

Universidad de Chile
Facultad de Ciencias Químicas y Farmacéuticas



**“Elucidando poblaciones de fagos marinos
portadores de Genes Metabólicos Auxiliares
(AMGs) involucrados en los ciclos
biogeoquímicos de Antártica”**

**Memoria presentada a la Universidad de Chile para
optar al Título Profesional de Bioquímico por:**

WILSON ELÍAS CASTILLO INAIPII

Directoras de Memoria

Dra. Beatriz Díez

Dra. Angélica Casanova-Katny

Profesor Patrocinante:

Dr. Carlos Santiviago

Santiago-Chile

Diciembre 2021

Esta memoria fue desarrollada en el Laboratorio de Ecología Microbiana y Microbiología Ambiental de Sistemas Extremos, de la Facultad de Ciencias Biológicas de la Pontificia Universidad Católica de Chile, bajo la dirección de la Dra. Beatriz Díez en conjunto con la Dra. Angélica Casanova-Katny y con el patrocinio del Dr. Carlos Santiviago.

La memoria contó con el financiamiento del proyecto FONDECYT N°118745 dirigido por la Dra. Angélica Casanova-Katny y el proyecto INACH RT_04_19 dirigido por la Dra. Beatriz Díez.

“¿Que cómo me siento?

*Bueno, no tengo palabras. Me siento, me siento...
me siento como la primavera después del invierno,
y el sol sobre el follaje; ¡y como todas las
trompetas y las arpas y todas las canciones que
he escuchado en mi vida”*

.- Sam Gamyi

Agradecimientos

Le doy gracias a Dios, al amor y a la música por acompañarme en este viaje. En este proceso me han acompañado personas admirables, y si de personas admirables hablo, las primeras son mis padres: Carola y Wilson. Gracias a ellos estoy aquí, muchas gracias por su soporte y por todo el amor que me han entregado a cada momento. También pienso en mi mejor amigo, que tengo la fortuna de que sea también mi hermano, gracias Agustín por demostrarme que con esfuerzo los sueños se logran, y por estar ahí cada vez que lo necesité. Muchas gracias a la familia Inaipil, les extraño y espero que pronto pase esta pandemia para que volvamos a reír juntos.

También le agradezco a la familia que encontré en el camino, al Pete y al Toto que con su cariño y apoyo me han hecho superar los momentos más crudos de la carrera. Gracias también a Fabián, Esteban y David, que gracias a ellos pude encontrar nuevamente valor en la amistad y me han acompañado en todo el proceso a punta de música y risas. Gracias a los incondicionales, a Paulo, Martín, Fabián, León y Choclo. Que, conociéndome desde pequeño, y con todos mis errores, aún así decidieron por mi amistad. Gracias a la Meche, pues aunque no lo sepas, tu vida me inspira a dar lo mejor en lo que me gusta. Agradezco las amistades de la Feña, Francyne y Ofé que han sido trascendentales en tan poco tiempo y llevo mucho de ellas en cada paso.

Agradezco a quienes me han inspirado y que gracias por su amor a la ciencia, me impulsaron a encontrar la profesión que amo. En esto quiero dar gracias a mis directoras de memoria, a la Dra. Beatriz por abrirme las puertas de su laboratorio y que gracias a su amor por lo que hace, me ha inspirado a investigar, gracias a usted pude trabajar con investigadores e investigadoras de excelencia. También agradezco a la Dra. Angélica por la confianza que depositó en mí. Agradezco a los profesores y profesoras que me animaron a seguir, y que con sus clases encendían una apasionante curiosidad. Gracias a los profesores de la facultad, en especial a la profesora María Antonieta y al profe Santiviago. También gracias a la profe Jeshu, que en el colegio me dio la oportunidad de conocer la carrera de bioquímica, sin ella no estaría aquí.

Gracias también a los amigos que han sido también mis tutores (algunos sin saberlo). Gracias al Jota, Tomás, Val, Cami, Negro, Negra, Pelao, Diegú, Andrea, Dani, Jorge, Lore, Melissa y Pastizal. Muchas gracias a mis compañeros de lab: Marianne, Sergio, Pablo, Jaime, Javier, Johanna, JB, Tomás, Carla, Christina, Karen y a todos los que han llegado a este maravilloso ambiente y cuna de grandes científicos y científicas. Agradezco en especial a Jero, que ha sido mi tutor desde el primer segundo en el lab y en cada momento me transmitió su pasión por la ciencia, también a Diego que ha sido mi amigo incondicional durante todo este tiempo.

El espacio no alcanza para agradecer a todas las personas que se me vienen a la mente, pero cada una de ellas sabe lo feliz que me hace el haberles encontrado en este proceso universitario. Muchas gracias a cada uno/a de ellos y ellas.

PRESENTACIÓN A CONGRESOS

Castillo-Inaipil W., Guajardo S., Buscaglia M., Cifuentes-Anticevic J. & Díez B. (2021). Elucidating populations of marine phages carrying AMGs involved in the Antarctica biogeochemical cycles. ASLO 2021 Aquatic Sciences Meeting. Virtual meeting, del 22 al 27 de junio del 2021.

Castillo-Inaipil W., Guajardo S., Buscaglia M., Cifuentes-Anticevic J., Casanova-Katny A. & Díez B. (2021). Elucidando poblaciones de fagos marinoportadores de AMGs involucrados en los ciclos biogeoquímicos de Antártica. XXXVII Congreso Nacional de Estudiantes de Biquímica y Carreras Afines. Congreso online, del 12 al 20 de agosto del 2021.

Castillo-Inaipil W., Guajardo S., Buscaglia M., Cifuentes-Anticevic J., Alcamán-Arias M. E., Casanova-Katny A., M.E., Farías, L. & Díez B. (2021). Comparación e interacción de comunidades de bacteriófagos del hielo glaciar y aguas costeras de Bahía Chile en Antártica. X Congreso Antártico. Congreso online, del 4 al 8 de octubre del 2021

ÍNDICE DE CONTENIDOS

PRESENTACIÓN A CONGRESOS	5
ÍNDICE DE CONTENIDOS	6
ÍNDICE DE FIGURAS	8
ÍNDICE DE TABLAS	9
ABREVIATURAS	10
RESUMEN	12
SUMMARY	14
1. INTRODUCCIÓN	16
1.1 Microorganismos en el sistema marino antártico	16
1.2 Virus en el ecosistema marino	17
1.3 Virus en el Océano Austral/Antártico.....	20
1.4 Aguas costeras de Bahía Chile	22
2. Hipótesis	26
3. Objetivo General.....	26
4. Objetivos específicos	26
5. Materiales y Metodologías.....	27
5.1 Sitio de muestreo y procesamiento de muestras utilizadas en esta memoria	27
5.2 Obtención de vOTUs a partir de metagenomas	32
5.4 Mapeo de lecturas contra vOTUs y construcción de tabla de conteos.....	38
5.5 Anotación y validación de AMGs.....	39
5.6 Análisis estadísticos y construcción de gráficos.....	41
6. Resultados	42
6.1 Identificación de vOTUs a partir de metagenomas de muestras de mar de la Bahía Chile y de hielo del glaciar Fuerza Aérea	42

6.2 Asignación taxonómica de vOTUs mediante redes de genes compartidos...	43
6.3 Asignación taxonómica de agrupamientos en la red sin asociación a referencia conocida	49
6.4 Distribución temporal y abundancia relativa de vOTUs en Bahía Chile y en hielo glaciar de Fuerza Aerea	51
6.5 Búsqueda de AMGs presentes en la comunidad viral marina de Bahía Chile y en muestra de hielo del glaciar Fuerza Aérea	54
7. Discusión	57
7.1 Ubicuidad y abundancia de familias de caudovirales en Bahía Chile similar a otros ambientes marinos	59
7.2 Presencia y potencial función ecológica de la familia <i>Phycodnaviridae</i> en aguas marinas de Bahía Chile	64
7.3 Potencial función ecológica viral asociada a la identificación de AMGs putativos en comunidades virales de Bahía Chile	67
8. Conclusiones.....	73
9. Referencias.....	74

ÍNDICE DE FIGURAS

Figura 1.- Mapa de punto de estudio, ubicado en Bahía Chile, Isla Greenwich, Islas Sheetland del Sur, Península Antártica.....	27
Figura 2.- Diagrama representando flujo de trabajo para el análisis de secuencias virales desde metagenomas.	34
Figura 3.- Generación de redes con VConTACT2.....	36
Figura 4.- Diagrama de Venn con la comparación de vOTUs presentes en muestras marinas de Bahía Chile y la muestra de hielo del glaciar Fuerza Aérea. 43	
Figura 5.- Red de genes compartidos en muestras de agua de mar de Bahía Chile y una muestra de hielo del glaciar Fuerza Aérea.....	46
Figura 6.- Abundancia relativa de las familias virales presentes en muestras de agua marina de Bahía Chile y de hielo del glaciar Fuerza Aérea.....	48
Figura 7.- Asignación de lecturas mapeadas contra el Grupo 2 de vOTUs desconocidos.....	50
Figura 8.- <i>Heatmap</i> de abundancias relativas a nivel de vOTUs entre muestras utilizadas en el estudio.....	53
Figura 9.- Abundancias de los vOTUs portadores de AMGs en los metagenomas marinos de Bahía Chile y en la muestra de hielo del glaciar Fuerza Aérea.....	56

ÍNDICE DE TABLAS

Tabla 1.	Muestras utilizadas en el estudio.....	28
Tabla 2.	Variables ambientales asociadas a metagenomas.....	29
Tabla 3.	Asignación a nivel de género por VConTACT2	47
Tabla 4.	Resultados de BLASTn sobre vOTUs reconocidos como NCLDV.51	

ABREVIATURAS

°C	Celsius
µm	Micrómetros
µM	Micromolar
ADN	Ácido desoxiribonucleico
AMG	<i>Auxiliary Metabolic Gene</i> (Gen Metabólico Auxiliar)
ARN	Ácido ribonucleico
Chla	Clorofila a
dCTP	<i>Deoxycytidine triphosphate</i> (Deoxicitidina trifosfato)
DMS	Sulfuro de dimetilo
DMSP	Dimetilsulfoniopropionato
DOM	Materia orgánica disuelta
dUMP	<i>Deoxyuridine monophosphate</i> (Deoxiuridina monofosfato)
dUTP	<i>Deoxyuridine triphosphate</i> (Deoxiuridina trifosfato)
h	Horas
ICTV	<i>International Committee on Taxonomy of Viruses</i> (Comité Internacional de Taxonomía de Virus)
L	Litros
m	Metros
M	Molar
mg	Miligramos
min	Minutos

mL	Mililitros
mM	Milimolar
n°	Número
NCLDV	<i>Nucleocitoviricota</i>
NR	<i>Non-redundant</i> (Base de datos de proteínas no redundante del NCBI)
NT	<i>Non-redundant nucleotide</i> (Base de datos de nucleótidos no redundante del NCBI)
PaV	Virus de <i>Phaeocystis antarctica</i>
PC	Agrupaciones protéicas
pb	Pares de bases
POM	Materia orgánica particulada
POV	<i>Pacific Ocean Virome</i>
rpm	Revoluciones por minuto
SO	<i>Southern Ocean</i> (Océano Austral/Antártico)
TMM	<i>Trimmed mean of M values</i> (media truncada de valores M)
TPM	<i>Transcripts Per Million</i> (Transcritos Por Millón)
v	Versión
VBR	<i>Virus to bacteria ratio</i> (Razón virus a bacteria)
VC	<i>Viral Cluster</i> (Conglomerado viral)
vOTU	<i>Viral Operational Taxonomic Unit</i> (Unidad Taxonómica Operacional viral)
WAP	<i>Western Antarctic Peninsula</i> (Península Antártica Occidental)
WtP	<i>What the Phage</i>

RESUMEN

La Península Antártica Occidental (WAP; por sus siglas en inglés) ha sufrido un aumento de temperatura durante los últimos años, afectando directa o indirectamente la producción primaria en el ambiente marino. El efecto de factores abióticos en las comunidades microbianas de Antártica ha sido estudiado profundamente, sin embargo, el efecto de los factores bióticos como los virus ha sido muy poco explorado. Bahía Chile es un ecosistema representativo de los ambientes costeros de la WAP, influenciado por la entrada de agua dulce debido al derretimiento de glaciares cercanos, y en la cual las comunidades microbianas cambian anualmente debido a diferentes escenarios ambientales. El objetivo de esta memoria es elucidar la estructura, composición y potencial función de la comunidad de bacteriófagos, como agentes relevantes que impactan la estructura y composición de la comunidad bacteriana, a una escala de varios años con condiciones productivas distintas en Bahía Chile. En esta memoria se analizó la composición taxonómica de fagos de ocho metagenomas obtenidos en el verano austral del 2014 al 2019. Además, se realizó la búsqueda y evaluación de la relevancia de esta comunidad de fagos como portadores de genes metabólicos auxiliares (AMGs; por sus siglas en inglés) que pueden modular cambios en el metabolismo de bacterias hospederas involucradas en los ciclos biogeoquímicos de Bahía Chile. Los resultados demuestran una ubicua y marcada presencia de comunidades virales del tipo ADN (1966 vOTUs) pertenecientes a las familias *Myoviridae* (la más abundante al igual que en muchas otras regiones marinas del mundo), *Siphoviridae*, *Podoviridae* infectando al bacterioplancton, y

Phycodnaviridae infectando a miembros del fitoplancton formadores de florecimientos relevantes en estas aguas polares. Ningún vOTU fue compartido entre las aguas marinas y el hielo de glaciación Fuerza Aérea cercano a la Bahía, por lo que se sugiere un bajo efecto de la comunidad viral por efecto del deshielo sobre las comunidades marinas en la bahía. El perfil viral proteico obtenido, muestra la presencia de 7 AMGs previamente descritos en la literatura, y 3 posibles AMGs nunca anteriormente descritos en ningún otro estudio, y que estarían involucrados en los ciclos del carbono (GT2, GT102, *dut*, *thyA* y *dcd*), biosíntesis de cobalamina (*cobS*), adquisición de nitrógeno orgánico (M17), ciclo del azufre (*dcm*), y estructura del ribosoma (*rpsA* y *rpsU*). Todo lo anterior sugiere, que estos virus podrían estar regulando la dinámica y el fitness de sus comunidades hospederas, modificando durante la infección procesos ecológicos relevantes como el ciclado de C, N y S en esta región. Esta memoria permite un mejor entendimiento del rol viral en el ambiente costero de Antártica, una región del planeta altamente afectada por el calentamiento global.

Elucidating populations of marine phages carrying Auxiliary Metabolic Genes (AMGs) involved in the Antarctica biogeochemical cycles

SUMMARY

The Western Antarctic Peninsula (WAP) has undergone an increase in temperature during the last few years, directly or indirectly affecting primary production in the marine environment. The effect of abiotic factors on Antarctic microbial communities has been studied in depth, however, the effect of biotic factors such as viruses has been poorly explored. Chile Bay is a representative ecosystem of the coastal environment of the WAP, influenced by freshening due to the melting of nearby glaciers, and in which summer microbial communities change annually due to different environmental scenarios. The aim of this study is to elucidate the structure, composition and potential function of the bacteriophage community, as relevant agents that impact the structure and composition of the bacterial community, at a scale of several years with different productive conditions in Chile Bay. In this study, the phages taxonomic composition of eight metagenomes obtained in the austral summer from 2014 to 2019 was analyzed. In addition, we searched for and evaluated the relevance of this phage community as carriers of auxiliary metabolic genes (AMGs) that can modulate changes in the metabolism of host bacteria involved in the biogeochemical cycles of Chile Bay. The results demonstrate a ubiquitous and marked presence of DNA viral communities (1966 vOTUs) belonging to the families *Myoviridae*, *Siphoviridae*, *Podoviridae* infecting bacterioplankton, and *Phycodnaviridae* infecting phytoplankton members that form relevant blooms in these polar waters. No vOTUs were shared between the marine waters and the Fuerza Aerea glacier ice near the Bay, suggesting a low effect of the viral community

due to the effect of melting ice on the marine communities in the bay. The viral protein profile obtained shows the presence of 7 AMGs previously described in the literature, and 3 possible AMGs never previously described in any other study, and that would be involved in the carbon cycle (GT2, GT102, *dut*, *thyA* y *dcd*), organic nitrogen acquisition (M17), sulfur cycle (*dcm*) and ribosome structure (*rpsA* y *rpsU*). All of the above suggests that these viruses could be regulating the dynamics and fitness of their host communities, modifying during infection relevant ecological processes within the C, N and S cycles in this region. This study allows a better understanding of the viral role in the coastal environment of Antarctica, a region of the planet highly affected by global warming.

1. INTRODUCCIÓN

1.1 Microorganismos en el sistema marino antártico

El sistema marino cubre aproximadamente el 70% de la superficie terrestre (Yau & Seth-Pasricha, 2019). Los microorganismos constituyen al 90% de la biomasa celular en el mar, jugando un rol preponderante en los ciclos energéticos y de nutrientes (Suttle, 2007). En los océanos, el proceso de fotosíntesis y producción primaria es llevado a cabo por bacterias autótrofas y por el fitoplancton eucarionte, a través de la fijación del dióxido de carbono (CO₂), dejando al carbono en una forma asimilable por todos los organismos (Pomeroy et al., 2007).

El Océano Austral/Antártico (SO; por sus siglas en inglés) está ubicado al sur del paralelo 60 Sur (IHO, 2000) y es considerado una de las regiones más productivas del planeta, contribuyendo aproximadamente con el 20% de la fijación de carbono a nivel global (Takahashi et al., 2002). La diversidad microbiana en aguas antárticas depende de las variaciones temporales, influenciada por la temperatura, luz, las corrientes y vientos, así como la salinidad (Yau & Seth-Pasricha, 2019). La Península Antártica Occidental (WAP; por sus siglas en inglés) es conocida como la región con los cambios más extremos en el hemisferio sur, la cual ha sufrido un calentamiento sustancial durante los últimos 50 años (Turner et al., 2005). Las aguas marinas en WAP están dominadas por diatomeas en el verano austral, mientras que otros eucariontes como criptófitas y haptófitas pueden formar los principales florecimientos desde primavera a otoño (Yau & Seth-Pasricha, 2019). Estos florecimientos de fitoplancton se asocian directamente con un aumento en la

concentración de clorofila *a* (Chl_a). Se ha descrito que cambios en la abundancia y composición de fitoplancton afecta a la abundancia y composición del bacterioplancton y del virioplancton (Evans et al., 2017; Luria et al., 2016). Sin embargo, en esta región el virioplancton no ha sido caracterizado en profundidad.

1.2 Virus en el ecosistema marino

Los virus son las entidades biológicas más abundantes del planeta, independientemente de la región geográfica o el ecosistema. Además, tienen una alta diversidad, siendo ésta mayor en los océanos (Suttle, 2005; Yau & Seth-Pasricha, 2019). Hace más de tres décadas se describió que en un mililitro de agua marina se puede encontrar aproximadamente 10^8 virus; sin embargo, esto depende de la localización y profundidad de la región marítima, siendo las aguas costeras las que han descrito una mayor abundancia viral (Bergh et al., 1989; Suttle, 2007; Yau & Seth-Pasricha, 2019). Por otro lado, las bacterias son los organismos celulares más abundantes en el ambiente marino, por lo que se piensa que la mayoría de los virus en los océanos corresponden a virus que infectan a bacterias, es decir bacteriófagos, también llamados fagos (Breitbart et al., 2018). Los fagos del sistema marino tienen dos ciclos de vida principales, el ciclo lítico y el lisogénico (Weinbauer, 2004). En el ciclo lítico, el fago inyecta su material genético en la célula hospedera, lo cual dirige inmediatamente a la célula a replicar el material genético del fago, formando partículas virales dentro de la célula hasta producir la lisis celular, liberando las partículas virales hacia el exterior. Por otro lado, en el ciclo lisogénico, el fago al inyectar su material genético dirige su integración en el cromosoma bacteriano (o en algunos casos puede ser mantenido como un elemento

extracromosomal) y de esta forma se puede replicar como profago junto con la célula, hasta que las condiciones ambientales y celulares permitan el paso del ciclo lisogénico al ciclo lítico (Breitbart et al., 2018). A pesar que los virus componen una pequeña parte de la biomasa total en el océano, pueden lisar aproximadamente el 20% de la biomasa celular cada día, cumpliendo un rol muy importante en los ciclos energéticos y de nutrientes (Suttle, 2007). La contribución exacta de los fagos en la mortalidad bacteriana depende de las condiciones ambientales locales y suele variar entre regiones geográficas, donde los estudios disponibles sugieren que los fagos ejercen un importante control sobre el número de bacterias (Fuhrman, 1999; Weitz et al., 2015).

En el caso de las bacterias autótrofas, existen dos mecanismos bióticos principales que generan su mortalidad: la depredación por protistas y la lisis provocada por virus (Wilhelm & Suttle, 1999). Por medio de la depredación, el consumo de bacterias por protozoos permite el movimiento del carbono hacia niveles tróficos superiores, produciendo un aporte a la materia orgánica particulada (POM; por sus siglas en inglés), la cual se hunde en las profundidades del océano secuestrando el carbono transferido, en el proceso denominado “bomba biológica” (Ducklow et al., 2001). Sin embargo, a través de la lisis viral la materia orgánica que forma la biomasa celular queda disponible para ser reutilizada en un nuevo ciclo, pasando a ser parte de la materia orgánica disuelta (DOM; por sus siglas en inglés) en el proceso llamado *viral shunt* (Wilhelm & Suttle, 1999). El proceso de *viral shunt* no solamente contribuye a la recirculación del carbono, sino que está involucrado en otros

procesos biológicos como, por ejemplo, en la disponibilidad de hierro para las comunidades del plancton (Poorvin et al., 2004).

Por otro lado, los virus al infectar a su hospedero pueden generar varios cambios metabólicos en la célula infectada, que pueden asociarse con diferentes mecanismos, los cuales son distintos para cada virus infectivo (Howard-Varona et al., 2020). Uno de los mecanismos utilizados por los virus es la utilización de genes metabólicos auxiliares (AMGs; por sus siglas en inglés), los cuales son genes del virus (ya sea lítico o lisogénico) que se inducen en el hospedero y se expresan generando un cambio en el metabolismo celular (Breitbart et al., 2007). En un inicio, los primeros AMGs descritos estaban involucrados en procesos como la fotosíntesis, el metabolismo de nucleótidos y el ciclo del azufre (Breitbart, 2011). Sin embargo, hoy en día gracias al avance en el análisis de metagenomas virales, la lista de AMGs es bastante extensa e incluso se han propuesto clasificaciones de éstos (Hurwitz & U'Ren, 2016). Debido a las presiones de selección, los altos costos energéticos y las condiciones fisiológicas, se presume que sólo se mantienen los AMG más beneficiosos en las poblaciones virales (Breitbart et al., 2018). Un estudio realizado en 60 genomas de cianofagos marinos, propone que los AMGs más comunes están asociados a presiones de selección estables, mientras que los AMGs raros son producto de la ganancia o la pérdida a través de presiones de selección fluctuantes (Crummett et al., 2016). Hoy en día las herramientas disponibles para el análisis de AMGs han permitido la expansión de su estudio y durante los últimos años varios grupos de investigación se han enfocado en la

búsqueda de nuevos AMGs (Huang et al., 2021; Kieft et al., 2021; Mara et al., 2020; Mizuno et al., 2019; Warwick-Dugdale et al., 2019).

1.3 Virus en el Océano Austral/Antártico

Estudios realizados en el SO han demostrado que el proceso de *viral shunt* contribuye de manera significativa al ciclo del carbono en esta región (Evans et al., 2009). En el SO la razón virus a bacteria (VBR; por sus siglas en inglés), está influenciada por una gran cantidad de factores que controlan la producción y pérdida de virus y bacterias, siendo útil para describir el efecto de la infección viral en los hospederos bacterianos. Las comunidades acuáticas en esta región tienen una VBR considerablemente baja (0.7-6) respecto a lo observado en el resto de los océanos (1-50) (Wommack & Colwell, 2000). Junto con esto, se piensa que los virus son la mayor causa de mortalidad bacteriana en el SO, incluso superando a la actividad bacteriovórica (Guixa-Boixereu et al., 2002; Vaqué et al., 2017).

Actualmente, existen pocos análisis de las comunidades virales en el SO, puesto que el estudio de éstas sólo es posible a medida que aumentan las herramientas disponibles y las posibilidades de muestreos más intensivos en la zona. El estudio a gran escala del *Tara Ocean's GOV 2.0* muestra que las comunidades virales marinas en el mundo pueden ser separadas en cinco zonas ecológicas, siendo una de ellas la Antártica (Gregory et al., 2019). En el estudio de Brum y colaboradores (2016), se estimó que aproximadamente el 90% de los virus de tipo ADN de doble hebra pertenecen al orden *Caudovirales* en la transición de primavera a verano en el SO. Sin embargo, ese estudio tuvo entre un 93% y 95% de lecturas que no fueron asignadas taxonómicamente, hecho que los mismos autores comentan sobre la

necesidad de aumentar los genomas de referencia para perfeccionar este tipo de análisis. En otro estudio realizado por Miranda y colaboradores (2016) se enfocaron en dilucidar la abundancia relativa de los virus del tipo ARN, cuyo resultado reveló que este tipo de virus contribuye entre el 8% y 65% del virioplancton en periodos de florecimiento fitoplanctónico en el verano austral. En este mismo estudio se describe que la mayoría de los virus del tipo ARN en los metagenomas de Antártica estudiados tienen genomas del tipo ARN de hebra simple (+) y son relativamente cercanos a los virus del orden *Picornavirales*. Por otro lado, en otro estudio se observó que la abundancia viral estaba más directamente correlacionada con la dinámica y diversidad del bacterioplancton que con la abundancia de fitoplancton (Evans et al., 2017).

En el ambiente marino de la Antártica existe poca información acerca de los AMGs de las comunidades virales; sin embargo, se sugiere que la presencia de AMGs en ambientes de baja temperatura podría ser una estrategia que asegure el ciclo de vida viral en ambientes extremos (Alarcón-Schumacher et al., 2019). En el estudio de Brum y colaboradores (2016) también se hace una comparación de las secuencias anotadas en la base de datos Pfam como secuencias únicas y categorizadas por función, describiéndose que el 9% de las secuencias que corresponden a AMGs de virus son específicas de la WAP, respecto a todas las secuencias de AMGs en la base de datos de viromas del Océano Pacífico, la cual es específica para metagenomas virales del sistema marino (Brum et al., 2016; Hurwitz & Sullivan, 2013). Por tanto, un estudio de las comunidades virales en el

ambiente antártico puede contribuir a la descripción de la función de los AMGs en este ecosistema.

1.4 Aguas costeras de Bahía Chile

Bahía Chile (62°27'6" S; 59°40'6" O), en las Islas Shetland del Sur, es una región representativa de los ambientes costeros en la WAP (**Figura 1**). En este lugar, ya se han realizado trabajos por nuestro grupo de investigación (liderado por la Dra. Díez), los cuales han tratado de dilucidar la composición de las comunidades eucariontes, procariontes (Alcamán-Arias et al., 2018; Fuentes et al., 2019) y más recientemente virales (Alarcón-Schumacher et al., 2019; Alarcón-Schumacher et al., 2021). En el estudio de Alarcón-Schumacher y colaboradores (2019) se describió la composición de la comunidad viral y los virus con mayor actividad durante un florecimiento fitoplanctónico y en ausencia de éste, en el verano tardío del año 2014. Este estudio se realizó usando metagenomas y metatranscriptomas celulares, lo que permitió analizar las asociaciones entre virus y potenciales hospederos presentes en las muestras. Durante periodos de baja producción primaria, medida a través de la concentración de Chla en el agua de mar, se describió que la familia viral *Myoviridae* era la que domina y que al contrastarlo con periodos de alta Chla, la familia viral *Phycodnaviridae* pasa a ser la dominante. Sin embargo, se determinó que la mayor actividad en ambos periodos es dominada por la familia *Phycodnaviridae*. En este mismo estudio se logró identificar el genoma del fago Pp_CBA, la presencia de genes que codifican una proteína *anti-freezing*, y genes que codifican transportadores de fósforo, lo que podría estar asociado a la adaptación por presencia de AMGs en este ambiente extremo.

Un trabajo posterior realizado por nuestro grupo de investigación comparó la comunidad viral del SO (incluyendo las muestras del verano austral del 2014 de Bahía Chile) con las comunidades virales que fueron extraídas desde la expedición *Tara Oceans GOV 2.0*, las cuales provienen de muestras de distintos ecosistemas marinos alrededor del mundo (Brum et al., 2015). En este estudio se describieron un total de 186 géneros endémicos del SO, demostrando que la comunidad viral en esta zona es divergente genéticamente respecto a las comunidades de otras zonas del mundo (Alarcón-Schumacher et al., 2021). Además en este mismo estudio se describieron distintas adaptaciones de la comunidad viral a temperaturas frías presentes en este océano, como los patrones de hidrofobicidad y las frecuencias de aminoácidos al comparar los perfiles protéicos de las comunidades virales de los distintos ambientes (Alarcón-Schumacher et al., 2021).

En la WAP, el deshielo de glaciares debido al constante aumento de las temperaturas ha impactado en las comunidades microbianas de la zona (Deppeler & Davidson, 2017). Sin embargo, el impacto sobre la comunidad viral marina a través del persistente aumento de temperaturas no ha sido explorado, así como tampoco la comunidad viral presente en los hielos de los glaciares de esta región, los cuales podrían afectar directa o indirectamente a la comunidad microbiana en las aguas circundantes a través de su llegada debido al derretimiento de estos glaciares.

Con todos los antecedentes entregados, y siendo tan importante la función pero tan pocos los estudios de la comunidad viral en el SO, es que se hace necesario realizar nuevos estudios a escala temporal para comprender su dinámica en el ambiente

marino antártico. En este sentido, el laboratorio de la Dra. Beatriz Díez ha recolectado muestras en un mismo punto en la Bahía Chile (punto P3; 62°27'6" S, 59°40'6" O; **Figura 1**) durante cinco años consecutivos (2014, 2016, 2017, 2018 y 2019), desde donde se obtuvieron varios metagenomas. La posibilidad de realizar un estudio a escala temporal de la comunidad viral en el SO es lo que impulsa por tanto esta memoria, para poder así identificar secuencias de origen viral desde metagenomas disponibles de Bahía Chile que nos entreguen información útil y novedosa sobre estas comunidades virales en la costa de la WAP. Actualmente ya se cuenta con una serie de herramientas especializadas para la extracción de secuencias virales desde metagenomas, por lo que visitar muestras ya analizadas con anterioridad utilizando estas nuevas herramientas podría otorgarnos nueva información importante acerca de la comunidad viral. Es por esto que para esta memoria se utilizaron los mismos metagenomas analizados por Alarcón-Schumacher y colegas (2019).

Por otro lado, debido al contexto de calentamiento global que afecta cada verano de manera distinta, un estudio a escala temporal como este podría ayudarnos a entender la dinámica de las comunidades virales según las condiciones de cada verano austral. Además, para evaluar el impacto que podría tener la comunidad viral presente en glaciares aledaños a la Bahía Chile sobre la comunidad microbiana marina, se identificó la comunidad viral presente en el glaciar Fuerza Aérea, el cual se encuentra en las cercanías de esta bahía. Finalmente, al identificar la comunidad viral y su abundancia a escala temporal en Bahía Chile, se abre la posibilidad de identificar genes de importancia para los ciclos biogeoquímicos que estén presentes

en los genomas virales identificados, y de esta manera poder proponer AMG's que puedan estar presentes en este ambiente marino Antártico y su relevancia en la respuesta a los cambios ambientales a los que se tiene que adaptar la comunidad microbiana en estas aguas marinas.

Debido a que desde las muestras disponibles se obtuvieron metagenomas, desde ellos se puede describir la comunidad de virus de doble hebra del tipo ADN y ARN, sin embargo, para la descripción de virus del tipo ARN se necesitaría complementar con metatranscriptomas de estas muestras, por lo que esta memoria se enfocó en la descripción de la comunidad viral del tipo ADN presente en agua marina de Bahía Chile y de hielo glacial.

2. Hipótesis

Las condiciones ambientales y productivas de cada verano (2014-2019) afectan la diversidad de comunidades de virus de ADN de aguas costeras de Bahía Chile (Antártica) y ésta se diferencia de la comunidad viral de hielo glaciar.

La comunidad viral de Bahía Chile porta tanto AMGs nuevos como conocidos, los que están involucrados en los ciclos biogeoquímicos de Antártica y son afectados por las condiciones ambientales y productivas de cada verano (2014-2019).

3. Objetivo General

Determinar la diversidad y composición de las comunidades de ADN virales marinas de Bahía Chile y de hielo del glaciar Fuerza Aérea cercano a la bahía, así como identificar AMGs presentes en sus virus bajo diferentes condiciones presentes a escala temporal interanual entre los veranos de 2014 y 2019.

4. Objetivos específicos

1. Determinar y comparar la diversidad y composición taxonómica de los virus de tipo ADN de aguas marinas de Bahía Chile presentes en cada verano entre 2014-2019, y en hielo del glaciar Fuerza Aérea, a través del análisis de metagenomas.
2. Identificar AMGs presentes en secuencias virales obtenidas desde metagenomas de muestras marinas de Bahía Chile de los veranos del 2014-2019 y en un metagenoma de hielo del glaciar Fuerza Aérea.

5. Materiales y Metodologías

5.1 Sitio de muestreo y procesamiento de muestras utilizadas en esta memoria

Con el fin de obtener muestras representativas de la época estival durante condiciones ambientales contrastantes a escala temporal, se analizaron 11 muestras de agua costera proveniente del punto P3 ($62^{\circ}27'6''$ S; $59^{\circ}40'6''$ O; **Figura 1**) durante los veranos del 2014, 2016, 2017, 2018 y 2019. Adicionalmente, una muestra de hielo del glaciar “Fuerza Aerea” en la bahía fue analizada para evaluar la influencia de la comunidad viral a través del deshielo en la comunidad marina de Bahía Chile. Esta muestra de hielo glaciar se colectó como se describe en el trabajo de Alcamán-Arias y colegas (2021) durante el verano del 2016 (**Tabla 1**).

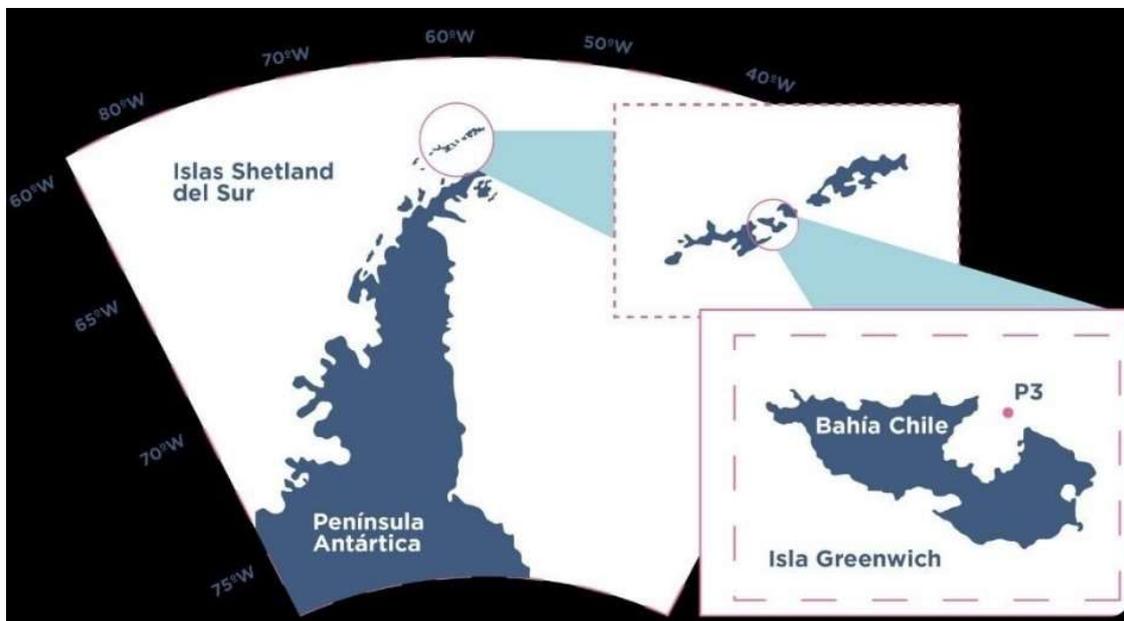


Figura 1.- Mapa de punto de estudio, ubicado en Bahía Chile, Isla Greenwich, Islas Shetland del Sur, Península Antártica.

Se indica el punto de muestreo P3 al interior de la bahía ($62^{\circ}27'6''$ S; $59^{\circ}40'6''$ O).

Tabla 1. **Muestras utilizadas en el estudio**

Muestra	Fecha de muestreo	Punto de muestreo	Profundidad [m]	Periodo	Fuente	Metagenoma	Ensamble	Parte de Coensamble (Ant_CA_1619)
P3-35	14/02/2014	P3	2	Día	Agua marina	Ant_st8_2014	Ant_st8_2014	No
P3-48	04/03/2014	P3	2	Noche	Agua marina	Ant_st22_2014	Ant_st22_2014	No
P3-59	25/02/2016	P3	2	Día	Agua marina	Ant_2_2016	Ant_2_2016	Si
P3-73	21/02/2017	P3	2	Día	Agua marina	Ant_2_2017	Ant_2_2017	Si
P3-75	23/02/2017	P3	2	Noche	Agua marina			
P3-89	03/03/2018	P3	2	Día	Agua marina	Ant_2_2018	Ant_2_2018	Si
P3-90	07/03/2018	P3	2	Día	Agua marina			
P3-113	05/03/2019	P3	2	Día	Agua marina	Ant_2_2019	Ant_2_2019	Si
P3-115	07/03/2019	P3	2	Día	Agua marina			
P3-114	05/03/2019	P3	30	Día	Agua marina	Ant_30_2019	Ant_30_2019	Si
P3-116	07/03/2019	P3	30	Día	Agua marina			
H-03	23/02/2016	Glaciar Fuerza Aérea	N.A.	S.I.	Hielo glaciar	Ant_ICE_2016	Ant_ICE_2016	No

S.I. = Sin información, N.A. = No aplica

Las muestras de agua costera de Bahía Chile se tomaron en superficie a 2 m de profundidad para cada año y adicionalmente se tomaron dos muestras a 30 m de profundidad en el año 2019. Para estas muestras se midieron las variables de temperatura y salinidad con un sensor multiparámetros OAKTON PCD650 para las muestras del 2014 y 2016, y con un sensor multiparámetros CTD *SeaBird19 plus* para las muestras del 2017 al 2019. La variable biológica clorofila *a* (Chla) fue obtenida desde 1 L de agua marina bombeada a través de una red de 150 µm para excluir organismos grandes. Luego se filtró a través de un filtro GF/F de fibra de vidrio de 0,7 µm y fue congelada hasta su procesamiento con los protocolos de extracción con metanol para las muestras del año 2014 (Chorus & Bartman, 1999) o acetona para las muestras del 2016 al 2019 (Strickland & Parsons, 1972), siendo detectada por espectrometría o fluorimetría, respectivamente. Los nutrientes nitrito (NO₂⁻), nitrato (NO₃⁻), fosfato (PO₄⁻³) y ácido silícico (Si(OH)₄) fueron muestreados

del mismo modo que para la Chla y guardados en frascos de polietileno de 15 mL congelados a -20°C en oscuridad hasta su análisis. Las concentraciones de estos nutrientes fueron medidas por bombeo a través de un analizador automático de flujo continuo Technicon AutoAnalyzer ® (AA3 Seal Analytical) como se describe en Alcamán-Arias y colegas (2018). Las variables fisicoquímicas asociadas a todos los metagenomas se muestran en la **Tabla 2**.

Tabla 2. **Variables ambientales asociadas a metagenomas**

Metagenoma	Temperatura [°C]	pH	Salinidad	Chla [mg/L]	NO ₂ [µM]	NO ₃ [µM]	PO ₄ [µM]	Si(OH) ₄ [µM]	NH ₄ [µM]
Ant_st8_2014	-0,1	7,47	33,2	0,36	0,16	19,52	1,61	43,62	S.I.
Ant_st22_2014	0,3	6,44	33,86	2,53	0,098	17,7	1,28	37,9	S.I.
Ant_2_2016	S.I.	6,98	S.I.	2,13	0,329	21,101	1,788	60,76	S.I.
Ant_2_2017	1,695	S.I.	33,88	0,388	0,196	20,455	1,46	43,28	0,775
Ant_2_2018	1,1784	S.I.	34,08205	7,26	0,335	15,94	0,63	29,215	0,98
Ant_2_2019	1,3242	6,6615	34,21285	5,61	0,1905	13,655	1,0615	51,725	2,77
Ant_30_2019	1,17215	6,6605	34,21755	2,52	0,175	17,2945	1,3615	55,35	4,325
Ant_ICE_2016	S.I.	S.I.	S.I.	S.I.	S.I.	S.I.	S.I.	S.I.	S.I.

S.I. = Sin información

Para los análisis de la comunidad microbiana y viral, las muestras de agua marina del año 2014 fueron filtradas por una red de 200 µm y las obtenidas en otros años por una red de 150 µm para excluir partículas y organismos grandes. Luego en el laboratorio, para cada año se filtraron entre 2 y 4 L de esta agua de mar por filtros de poro de tamaño 0,22 µm utilizando una bomba peristáltica (Cole Palmer System Modelo n°7553-70; 6-600 rpm). Luego de la filtración, cada filtro fue inmediatamente congelado a -20°C y mantenidos a -80°C en el laboratorio hasta la extracción de

ADN. La muestra de hielo glaciar fue trasladada al laboratorio en Bahía Chile en una caja estéril y mantenida a temperatura ambiente ($\pm 4^{\circ}\text{C}$) por 24 h hasta su derretimiento completo, tras lo cual fue procesada como se describe arriba para las muestras marinas.

La extracción de ADN de las muestras del 2014 fue realizada según el protocolo de Alcamán-Arias y colegas (2018), mientras que para las demás muestras (incluyendo la de hielo glaciar) se utilizó el protocolo descrito por Cifuentes-Anticevic y colegas (2021). Brevemente, los filtros del año 2014 fueron resuspendidos en amortiguador de lisis (Tris-HCl 10 mM pH8,0, EDTA 1mM, NaCl 0,15 M, 1% dodecilsulfato sódico, 0,1 mg ml⁻¹ de proteinasa K) y perlas estériles de vidrio. Los filtros del resto de los años, se resuspendieron en amortiguador etil xantogenato de potasio (Etil xantogenato de potasio al 1% (Sigma-Aldrich, E.E.U.U.), Tris-HCl 100 mM pH7,4, EDTA 20 mM pH8,0, acetato de amonio 800 mM) y perlas estériles de vidrio. El mix de las muestras del 2014 se incubó a 37°C por 1 h con agitación por vórtex cada 15 min. Luego se agregó NaCl y bromuro de cetrimonio hasta alcanzar una concentración final de 0,6 M y 1%, respectivamente, y luego se incubó a 65°C por 10 min. Mientras que el mix para las muestras de los años 2016 al 2019 fue incubado a 65°C por 2 h. El ADN para todas las muestras se extrajo con fenol cloroformo-alcohol isoamílico (25:24:1) y el fenol residual fue eliminado con cloroformo-alcohol isoamílico (24:1). Se limpió el extracto con dos precipitaciones sucesivas con 1 volumen de isopropanol frío y etanol al 70%.

Las muestras de ADN del año 2014 fueron secuenciadas con tecnología Illumina Hi-seq, para la obtención de fragmentos cortos (150 pb). Para esto, se prepararon las librerías de ADN con el kit NEXTflex® Rapid DNA-Seq (Bioo Scientific

Corporation Austin, TX, E.E.U.U.) según las especificaciones del fabricante. Mientras, las muestras de ADN de los años 2016 al 2019 fueron secuenciadas con tecnología Illumina NovaSeq (Roy J. Carver Biotechnology Center, Illinois, E.E.U.U.) para la obtención de fragmentos cortos (150 pb). Las librerías fueron preparadas con el kit KAPA Hyper Library construction (Kapa Biosystems, Roche).

Una vez obtenidas las secuenciaciones de los metagenomas (**Tabla 1**), los archivos fastq fueron generados y la clasificación de las secuencias obtenidas en función de los oligonucleótidos flanqueantes utilizados en la secuenciación (demultiplexado) con el *software* bcl2fastq v2.20 Conversion (Illumina). Luego se realizó un control de calidad de las lecturas crudas de metagenomas, utilizando el *software* FastQC (Andrews, 2010) y las secuencias fueron sometidas a un recorte de calidad con el *software* PrinSeq (Schmieder & Edwards, 2011) para los metagenomas del 2014, y para los metagenomas de los años 2016-2019 se realizó con el *software* Cutadapt (Martin, 2011), en el que se realizó un corte de las primeras nueve bases del extremo 5' y un corte de calidad en el extremo 3' con una calidad mínima de 30 de Phred. Además, las bases ambiguas y las secuencias de baja complejidad fueron removidas. Desde las lecturas limpias se realizó un ensamble de cada muestra secuenciada utilizando el *software* SPAdes (Bankevich et al., 2012) para los metagenomas del 2014 (Ant_st8_2014 y Ant_st22_2014) y metaSPAdes (Nurk et al., 2017) para el resto de metagenomas (Ant_ICE_2016, Ant_2_2016, Ant_2_2017, Ant_2_2018, Ant_2_2019 y Ant_30_2019). La finalidad de realizar los ensamblajes fue obtener regiones de consenso de ADN a partir de los fragmentos cortos obtenidos desde la secuenciación. Adicional a esto, con los seis metagenomas provenientes de agua marina de los años 2016 al 2019 (**Tabla 1**) se

realizó un coensamble (Ant_CA_1619) utilizando el *software* MEGAHIT (Li et al., 2015) con los parámetros por defecto.

5.2 Obtención de vOTUs a partir de metagenomas

Con el fin de conseguir secuencias virales de cada muestra se utilizaron los ensamblajes y coensamble de metagenomas (**Figura 2**, **Tabla 2** incluye metadata de cada metagenoma). Para cada uno de los ensamblajes se realizó un filtro de tamaño de 10000 pares de bases (pb), para trabajar con aquellos *contigs* de mayor tamaño y mayor grado de confianza de las secuencias virales que se pudieran extraer. Luego, a cada metagenoma se le realizó un análisis a través del flujo de *What the Phage* v.0.8.0 (WtP) (Marquet et al., 2020). Este programa utiliza nueve herramientas de identificación de secuencias de bacteriófagos (VirFinder v1.1, PPR-Meta v1.1, VirSorter v1.0.6, DeepVirFinder v1.0, Metaphinder sin versión de lanzamiento, MARVEL v0.2, sourmash v2.0.1, Vibrant v1.2.1 y VirNet v0.1), en la que cada una de estas nueve herramientas utiliza distintas estrategias para la identificación viral.

Por ejemplo, VirFinder utiliza *machine learning* para identificar frecuencias k-mer, las cuales son subsecuencias de tamaño k contenidas en una cadena nucleotídica. Las frecuencias k-mer virales son perceptiblemente diferentes a las del hospedero (Ren et al., 2017). DeepVirFinder utiliza algoritmos de *deep learning* que permiten la comparación con secuencias virales ya reportadas desde metaviomas ambientales (Ren et al., 2020). PPR-Meta se basa en un algoritmo de *deep learning* capaz de reconocer secuencias cortas, de esta forma es capaz de identificar secuencias de fagos con una mayor sensibilidad respecto a otras estrategias (Fang

et al., 2019). VirSorter utiliza un modelo probabilístico en el contenido génico y características genómicas para clasificar secuencias en distintas categorías de confianza, dependiendo de la probabilidad de identidad viral (Roux et al., 2015). MARVEL tiene un enfoque parecido a VirSorter, utilizando el contenido génico y características genómicas, pero utilizando la estrategia de *machine learning* en vez de un modelo probabilístico (Amgarten et al., 2018). Metaphinder está basado en la comparación de las secuencias contra una base de datos de bacteriófagos de genoma completo (Jurtz et al., 2016). Sourmash utiliza bocetos genómicos MinHash, los cuales son estructuras de resumen aleatorias que permiten la comparación entre ellas utilizando el índice de similitud de Jaccard, otorgándole un puntaje según su similitud a secuencias virales (Brown & Irber, 2016). VIBRANT utiliza un método híbrido de *machine learning* y similitud de proteínas para la recuperación y anotación de secuencias virales (Kieft et al., 2020). Finalmente, VirNet utiliza un modelo de *deep learning* para la identificación de secuencias virales cortas (Abdelkareem et al., 2018). El resultado de este flujo de trabajo nos entrega los contigs positivos para fagos identificados por cada una de las herramientas **(Figura 2)**.

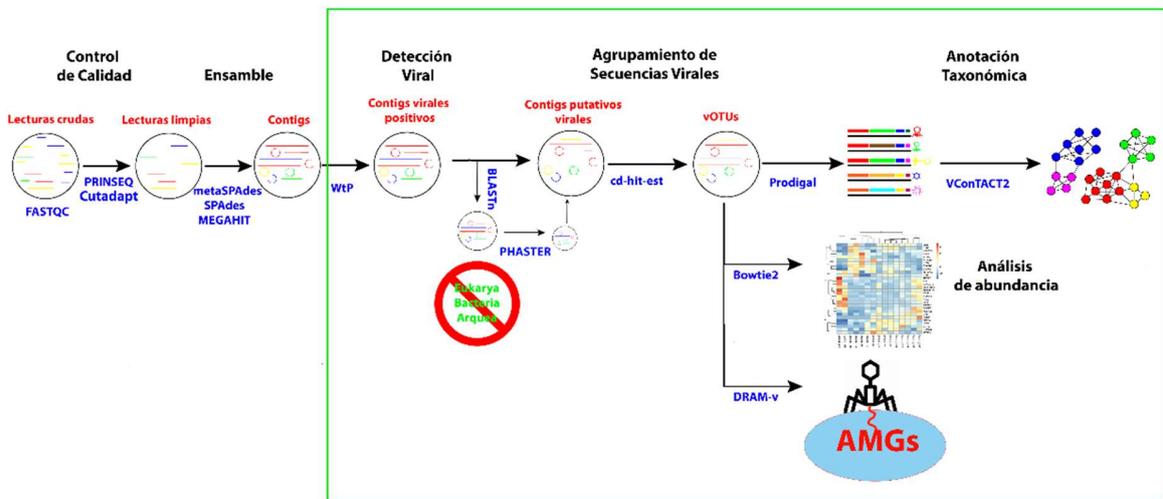


Figura 2.- Diagrama representando flujo de trabajo para el análisis de secuencias virales desde metagenomas.

En negro se titulan los procesos importantes en el flujo, en rojo la naturaleza de los datos en cada paso y en azul se muestran los programas principales para el procesamiento de datos. En el recuadro verde se marcaron los procedimientos realizados en el desarrollo de esta memoria.

Para realizar un control de calidad de las secuencias, se identificaron las secuencias que alinean con alguno de los 3 dominios de la vida. Para esto, solamente a las secuencias reconocidas por WtP como fagos líticos se les realizó un alineamiento contra secuencias celulares usando BLASTn (Camacho et al., 2009) y la base de datos nucleotídica no redundante (NT) del NCBI. De esta forma, se filtraron los *contigs* que alinean con un *e-value* menor a 10^{-7} y con un porcentaje de cobertura mayor al 5% hacia secuencias pertenecientes a los dominios Archaea, Eukarya o Bacteria. Para evitar que se filtraran secuencias insertas en genomas de hospederos, previamente no reconocidas como fagos lisogénicos por WtP, se realizó una búsqueda con el *software* especializado en la identificación de secuencias de profagos PHASTER (Arndt et al., 2016). Éste se basa en la comparación con bases de datos de profagos y la identificación de marcadores de

fagos lisogénicos. De esta forma, los *contigs* identificados por PHASTER no fueron eliminados.

Posteriormente, se concatenaron los archivos fasta que contenían las secuencias virales de todos los ensamblajes analizados, y a la totalidad de *contigs* virales se les realizó un agrupamiento de secuencias que tuvieran un 95% de identidad nucleotídica y un 80% de cobertura de la secuencia más corta hacia la secuencia más larga del grupo utilizando el *software* cd-hit-est (Li & Godzik, 2006). Con esto se obtuvo el set de las “unidades taxonómicas operativas virales” (vOTUs; por sus siglas en inglés) y la secuencia más larga se utilizó como la representativa de cada grupo.

5.3 Asignación taxonómica de vOTUs

Debido a la falta de marcadores filogenéticos universales para virus, durante los últimos años se ha desarrollado una nueva metodología para la asignación taxonómica viral, la consiste en la creación de redes de genes compartidos, basada en agrupaciones de proteínas compartidas entre genomas virales. Una de las herramientas diseñadas para crear redes de genes virales compartidos es VConTACT2 (Bin Jang et al., 2019), la cual tiene como primer paso agrupar las secuencias proteicas, que luego son ordenadas en cada genoma para asociar genomas virales en conglomerados (VC; por sus siglas en inglés). Cuantas más proteínas compartidas entre dos genomas, mayor es la probabilidad de que los dos genomas estén filogenéticamente relacionados (**Figura 3**). En base al número de proteínas compartidas, VConTACT2 combina estrategias de puntuación de confianza para entregar un valor estadístico a la conexión entre secuencias virales.

Estas estrategias de puntuación son dos: la primera otorga un puntaje de confianza basado en la topología del conglomerado y la segunda otorga un puntaje según la probabilidad de que cada grupo corresponda a un género aprobado por el Comité Internacional de Taxonomía de Virus (ICTV; por sus siglas en inglés) (Bin Jang et al., 2019).

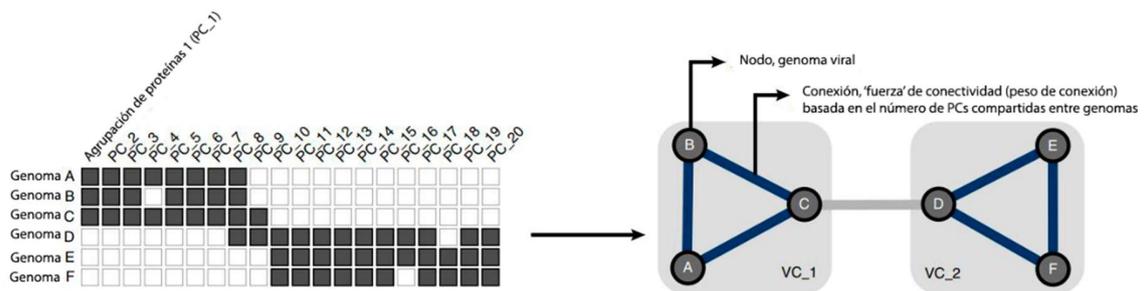


Figura 3.- Generación de redes con VConTACT2.

En la figura se muestra el proceso de generación de agrupaciones virales (VC) a través del análisis de conexiones basadas en el número de agrupaciones proteicas compartidas (PCs). *Imagen modificada de Bin Jang et al., 2019.*

Con el fin de crear esta red, al set de vOTUs se le realizó la predicción de proteínas a través del *software* Prodigal v2.6.3 utilizando la opción meta y sin tomar en cuenta las secuencias Shine-Dalgarno (Hyatt et al., 2010). Una vez obtenidas las secuencias proteicas de cada set de vOTUs, se utilizó el *software* VConTACT2 v9.19 (Bin Jang et al., 2019) en el formato de aplicación de la plataforma CyVerse (Merchant et al., 2016), siguiendo el protocolo disponible en la plataforma, bajo los parámetros por defecto y se utilizó como referencia la base de datos de virus de arqueas y bacterias del RefSeq v.201 con la taxonomía del ICTV junto con el NCBI.

La red de genes compartidos fue visualizada con el *software* Cytoscape (Shannon et al., 2003), en el cual cada nodo representaba a un vOTU (perteneciente al set de metagenomas de Bahía Chile o a alguna de las referencias de la base de datos) y

las conexiones fueron representadas por líneas que conectan entre nodos. Se utilizó el diseño por defecto otorgándole la conexión según el valor estadístico de cercanía entre vOTUs que nos entrega VConTACT2.

Para la asignación taxonómica a nivel de género, se buscaron los VC que contuvieran tanto nodos de vOTUs provenientes de los metagenomas de Bahía Chile como nodos de referencia, asignándolos al mismo género a la cual pertenecía la referencia presente en el VC. Tanto para los vOTUs pertenecientes a VCs que no contuvieran referencias, como para los que no fueron incluidos dentro de algún VC específico (pero que se agrupaban visualmente con otros vOTUs donde podría haber referencias), se realizó una asignación taxonómica a nivel de familia con la herramienta de difusión de calor de Cytoscape para encontrar el nodo estadísticamente más cercano que representara a una referencia. De esta manera, se asignó taxonómicamente al vOTU a la misma familia a la que perteneciera la referencia más cercana al mismo.

Por otra parte, para poder identificar el conglomerado de nodos que no contenía ninguna secuencia de referencia y poder asociarlo a algún grupo viral, se realizó un mapeo de las lecturas de los metagenomas contra estos *contigs* con el *software* Bowtie2 (Langmead & Salzberg, 2012). Luego se extrajeron las lecturas *forward* y a estas lecturas se les realizó una anotación contra la base de datos proteica no redundante (NR) del NCBI utilizando DIAMOND (Buchfink et al., 2015) con la opción --taxonlist 10239 para obtener la asignación hacia algún grupo viral. Este resultado fue visualizado a través de MEGAN (Huson et al., 2007).

Para corroborar la asignación a la familia *Phycodnaviridae*, se realizó una búsqueda de los vOTUs a través del *software* VirSorter2 (Guo et al., 2021) que permite el reconocimiento y procesamiento de secuencias pertenecientes al grupo de *Nucleocitoviricota* (NCLDV), a través de marcadores virales. Se asignaron solo los vOTUs que fueron reconocidos por VirSorter2 y que tuvieron al menos un *hallmark* para el grupo NCLDV. Su asignación fue realizada según el mejor *hit* con un *e-value* menor a 10^{-7} , comparando contra la base de datos NT del NCBI, utilizando la herramienta BLASTn.

5.4 Mapeo de lecturas contra vOTUs y construcción de tabla de conteos.

Con el fin de reconocer la abundancia de cada vOTU en cada muestra, se realizó un mapeo de las lecturas limpias de cada metagenoma sobre el set de vOTUs. Esto se realizó con el *software* Bowtie2, el cual genera archivos de salida con extensión .sam. Para manejar este tipo de archivos se utilizaron las herramientas samtools (Li et al., 2009) y bbmap (Bushnell, 2014). Con samtools se transformaron los archivos de salida a un archivo binario de extensión .bam para permitir su manejo, y desde éstos, se extrajo el número de lecturas de cada metagenoma, las que fueron mapeadas contra cada vOTU. Con bbmap y la opción pileup.sh se extrajeron las estadísticas del mapeo.

Con la información del mapeo se construyó una tabla de conteos con R (R Core Team, 2020). Considerando un conteo como una lectura sobre un vOTU, primero fueron filtrados los conteos que no tuvieran un porcentaje de cobertura mayor al 75% de cada vOTU. Luego, se realizó una fusión de las tablas de conteos de cada metagenoma, agregando la información del largo de cada vOTU, para obtener así

la tabla de conteos cruda. A modo de normalización, se utilizó la media truncada de valores M (TMM; por sus siglas en inglés) para eliminar los sesgos de composición entre librerías (Robinson & Oshlack, 2010). Por otro lado, posteriormente a la normalización por TMM se normalizaron los conteos de lecturas por el largo de cada vOTU y posteriormente por la profundidad de la secuenciación, obteniendo la normalización de transcritos por millón de kilobases (TPM; por sus siglas en inglés). Todas las normalizaciones se realizaron con el paquete estadístico edgeR (Robinson et al., 2010).

5.5 Anotación y validación de AMGs

La asignación de AMGs y análisis funcional de fagos asociado al metabolismo microbiano, no es fácil de abordar por tres razones: (1) La disponibilidad de secuencias proteicas es variable según las bases de datos, (2) puede existir un sesgo por contaminación de secuencias bacterianas dentro de las secuencias virales, debido a que algunas herramientas de extracción de secuencias virales pueden dejar restos de secuencias celulares y (3) el consenso de las características mínimas para considerar una proteína dentro de una secuencia viral como AMG aún se encuentra en discusión.

Para abordar esta problemática y generar una herramienta que pudiera procesar una gran cantidad de información en secuencias, el año pasado se publicó DRAM (*Distilled and Refined Annotation of Metabolism*). Esta herramienta se especializa en la caracterización de comunidades microbianas y virales, gracias a la extensión DRAM-v específica para la búsqueda de AMGs en vOTUs (Shaffer et al., 2020).

Esta herramienta es capaz de predecir las proteínas presentes en un set de vOTUs y comparar sus secuencias aminoacídicas contra las bases de datos KEGG, UniRef90, MEROPS y la base de datos de proteínas virales del NCBI. También es capaz de realizar búsquedas de perfiles ocultos de Markov (HMM; por sus siglas en inglés) y realizar la búsqueda de los mismos en las bases de datos del Pfam, KOfam, VOGDB y CAZy. Luego de la anotación, se asigna un puntaje auxiliar del 1 al 5 a cada gen dentro de un vOTU. De esta forma, un puntaje igual a 1 representa un gen que tiene una alta probabilidad de estar codificado en un virus; en cambio, un puntaje de 4 o 5 representa un gen para el que hay que tener precaución para designarlo como viral. Este puntaje dependerá del contexto génico de cada gen, específicamente de la presencia o no de *hallmarks* virales y/o genes del tipo viral flanqueando al gen de interés. Finalmente, dependiendo de las características del potencial AMG, se compara con AMGs ya validados en la literatura para realizar el proceso de “destilado”, que consiste en crear un resumen de vOTUs con todos los potenciales AMGs asociados a un proceso metabólico, sea conocido o no.

Para aplicar esta metodología dentro del set de vOTUs utilizados en la presente memoria, fue necesario primero utilizar el *software* VirSorter2 para generar categorías de vOTUs y de esta forma obtener los archivos de entrada necesarios para el proceso de anotación y destilado en DRAM-v (que solo utiliza estos archivos de entrada).

5.6 Análisis estadísticos y construcción de gráficos

Todos los análisis estadísticos fueron realizados en R, utilizando el paquete microbiome (Leo Lahti & Shetty, 2012), mientras que los gráficos fueron realizados en R con el paquete ggplot2 (Wickham, 2016).

6. Resultados

6.1 Identificación de vOTUs a partir de metagenomas de muestras de mar de la Bahía Chile y de hielo del glaciar Fuerza Aérea

Las muestras obtenidas desde Bahía Chile y la muestra obtenida desde hielo glaciar fueron sometidas a extracción de ADN y posterior secuenciación por Illumina, obteniéndose secuencias cortas (150 pb), las cuales fueron limpiadas y ensambladas para obtener las secuencias que finalmente fueron utilizadas en esta memoria. Desde los ensambles de 7 metagenomas de agua de mar y uno de hielo glaciar, y el coensamble de todos los metagenomas de las muestras de agua de mar entre 2016 y 2019, se consiguieron extraer 2842 *contigs* reconocidos como virales por las herramientas del flujo de trabajo WtP (*contigs* virales positivos). De éstos, 372 *contigs* alinearon contra secuencias de los dominios Eukarya, Bacteria o Archaea, y tras la evaluación de presencia de fagos lisogénico con PHASTER, se reconocieron 21 *contigs* virales como profagos. De esta manera, en total se recuperaron 2491 *contigs* putativos virales que posteriormente fueron agrupados por identidad y cobertura para obtener el set de vOTUs final con el que se trabajó en esta memoria. En total, se obtuvo un set de 1966 vOTUs, a los cuales se les realizó el mapeo de las lecturas limpias para estimar la abundancia en cada metagenoma. De la totalidad de vOTUs se pudieron diferenciar aquellos provenientes de agua marina de los de hielo glaciar, esto se muestra en un diagrama de Venn donde hay un total de 225 vOTUs específicos de hielo glaciar y 1741 vOTUs específicos de los ensambles de las diferentes muestras de Bahía

Chile entre los veranos del 2014 al 2019 (**Figura 4**). Ningún vOTU fue compartido entre ambos ambientes (hielo glaciar y agua de mar).

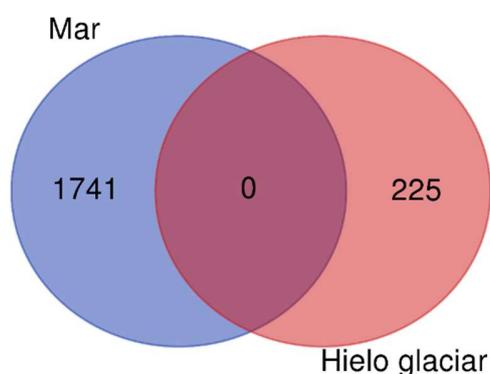


Figura 4.- Diagrama de Venn con la comparación de vOTUs presentes en muestras marinas de Bahía Chile y la muestra de hielo del glaciar Fuerza Aérea.

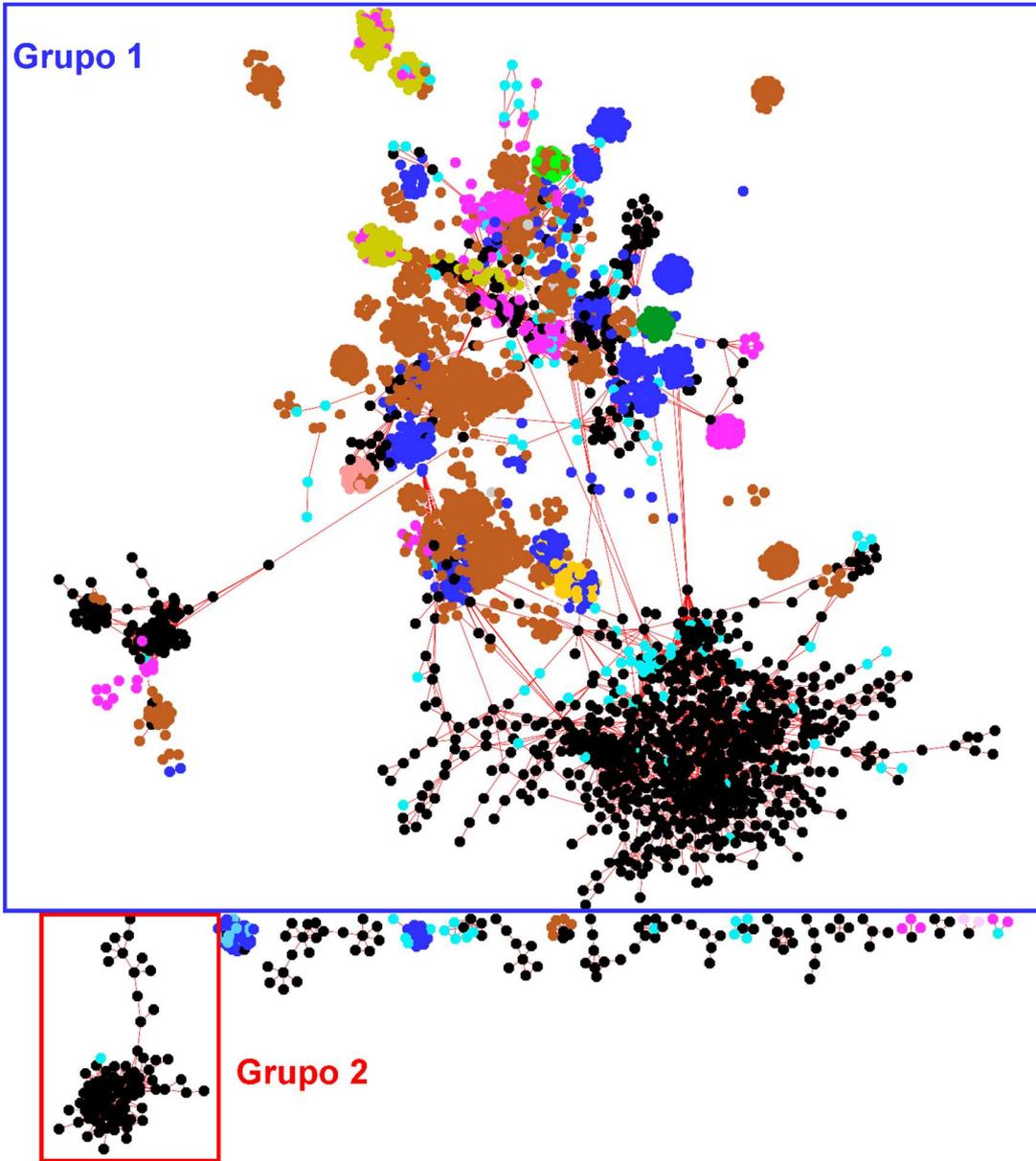
En azul se muestra el número de vOTUs presentes en las muestras marinas de Bahía Chile entre el 2014 y el 2019, mientras que en rojo se muestran los vOTUs presentes en la muestra de hielo del glaciar Fuerza Aérea.

6.2 Asignación taxonómica de vOTUs mediante redes de genes compartidos

Para asignar taxonómicamente a los vOTUs, se realizó la predicción de las proteínas contenidas en cada vOTU para crear una red de genes compartidos, la cual consiste en otorgar una conexión de cercanía a cada vOTU hacia el resto, usando referencias obtenidas de la versión 201 de la base de datos de virus de Arqueas y Bacterias del NCBI. Se espera que cuando más proteínas son compartidas, mayor es la cercanía en la red, y por lo tanto mayor es la probabilidad de que los vOTUs estén relacionados taxonómicamente. Desde el resultado de la red de genes compartidos del set de vOTUs extraídos desde los 8 ensamblajes (mar y la muestra única de hielo) y el coensamble (solo marino), fue posible asignar a nivel de familia 1091 vOTUs (**Figura 5**), mientras que los 875 vOTUs restantes no lograron ser asignados a ningún nivel taxonómico. En la red se obtuvo una agrupación mayoritaria de géneros virales que fue designada como Grupo I.

Además, se observó la presencia de una agrupación de 91 vOTUs que no tienen parecido a ningún virus conocido de las bases de datos y que fue designado en esta memoria como Grupo 2. En la red también existieron muchos vOTUs que no hicieron conexión con ningún otro vOTU, éstos se consideraron como *singletons* y fueron eliminados de la red final ya que no otorgan información (**Figura 5a**). Otros pequeños módulos de la red solo contenían vOTUs de Bahía Chile, que se agrupan solo entre ellos, sin incluir ningún vOTU de referencia, estos nodos quedaron sin asignación taxonómica.

(a)



- vOTUs Bahía Chile
- vOTUs hielo glaciar
- *Ackermannviridae*
- *Autographviridae*
- *Corticoviridae*
- *Demereciviridae*
- *Drexleriviridae*
- *Herelleviridae*
- *Myoviridae*
- *Podoviridae*
- *Siphoviridae*
- Unassigned



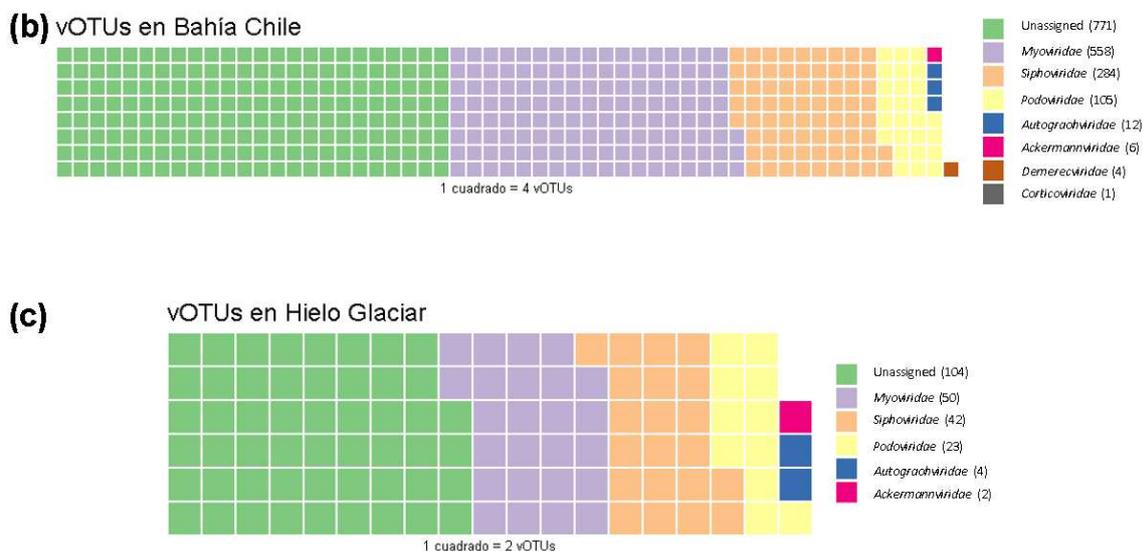


Figura 5.- Red de genes compartidos en muestras de agua de mar de Bahía Chile y una muestra de hielo del glaciar Fuerza Aérea.

(A) Cada círculo en la red representa un nodo, el cual representa al genoma de un virus (sea de referencia o de las muestras de este estudio de Bahía Chile o del glaciar Fuerza Aérea). Cada línea roja de unión entre dos nodos representa la conexión entregada según la función de probabilidad realizada por VConTACT2. En la imagen solo se muestran las conexiones que involucran nodos de Bahía Chile o del Glaciar Fuerza Aérea. Las referencias se colorearon según la familia a la cual pertenecían, mientras que las muestras provenientes del estudio en Bahía Chile se colorearon de negro y las del Glaciar Fuerza Aérea se colorearon de celeste. Encerrado en un rectángulo azul se encuentra el conglomerado más grande, denominado Grupo 1. Encerrado en un rectángulo rojo se muestra el agrupamiento designado como Grupo 2 en esta memoria y que contiene 91 vOTUs desconocidos, ya que no agrupa con genomas de referencia. Se resumieron las asignaciones taxonómicas de los vOTUs provenientes de **(B)** muestras marinas de Bahía Chile y de **(C)** vOTUs provenientes de hielo del Glaciar Fuerza Aérea, rescatadas desde la red de genes compartidos.

En la red, las familias representadas en el set de vOTUs marinos (nodos coloreados de negro; **Figura 5a**) fueron: *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Ackermannviridae*, *Autographviridae*, *Podoviridae*, *Demereciviridae* y *Corticoviridae* (**Figura 5a y 5b**). Por otra parte los vOTUs específicos de la muestra de hielo glaciar (nodos coloreados celeste, **Figura 5a**) se vieron representados solo por las familias: *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Autographviridae* y *Ackermannviridae* (**Figura 5a y 5c**). En casi su totalidad, estas familias están representando al orden

Caudovirales solamente la familia *Corticoviridae* pertenece al orden *Vinavirales*. De esta última familia solo se pudo asignar un vOTU marino del total de vOTUs asignados.

En cuanto a la asignación a nivel de género, solo pudieron ser asignados 15 vOTUs, 13 de ellos a géneros marinos como *Canoevirus*, *Rigallicivirus*, *Vhulanivirus*, *Aresaunavirus* y dos géneros adicionales no clasificados dentro de la familia *Siphoviridae*. Otros tres géneros de la familia *Podoviridae* no pudieron ser clasificados (**Tabla 3**). Dos de los vOTUs asignados a los géneros *Myxovirus* y *Atuphduovirus*, provienen de hielo glaciar.

Tabla 3. **Asignación a nivel de género por VCONTACT2**

vOTUs	Género	Viral Cluster (VC)	Origen
BC_st22_2014_C_22	<i>Canoevirus</i>	VC_103_0	Marino
BC_st22_2014_C_311	<i>Unclassified Siphoviridae</i>	VC_185_0	Marino
BC_2_2018_C_118	<i>Unclassified Podoviridae</i>	VC_238_0	Marino
BC_st22_2014_C_277	<i>Rigallicivirus</i>	VC_271_0	Marino
BC_30_2019_C_183; BC_CA_1619_C_496; BC_2_2018_C_70	<i>Unclassified Podoviridae</i>	VC_275_0	Marino
BC_st22_2014_C_189	<i>Unclassified Siphoviridae</i>	VC_401_0	Marino
BC_st22_2014_C_266; BC_st22_2014_C_178; BC_st22_2014_C_296	<i>Vhulanivirus</i>	VC_451_0	Marino
BC_ICE_2016_C_210	<i>Myxovirus</i>	VC_591_0	Hielo glaciar
BC_st22_2014_C_211	<i>Aresaunavirus</i>	VC_60_0	Marino
BC_ICE_2016_C_78	<i>Atuphduovirus</i>	VC_86_0	Hielo glaciar
BC_st22_2014_C_314	<i>Unclassified Siphoviridae</i>	VC_90_0	Marino

Por otro lado, para conocer la abundancia de cada familia en cada metagenoma, se mapearon las lecturas limpias de cada metagenoma al set de vOTUs. Los conteos se agruparon a nivel de familia viral (**Figura 6**), verificándose la presencia mayoritaria de las familias *Myoviridae* (entre 50.7% y 61.3%), *Siphoviridae* (entre 28.1% y 41%) y *Podoviridae* (entre 0.1% y 18.5%) pertenecientes al orden

Caudovirales, en el total de las muestras provenientes de Bahía Chile (ver **Figura 6**). *Myoviridae* y *Siphoviridae* fueron las más abundantes, inclusive en la muestra de hielo glaciario (44.1% *Myoviridae*, 35.5% *Siphoviridae* y 17.6% *Podoviridae*). En particular la familia *Podoviridae* fue menos abundante en las muestras del 2014 y 2017.

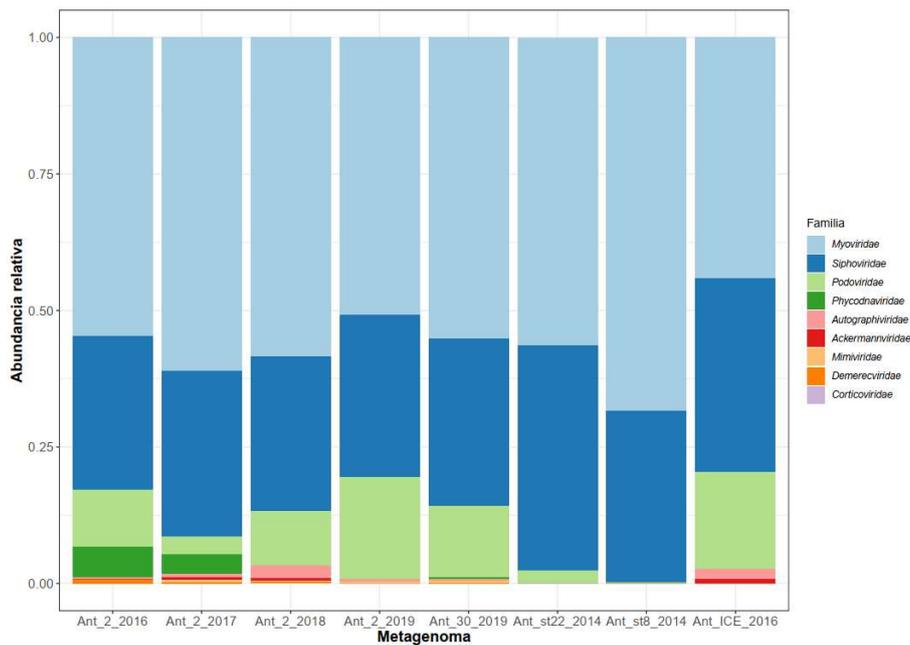


Figura 6.- Abundancia relativa de las familias virales presentes en muestras de agua marina de Bahía Chile y de hielo del glaciario Fuerza Aérea.

Se agruparon los conteos por familias y se realizó un gráfico de composición a partir de las secuencias que fueron asignadas a este nivel taxonómico.

Adicionalmente a estas familias más dominantes, otras familias dentro de los caudovirales aparecen también representadas en algunas de las muestras como son *Autographviridae* (entre 0.22% y 2.2%), *Ackermannviridae* (entre 0.01% y 0.4%), ambas ausentes en las muestras del año 2014, y *Demereciviridae* (entre 0.01% y 0.7%) ausente en las muestras del 2014 y del 2019 a 2 m de profundidad. De estas familias, en la muestra de hielo glaciario solo se encuentran presentes

Autographviridae y *Ackermanviridae* (1.9% y 0.8%, respectivamente). Por otro lado, la familia *Corticoviridae*, perteneciente al orden *Vinavirales*, estuvo presente en la muestra de alta Chla (Ant_st22_2014) del 2014 (0.02%).

Cabe destacar la presencia adicional de miembros de la familia *Phycodnaviridae* (0.1%-5.5%), representado mayoritariamente por un género no clasificado de *Pymnesiovirus*, el cual se encuentra presente en las muestras marinas entre los años 2016 y 2019. Algunos conteos fueron también asociadas a otros virus gigantes de la familia *Mimiviridae* (0.2%-0.4%) entre los años 2017 y 2019.

6.3 Asignación taxonómica de agrupamientos en la red sin asociación a referencia conocida

Para identificar la afiliación de los 91 vOTUs que pertenecen al agrupamiento más grande observado entre los vOTUs que no agrupan con ninguna referencia en la red (Grupo 2), se realizó un mapeo de las lecturas de estos 91 vOTUs y se asignaron a través de BLASTn, permitiendo una asignación total del 37.9% de las lecturas mapeadas (**Figura 7**).

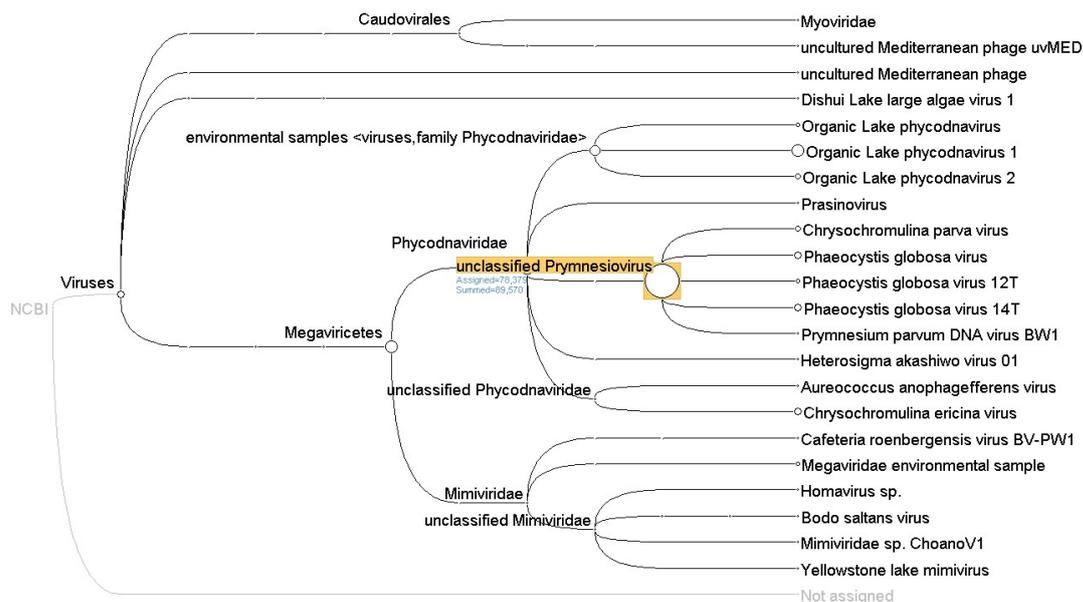


Figura 7.- Asignación de lecturas mapeadas contra el Grupo 2 de vOTUs desconocidos. La visualización de la asignación de lecturas fue realizada a través de MEGAN. El tamaño de cada nodo representa el número de lecturas asignadas al nivel. En naranja se marca el nivel taxonómico que se lleva la mayor parte de lecturas asignadas.

Según la asignación realizada, la mayor parte de las lecturas del Grupo 2 estarían asociadas a un género no clasificado de *Prymnesiovirus*, al cual pertenecen las especies de virus infectivas de *Phaeocystis globosa*, perteneciente al fitoplancton y que ha sido ampliamente descrita en aguas marinas antárticas (Figura 7). Debido a que la mayor parte de estas lecturas son reconocidas como pertenecientes al filo NCLDV, también conocidos como virus gigantes, se utilizó la herramienta VirSorter2 para corroborar que los 91 vOTUs podrían ser asignados a este grupo. A través de esta herramienta, se reconocieron 89 secuencias completas de las 91 como pertenecientes al grupo NCLDV. Sin embargo, solo 24 secuencias contenían al menos un *hallmark* de NCLDV, por lo que solo estas 24 fueron utilizados para su asignación final, en la cual el resultado del mejor *hit* contra la base de datos NT fue obtenido para cada uno, como se muestra en la Tabla 4. Cabe destacar que los 24

vOTUs asignados pertenecen al set de vOTUs marinos provenientes de Bahía Chile.

Tabla 4. **Resultados de BLASTn sobre vOTUs reconocidos como NCLDV**

vOTU	ID mejor hit	Mejor Hit
BC_2_2017_C_241	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_CA_1619_C_89	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_2_2016_C_51	HQ704802.1	Organic Lake phycodnavirus 1 genomic sequence
BC_2_2018_C_81	KT445356.1	Megaviridae environmental sample clone WC2-12 major capsid protein gene
BC_CA_1619_C_553	KT445356.1	Megaviridae environmental sample clone WC2-12 major capsid protein gene
BC_CA_1619_C_655	HQ704802.1	Organic Lake phycodnavirus 1 genomic sequence
BC_CA_1619_C_415	XM_023476618.1	<i>Eurytemora affinis</i> zinc finger protein 665-like
BC_CA_1619_C_453	NA	NA
BC_2_2017_C_30	KT445759.1	Megaviridae environmental sample clone Q1-B2-13 major capsid protein gene
BC_2_2017_C_105	LT635614.2	<i>Plasmodium vivax</i> genome assembly, chromosome
BC_30_2019_C_174	MN448299.1	Megaviridae environmental sample isolate oPacV-421 clone oPacV-421_c10
BC_30_2019_C_193	KT445759.1	Megaviridae environmental sample clone Q1-B2-13 major capsid protein gene
BC_CA_1619_C_243	KJ558371.1	<i>Prymnesium kappa</i> virus isolate PkV_RF02 major capsid protein (MCP) gene
BC_CA_1619_C_673	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_CA_1619_C_511	NA	NA
BC_CA_1619_C_548	XM_017938062.1	<i>Habropoda laboriosa</i> ribonucleoside-diphosphate reductase subunit M2 B
BC_CA_1619_C_876	HQ704802.1	Organic Lake phycodnavirus 1 genomic sequence
BC_2_2017_C_34	CP018793.1	<i>Campylobacter</i> sp. RM12175
BC_2_2017_C_286	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_2_2019_C_142	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_CA_1619_C_148	XM_001454265.1	<i>Paramecium tetraurelia</i> hypothetical protein
BC_CA_1619_C_317	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_CA_1619_C_726	KC662249.1	<i>Phaeocystis globosa</i> virus strain 16T
BC_CA_1619_C_243	KJ558371.1	<i>Prymnesium kappa</i> virus isolate PkV_RF02 major capsid protein (MCP) gene

*N.A. = Ninguna asignación

6.4 Distribución temporal y abundancia relativa de vOTUs en Bahía Chile y en hielo glaciar de Fuerza Aerea

Para evaluar la abundancia de los vOTU obtenidos por cada muestra de agua de mar y en la de hielo, se realizó un mapeo de las lecturas limpias provenientes de la secuenciación hacia el set de vOTUs. Para este análisis solo se incluyeron vOTUs de >1% de abundancia relativa entre muestras (209 vOTUs), incluyendo los vOTUs

que no fueron asignados taxonómicamente (18 vOTUs de los 209 vOTUs mayores al 1%).

Este análisis de abundancias a nivel de vOTUs mostró que existen diferencias entre muestras, representadas en *clusters* diferentes en la **Figura 8**. En especial hubo una clara separación de las muestras de agua de mar y la proveniente de hielo glaciar, donde todas las lecturas de esta última no presentan abundancia alguna en los vOTUs del resto de las muestras, indicando su ausencia en las muestras de agua de mar de Bahía Chile analizadas en este estudio (**Figura 8**). Las dos muestras del año 2014 se asociaron por similitud, separándose del resto de las muestras de otros años, aunque también se detectaron algunas variaciones entre sus poblaciones de vOTUs. Por otro lado, las muestras de mar de los años 2016 y 2017 fueron más similares entre ellas y comparten ciertas poblaciones de virus que no aparecen en el resto de los años estudiados.

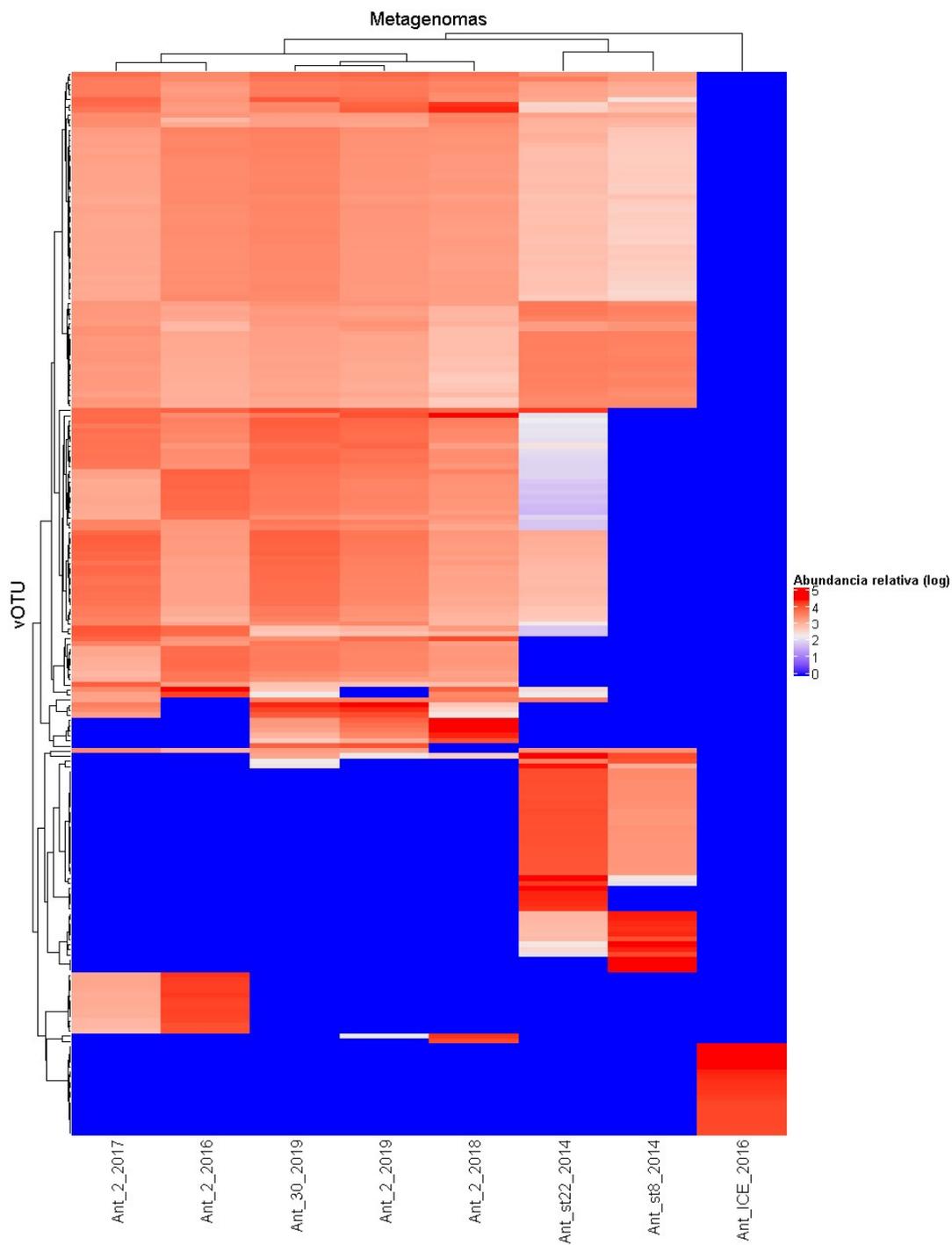


Figura 8.- Heatmap de abundancias relativas a nivel de vOTUs entre muestras utilizadas en el estudio.

Se compararon las abundancias de los vOTUs que tuvieran >1% de abundancia relativa entre muestras, incluyendo los vOTUs que no fueron asignados. Se realizó un agrupamiento jerárquico de metagenomas y de vOTUs mostrado en la parte superior y a la izquierda de la figura.

6.5 Búsqueda de AMGs presentes en la comunidad viral marina de Bahía Chile y en muestra de hielo del glaciar Fuerza Aérea

En la búsqueda de potenciales AMGs a través de la herramienta DRAM-v, este trabajo consideró solamente a las proteínas anotadas con un puntaje igual a 1. Este puntaje considera únicamente a los genes que estuvieran flanqueados por al menos un *hallmark* viral a cada lado del gen de interés.

En la totalidad de las muestras, tanto marinas como de hielo glaciar, se identificaron 10 AMGs distribuidos en 14 vOTUs, los cuales cumplen distintas funciones a nivel celular (**Figura 9**). Se describió la presencia de la familia M17 de aminopeptidasas en 3 vOTUs pertenecientes a la familia *Myoviridae*. También se identificó el gen *dcm* que codifica la metiltransferasa de citosina. Este AMG aparece tanto en muestras marinas como en la muestra de hielo glaciar, pero en esta última no fue posible una asignación taxonómica a nivel de familia. Para el caso de las muestras de mar, éste fue afiliado a dos vOTUs, uno de la familia *Myoviridae* y otro a la familia *Siphoviridae* (**Figura 9**).

Además, se identificaron dos genes pertenecientes a dos familias distintas de glicosiltransferasas, GT2 y GT102. La primera fue identificada en dos vOTUs, uno perteneciente a la familia *Phycodnaviridae* y el otro a la familia *Myoviridae*. Por otro lado, la familia de glicosiltransferasas GT102, fue identificada en un vOTU sin asignación taxonómica. Otros tipos de AMGs identificados fueron tres genes involucrados en la síntesis de nucleótidos, específicamente en la producción de timidina. Los tres genes fueron identificados en vOTUs pertenecientes a distintas familias. El gen *dut*, que codifica la enzima deoxiuridina trifosfatasa, fue encontrado

en un vOTU de la familia *Phycodnaviridae*. Por su parte, el gen *dcd*, que codifica la deaminasa dCTP, fue identificado en un vOTU de la familia *Siphoviridae* y en un vOTU que no pudo ser asignado taxonómicamente. El último gen involucrado en la síntesis de timidina encontrado en vOTUs de Bahía Chile, fue el gen *thyA*, el cual codifica la enzima timidina sintentasa y fue identificado en 2 vOTUs pertenecientes a la familia *Myoviridae* y *Podoviridae* (**Figura 9**).

Por otro lado, se identificaron dos AMGs involucrados en la estructura de la subunidad menor del ribosoma, *rpsA* y *rpsU*, los cuales codifican para las proteínas RP-S1 y RP-21, respectivamente. Cada uno forma parte de un vOTU distinto, ambos pertenecientes a la familia *Myoviridae*. Finalmente, también se identificó el gen *cobS* que codifica la quelatasa de cobalto, involucrada en la síntesis de cobalamina. Este gen fue identificado en un vOTU de la familia *Myoviridae*.

Cabe destacar que la gran mayoría de vOTUs portadores de AMGs solo contienen un único AMG. Sin embargo, se identificaron dos vOTUs pertenecientes a la familia *Myoviridae* que contienen más de un AMG, uno de ellos contiene genes que codifican la glicosiltransferasa de la familia GT2 y el gen *rpsA*. Por su parte, el otro contiene un gen que codifica una proteína de la familia de aminopeptidasas M17 y los genes *rpsU* y *cobS*.

Para conocer la abundancia de cada vOTU portador de AMG, se utilizó el mapeo de lecturas limpias hacia la totalidad del set de vOTUs para calcular la abundancia relativa de cada uno de los vOTUs aquí descritos (**Figura 9**). Respecto al resultado de abundancias, las muestras del 2018 contienen la mayor cantidad de vOTUs portadores de AMGs, mientras que en la muestra de baja Ch1a del 2014 (Ant_st8_2014) no se encontró ningún vOTU portador de AMGs.

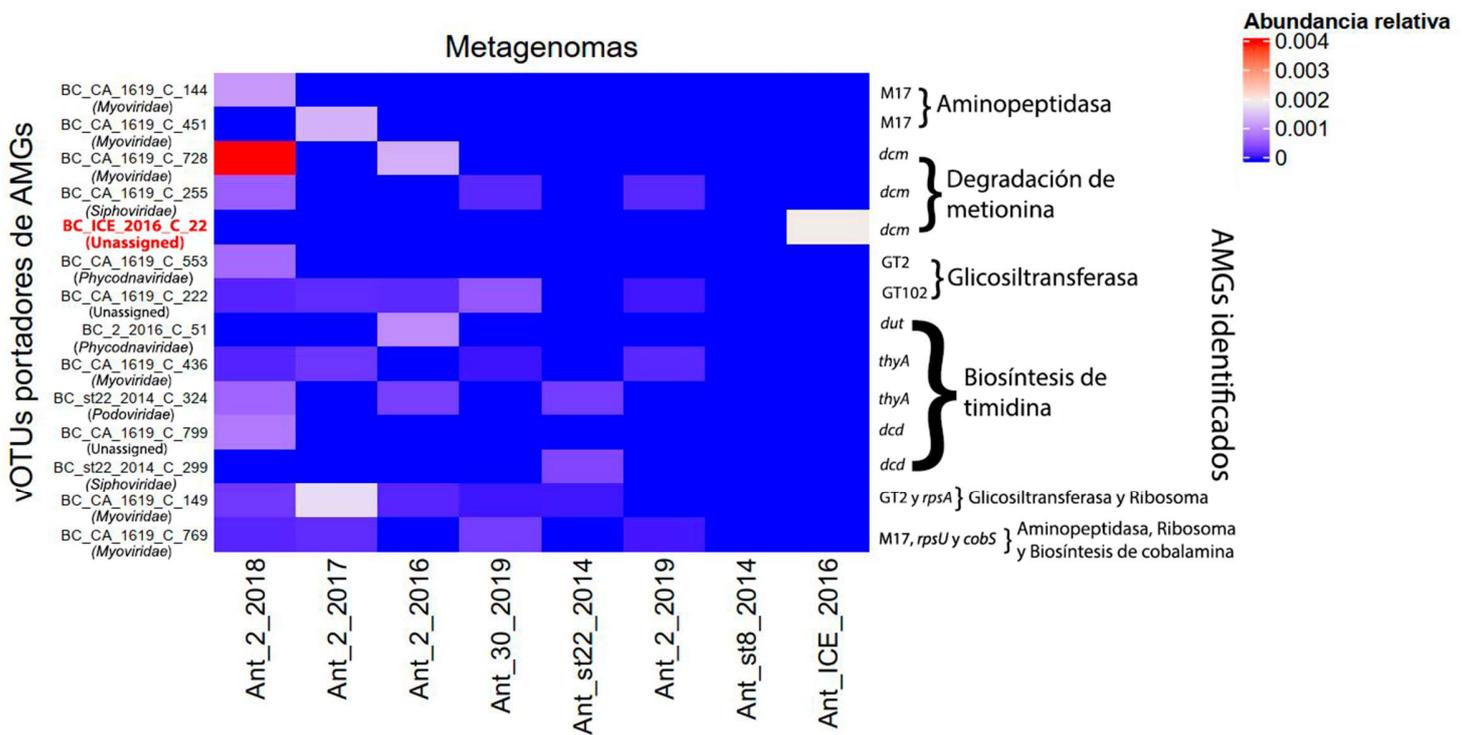


Figura 9.- Abundancias de los vOTUs portadores de AMGs en los metagenomas marinos de Bahía Chile y en la muestra de hielo del glaciar Fuerza Aérea.

En la imagen se muestran las abundancias relativas a la totalidad de vOTUs presentes en cada metagenoma que portan AMGs descritos en esta memoria. A la izquierda se muestran los vOTUs portadores de AMGs y en paréntesis la familia a la cual pertenecen. En rojo se muestra el vOTU perteneciente a la muestra de hielo glaciar. A la derecha se muestran los AMGs identificados y su función.

7. Discusión

El análisis de las comunidades virales en sistemas naturales ha ido creciendo durante la última década, y la generación de nuevas estrategias para identificación de secuencias virales mediante herramientas de secuenciación de alto rendimiento ha facilitado mucho su estudio. Desde esta premisa, se han realizado varios proyectos de análisis de viomas a gran escala, como por ejemplo *Malaspina Expedition*, *Global Ocean Sampling Expedition*, *Pacific Ocean Virome* y *Tara Ocean's Global Ocean Virome*, enfocados en el estudio del ambiente marino (Brum et al., 2015; Gregory et al., 2019; Hurwitz & Sullivan, 2013; Roux et al., 2016; Williamson et al., 2008). Estos estudios han facilitado de gran manera la disponibilidad de secuencias virales a la fecha en las bases de datos, y en particular para el ambiente marino.

Cuando hablamos de virus, no podemos dejar de hablar de sus hospederos. Tan solo hace unos años no existían muchas estrategias de asignación de hospederos, sin embargo, el avance de la tecnología ha permitido su estudio. La combinación de secuenciación *single-cell*, mejoramientos en las técnicas de muestreo, el avance de las tecnologías de secuenciación y las nuevas herramientas de manejo de datos han permitido el descubrimiento de virus que infectan a algunos grupos importantes de microorganismos marinos (Ahlgren et al., 2017; Labonté et al., 2015; Martínez-García et al., 2014; Vik et al., 2017).

Dada la información conocida para sistemas costeros como los de Bahía Chile en Antártica, nuevos estudios de la comunidad viral deben estar ahora enfocados en dilucidar los hospederos de los bacteriófagos descritos previamente y como parte

de esta memoria, para poder así evaluar de mejor manera su rol ecológico, tanto por su efecto lítico, como por el manejo metabólico de sus hospederos a través de AMGs.

En esta memoria, se logró rescatar un gran número de vOTUs desde metagenomas marinos previamente disponibles en el laboratorio de la Dra. Díez, que fueron obtenidos desde muestras de agua de mar superficial costera en la WAP, como son las aguas de Bahía Chile, así como vOTUs desde un metagenoma de hielo glaciar cercano a esta bahía. Las fracciones de tamaño utilizadas en este estudio corresponden a las fracciones celulares $>0.2 \mu\text{m}$, donde las señales virales pueden provenir de la secuenciación de virus lisogénicos, células infectadas por fagos líticos, virus gigantes o partículas virales asociadas a partículas más grandes (Williamson et al., 2008). Las aguas de Bahía Chile representan la condición típica de las zonas costeras de la Península Antártica, presentando una alta productividad durante el periodo de verano debido a la concentración de nutrientes superficial, y gracias al enriquecimiento de hierro gracias a sedimentos costeros y el ascenso de las profundas corrientes circumpolares, lo cual promueve la productividad fitoplanctónica (Deppeler & Davidson, 2017). Por otro lado, las aguas de Bahía Chile están influenciadas por el aporte de agua dulce de deshielo desde glaciares aledaños como es el glaciar Fuerza Aérea, las cuales pueden afectar las comunidades microbianas y virales en estas aguas costeras antárticas. Por tanto, en esta memoria, se compararon vOTUs de metagenomas marinos costeros como de hielo glaciar de Bahía Chile taxonómicamente gracias a la actual técnica de redes de genes compartidos. A través de esta metodología, se pudo asignar un 55% de los vOTUs a nivel de familia, lo cual supera a los estudios anteriores realizados en

la WAP (Alarcón-Schumacher et al., 2019; Brum et al., 2016; Miranda et al., 2016). Sin embargo, cabe destacar que todavía hay un gran número de secuencias que no pudieron ser asignadas y que actualmente son desconocidas.

Los resultados demostraron que las comunidades virales de hielo glaciar aledaño a la Bahía Chile son diferentes a las de aguas de la bahía, no compartiendo ningún vOTU entre ellas (**Figura 4**). Esto significa que las comunidades de bacteriófagos de glaciares no estarían *a priori* interaccionando, ni por tanto afectando directamente a las comunidades bacterianas marinas en la bahía. Este resultado es relevante, debido al creciente aumento de la deglaciación debido al calentamiento global, sobre las aguas marinas antárticas del WAP, que está afectando a la productividad primaria de la zona (Kavanaugh et al., 2015).

7.1 Ubicuidad y abundancia de familias de caudovirales en Bahía Chile similar a otros ambientes marinos

Estudios previos de metagenómica en la Bahía Chile (WAP), realizados por nuestro grupo de investigación, han mostrado una dominancia de fagos del orden *Caudovirales* (Alarcón-Schumacher et al., 2019) para el verano austral del 2014. Durante un periodo de baja productividad en la bahía, estimado por una baja concentración de Chla, se determinó la dominancia de la familia *Myoviridae* con un ~82% de las asignaciones, seguido por un 9% de la familia *Phycodnaviridae*, un ~8% de la familia *Inoviridae* y finalmente un ~1% de la familia *Siphoviridae*. En cambio, en una muestra colectada bajo una condición de alta Chla para ese mismo año en la bahía, se describió una asignación del ~93% a la familia *Phycodnaviridae*

y solo un ~3% asignado a la familia *Myoviridae* o a la familia *Siphoviridae* con un ~3.5% de asignación.

Tras este primer estudio de Alarcón-Schumacher y colaboradores, se llevaron a cabo nuevas expediciones a Bahía Chile entre los veranos del 2016 y 2019, desde donde se obtuvieron los nuevos metagenomas de muestras de aguas superficiales de la bahía. En esta memoria se reanalizó la comunidad viral de los metagenomas del 2014 (mismos usados por Alarcón-Schumacher et al., 2019), junto con los nuevos metagenomas obtenidos para las muestras del 2016 al 2019. Este análisis demostró una vez más, que en aguas costeras de Bahía Chile durante el verano hay dominancia del orden *Caudovirales* (fagos con cola), siendo éstos asociados a 3 familias principalmente, dos de ellas anteriormente descritas en el primer estudio: *Myoviridae* (con cola contráctil), *Siphoviridae* (con cola no contráctil) y una nueva familia que también fue abundante y que anteriormente no fue descrita en Bahía Chile: *Podoviridae* (con cola corta).

Las principales diferencias encontradas entre el análisis de Alarcón-Schumacher y colaboradores (2019) para el año 2014 y este nuevo análisis a lo largo de varios años y condiciones de productividad en este sistema marino costero es la predominancia de la familia *Myoviridae* durante los años, independiente del aumento en la productividad primaria (Chla) durante cada periodo (**Figura 6**). Esto se muestra incluso en las mismas muestras del 2014, a pesar de que existe una leve disminución en la abundancia de esta familia en el periodo de alta Chla (de ~68% a ~56%; **Figura 5**), la disminución es mucho menor que el estudio anteriormente mencionado (de ~82% a ~3%). Esta diferencia en la disminución de

la familia *Myoviridae* podría deberse a un apantallamiento debido al aumento considerable de virus que infectan a eucariotes, como los de la familia *Phycodnaviridae*, determinado anteriormente con una metodología menos específica durante los periodos de alta Chla (Alarcón-Schumacher et al., 2019), y que se podría encontrar en menor cantidad en la presente memoria debido al tipo de técnicas utilizadas. Además, es relevante el aumento de la abundancia de la familia *Siphoviridae*, la cual fue 30 veces más abundante en la muestra de baja Chla (Ant_st8_2014) y 13 veces más abundante en la muestra de alta Chla (Ant_st22_2014). Finalmente, llama la atención la presencia de la familia *Podoviridae* durante el 2014 así como la ausencia de la familia *Inoviridae*, resultados contrastados con los del trabajo de Alarcón-Schumacher y colaboradores (2019), donde no se encontró abundancia alguna de la familia *Podoviridae*, pero se determinó una abundancia de la familia *Inoviridae*. Todas estas diferencias pueden ser explicadas por el uso de herramientas especializadas para extracción de secuencias de bacteriófagos, y el aumento de las bases de datos virales durante los últimos años. De esta forma se pudo realizar un análisis más completo de las asignaciones virales específicas para bacteriófagos, contrastado con el estudio realizado hace algunos años sobre las mismas muestras de esta bahía. Cabe destacar que todas las diferencias mencionadas entre los metagenomas provenientes del año 2014 con el resto de los años, puede deberse a un sesgo debido a la metodología de extracción de ADN y de posterior secuenciación utilizadas en años posteriores.

Por otro lado, la comparación de las comunidades virales encontradas en Bahía Chile en esta memoria a escala temporal en el verano austral, demuestra similitudes

con lo encontrado para otros ambientes marinos del mundo. Por ejemplo, estudios realizados en aguas superficiales de mar abierto en el giro del Pacífico norte, muestran una gran abundancia de las familias *Myoviridae* y *Podoviridae*, donde la familia *Siphoviridae*, a pesar de estar presente, fue menos abundante (Aylward et al., 2017). La misma dominancia de *Myoviridae* y una gran abundancia de la familia *Podoviridae* se describe en estudios realizados en otras zonas del Océano Pacífico, específicamente en muestras de mar abierto de Bahía Monterrey (California, E.E.U.U.) y en aguas superficiales de la Gran Barrera de Coral (Australia) (Hurwitz & Sullivan, 2013). En el mar Báltico también se ha descrito una dominancia de la familia *Podoviridae*, seguido por la familia *Myoviridae* (Zeigler Allen et al., 2017).

Estudios realizados en ambientes marinos de aguas profundas del Océano Atlántico y del Mar Mediterráneo (5200 m y 2400 m de profundidad, respectivamente), han descrito una dominancia de la familia *Myoviridae* seguida por la familia *Siphoviridae*, y finalmente una menor abundancia de *Podoviridae* (Winter et al., 2014), muy similar a la tendencia encontrada en Bahía Chile para periodos de alta Chla.

Si enfocamos la comparación de Bahía Chile con otras zonas costeras del mundo, como la Bahía Delaware y la Bahía Chesapeake en el Océano Atlántico, en éstas se ha descrito una dominancia de las tres familias mencionadas, encontrando una mayor abundancia de la familia *Myoviridae* frente a las demás. Esto concuerda con lo que sucede en Bahía Chile, habiendo diferencias en la abundancia de las otras dos grandes familias. Se ha descrito que la familia *Siphoviridae* presenta una abundancia mucho menor que la familia *Podoviridae* (Sun et al., 2021). Algo parecido sucede en aguas costeras del Mar del Norte, así como también en el

Océano Atlántico; sin embargo, en las zonas más cercanas a la costa, la abundancia de la familia *Podoviridae* es muy similar a la de la familia *Siphoviridae* (Garin-Fernandez et al., 2018).

Un estudio realizado en el WAP, específicamente en la estación Palmer, comparó muestras provenientes de agua costera superficial, en la cual se encontró una dominancia de la familia *Myoviridae*, seguida por *Podoviridae* y finalmente *Siphoviridae* en fagos líticos. Una situación diferente sucedió para la comunidad de fagos lisogénicos, en la cual se describe una dominancia de la familia *Podoviridae*, seguida en una proporción similar por las familias *Myoviridae* y *Siphoviridae* (Brum et al., 2016). Dicho estudio por tanto sugiere que las familias dominantes dependerán del tipo de ciclo de vida que predomine, siendo necesario seguir indagando al respecto, debido a la baja cantidad de secuencias que pudieron ser asignadas taxonómicamente en ese estudio. Al otro extremo de Antártica, específicamente en la Bahía Prydz, zona marina adyacente a la barrera de hielo Amery, Gong y colaboradores determinaron que la familia más abundante de la zona era *Siphoviridae*, seguida por *Myoviridae* y finalmente por *Podoviridae*, lo que demuestra al contrastar con lo expuesto en esta memoria, que las comunidades virales pueden variar en sus abundancias a nivel de familia dependiendo de la zona del SO en la que nos encontremos (Gong et al., 2018).

En resumen, en Bahía Chile se observa una dominancia de la familia *Myoviridae*, concordando con muchas otras zonas marinas mencionadas. Sin embargo, en Bahía Chile ésta es seguida en abundancia por la familia *Siphoviridae*, lo cual es una distinción importante respecto a otras zonas costeras, que por lo general muestran una mayor abundancia de la familia *Podoviridae*. Sin embargo, en Bahía

Chile se puede observar una menor abundancia de la familia *Podoviridae* (**Figura 6**) en las muestras bajo una condición de menor Chla (Ant_2_2017 y Ant_st8_2014), sugiriéndose que la comunidad fitoplanctónica (reflejo de la potencial producción primaria en este sistema) podría modular indirectamente la presencia de familias específicas virales en aguas costeras de Antártica. Sin embargo, esta hipótesis requiere de un análisis mayor con nuevos estudios, ya que una abundancia similar fue obtenida en una de las muestras de alta Chla en el 2014 (Ant_st22_2014).

7.2 Presencia y potencial función ecológica de la familia *Phycodnaviridae* en aguas marinas de Bahía Chile

A pesar que las herramientas utilizadas en esta memoria están descritas para recuperar específicamente bacteriófagos, en nuestro análisis de redes de genes compartidos (**Figura 5**), se pudo identificar un agrupamiento que se separaba completamente y que fue identificado como un set de vOTUs perteneciente al grupo NCLDV. Esto muestra una asignación no específica desde las herramientas especializadas para la identificación de bacteriófagos empleada en esta memoria como es WtP (Marquet et al., 2020). Sin embargo, la presencia de este grupo en aguas marinas es muy interesante, ya que en Bahía Chile ya se había descrito una dependencia positiva de la presencia de la familia *Phycodnaviridae* con respecto a los florecimientos de fitoplancton durante el año 2014 (Alarcón-Schumacher et al., 2019). Los resultados expuestos en la presente memoria muestran la presencia de este grupo de virus gigantes en la Bahía Chile a escala interanual, y posiblemente en aguas del WAP, mostrando una mayor abundancia de esta familia durante los años 2016 y 2017 (**Figura 6**). Nuevos estudios usando metodologías especializadas

para el rescate de secuencias del grupo NCLDV, son necesarios para entender mejor la dinámica de este grupo particular de virus en aguas antárticas.

Respecto a la identidad de estas secuencias, llama en especial la atención la identificación de lecturas que alinearon contra un género no clasificado de *Prymnesiovirus* (**Figura 7**) o contra el virus de *Phaeocystis globosa* 16T (**Tabla 4**). Este último, también está relacionado con el género *Prymnesiovirus*, pero aún no ha sido aprobado como especie perteneciente a este género (Santini et al., 2013). En el estudio de Alarcón-Schumacher y colaboradores (2019) se realizó una filogenia con el gen marcador *polB* para la particular identificación del grupo NCLDV en Bahía Chile durante el 2014. Estos resultados también arrojaron una cercanía de las secuencias estudiadas con virus de *Phaeocystis globosa* y *Phaeocystis pouchetti*.

Las especies hospederas de estos dos tipos de virus corresponden a los principales componentes de florecimientos fitoplanctónicos en las costas marinas alrededor del mundo. *P. globosa* tiene una tendencia a presentarse en aguas templadas, mientras que *P. pouchetti* y *P. antarctica* se adaptan mucho mejor a ambientes fríos, específicamente en Ártico y Antártica, respectivamente (Schoemann et al., 2005). Los florecimientos de la especie haptófito *P. antarctica* en el SO son acompañados de una comunidad específica y estable de bacterias (Delmont et al., 2014). Esta especie de alga no solo tiene un efecto sobre la comunidad marina, sino que también se ha descrito su influencia sobre el ciclo del azufre, específicamente por ser una de las mayores productoras de dimetilsulfoniopropionato (DMSP) (Asher et al., 2017). El DMSP es un compuesto sulfurado que al ser liberado al ambiente marino tras la lisis celular, es un precursor de sulfuro de dimetilo (DMS) que se

produce por descomposición enzimática del DMSP producida por liasas bacterianas o de algas (Cantoni et al., 1956; Stoer, 2017; Wagner & Stadtman, 1962). Se ha descrito que en regiones costeras de la Antártica se produce una mayor concentración de DMS anual respecto a otras regiones del mundo (Wang et al., 2018).

Una vez producido el DMS, debido a su volatilidad es liberado a la atmósfera, donde debido a la radiación se oxida y forma aerosoles de productos sulfurados que van a permitir la formación de núcleos de condensación de nubes (Veres et al., 2020). De esta forma se incrementa la densidad de las nubes en el cielo, produciéndose una disminución de la radiación y temperatura, lo que provoca que al mar llegue una menor cantidad de luz solar y un clima más atenuado. Esta nueva condición provoca a su vez una disminución de la productividad primaria del fitoplancton por la disminución de la energía lumínica (Stoer, 2017). Debido a su importante rol en la modulación de la comunidad de microorganismos y la relación con el ciclo del azufre, es de especial interés la descripción de la especie viral que podría estar infectando y modulando la abundancia de la especie *P. antarctica* en este sistema marino polar. A la fecha, no existe ningún genoma disponible de esta especie viral, sin embargo, se ha sugerido la presencia del virus de *P. antarctica* (PaV) en Bahía Chile durante el verano del 2014 (Alarcón-Schumacher et al., 2019).

En la presente memoria, se pudo determinar la presencia de secuencias con alta probabilidad de ser identificadas como PaV, especialmente durante los años 2016 y 2017 (**Figura 6**). Cabe destacar, que la poca abundancia de este grupo viral podría deberse a la metodología utilizada en esta memoria, la cual estuvo especializada en el rescate de secuencias de bacteriófagos, lo que hace ahora necesario utilizar

otro tipo de metodologías para el rescate de secuencias del grupo NCLDV con el fin de caracterizar el genoma de PaV.

7.3 Potencial función ecológica viral asociada a la identificación de AMGs putativos en comunidades virales de Bahía Chile

Los bacteriófagos pueden afectar a la comunidad microbiana portando genes que alteren el metabolismo celular para favorecer el ciclo de vida del fago. A los genes portados por los virus que cumplan la función de modificar el metabolismo celular se les llama genes metabólicos auxiliares (AMGs). Se han descrito una variedad de AMGs con distintas ventajas para la reproducción viral (Warwick-Dugdale et al., 2019). Sin embargo, Antártica es una región con poca información al respecto.

En los pocos estudios realizados en aguas antárticas, se han descrito AMGs que son muy comunes en las comunidades virales a nivel mundial, tales como genes de transportadores de membrana (Brum et al., 2016). Sin embargo, se especula que las comunidades virales en aguas antárticas pueden ser una fuente de AMGs desconocidos, debido a las presiones ambientales y la limitación de nutrientes. Un ejemplo de ello, es que los fagos de aguas antárticas podrían otorgar genes relacionados con el estrés por frío, dándole una ventaja de supervivencia al hospedero durante el periodo de infección (Alarcón-Schumacher et al., 2019; Brum et al., 2016).

En esta memoria, se encontró la presencia de una aminopeptidasa de la familia M17 en fagos miembros de la familia *Myoviridae*, la cual nunca antes había sido reportada como AMG en ningún sistema natural. Esta familia de aminopeptidasas está involucrada en diversos procesos en bacterias marinas, como la recombinación

sitio específica de ADN y la regulación de factores de virulencia (Alonso et al., 2020). Debido a la gran diversidad de funciones para este tipo de aminopeptidasas, es necesario ahora complementar con el contexto del hospedero para entender el efecto que podría tener este AMG en la comunidad bacteriana de Bahía Chile. Otro de los potenciales AMGs observados en estas muestras de agua de mar pero también en el hielo glaciar, es el gen *dcm*, el cual codifica la metiltransferasa de citosina, que ya ha sido reportado en estudios anteriores para diversos ecosistemas, incluido el marino, y que está involucrado indirectamente en la producción de azufre mediante la degradación de compuestos orgánicos como metionina y cisteína, lo que indicaría que la comunidad viral podría aportar en la producción de azufre durante la infección del hospedero (Kieft et al., 2021). Se ha demostrado que el aumento de la concentración de azufre en el ambiente marino afecta el crecimiento de las comunidades bacterianas involucradas, pero también las que no están involucradas en el ciclo del azufre (Mirzoyan & Schreier, 2014). Así, la ventaja que podría obtener la comunidad de fagos portadora del gen *dcm* es promover el crecimiento específico de una comunidad bacteriana azufre tolerante, mientras ésta está infectada. Sin embargo, es necesario ahondar en el tipo de hospedero que tiene este tipo de virus, ya que cada tipo de bacteria se comporta de una manera distinta ante un contexto de elevada concentración en azufre (Mirzoyan & Schreier, 2014).

Respecto a la codificación de glicosiltransferasas en virus, es un fenómeno estudiado desde hace varios años, en el cual se han descrito glicosiltransferasas codificadas por virus que se describen con múltiples funciones que afectan la interacción de proteínas en el hospedero (Markine-Goriaynoff et al., 2004). En esta

memoria se ha descrito la familia de glicosiltransferasas GT2 presente en virus de Bahía Chile. Esta familia de glicosiltransferasas se ha reportado con anterioridad en arqueas, bacterias, eucariontes y virus, y es considerada una de las familias de glicosiltransferasas más representativas de las 90 totales descritas, que podría ser una de las más ancestrales representando múltiples actividades enzimáticas en los miembros de esta familia (Coutinho et al., 2003). Faltaría por tanto más información para describir una función específica para este AMG encontrado en Bahía Chile. Por otra parte, la familia de glicosiltransferasas GT102, también encontrada en virus de Bahía Chile durante el presente estudio, no ha sido descrita hasta el momento como AMG en ningún sistema natural. Este tipo de glicosiltransferasas se han caracterizado como enzimas fundamentales para la biosíntesis de antígeno-O de la membrana externa de las bacterias Gram negativas (Mills et al., 2004). Respecto a la membrana externa bacteriana en el ambiente marino, se ha descrito que una especie de cianofagos utiliza la interacción con el antígeno-O en el hospedero para lograr la infección (Xiong et al., 2019). Es interesante la modificación de una estructura que interactúa con el medio extracelular, en especial si es posible que sea una estructura que interacciona directamente con el virus para lograr la infección. Por esto, es interesante y necesario investigar más respecto a este AMG, para lograr identificar la ventaja otorgada al ciclo infeccioso.

La presencia de genes ribosomales en secuencias virales, previamente asociado a contaminaciones en la extracción de secuencias virales desde metagenomas, es ahora reconocido para una gran variedad de virus. En general, las estrategias virales para la reproducción son muchas y al parecer la adquisición de proteínas ribosomales es muy común (Mizuno et al., 2019). En esta memoria se encontraron

dos proteínas que son parte estructural de la subunidad menor del ribosoma, encontradas en vOTUs miembros de la familia *Myoviridae*. La primera es la proteína RP-21 (codificada por el gen *rpsU*), la cual se ha descrito anteriormente en virus marinos y se ha sugerido que podría sustituir a la RP-21 del hospedero favoreciendo así la traducción de las proteínas virales (Mizuno et al., 2019). En la comunidad viral de Bahía Chile también se encontró la proteína RP-S1 (codificada por el gen *rpsA*) y que es descrita por primera vez como un AMG. Más estudios son ahora necesarios para entender mejor el potencial del uso de esta estrategia para la reproducción viral marina a escala global.

Durante la infección viral, la síntesis de nucleótidos es un proceso muy relevante y por tanto no es extraño encontrar también AMGs involucrados en este proceso. De los descritos en el presente trabajo en Bahía Chile, se encontraron 3 genes involucrados en la síntesis de timidina trifosfato, uno de los cuatro nucleótidos trifosfato que se utilizan en la síntesis de ADN. Uno de estos 3 genes es el ya reportado como AMG *dcd*, involucrado en la desaminación de dCTP para producir dUMP, el cual es sustrato de la enzima timidina sintetasa (Huang et al., 2021). También se encontró la presencia del gen *dut*, que codifica la enzima deoxiuridina trifosfatasa, la cual ya ha sido reportada como un AMG en viromas de océano profundo del Pacific Ocean Virome (POV) sin análisis taxonómico viral (Warwick-Dugdale et al., 2019). Esta enzima está involucrada en la conversión de dUTP en dUMP (producto común con la desaminasa de dCTP). En Bahía Chile también se observó la presencia en vOTUs de la timidina sintetasa codificada por el gen *thyA*, la cual tiene homólogos virales (Arvizu-Flores et al., 2009; Graziani et al., 2004). La función de esta sintetasa es la biosíntesis del nucleósido de timidina (Maley & Maley,

1988). A pesar que la síntesis de nucleótidos es una estrategia común para potenciar la síntesis de ADN durante la infección viral, resulta curioso que se identificaran tres genes involucrados en la síntesis de un solo nucleótido y no de algún otro, esto en distintos virus de distintas familias (*Myoviridae*, *Siphoviridae* y *Phycodnaviridae*). Bajo este contexto, se ha demostrado que los genes *dcd* y *dut* están directamente relacionados con la prevención de la uracilación del ADN debido a la acumulación de dUTP, la cual produce mortalidad (Hirmondo et al., 2017). Por lo que la presencia de los AMGs involucrados en esta vía podrían cumplir un rol protector del ADN en el hospedero durante la infección viral, permitiendo la replicación del virus.

Finalmente, el gen *cobS*, ampliamente estudiado y reconocido como AMG y que expresa la quelatasa de cobalamina, involucrada en la biosíntesis de porfirina en bacterias, fue también observado en miembros virales de la familia *Myoviridae* de Bahía Chile. Este gen es de hecho utilizado como un *hallmark* viral para cianofagos pertenecientes a la familia *Myoviridae* (Adriaenssens & Cowan, 2014), lo cual es extraño específicamente en aguas antárticas, donde la presencia de cianobacterias es comunmente baja (Vincent, 2002). Por esto, es necesario ahondar en el rango de la comunidad cubierta por este *hallmark* viral.

Como se ha descrito, la mayoría de los AMGs identificados en esta memoria se encontraron en vOTUs miembros de la familia *Myoviridae*, la cual es la familia más abundante en estas y otras aguas marinas de todo el mundo, por lo que el hecho de reconocer AMGs en estos vOTUs puede deberse a la facilidad para recuperar secuencias debido a su gran abundancia en los set virales, y no a que esta familia porte más AMGs que las demás. A pesar de que se encontraron una gran variedad

de AMGs, éstos fueron identificados en pocos vOTUs que a nivel de abundancias representan a una baja proporción de la comunidad viral durante los años 2014 al 2019 en Bahía Chile (**Figura 9**).

El estudio de AMGs es relevante para poder comprender de mejor manera como las comunidades virales están interaccionando con sus hospederos en este sistema marino polar, donde las estrategias de supervivencia y adaptaciones tanto de comunidades microbianas como virales mantienen la funcionalidad ecológica en este sistema extremo de tan alta relevancia en el secuestro global de CO₂ en el planeta.

8. Conclusiones

- En Bahía Chile existen comunidades virales dominadas por la familia *Myoviridae*, seguido por *Siphoviridae*, las cuales se mantienen en proporción similar durante los años 2014-2019. También existe una menor abundancia de la familia *Podoviridae* que fluctúa entre los veranos australes de cada año.
- La comunidad viral de hielo glaciar es totalmente diferente a la comunidad viral marina de Bahía Chile al comparar la cantidad de vOTUs presentes. No así, si comparamos las diferencias a nivel taxonómico de familia, en donde se encuentran las mismas familias dominantes.
- En Bahía Chile hay presencia de virus pertenecientes a la familia *Phycodnaviridae* muchos de ellos identificados como PaV, virus previamente descritos en estas aguas con potencial de infectar a la haptófita formadora de florecimientos *Phaeocystis antarctica*.
- La comunidad viral de Bahía Chile contiene AMGs ya descritos en aguas marinas, pero también nuevos AMGs nunca descritos como AMGs involucrados en el metabolismo del carbono (glicosiltransferasas GT2 y GT102, *dut*, *thyA* y *dcd*) y de elementos orgánicos nitrogenados (aminopeptidasa M17), en el ciclo del azufre (*dcm*), biosíntesis de cobalamina (*cobS*) y en la estructura ribosomal (*rpsA*, *rpsU*).
- Las abundancias de los vOTUs portadores de AMGs es baja y fluctúa entre los veranos australes desde el 2014 al 2019.

9. Referencias

- Abdelkareem, A. O., Khalil, M. I., Elaraby, M., Abbas, H., & Elbehery, A. H. A. (2018). VirNet: Deep attention model for viral reads identification. *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, 623-626. <https://doi.org/10.1109/ICCES.2018.8639400>
- Adriaenssens, E. M., & Cowan, D. A. (2014). Using Signature Genes as Tools To Assess Environmental Viral Ecology and Diversity. *Applied and Environmental Microbiology*, *80*(15), 4470-4480. <https://doi.org/10.1128/AEM.00878-14>
- Ahlgren, N. A., Chen, Y., Needham, D. M., Parada, A. E., Sachdeva, R., Trinh, V., Chen, T., & Fuhrman, J. A. (2017). Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environmental Microbiology*, *19*(6), 2434-2452. <https://doi.org/10.1111/1462-2920.13768>
- Alarcón-Schumacher, T., Guajardo-Leiva, S., Antón, J., & Díez, B. (2019). Elucidating Viral Communities During a Phytoplankton Bloom on the West Antarctic Peninsula. *Frontiers in Microbiology*, *10*, 1014. <https://doi.org/10.3389/fmicb.2019.01014>
- Alarcón-Schumacher, T., Guajardo-Leiva, S., Martínez-García, M., & Díez, B. (2021). Ecogenomics and Adaptation Strategies of Southern Ocean Viral Communities. *mSystems*, *6*(4). <https://doi.org/10.1128/mSystems.00396-21>
- Alcamán-Arias, M. E., Farías, L., Verdugo, J., Alarcón-Schumacher, T., & Díez, B. (2018). Microbial activity during a coastal phytoplankton bloom on the

- Western Antarctic Peninsula in late summer. *FEMS Microbiology Letters*, 365(10), 10.1093/femsle/fny090. <https://doi.org/10.1093/femsle/fny090>
- Alcamán-Arias, M. E., Fuentes-Alburquenque, S., Vergara-Barros, P., Cifuentes-Anticevic, J., Verdugo, J., Polz, M., Farías, L., Pedrós-Alió, C., & Díez, B. (2021). Coastal Bacterial Community Response to Glacier Melting in the Western Antarctic Peninsula. *Microorganisms*, 9(1), 88. <https://doi.org/10.3390/microorganisms9010088>
- Alonso, I. P., Méndez, L. R., Almeida, F., Tresanco, M. E. V., Sánchez, Y. A., Hernández-Zanuy, A., Álvarez-Lajonchere, L., Díaz, D., Sánchez, B., Florent, I., Schmitt, M., Cisneros, M., & Charli, J. L. (2020). Marine organisms: A source of biomedically relevant metallo M1, M2 and M17 exopeptidase inhibitors. *Revista Cubana de Ciencias Biológicas* 8(2), 1-36.
- Amgarten, D., Braga, L. P. P., da Silva, A. M., & Setubal, J. C. (2018). MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Frontiers in Genetics*, 9, 304. <https://doi.org/10.3389/fgene.2018.00304>
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data* – *ScienceOpen*. <https://www.scienceopen.com/document?vid=de674375-ab83-4595-afa9-4c8aa9e4e736>
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 44(W1), W16-21. <https://doi.org/10.1093/nar/gkw387>
- Arvizu-Flores, A. A., Aispuro-Hernandez, E., Garcia-Orozco, K. D., Varela-Romero, A., Valenzuela-Soto, E., Velazquez-Contreras, E. F., Rojo-Domínguez, A.,

- Yepiz-Plascencia, G., Maley, F., & Sotelo-Mundo, R. R. (2009). Functional identity of the active sites of crustacean and viral thymidylate synthases. *Comparative Biochemistry and Physiology. Toxicology & Pharmacology: CBP*, 150(3), 406-413. <https://doi.org/10.1016/j.cbpc.2009.06.008>
- Asher, E. C., Dacey, J. W. H., Stukel, M., Long, M. C., & Tortell, P. D. (2017). Processes driving seasonal variability in DMS, DMSP, and DMSO concentrations and turnover in coastal Antarctic waters: DMS, DMSP, and DMSO in coastal Antarctic waters. *Limnology and Oceanography*, 62(1), 104-124. <https://doi.org/10.1002/lno.10379>
- Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., Romano, A. E., & DeLong, E. F. (2017). Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences*, 114(43), 11446-11451. <https://doi.org/10.1073/pnas.1714821114>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5), 455-477. <https://doi.org/10.1089/cmb.2012.0021>
- Bergh, Ø., Børsheim, K. Y., Bratbak, G., & Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature*, 340(6233), 467-468. <https://doi.org/10.1038/340467a0>

- Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., Brister, J. R., Kropinski, A. M., Krupovic, M., Lavigne, R., Turner, D., & Sullivan, M. B. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology*, 37(6), 632-639. <https://doi.org/10.1038/s41587-019-0100-8>
- Breitbart, M. (2011). Marine Viruses: Truth or Dare. *Annual Review of Marine Science*, 4(1), 425-448. <https://doi.org/10.1146/annurev-marine-120709-142805>
- Breitbart, M., Bonnain, C., Malki, K., & Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nature Microbiology*, 3(7), 754-766. <https://doi.org/10.1038/s41564-018-0166-y>
- Breitbart, M., Thompson, L., Suttle, C., & Sullivan, M. (2007). Exploring the Vast Diversity of Marine Viruses. *Oceanography*, 20(2), 135-139. <https://doi.org/10.5670/oceanog.2007.58>
- Brown, C. T., & Irber, L. (2016). sourmash: A library for MinHash sketching of DNA. *Journal of Open Source Software*, 1(5), 27. <https://doi.org/10.21105/joss.00027>
- Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W., & Sullivan, M. B. (2016). Seasonal time bombs: Dominant temperate viruses affect Southern Ocean microbial dynamics. *The ISME Journal*, 10(2), 437-449. <https://doi.org/10.1038/ismej.2015.125>
- Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., Chaffron, S., Cruaud, C., Vargas, C. de, Gasol, J. M., Gorsky, G., Gregory, A. C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S.,

- Poulos, B. T., Bonnie, T., Schwenck, S. M., Speich, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Tara Oceans Coordinatirs, Bork, P., Bowler, C., Sunagawa, S., Wincker, P., Karsenti, E. & Sullivan, M. B. (2015). Patterns and ecological drivers of ocean viral communities. *Science*, 348(6237). <https://doi.org/10.1126/science.1261498>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59-60. <https://doi.org/10.1038/nmeth.3176>
- Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner* (LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). <https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Cantoni, G. L., Anderson, D. G., & Rosenthal, E. (1956). Enzymatic cleavage of dimethylpropiothetin by *Polysiphonia lanosa*. *Journal of Biological Chemistry*, 222(1), 171-177. [https://doi.org/10.1016/S0021-9258\(19\)50782-7](https://doi.org/10.1016/S0021-9258(19)50782-7)
- Chorus & Bartman. (1999). Toxic Cyanobacteria in Water—A guide to their public health consequences, monitoring and management. <https://doi.org/10.1201/9781003081449>
- Cifuentes-Anticevic, J., Alcamán-Arias, M. E., Alarcón-Schumacher, T., Tamayo-Leiva, J., Pedrós-Alió, C., Farías, L., & Díez, B. (2021). Proteorhodopsin

- Phototrophy in Antarctic Coastal Waters. *MSphere*, 6(4), e0052521.
<https://doi.org/10.1128/mSphere.00525-21>
- Coutinho, P. M., Deleury, E., Davies, G. J., & Henrissat, B. (2003). An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology*, 328(2), 307-317. [https://doi.org/10.1016/S0022-2836\(03\)00307-3](https://doi.org/10.1016/S0022-2836(03)00307-3)
- Crummett, L. T., Puxty, R. J., Weihe, C., Marston, M. F., & Martiny, J. B. H. (2016). The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology*, 499, 219-229.
<https://doi.org/10.1016/j.virol.2016.09.016>
- Delmont, T. O., Hammar, K. M., Ducklow, H. W., Yager, P. L., & Post, A. F. (2014). *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the Amundsen Sea polynya. *Frontiers in Microbiology*, 5, 646.
<https://doi.org/10.3389/fmicb.2014.00646>
- Deppeler, S. L., & Davidson, A. T. (2017). Southern Ocean Phytoplankton in a Changing Climate. *Frontiers in Marine Science*, 4, 40.
<https://doi.org/10.3389/fmars.2017.00040>
- Ducklow, H., Steinberg, D., & Buesseler, K. (2001). Upper Ocean Carbon Export and the Biological Pump. *Oceanography*, 14(4), 50-58.
<https://doi.org/10.5670/oceanog.2001.06>
- Evans, C., Brandsma, J., Pond, D. W., Venables, H. J., Meredith, M. P., Witte, H. J., Stammerjohn, S., Wilson, W. H., Clarke, A., & Brussaard, C. P. D. (2017). Drivers of interannual variability in virioplankton abundance at the coastal western Antarctic peninsula and the potential effects of climate change.

Environmental Microbiology, 19(2), 740-755. <https://doi.org/10.1111/1462-2920.13627>

Evans, C., Pearce, I., & Brussaard, C. P. D. (2009). Viral-mediated lysis of microbes and carbon release in the sub-Antarctic and Polar Frontal zones of the Australian Southern Ocean. *Environmental Microbiology*, 11(11), 2924-2934. <https://doi.org/10.1111/j.1462-2920.2009.02050.x>

Fang, Z., Tan, J., Wu, S., Li, M., Xu, C., Xie, Z., & Zhu, H. (2019). PPR-Meta: A tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience*, 8(6), giz066. <https://doi.org/10.1093/gigascience/giz066>

Fuentes, S., Arroyo, J. I., Rodríguez-Marconi, S., Masotti, I., Alarcón-Schumacher, T., Polz, M. F., Trefault, N., De la Iglesia, R., & Díez, B. (2019). Summer phyto- and bacterioplankton communities during low and high productivity scenarios in the Western Antarctic Peninsula. *Polar Biology*, 42(1), 159-169. <https://doi.org/10.1007/s00300-018-2411-5>

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, 399(6736), 541-548. <https://doi.org/10.1038/21119>

Garin-Fernandez, A., Pereira-Flores, E., Glöckner, F. O., & Wichels, A. (2018). The North Sea goes viral: Occurrence and distribution of North Sea bacteriophages. *Marine Genomics*, 41, 31-41. <https://doi.org/10.1016/j.margen.2018.05.004>

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.
<https://doi.org/10.3389/fmicb.2018.02981>
- Graziani, S., Xia, Y., Gurnon, J. R., Van Etten, J. L., Leduc, D., Skouloubris, S., Myllykallio, H., & Liebl, U. (2004). Functional analysis of FAD-dependent thymidylate synthase ThyX from *Paramecium bursaria* Chlorella virus-1. *The Journal of Biological Chemistry*, 279(52), 54340-54347.
<https://doi.org/10.1074/jbc.M409121200>
- 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., Domínguez-Huerta, G., Ferland, J., Kandels, S., Liu, Y., Marec, C., Pesant, S., Picheral, M., Pisarev, S., Poulain, J., Tremblay, J., Vik, D., Iudicone, T.D., Karp-Boss, L., Roux, S., Sunagawa, S. & Sullivan, M. B. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, 177(5), 1109-1123.e14. <https://doi.org/10.1016/j.cell.2019.03.040>
- Guixa-Boixereu, N., Vaqué, D., Gasol, J. M., Sánchez-Cámara, J., & Pedrós-Alió, C. (2002). Viral distribution and activity in Antarctic waters. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(4), 827-845.
[https://doi.org/10.1016/S0967-0645\(01\)00126-6](https://doi.org/10.1016/S0967-0645(01)00126-6)
- Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: A multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, 9(1), 37. <https://doi.org/10.1186/s40168-020-00990-y>

- Hirmondo, R., Lopata, A., Suranyi, E. V., Vertessy, B. G., & Toth, J. (2017). Differential control of dNTP biosynthesis and genome integrity maintenance by the dUTPase superfamily enzymes. *Scientific Reports*, 7(1), 6043. <https://doi.org/10.1038/s41598-017-06206-y>
- Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H., Andreopoulos, B., Brewer, H. M., Glavina del Rio, T., Adkins, J. N., Paul, S., Sullivan, M. B., & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *The ISME Journal*, 14(4), 881-895. <https://doi.org/10.1038/s41396-019-0580-z>
- Huang, X., Jiao, N., & Zhang, R. (2021). The genomic content and context of auxiliary metabolic genes in roseophages. *Environmental Microbiology*, 23(7), 3743-3757. <https://doi.org/10.1111/1462-2920.15412>
- Hurwitz, B. L., & Sullivan, M. B. (2013). The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *PLOS ONE*, 8(2), e57355. <https://doi.org/10.1371/journal.pone.0057355>
- Hurwitz, B. L., & U'Ren, J. M. (2016). Viral metabolic reprogramming in marine ecosystems. *Current Opinion in Microbiology*, 31, 161-168. <https://doi.org/10.1016/j.mib.2016.04.002>
- Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3), 377-386. <https://doi.org/10.1101/gr.5969107>
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site

- identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- IHO (International Hydrographic Organization). (2000). Report of the International Hydrographic Organization. Working paper No. 57 (WP 57). *20th Session of the United Nations Group of Experts on Geographical Names, New York.* <https://iho.int/uploads/user/pubs/standards/s-57/31Main.pdf>
- Jurtz, V. I., Villarroel, J., Lund, O., Voldby Larsen, M., & Nielsen, M. (2016). MetaPhinder—Identifying Bacteriophage Sequences in Metagenomic Data Sets. *PLOS ONE*, 11(9), e0163111. <https://doi.org/10.1371/journal.pone.0163111>
- Kavanaugh, M. T., Abdala, F. N., Ducklow, H., Glover, D., Fraser, W., Martinson, D., Stammerjohn, S., Schofield, O., & Doney, S. C. (2015). Effect of continental shelf canyons on phytoplankton biomass and community composition along the western Antarctic Peninsula. *Marine Ecology Progress Series*, 524, 11-26. <https://doi.org/10.3354/meps11189>
- Kieft, K., Breister, A. M., Huss, P., Linz, A. M., Zanetakos, E., Zhou, Z., Rahlff, J., Esser, S. P., Probst, A. J., Raman, S., Roux, S., & Anantharaman, K. (2021). Virus-associated organosulfur metabolism in human and environmental systems. *BioRxiv*, 2021.01.05.425418. <https://doi.org/10.1101/2021.01.05.425418>
- Kieft, K., Zhou, Z., & Anantharaman, K. (2020). VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*, 8(1), 90. <https://doi.org/10.1186/s40168-020-00867-0>

- Labonté, J. M., Swan, B. K., Poulos, B., Luo, H., Koren, S., Hallam, S. J., Sullivan, M. B., Woyke, T., Eric Wommack, K., & Stepanauskas, R. (2015). Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *The ISME Journal*, 9(11), 2386-2399. <https://doi.org/10.1038/ismej.2015.48>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. <https://doi.org/10.1038/nmeth.1923>
- Leo Lahti, & Shetty, S. (2012). *Microbiome R package*. <https://bioconductor.org/packages/release/bioc/html/microbiome.html>
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, 31(10), 1674-1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, 22(13), 1658-1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Luria, C. M., Amaral-Zettler, L. A., Ducklow, H. W., & Rich, J. J. (2016). Seasonal Succession of Free-Living Bacterial Communities in Coastal Waters of the

- Western Antarctic Peninsula. *Frontiers in Microbiology*, 7, 1731.
<https://doi.org/10.3389/fmicb.2016.01731>
- Maley, G. F., & Maley, F. (1988). Properties of a defined mutant of *Escherichia coli* thymidylate synthase. *The Journal of Biological Chemistry*, 263(16), 7620-7627.
- Mara, P., Vik, D., Pachiadaki, M. G., Suter, E. A., Poulos, B., Taylor, G. T., Sullivan, M. B., & Edgcomb, V. P. (2020). Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *The ISME Journal*, 14(12), 3079-3092.
<https://doi.org/10.1038/s41396-020-00739-3>
- Markine-Goriaynoff, N., Gillet, L., Van Etten, J. L., Korres, H., Verma, N., & Vanderplasschen, A. (2004). Glycosyltransferases encoded by viruses. *Journal of General Virology*, 85(10), 2741-2754.
<https://doi.org/10.1099/vir.0.80320-0>
- Marquet, M., Hölzer, M., Pletz, M. W., Viehweger, A., Makarewicz, O., Ehrlich, R., & Brandt, C. (2020). *What the Phage: A scalable workflow for the identification and analysis of phage sequences*. *BioRxiv*, 2020.07.24.219899.
<https://doi.org/10.1101/2020.07.24.219899>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10-12.
<https://doi.org/10.14806/ej.17.1.200>
- Martínez-García, M., Santos, F., Moreno-Paz, M., Parro, V., & Antón, J. (2014). Unveiling viral–host interactions within the ‘microbial dark matter’. *Nature Communications*, 5(1), 4542. <https://doi.org/10.1038/ncomms5542>

- Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D., & Antin, P. (2016). The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLOS Biology*, *14*(1), e1002342. <https://doi.org/10.1371/journal.pbio.1002342>
- Mills, J. A., Motichka, K., Jucker, M., Wu, H. P., Uhlik, B. C., Stern, R. J., Scherman, M. S., Vissa, V. D., Pan, F., Kundu, M., Ma, Y. F., & McNeil, M. (2004). Inactivation of the Mycobacterial Rhamnosyltransferase, Which Is Needed for the Formation of the Arabinogalactan-Peptidoglycan Linker, Leads to Irreversible Loss of Viability. *Journal of Biological Chemistry*, *279*(42), 43540-43546. <https://doi.org/10.1074/jbc.M407782200>
- Miranda, J. A., Culley, A. I., Schvarcz, C. R., & Steward, G. F. (2016). RNA viruses as major contributors to Antarctic virioplankton. *Environmental Microbiology*, *18*(11), 3714-3727. <https://doi.org/10.1111/1462-2920.13291>
- Mirzoyan, N., & Schreier, H. J. (2014). Effect of sulfide on growth of marine bacteria. *Archives of Microbiology*, *196*(4), 279-287. <https://doi.org/10.1007/s00203-014-0968-0>
- Mizuno, C. M., Guyomar, C., Roux, S., Lavigne, R., Rodriguez-Valera, F., Sullivan, M. B., Gillet, R., Forterre, P., & Krupovic, M. (2019). Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nature Communications*, *10*(1), 752. <https://doi.org/10.1038/s41467-019-08672-6>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824-834. <https://doi.org/10.1101/gr.213959.116>

- Pomeroy, L., leB. Williams, P., Azam, F., & Hobbie, J. (2007). The Microbial Loop. *Oceanography*, 20(2), 28-33. <https://doi.org/10.5670/oceanog.2007.45>
- Poorvin, L., Rinta-Kanto, J. M., Hutchins, D. A., & Wilhelm, S. W. (2004). Viral release of iron and its bioavailability to marine plankton. *Limnology and Oceanography*, 49(5), 1734-1741. <https://doi.org/10.4319/lo.2004.49.5.1734>
- R Development Core Team. (2010). *R: a language and environment for statistical computing: Reference index*. R Foundation for Statistical Computing. <http://softlibre.unizar.es/manuales/aplicaciones/r/fullrefman.pdf>
- Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, 5(1), 69. <https://doi.org/10.1186/s40168-017-0283-5>
- Ren, J., Song, K., Deng, C., Ahlgren, N. A., Fuhrman, J. A., Li, Y., Xie, X., Poplin, R., & Sun, F. (2020). Identifying viruses from metagenomic data using deep learning. *Quantitative Biology*, 8(1), 64-77. <https://doi.org/10.1007/s40484-019-0187-4>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <https://doi.org/10.1186/gb-2010-11-3-r25>

- Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., Poulos, B. T., Solonenko, N., Lara, E., Poulain, J., Pesant, S., Kandels-Lewis, S., Dimier, C., Picheral, M., Searson, S., Cruaud, C., Alberti, A., Duarte, C. M., Gasol, J. M., Vaqué, D., Tara Oceans Coordinators, Bork, P., Acinas, S., Wincker, P. & Sullivan, M. B. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, *537*(7622), 689-693. <https://doi.org/10.1038/nature19366>
- Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: Mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. <https://doi.org/10.7717/peerj.985>
- Santini, S., Jeudy, S., Bartoli, J., Poirot, O., Lescot, M., Abergel, C., Barbe, V., Wommack, K. E., Noordeloos, A. A. M., Brussaard, C. P. D., & Claverie, J.-M. (2013). Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proceedings of the National Academy of Sciences*, *110*(26), 10800-10805. <https://doi.org/10.1073/pnas.1303251110>
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*, *27*(6), 863-864. <https://doi.org/10.1093/bioinformatics/btr026>
- Schoemann, V., Becquevort, S., Stefels, J., Rousseau, V., & Lancelot, C. (2005). *Phaeocystis* blooms in the global ocean and their controlling mechanisms: A review. *Journal of Sea Research*, *53*(1-2), 43-66. <https://doi.org/10.1016/j.seares.2004.01.008>

- Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, *48*(16), 8883-8900. <https://doi.org/10.1093/nar/gkaa621>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498-2504. <https://doi.org/10.1101/gr.1239303>
- Stoer, A. (2017). Phytoplankton Climate Regulation in Positive and Negative Feedback Systems: The CLAW and anti-CLAW hypotheses. *Oceans First*, *4*, 43-52.
- Strickland, J. D. H., & Parsons, T. R. (1972). A Practical Handbook of Seawater Analysis. *FISHERIES RESEARCH BOARD OF CANADA*, *167*. https://epic.awi.de/id/eprint/39262/1/Strickland-Parsons_1972.pdf
- Sun, M., Zhan, Y., Marsan, D., Páez-Espino, D., Cai, L., & Chen, F. (2021). Uncultivated Viral Populations Dominate Estuarine Viromes on the Spatiotemporal Scale. *MSystems*, *6*(2), e01020-20 <https://doi.org/10.1128/mSystems.01020-20>
- Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356-361. <https://doi.org/10.1038/nature04160>
- Suttle, C. A. (2007). Marine viruses—Major players in the global ecosystem. *Nature Reviews Microbiology*, *5*(10), 801-812. <https://doi.org/10.1038/nrmicro1750>

- Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A., & Iagovkina, S. (2005). Antarctic climate change during the last 50 years. *International Journal of Climatology*, 25(3), 279-294. <https://doi.org/10.1002/joc.1130>
- Vaqué, D., Boras, J. A., Torrent-Llagostera, F., Agustí, S., Arrieta, J. M., Lara, E., Castillo, Y. M., Duarte, C. M., & Sala, M. M. (2017). Viruses and Protists Induced-mortality of Prokaryotes around the Antarctic Peninsula during the Austral Summer. *Frontiers in Microbiology*, 8, 241. <https://doi.org/10.3389/fmicb.2017.00241>
- Veres, P. R., Neuman, J. A., Bertram, T. H., Assaf, E., Wolfe, G. M., Williamson, C. J., Weinzierl, B., Tilmes, S., Thompson, C. R., Thames, A. B., Schroder, J. C., Saiz-Lopez, A., Rollins, A. W., Roberts, J. M., Price, D., Peischl, J., Nault, B. A., Møller, K. H., Miller, D. O., Meinardo, S., Li, Q., Lamarque, J., Kupc, A., Kjaergaard, H., Kinnison, D., Jimenez, J. L., Jernigan, C., Hornbrook, R., Hills, A., Dollner, M., Day, D. A., Cuevas, C. A., Campuzano-Jost, P., Burkholder, J., Bui, T. P., Brune, W.H., Brown, S. S., Brock, C. A., Bourgeois, I., Blake, D. R., Apel, E. C. & Ryerson, T. B. (2020). Global airborne sampling reveals a previously unobserved dimethyl sulfide oxidation mechanism in the marine atmosphere. *Proceedings of the National Academy of Sciences*, 117(9), 4505-4510. <https://doi.org/10.1073/pnas.1919344117>
- Vik, D. R., Roux, S., Brum, J. R., Bolduc, B., Emerson, J. B., Padilla, C. C., Stewart, F. J., & Sullivan, M. B. (2017). Putative archaeal viruses from the mesopelagic ocean. *PeerJ*, 5, e3428. <https://doi.org/10.7717/peerj.3428>

- Vincent, W. F. (2002). Cyanobacterial Dominance in the Polar Regions. En B. A. Whitton & M. Potts (Eds.), *The Ecology of Cyanobacteria: Their Diversity in Time and Space* (pp. 321-340). Springer Netherlands. https://doi.org/10.1007/0-306-46855-7_12
- Wagner, C., & Stadtman, E. R. (1962). Bacterial fermentation of dimethyl- β -propiothetin. *Archives of Biochemistry and Biophysics*, 98(2), 331-336. [https://doi.org/10.1016/0003-9861\(62\)90191-1](https://doi.org/10.1016/0003-9861(62)90191-1)
- Wang, S., Maltrud, M., Elliott, S., Cameron-Smith, P., & Jonko, A. (2018). Influence of dimethyl sulfide on the carbon cycle and biological production. *Biogeochemistry*, 138(1), 49-68. <https://doi.org/10.1007/s10533-018-0430-5>
- Warwick-Dugdale, J., Buchholz, H. H., Allen, M. J., & Temperton, B. (2019). Host-hijacking and planktonic piracy: How phages command the microbial high seas. *Virology Journal*, 16(1), 15. <https://doi.org/10.1186/s12985-019-1120-1>
- Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, 28(2), 127-181. <https://doi.org/10.1016/j.femsre.2003.08.001>
- Weitz, J. S., Stock, C. A., Wilhelm, S. W., Bourouiba, L., Coleman, M. L., Buchan, A., Follows, M. J., Fuhrman, J. A., Jover, L. F., Lennon, J. T., Middelboe, M., Sonderegger, D. L., Suttle, C. A., Taylor, B. P., Frede Thingstad, T., Wilson, W. H., & Eric Wommack, K. (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *The ISME Journal*, 9(6), 1352-1364. <https://doi.org/10.1038/ismej.2014.220>
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York.*, ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>

- Wilhelm, S. W., & Suttle, C. A. (1999). Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. *BioScience*, 49(10), 781-788. <https://doi.org/10.2307/1313569>
- Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., Andrews-Pfannkoch, C., Fadrosh, D., Miller, C. S., Sutton, G., Frazier, M., & Venter, J. C. (2008). The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples. *PLOS ONE*, 3(1), e1456. <https://doi.org/10.1371/journal.pone.0001456>
- Winter, C., Garcia, J. A. L., Weinbauer, M. G., DuBow, M. S., & Herndl, G. J. (2014). Comparison of Deep-Water Viromes from the Atlantic Ocean and the Mediterranean Sea. *PLoS ONE*, 9(6), e100600. <https://doi.org/10.1371/journal.pone.0100600>
- Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews*, 64(1), 69-114. <https://doi.org/10.1128/MMBR.64.1.69-114.2000>
- Xiong, Z., Wang, Y., Dong, Y., Zhang, Q., & Xu, X. (2019). Cyanophage A-1(L) Adsorbs to Lipopolysaccharides of *Anabaena* sp. Strain PCC 7120 via the Tail Protein Lipopolysaccharide-Interacting Protein (ORF36). *Journal of Bacteriology*, 201(3), e00516-18. <https://doi.org/10.1128/JB.00516-18>
- Yau, S., & Seth-Pasricha, M. (2019). Viruses of Polar Aquatic Environments. *Viruses*, 11(2), 189. <https://doi.org/10.3390/v11020189>
- Zeigler Allen, L., McCrow, J. P., Ininbergs, K., Dupont, C. L., Badger, J. H., Hoffman, J. M., Ekman, M., Allen, A. E., Bergman, B., & Venter, J. C. (2017). The Baltic

Sea Virome: Diversity and Transcriptional Activity of DNA and RNA Viruses.

mSystems, 2(1), e00125-16. <https://doi.org/10.1128/mSystems.00125-16>