislamiento de DNA PowerSoil (MO BIO-QIAGEN).	En total, seis muestras tuvieron rendimiento alto de DNA, tres de agua (D18_MAX, A04_MIN y A04_AAIW) y seis de sedimento (D18_SED1-3, D18_SED, A04_SED y E03_SED) y se usaron para llevar a cabo secuenciación de metagenoma completo. Las librerías se prepararon usando el kit Truseq DNA PCR-Free HT Library Prep (Illumina). Las librerías se secuenciaron con la plataforma Illumina NextSeq550 obteniendo secuencias pareadas de 75 o 150 pb de longitud.
DWH_W (Muestras de agua del DWH)	(Mason <i>et al.</i> , 2012)	Las muestras se tomaron de mayo a junio del 2010. De cada estación se tomaron de 1-5 litros de agua de mar con botellas Niskin que después fueron filtrados a través de un filtro con 0.22-µm de diámetro de poro e inmediatamente se congelaron y conservaron a -20 °C	El DNA se extrajo a partir de los filtros usando una versión modificada del método de Miller, el cual utiliza lisis por SDS, buffer de fosfatos y fenol:cloroformo:isoamyl alcohol. Después, para limpiar las muestras se utilizó el kit de extracción de MoBio Soil DNA (MoBio, Carlsbad, CA, USA)	El DNA se fragmentó, se le repararon los extremos y se le ligaron los adaptadores 1 y 2 de Illumina. Las librerías se secuenciaron usando la plataforma Illumina GAIIx obteniendo secuencias pareadas de 114 pb de longitud.

		hasta que fueran procesadas.		
DWH_S (Muestras de sedimento del DWH)	(Mason <i>et al.</i> , 2014)	Los núcleos de sedimento se colectaron del 16 de septiembre al 20 de octubre del 2010. Las muestras se congelaron a -80 °C y se seccionaron y se seleccionaron las capas superficiales de 0-1 cm de cada núcleo.	Se tomaron 0.25 g de sedimento a partir de los cuales se extrajo DNA siguiendo el protocolo de <i>Earth Microbiome Project DNA</i> , el cual usa una versión modificada del kit de aislamiento de DNA PowerSoil-htp de 96 pozos con la plataforma robótica epMotion 5075.	Las librerías de secuenciación metagenómica se prepararon y secuenciaron en Argonne National Laboratory. Para cada muestra, se utilizó 1µg de NA genómico para la preparación de librerías usando el kit de Illumina TruSeq. Las librerías se secuenciaron con la plataforma Illumina HiSeq obteniendo secuencias pareadas de 100 pb de longitud.
ZHAO_S (muestras de sedimento del noroeste del GoM)	(Zhao <i>et al.</i> , 2020)	Las muestras se obtuvieron con un nucleador multiple operado vía remota ayudado con cámara de video. Las secciones de sedimento de que estaban ente los 10 y 15 cm por debajo del suelo marino de cada núcleo se almacenaron en un tubo falcon y se preservaron a -80 °C hasta que fueran utilizadas.	El DNA se extrajo a partir de 0.5 g de sedimento usando el kit de extracción de DNA MoBio PowerSoil siguiendo el protocolo del producto.	Las librerías de DNA se prepararon usando el kit de preparación de librerías NEBNext UltraN II DNA y se secuenciaron con la plataforma de Illumina HiSeq obteniendo secuencias pareadas de 250 pb de longitud.
DONG_S (muestras de sedimento del noreste del GoM)	(Dong <i>et al.</i> , 2019)	Las tres muestras de sedimento marino se tomaron con nucleadores entre 5-6 m debajo del suelo marino y se seccionaron en intervalos de 20 cm.	El DNA se extrajo a partir de 10 g de sedimento usando el kit de aislamiento de DNA PowerMax Soil (12988-10, QIAGEN) siguiendo el protocolo establecido.	Las librerías de DNA se prepararon al fragmentar el DNA genómico usando un sonicador y el kit de preparación de librerías de DNA NEBNext Ultra II. Las librerías se secuenciaron con la plataforma de Illumina NextSeq 500 obteniendo secuencias pareadas de 150 pb.

Tabla A2. Comandos utilizados para correr los programas.

Programa	Comando
cutadapt	--pair-filter=any -q 20 -b "G{15}" -B "G{15}" -b "T{15}" -B "T{15}" -b "C{15}" -B "C{15}" -b "A{15}" -B "A{15}" -n 10 --trim-n --max-n 20 -m 40:40
bbmap	k=13 usemodulo=t slow=t minid=0.90
DIAMOND blastp/blastx	--more-sensitive --outfmt 100 --salltitles --max-target-seqs 15 -- query-cover 70 --max-hsps 1 -e 0.001 --index-chunks 1
MEGAN	minscore=50.0/100.0, maxExpected=0.001/0.00001 topPercent=10.0 minSupportPercent=0.0 minSupport=4 lcaAlgorithm=weighted lcaCoveragePercent=80.0 readAssignmentMode=readMagnitude

Tabla A3. Visión general de la calidad de las muestras, organizadas por el tipo de muestra (agua o sedimento) y por estudio del que provienen. Número total de secuencias obtenidas por medio de secuenciación masiva (suma de las secuencias pareadas) y secuencias válidas después de los filtros de calidad.

Muestra	Sustrato	Número de secuencias	Reads válidos	Porcentaje (%)
D18_MAX	Agua	77,396,766	50,921,708	66
A04_MIN	Agua	83,327,042	22,485,176	27
A04_AAIW	Agua	144,635,708	95,570,256	66
DWH_1	Agua	61,090,938	56,186,406	92
DWH_2	Agua	61,738,773	56,476,504	91
DWH_3	Agua	71,339,076	69,224,500	97
DWH_4	Agua	73,669,370	71,015,856	96
DWH_5	Agua	60,335,450	52,692,362	87
DWH_6	Agua	59,845,240	52,459,030	88
D18_SED1	Sedimento	106,220,230	91,050,952	86
D18_SED2	Sedimento	89,577,074	75,862,692	85
D18_SED3	Sedimento	86,885,754	72,613,410	84
D18_SED	Sedimento	162,961,050	154,980,376	95
A04_SED	Sedimento	147,151,128	141,418,102	96
E03_SED	Sedimento	208,581,878	204,887,120	98
ZHAO_S1	Sedimento	23,102,982	23,073,732	100
ZHAO_S2	Sedimento	42,616,380	42,542,242	100
ZHAO_S3	Sedimento	65,926,174	65,796,894	100
ZHAO_S4	Sedimento	39,202,492	39,154,666	100
ZHAO_S5	Sedimento	70,809,448	70,657,192	100
ZHAO_S6	Sedimento	68,096,796	67,941,690	100
ZHAO_S7	Sedimento	64,501,938	64,366,376	100
DWH_S1	Sedimento	165,880,544	130,188,062	78
DWH_S2	Sedimento	274,876,398	208,689,654	76
DWH_S3	Sedimento	152,771,600	130,776,672	86
DWH_S4	Sedimento	186,641,968	151,078,800	81
DWH_S5	Sedimento	203,867,834	179,164,912	88
DWH_S6	Sedimento	265,712,048	248,249,954	93
DWH_S7	Sedimento	181,464,634	129,841,290	72
DWH_S8	Sedimento	144,006,094	125,331,048	87
DWH_S9	Sedimento	284,616,506	240,363,620	84
DWH_S10	Sedimento	344,223,276	270,568,244	79
DWH_S11	Sedimento	368,517,966	319,182,878	87
DWH_S12	Sedimento	290,736,102	239,903,856	83
DWH_S13	Sedimento	354,848,004	269,145,482	76
DWH_S14	Sedimento	294,829,388	242,577,674	82
DONG_S1	Sedimento	86,682,566	76,123,366	88
DONG_S2	Sedimento	139,063,270	138,015,290	99
DONG_S3	Sedimento	149,152,910	148,149,030	99

Tabla A4. Estadísticas de los ensambles del total de las muestras. Muestras organizadas por el tipo de muestra (agua o sedimento) y por estudio del que provienen.

Muestra	Sustrato	Número de		N50	N75	L50	L75	GC (%)	# N's
		contigs (≥ 300 pb)	Contig más largo (pb)						
D18_MAX	Agua	114,027	60,843	1,072	583	23,276	54,691	50	0
A04_MIN	Agua	21,392	5,446	574	388	6,225	12,596	44	0
A04_AAIW	Agua	134,233	294,240	1,428	651	18,967	55,213	54	0
DWH_1	Agua	92,930	34,427	519	372	24,868	54,408	45	0
DWH_2	Agua	95,052	33,271	516	372	25,546	55,759	45	0
DWH_3	Agua	39,611	52,198	928	445	5,910	18,385	42	0
DWH_4	Agua	40,256	54,241	929	446	6,076	18,704	42	0
DWH_5	Agua	161,890	20,014	461	357	52,215	101,304	48	0
DWH_6	Agua	160,428	26,502	461	357	51,804	100,418	48	0
D18_SED1	Sedimento	2,934,885	71,675	456	361	966,580	1,851,347	59	0
D18_SED2	Sedimento	2,746,515	508,496	493	374	839,352	1,679,262	56	0
D18_SED3	Sedimento	2,478,081	197,040	481	370	770,943	1,527,740	49	0
D18_SED	Sedimento	2,739,866	192,723	500	359	747,520	1,627,438	54	0
A04_SED	Sedimento	2,401,218	96,944	437	349	736,790	1,494,555	64	0
E03_SED	Sedimento	470,222	51,244	572	387	128,804	272,575	60	0
ZHAO_S1	Sedimento	563,821	46,164	485	417	219,579	381,358	59	0
ZHAO_S2	Sedimento	1,003,523	40,116	537	438	363,532	655,988	54	0
ZHAO_S3	Sedimento	786,788	35,468	566	445	263,378	498,144	52	0
ZHAO_S4	Sedimento	914,148	66,527	511	428	344,108	608,040	57	0
ZHAO_S5	Sedimento	734,824	107,246	580	448	233,311	456,668	50	0
ZHAO_S6	Sedimento	666,084	49,096	587	450	212,532	413,219	51	0
ZHAO_S7	Sedimento	666,599	75,669	587	449	206,844	410,951	51	0
DWH_S1	Sedimento	744,927	81,966	501	366	214,178	446,790	41	0
DWH_S2	Sedimento	830,443	150,879	581	385	194,795	462,726	42	0
DWH_S3	Sedimento	736,084	73,965	474	359	219,175	449,146	43	0
DWH_S4	Sedimento	1,118,684	144,574	560	382	280,041	636,925	41	0
DWH_S5	Sedimento	1,205,033	152,060	501	365	336,480	717,525	44	0
DWH_S6	Sedimento	1,462,818	102,529	438	348	467,408	919,818	47	0
DWH_S7	Sedimento	297,574	30,361	411	340	104,945	193,889	52	0
DWH_S8	Sedimento	659,672	55,012	507	367	185,873	393,024	41	0
DWH_S9	Sedimento	663,327	92,103	472	357	197,598	404,997	45	0
DWH_S10	Sedimento	432,958	82,365	456	353	130,368	266,390	45	0
DWH_S11	Sedimento	419,048	52,010	439	347	133,712	263,330	48	0
DWH_S12	Sedimento	303,443	45,394	417	341	101,745	194,557	45	0
DWH_S13	Sedimento	325,965	77,059	422	343	105,215	206,475	49	0
DWH_S14	Sedimento	348,184	35,408	400	337	127,767	230,261	52	0
DONG_S1	Sedimento	877,690	50,712	420	344	294,931	562,524	51	0
DONG_S2	Sedimento	3,350,940	414,750	504	370	912,324	1,981,082	46	0
DONG_S3	Sedimento	3,318,472	191,246	507	371	921,249	1,970,213	46	0

Tabla A5. Número de ORF's obtenidos por muestra y ORF's por kilobase del total de pb ensambladas. Muestras organizadas por el tipo de muestra (agua o sedimento) y por estudio del que provienen.

Nombre	Sustrato	ORF's	ORF's/kb
D18_MAX	Agua	118,657	1.2
A04_MIN	Agua	14,414	1.2
A04_AAIW	Agua	146,473	1.1
DWH_1	Agua	80,175	1.6
DWH_2	Agua	82,016	1.6
DWH_3	Agua	38,690	1.3
DWH_4	Agua	39,541	1.3
DWH_5	Agua	131,014	1.7
DWH_6	Agua	129,920	1.7
D18_SED1	Sedimento	2,206,133	1.6
D18_SED2	Sedimento	2,134,159	1.5
D18_SED3	Sedimento	1,864,032	1.5
D18_SED	Sedimento	2,497,059	1.7
A04_SED	Sedimento	2,697,155	2.3
E03_SED	Sedimento	460,940	1.7
ZHAO_S1	Sedimento	462,019	1.6
ZHAO_S2	Sedimento	852,500	1.5
ZHAO_S3	Sedimento	689,055	1.5
ZHAO_S4	Sedimento	760,826	1.6
ZHAO_S5	Sedimento	649,480	1.4
ZHAO_S6	Sedimento	590,121	1.4
ZHAO_S7	Sedimento	594,129	1.4
DWH_S1	Sedimento	577,752	1.5
DWH_S2	Sedimento	702,573	1.4
DWH_S3	Sedimento	554,459	1.5
DWH_S4	Sedimento	910,643	1.4
DWH_S5	Sedimento	946,850	1.5
DWH_S6	Sedimento	996,717	1.4
DWH_S7	Sedimento	212,483	1.6
DWH_S8	Sedimento	525,426	1.5
DWH_S9	Sedimento	497,250	1.5
DWH_S10	Sedimento	325,203	1.5
DWH_S11	Sedimento	290,745	1.5
DWH_S12	Sedimento	209,486	1.5
DWH_S13	Sedimento	218,958	1.4
DWH_S14	Sedimento	232,057	1.6
DONG_S1	Sedimento	623,812	1.5
DONG_S2	Sedimento	2,445,541	1.3
DONG_S3	Sedimento	2,413,764	1.4

Tabla A6. Géneros virales con mayor abundancia relativa promedio. Se muestran los primeros 20 géneros más abundantes debido a que la lista total es de 504 géneros.

Géneros	Familia	Abundancia	
		relativa promedio (%)	Desviación estándar
<i>Thornevirus</i>	<i>Myoviridae</i>	9.0	5.2
<i>Polybotosvirus</i>	<i>Myoviridae</i>	6.5	2.3
<i>Likavirus</i>	<i>Siphoviridae</i>	6.1	3.0
<i>Biquartavirus</i>	<i>Myoviridae</i>	6.0	2.2
<i>Minunavirus</i>	<i>Siphoviridae</i>	3.9	3.4
<i>Lederbergvirus</i>	<i>Podoviridae</i>	2.9	2.5
<i>Punavirus</i>	<i>Myoviridae</i>	2.7	2.1
<i>Casadabanvirus</i>	<i>Siphoviridae</i>	2.5	1.4
<i>Omegavirus</i>	<i>Siphoviridae</i>	2.1	1.4
<i>Kafunavirus</i>	<i>Podoviridae</i>	2.1	3.2
<i>Jenstvirus</i>	<i>Siphoviridae</i>	2.1	2.1
<i>Llyrvirus</i>	<i>Myoviridae</i>	2.1	1.5
<i>Delepquintavirus</i>	<i>Siphoviridae</i>	2.0	1.2
<i>Eneladusvirus</i>	<i>Myoviridae</i>	1.8	1.4
<i>Stanholtvirus</i>	<i>Stanholtvirus</i>	1.7	2.2
<i>Sinsheimervirus</i>	<i>Microviridae</i>	1.7	4.2
<i>Agricanavirus</i>	<i>Myoviridae</i>	1.6	1.4
<i>Magadivirus</i>	<i>Magadivirus</i>	1.4	1.2
<i>Bixzunavirus</i>	<i>Myoviridae</i>	1.4	1.4
<i>Chiangmaivirus</i>	<i>Myoviridae</i>	1.3	1.7

Tabla A7. Estadísticas de diversidad viral a nivel de especie del total de las muestras. Muestras organizadas por el tipo de muestra (agua o sedimento) y por estudio del que provienen.

Muestras	Sustrato	Abundancia viral (secuencias)	Porcentaje de secuencias virales respecto al total de secuencias (%)	Riqueza de especies virales	Índice de Simpson	Índice de Shannon
D18_MAX	Agua	763,398	3	408	0.99	5.9
A04_MIN	Agua	164,284	1.5	279	0.99	5.5
A04_AAIW	Agua	2,986,042	6.3	387	0.99	5.9
DWH_W1	Agua	1,916,713	3.6	306	0.99	5.9
DWH_W2	Agua	2,017,204	3.8	309	0.99	5.8
DWH_W3	Agua	2,653,535	4.1	229	0.99	5.5
DWH_W4	Agua	2,718,261	4	224	0.99	5.4
DWH_W5	Agua	650,153	1.4	603	0.99	6.3
DWH_W6	Agua	645,929	1.4	604	0.99	6.3
ZHAO_S1	Sedimento	818,431	4.7	603	0.99	6.3
ZHAO_S1	Sedimento	818,431	4.7	603	0.99	6.3
ZHAO_S2	Sedimento	804,583	3	816	0.99	6.6
ZHAO_S3	Sedimento	603,533	1.5	894	0.99	6.6
ZHAO_S4	Sedimento	1,251,616	5.1	784	0.99	6.5
ZHAO_S5	Sedimento	89,671	0.3	409	0.99	5.9
ZHAO_S6	Sedimento	264,940	0.7	598	0.99	6.2
ZHAO_S7	Sedimento	675,289	1.7	433	0.99	6
D18_SED1	Sedimento	170,701	0.4	506	0.99	6.1
D18_SED2	Sedimento	884,953	2.4	446	0.99	6
D18_SED3	Sedimento	813,048	2.3	391	0.99	5.9
D18_SED	Sedimento	780,621	1.1	409	0.99	5.9
A04_SED	Sedimento	1,490,405	2.2	476	0.99	6.1
E03_SED	Sedimento	1,271,544	1.3	352	0.99	5.7
DWH_S1	Sedimento	1,639,681	2.6	1158	0.99	7
DWH_S2	Sedimento	2,825,671	2.8	1059	0.99	6.9
DWH_S3	Sedimento	1,438,885	2.3	1191	0.99	7
DWH_S4	Sedimento	2,198,785	3	1097	0.99	6.9
DWH_S5	Sedimento	2,097,939	2.4	1151	0.99	7
DWH_S6	Sedimento	2,041,318	1.7	1164	0.99	7
DWH_S7	Sedimento	315,784	0.5	996	0.99	6.7
DWH_S8	Sedimento	1,420,478	2.3	1142	0.99	6.9
DWH_S9	Sedimento	2,166,833	1.9	1115	0.99	7
DWH_S10	Sedimento	1,686,851	1.3	1113	0.99	6.9
DWH_S11	Sedimento	1,256,049	0.8	1088	0.99	6.9
DWH_S12	Sedimento	931,952	0.8	1214	0.99	7
DWH_S13	Sedimento	624,716	0.5	1067	0.99	6.8
DWH_S14	Sedimento	492,227	0.5	1073	0.99	6.8
DONG_S1	Sedimento	340,992	0.9	491	0.99	6
DONG_S2	Sedimento	2,062,631	3	850	0.99	6.7
DONG_S3	Sedimento	2,566,983	3.5	992	0.99	6.8