High-throughput isolation and ecogenomics of phages infecting fastidious marine heterotrophic bacteria

Submitted by

Holger Heiko Buchholz

To the University of Exeter as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

In September 2021

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

(Signature) ……

Abstract

Marine Bacteria and their associated viruses are key players in shaping microbial community structures and global biogeochemical cycles. Cell lysis through viral predation is a crucial component for the recycling of carbon compounds and other nutrients, and interaction between viruses and hosts can also alter cellular functions via metabolic 'hijacking'. The culture-based study of phototrophic cyanobacterial virus-host systems has revealed that these constant co-evolutionary pressures in virus-host systems are further escalated by virally mediate horizontal gene transfer. Viruses of fastidious, heterotrophic bacteria are among the most abundant and ecologically significant virus-host systems in the oceans, but the dearth of cultured model systems has hampered the progression in testing hypotheses and interpretation of meta'omics data. The central goal of this thesis was to establish efficient and high-throughput methodologies for viral isolation, specifically targeting fastidious heterotrophic microbes. Using a novel isolate of the ubiquitous SAR11 clade as a model for highly abundant hosts, and three novel isolates of the methylotrophic OM43 clade for a lower-abundance host, I devised an optimised workflow based on established Dilution-to-Extinction culturing techniques that resulted in the isolation of over 117 viruses. The results show that among these novel viruses were the first known siphoviruses of the SAR11 clade and the first recorded viruses infecting members of the OM43 clade. Genomic evidence by the *Methylophilales* phage Melnitz indicated that inter-class host transitioning between streamlined heterotrophs may occur, and highlighted unusual viral features such as curli genes and glutamine riboswitches. Furthermore, utilising the power of metagenomics coupled with culture based host range experiments, the *Pelagibacter* phage Skadi was revealed as a highly abundant polar virus, and as an example for ecotypic niche specificity among the most abundant viruses on Earth. Culture-based identification of virus-host pairs combined with ecogenomic interpretation of metagenomic data will improve our understanding of ecological patterns and of viral strategies. The results from this thesis illuminated parts of the viral "dark matter" and highlighted possible novel viral strategies. This work can be adapted to many different systems, both of high-

and low abundance, and should be used to improve future viral isolation campaigns.

Acknowledgements

Working with some of the most difficult virus-host systems marine virology has to offer has been the greatest challenge of my life. Never have I experienced anything that was as frustrating but simultaneously as rewarding as undertaking my PhD. Without a doubt this has only been possible because of all the incredible people I had the pleasure to get to know and work with over the past years. I owe a tremendous amount of gratitude towards my primary supervisor Dr. Ben Temperton: His fierce loyalty towards his students and passion for the microbial realm has inspired me to keep going whenever I was struggling. I will miss our insomnia fuelled late night online-discussions about weird and wonderful scientific ideas, and all the crazy experiments we came up with (much to the dismay of Michelle who had to keep us grounded in reality). I hope we will get to work on many more wonky projects in the future!

I would also like to thank Professor Mike Allen, whose happiness is truly contagious, but foremost for all his grounded advice and support over the years. Speaking of calm advice, thanks also goes to Rachel Parsons in Bermuda, a wonderful scientist who has helped me make tough decisions.

To all the Tempertonians, thank you so much for being a part of my PhD experience! Especial thanks go to Ashely Bell and Luis Bolaños, who always helped me even with the dumbest bioinformatics questions. Thank you to Jo Warwick-Dugdale, for all the PhD-life advice. To Lucy Whitherall and Shayma Alathari, you were always there for me when I needed a friend. And thank you, Julie Fletcher, for being such a kind and supportive human. And of course, thank you so much Michelle Michelsen, you are the best friend and lab-wife anyone could possibly ask for! M^2, you always got my back, even when I was a pain, and I cannot overstate how thankful I am for everything you did for me, not just in the lab, but as my friend!

To Colin Börries back home: Thank you for being my oldest friend, brother. You are always there for me when I need to vent or just laugh about the vampire chef. I promise I will come back to Germany sometimes!

Lastly, to my incredible family: Needless to say, I love you all, and I could not have done this PhD without all the support and love you have given me. To my

sister Franka, thank you for always listening and talking to me when I feel lonely. My little brother Konrad, who better to talk to about all the nerdy stuff! Special thanks to my older borther Bertram and his family, you are always there when I need life advice (or just cute baby pictures)! And last but not least, to my wonderful parents: Thank you so much for installing a love for the natural environment in me from an early age, and encouraging and enabling me to pursue my dreams.

I would like to thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing SE and TE microscopy and imaging. I would also like to thank the crew of the *R/V Plymouth Quest* and our collaboratos at PML for collecting water samles, and the driver Magic for delivering water samples from Plymouth to Exeter. Thanks also goes to the crew of the *R/V Atlantic Explorer* and our collaborators at the Bermuda Institute of Ocean Sciences (BIOS).
This project is funded by the Natural Environment Research Council (NERC) Great Western Four+ (GW4+) Doctoral Training program. Additional support was given to Michelle Michelsen and Ben Temperton as part of the NERC (NE/R010935/1) grant and by the Simons Foundation BIOS-SCOPE program. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk) which is supported by the BBSRC (grant number BB/L024209/1). This project used equipment funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). Bioinformatic analyses were conducted using the high-performance computing, ISCA, provided by the University of Exeter. This project used publically available environmental data from the Western Channel Observatory, who are funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science, grant number NE/R015953/1

List of Contents

List of Figures and Tables

Chapter 1: Efficient Dilution-to-Extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria

Author’s Declaration

My thesis was prepared as three research papers, which were reformatted into chapters to allow for a uniform editorial and referencing style, with figures and tables embedded within the text. . I am the lead author on all three papers, one paper has been published, a second paper is under review and publically available as pre-print, the third paper is in preparation for submission. I conducted and planned the work, all contributions from co-authors and my supervisory team are acknowledged at the beginning of each chapter. All references in this thesis are listed at end of the thesis. Third party contributions are acknowledged in the Acknowledgements section of the thesis, as well as at the end of each chapter.

List of Definitions and Abbreviations

AMG/s	Auxiliary Metabolic Gene/s: host-derived viral genes
AAI	Average Amino Acid Identity
AMR	Antimicrobial Resistance
ANI	Average Nucleotide Identity
ANT	Antarctic/ Southern Ocean ecological zone
ARC	Arctic ecological zone
ASM	Artificial Seawater Media
ASM1	Artificial Seawater Medium defined by Carini *et al*., 2014
BATHY	Bathypelagic ecological zone
BATS	Bermuda Atlantic Time Series sampling station
Bp	Basepair
BQH	Black Queen Hypothesis; describes a mechanism by which metabolic needs of microbes can be met from compounds produces by other microbes
btuB	Vitamin B12 transporter gene
CsgA/ CsgB	Proteins that assemble into extracellular amyloid fibres
CsgF	Structural protein forming a secondary channel ~14.8Å in diameter at the neck of the CsgG beta barrel, part of curli operon
CsgG	Structural protein forming a beta barrel membrane channel ~12.9Å in diameter, part of curli operons
CsgGF	Protein complex formed by CsgG and CsgF proteins that make up an outer-membrane pore (part of the type VIII secretion system)
Curli	Set of proteins involved in amyloid fibre production for bacterial biofilm formation
Cyanophage/s	Virus/es infecting the phylum Cyanobacteria
DMS	Dimethyl sulfide
DMSP	Dimethylsulfoniopropionate

DnaA	Gene responsible for chromosomal replication initiation
DnaB	Gene encoding DNA helicases for DNA replication
DNAP	DNA polymerase
dNTPs	deoxyribonucleoside triphosphates (for DNA synthesis)
DOM	Dissolved organic matter
DtE	Dilution-to-Extinction; culturing technique involving a dilution series of viruses or microbes until theoretically only one particle/cell remains in a given volume of media
EM	Electron microscopy
EcfG	RNA polymerase sigma-factor in Alphaproteobacteria
Equal-Opportunity model	Describes the turnover of marine phages, where following the decline of the dominant bacterial population, all virus-host species have an equal chance to replace the empty niche
GF/D	Glass microfiber filter grade
glnA RNA motif	Glutamine riboswitch
glnA	Gene encoding glutamine synthases
GMQE	Global model Quality Estimate
GOV2	Global Ocean Virome (2): Dataset containing 145 viromes from five distinct global ecological zones
HCS	Humboldt Current System, cold-water current flowing from Antarctica parallel to the western South American coast
HMP	thiamine precursor 4-amino-5-hydroxymethyl-2-methylpyrimidine
HMW	High molecular weight
HPC	High-Performance Computing
Hsp20	Heatshock/ physical stress response protein
HVR/s	Hypervariable Region/s, highly variable genomic

	region within a viral or bacterial taxon
ICTV	International Committee on Taxonomy of Viruses
IPCC	Intergovernmental Panel on Climate Change
ITS	Internal Transcribed Spacer
L4	Coastal site of the Western Channel Observatory
Lysogenic cycle	Persistence and/or genome integration of the viral genetic material within the host
Lytic cycle	Viral replication cycle ending in cell lysis and progeny release
MAVGs	Metagenomically assembled viral genomes
mazG	Nucleoside triphosphate pyrophosphohydrolase; regulatory gene for controlled bacterial cell death
Mbp	Megabase pair
MeOH	Methanol
Microbial loop hypothesis	Describes the microbial link between primary production and dissolved organic carbon towards higher trophic levels
NCBI	National Center for Biotechnology Information
NERC	Natural Environment Research Council
nrdA/nrdE	Genes encoding ribonucleotide reductase alpha chain proteins
nrdB/nrdF	Genes encoding ribonucleotide reductase beta chain proteins
NSM	Natural Seawater based Media
ODV	Ocean Data Viewer, software package for visualisation of oceanographic data
ORF	Open Reading Frame
OTU/s	Operational Taxonomic Unit/s
Pac	Terminase binding site for phage DNA packaging
PC	Polycarbonate
Pelagiphage	Virus infecting members of the order Pelagibacterales
PEOD	Pacific Equatorial Divergence Province

Phage/s	Short for ‘Bacteriophage’ – virus/es infecting bacteria
phoH	Gene encoding a putative phosphate-starvation protein
PilQ	Secretin protein complex (Type IV secretion system)
Piggyback-the-Winner	A dynamic model which suggests that viral lysogenic replication is prevalent when host abundances and growth rates are high
polB	DNA polymerase gene
PMP-MAVG/s	Pelagimyophage Metagenomically assembled viral genome/s
Ppk/ppx	polyphosphate kinase
PQQ	pyrroloquinoline quinone
psbA	Gene encoding protein D1 associated with photosystem II
PTR-MS	Proton-transfer mass spectrometry (used for analysing gas composition)
PVDF	Polyvinylidene fluoride
recA	Gene involved in DNA repair and maintenance.
Red Queen Hypothesis	Evolutionary hypothesis which suggests that species must constantly adapt and compete against other species to remain competitive for resources
RNAP	RNA polymerase protein
RPKM	Reads recruited per kilobase of contig per million reads of a virome
RpoS	Sigma-factor in Gammaproteobacteria
rpsU	Gene encoding 30S ribosomal subunit 21S
Riboswitch	Regulatory RNA structure
Royal-Family Model	Describes the turnover of marine phages, where following the decline of the dominant bacterial population, the virus-host species claiming the empty niche will be related to the predecessor and

	possibly acquired resistances
S21	Ribosomal gene involved in translation initiation and mRNA binding
SAAV/s	Single-Amino Acid Variant/s
SagA	Microbial autolysin protein
S-layer	Cell surface layer-type protein
Viral Seed-Bank Model	Ecological model which states that the majority of marine viruses is present in all marine provinces but only viruses of the dominant microbes replicate and become abundant
smpB	Gene encoding RNA-binding protein for peptide-tagging of tmRNA
SNP/s	Single Nucleotide Polymorphism/s
SPSG	South Pacific Subtropical Gyre Province
ssrA	Gene encoding a bacterial two-piece transfer messenger RNA protein
TEM	Transmission Electron Microscopy
TerL	Gene encoding a large terminase subunit
TerS	Gene encoding a small terminase subunit
TFF	Tangential Flow Filtration
thyX	Gene encoding a thymidylate synthase
T4-like	Tailed dsDNA virus of the *Myoviridae* morphotype (order *Caudovirales)* similar to the model *E. coli* virus T4
T7-like	Tailed dsDNA virus of the *Podoviridae* morphotype (order *Caudovirales)* similar to the model *E. coli* virus T7
TonB	Abundant receptor protein that is part of outer-membrane substrate-specific transport
TT-EPI	Temperate to tropical epipelagic zone
TT-MES	Temperate to tropical mesopelagic zone
vConTACT2	Shared protein network based tool for taxonomic assignments of viral genera

Viral “dark matter”	Unidentifiable sequences in viral metagenomes and proteomes
Viral Shunt	Ecological mechanisms by which viral lysis prevents cell bound nutrients to move up trophic levels
Virocell	Conceptualisation of the intracellular stage of the viral replication cycle; highlights the infected cell as a biologically different entity compared to uninfected cells and free virions
VOC/s	Volatile Organic Compound/s
vSAG/s	viral Single-Amplified Genome/s
WCO	Western Channel Observatory
WEC	Western English Channel
XoxF	Methanol dehydrogenase

Chapter 1 Introduction

1.1. Viruses are an essential part of global marine ecosystems

Microorganisms are fundamental to all life on Earth – driving biogeochemical cycles and laying the foundations of global ecological food webs (Azam and Malfatti, 2007; Falkowski *et al.*, 2008, Jiao *et al.*, 2010). Bacterioplankton consume about 30-60% of the organic carbon produced by phytoplankton, which led to the 'microbial loop' hypothesis that describes the trophic link between bacterial incorporation of dissolved organic matter (DOM) to higher trophic levels in marine ecosystems (Pomeroy, 1974; Azam *et al.*, 1983). The virosphere is a fundamental part of all ecosystems and can be observed wherever life is present. More than 70% of our planet is covered by oceans, and since the 1970s (Torrella and Morita, 1979) we have known that marine viruses are exceedingly abundant, with reports of 10 to 100 million viruses per millilitre of seawater, or up to ~4 x 10^{30} marine viruses in total (Bergh *et al*., 1989; Weinbauer, 2004; Suttle, 2005). Marine virology studies often begin by putting these numbers into context, for example, the total of all marine viruses constitutes about 200 Mt carbon, equivalent to the biomass of 75 million blue whales, or, assuming an average length of 50 nm, lining up all viral particles would reach the Andromeda Galaxy, 2.5 million light years away (Suttle, 2005). Numbers like these are doubtlessly impressive, and place marine viruses as the most abundant biological entities in the oceans by at least an order of magnitude. However, such numbers offer little insight in terms of ecological understanding. Viruses are the second largest carbon pool after marine prokaryotes (comprised mostly of bacteria), and the majority of marine viruses infect bacteria (bacteriophages), hinting at their true ecological significance. Though extrapolations are prone to inaccuracy, bacteriophages ('phages') are thought to be responsible for approximately half of bacterial mortality in total (Noble and Fuhrman, 2000), killing between 15% and 40% of surface water bacteria every day (Suttle, 1994; Wommack *et al.*, 2000). This translates into an estimated 10^{28} infections per day which releases up to 10^{9} tons of cell-bound DOM back into the water column (Suttle, 2007; Breitbart, 2012), where it is available as nutrients for microbial uptake again. Approximately

a quarter of marine biomass is recycled through this ‘viral shunt’ (Wilhelm and Suttle, 1999); phages therefore significantly influence global biogeochemical cycles (Weitz and Wilhelm, 2012; Weitz *et al.*, 2015).

The ecology of marine viruses and bacteria is clearly intertwined, therefore the dynamics between virus and host populations are important when considering the impact of particular virus-host systems on marine biogeochemistry. Using a combination of metagenomes and culture-based studies of viral diversity, Breitbart et al. describe a global distribution of near identical double-stranded DNA (dsDNA) phages of both high and low abundances (Breitbart *et al.*, 2002, 2004). As viral genotypes are in constant flux depending on which host is abundant, i.e. which phage genotype produces high numbers of progeny, they suggest that the majority of low-abundance genotypes found at any location are not actively reproducing, but are part of a “seed bank” fraction – described as the viral Seed-Bank model (Breitbart and Rohwer, 2005). When host conditions become favourable, associated phages increase in abundance, whereas the previously active fraction begins to decrease in abundance and becomes part of a long tail of rare viral taxa. To help explain the turnover of marine phages in a niche with high diversity of viruses and hosts, two models were proposed (Breitbart *et al.*, 2018). The ‘Equal-Opportunity’ model states that all hosts and phages have an equal chance to become dominant after viral predation decimated a bacterial population; whereas the ‘Royal-Family’ model suggests that dominant virus-host systems are more likely to be succeeded by closely related taxa that have evolved resistance to their most abundant phages. Though a large percentage of phages are thought to exhibit high host specificity (Koskella and Meaden, 2013), a broad range of suitable hosts allows viruses to replicate even if another specific host type is insufficiently abundant (Ross *et al.*, 2016). Often broad host ranges are thought to be less efficient for viral replication, whereas specialisation on a narrow range of hosts allows the virus to maximise progeny production (Zborowsky and Lindell, 2019), thus phage life strategies are often partitioned into generalist and specialists. Another aspect of viral life strategies is the persistence continuum of purely lytic phages to lysogenic phages that integrate into the host genome (Warwick-Dugdale *et al.*, 2019). Because viruses are obligate molecular parasites that interact with host DNA in a variety

of ways, horizontal gene transfer between viruses and hosts is common, creating characteristically mosaic phage genomes which feed into the large reservoirs of viral genetic diversity in the oceans (Pedulla *et al.*, 2003; Mavrich and Hatfull, 2017). However, this vastness of phage diversity in the ocean, as well as the technical challenges of cultivating suitable hosts (Steen *et al.*, 2019), has limited experimental examination for many marine phages.

In recent years more attention has been paid to the influence of phages on the cell during infections, and how viruses shape microbial community structures beyond virally induced cell lysis. The power of culture-dependent experiments has been demonstrated through the study of phototrophic cyanobacteria and their associated viruses. Cyanobacteria of the diverse genera *Synechococcus* and *Prochlorococcus* are estimated to contribute about 25% of oceanic oxygen production or about 15% of the global oxygen production (Rappé *et al.*, 1997; Field *et al.*, 1998; Flombaum *et al.*, 2013). Subsequently, the study of viruses infecting this essential microbial group (cyanophages) has received much attention (Mann, 2003; Sullivan *et al.*, 2003). Selective pressure and horizontal gene transfer enable increased co-evolution and a constant microbial arms race between phages and bacteria, with both evolving to overcome host defences or evade infections, respectively (Weitz *et al.*, 2005; Roux *et al.*, 2014; Mavrich and Hatfull, 2017). Many phages carry host-derived metabolic gene encoding sequences that allow the phage to hijack the cellular machinery and express these Auxiliary Metabolic Genes (AMGs) to maximise the uptake and metabolism of compounds required for phage replication (Hurwitz *et al.*, 2016; Warwick-Dugdale *et al.*, 2019). This can drastically alter the ecological niche of infected cells, giving rise to the concept of “virocells” as functionally distinct entities compared to uninfected cells (Forterre, 2011, 2013; Howard-Varona *et al.*, 2020). During cyanophage infection viral AMGs have been shown to redirect metabolic and photosynthetic pathways of the cell, altering rates of energy production and consumption (Thompson *et al.*, 2011) and manipulate uptake-rates for compounds required for viral replication, such as nitrogen (Waldbauer *et al.*, 2019) and phosphate (Wilson *et al.*, 1996, 1998). Many cyanophages carry photosynthetic genes (Mann *et al.*, 2003; Sullivan *et al.*, 2005), with some encoding both photosystem I and II proteins (Fridman *et al.*, 2017). These genes

have been experimentally confirmed to translate into proteins during infection (Lindell *et al.*, 2005). Cyanophages can also inhibit CO_2 fixation in their host (Puxty *et al.*, 2016), or could prevent photo-inhibition when sunlight is too intense (Mann *et al.*, 2003). Experiments like these have therefore provided knowledge about how cyanophage infection in the most important phototrophs has profound impacts on their metabolic role in the environment, with major implications for global carbon fixation rates (Puxty *et al.*, 2018). However, for the majority of marine microbial taxa outside of phototrophic cyanobacteria, virus-host model systems are few and far between, either because the hosts continue to evade culturing, or because traditional plaque-based methods for viral isolation are unsuitable. Our understanding for the viral ecology involving such taxa is primarily based on culture-independent methods.

1.2 Advantages and disadvantages of computational marine virology

With the advent of high performance computing and ever-improving sequencing technology, culture-independent techniques such as metagenomics and single-cell/virus genomics have over the past decade become the primary tools of marine virology (Brum and Sullivan 2015; Hingamp *et al.*, 2013; Mizuno *et al.*, 2013; Roux *et al.*, 2016; Coutinho *et al.*, 2017; Gregory *et al.*, 2019; Martínez Martínez *et al.*, 2020). Indeed, these approaches have proven themselves as invaluable for investigating marine viruses, leading to hundreds of metagenomes, and the recovery of metagenomically assembled viral genomes (MAVGs) in the ten-thousands, resolving much of the viral biodiversity on a population level for large parts of all surface oceans (Mizuno *et al.*, 2013; Roux *et al.*, 2015; Hurwitz *et al.*, 2018). A plethora of computational methods to interpret existing datasets have been developed based on model virus-host systems, enabling for example the identification and classification of viral contigs and genes (e.g. VirSorter, VirFinder, MARVEL), taxonomic and phylogenetic analyses (Turner *et al.*, 2021) and quantification of viral population boundaries. However, only an estimated 10% of viral contigs in metagenomes are somewhat similar to viruses with a known or putatively assigned host, limiting ecological interpretations of genomic databases to viruses that have been brought into culture or are otherwise closely

related to previously identified virus-host systems (Hurwitz and Sullivan, 2013; Norman *et al.*, 2015). Further limiting the application possibilities of *in silico* studies is that these 10% of viruses are heavily biased by available model systems towards those whose hosts grow readily in laboratory conditions. From more than 13,830 unique cultured phages available across various platforms, e.g. the National Center for Biotechnology Information (NCBI, (Brister *et al.*, 2015)), Cook et al. point out that one third of these phages were isolated on *Mycobacterium* and *Gordonia* which were used as hosts by the successful SEA-PHAGES program (Hanauer *et al.*, 2017; Cook *et al.*, 2021). In contrast, only 32 phages of the abundant and ecologically critically important order *Pelagibacterales* exist (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Du *et al.*, 2021). This problem is further escalated by the difficulties in assembling genomes of viruses infecting many marine microbial phyla, as is the case with *Pelagibacter* phages due to their unique genomic properties (Warwick-Dugdale *et al.*, 2019). Subsequently, genomes of many abundant and ecologically important viral taxa, such as *Pelagibacter* phages, are woefully underrepresented in assembled genomes (Nayfach *et al.*, 2021). Therefore, computational methods such as metagenomics are evidently useful tools for identifying viruses, but current state-of-the-art technology is still limited by available references. Without cultured model systems metagenomics is unable to answer fundamental ecological questions such as what viruses are doing how in the environment. In other words, viral metagenomics has achieved an increase in 'known unknowns', but experimentation is indispensable to attain 'known knowns'; and for any *in vitro* experiments, cultured viruses are a prerequisite. Whilst culturing unknown viruses requires time and resources, the effort of acquiring model systems is iteratively rewarded because new systems and reference genomes facilitate the re-interpretation of all historic metagenomics data and provide improved ecological meaning. The improved understanding of data interpretations will in turn allow for the raising of new hypotheses, which can be tested experimentally, and will likely lead to improved cultivation concurrently, using reverse genomics for targeted isolation (Cross *et al.*, 2019), similar to how the isolation of SAR11 improved Dilution-to-Extinction (DtE) cultivation methods (Carini *et al.*, 2013; Giovannoni, 2017; Henson *et al.*, 2020). Therefore the benefits of model virus-

host systems isolation and cultivation are circular and are amplified on each turn of the wheel.

1.3. Genome Streamlining in oligotrophic bacteria

To isolate any phage, a cultured bacterial host is required. In terms of sheer abundance, the vast majority of heterotrophic pelagic bacteria belong to oligotrophic taxa populating the nutrient-poor pelagic environment (Rappé *et al.*, 1997; Schattenhofer *et al.*, 2009; Nelson *et al.*, 2014; Korlević *et al.*, 2015). The definition of *oligotrophy* refers to microbial growth adapted to low-nutrient environments such as most oceanic provinces. In contrast, the term *copiotrophy* applies to microbes that are defined by their rapid growth at high nutrient concentrations, which is typical for chemo-heterotrophic bacteria that readily grow in laboratory culture conditions (Lauro *et al.*, 2009). Related to the concepts of oligotrophs and copiotrophs are the terms *r* and *k* strategists, where *R* strategist refers to reproductive success through rapid reproduction, out-growing competitors in copiotrophic environments, whereas *k* strategist refers to bacteria that grow slowly but outcompete other microbes via efficient replication and resource management (Klappenbach *et al.*, 2000). According to genome streamlining theory, low-nutrient environments and other environmental stress leads to reductive evolution in microbes, selecting for efficient metabolism, smaller cells, and reduced genomes with fewer regulatory mechanisms (Giovannoni *et al.*, 2005; Giovannoni *et al.*, 2014; Feng *et al.*, 2021; Simonsen, 2021). In oligotrophic oceans, competition for low levels of nutrients is extreme and requires efficient high-affinity transporters (Hirsch *et al.*, 1979). The effective competition for rare nutrients by oligotrophs has been addressed by Don Button's uptake kinetic theories (Button 1994). In order to exploit rare "pockets" of nutrients in the water before the competition from other microbes can, the reduced regulatory mechanisms allows many oligotrophs to continuously take up all substrates without the delay from transcriptional regulation (Sowell *et al.*, 2009). In addition, many oligotrophic bacteria have a reduced genome size, which maximises reproduction with minimal nutrient requirements. The Black Queen Hypothesis (BQH) suggests that losing genomic features may give a selective

advantage via reduced metabolic costs for cell division, provided the genomic losses can be compensated for, and the cell remains functional (Morris *et al.*, 2012). In marine oligotrophic environments this may be achieved through coevolution, where essential metabolic requirements after the loss of a metabolic function via adaptive gene loss can be substituted for with a different nutrient source i.e. compounds that were produced by other microbes and released as communal goods (Tripp *et al.*, 2008; Morris *et al.*, 2011). Whilst advantageous in oligotrophic environments, chemo-heterotrophic oligotrophs with streamlined, reduced genomes are extremely challenging to work with in the laboratory (Giovannoni *et al.*, 2005).

1.4 Cultivation challenges for oligotrophic bacterial hosts

The “great plate count anomaly” was described in 1985 (Staley and Konopka, 1985), addressing the discrepancies between the numbers of colony forming bacteria that could be grown on agar plates and numbers of cells that previous studies observed through microscopy in samples from the natural environment. This effect is particularly well known in marine environments, where only a small fraction of bacteria grow on agar plates (Kogure *et al.*, 1979, Joint *et al.*, 2010). To increase cultured representation of marine phyla, Dilution-to-Extinction (DtE) methods were developed with natural seawater based media (NSM), allowing for the isolation of highly diverse and abundant marine microbes without the need for plates (Connon and Giovannoni, 2002). However, nutrient concentrations of natural seawater vary in composition depending on seasons and locations. Hence, studies using NSMs are challenging to reproduce with exact growth conditions. Under BQH dynamics, some oligotrophs might still not be able to grow without the additional supply of compounds that are produced by other microbes in the environment. Consequently, researchers have used culture experiments and genomes of model microbes isolated on NSM to identify nutrient requirements for designing artificial seawater media (ASM) targeting specific bacterial clades (Carini *et al.*, 2013; Sosa *et al.*, 2015). These nutrient requirements can be difficult to fulfil in ASM, because streamlined genomes often lack the metabolic pathways to synthesize required compounds, for example,

thiamine (Vitamin B_1) is essential for growth of *Candidatus Pelagibacter ubique*, but the bacteria cannot use exogenous thiamine, and requires 0.7 zeptomoles of the thiamine precursor 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) per cell (Carini *et al.*, 2014). In addition, oligotrophs are highly sensitive to inorganic cofactors such as metals and carbon compounds (Beier *et al.*, 2015). This is challenging in laboratory conditions, because even nanomolar disparities in nutrient concentrations or accidental addition of compounds (e.g. “leaking” carbon atoms from polypropylene plastic-ware or syringe filters, or iron residue from autoclaves) will cause cultures to fail. The sensitivity to carbon compounds also causes many oligotrophic bacteria to succumb to high-nutrient stress from agar plates, in addition many marine phyla lack the ability to form colonies (Rappé *et al.*, 2002; Simu *et al.*, 2005), and thus agar plates are ill-suited for the study of the marine microbiota. Most oligotrophic bacteria are slow growing with generation times of up to two days (reviewed in (Kirchman, 2016), resulting in up to several months of incubation time required for DtE experiments. Furthermore, many bacteria under nutrient limitation stress use transcriptional regulation to enter a dormant state, e.g. the sigma factor RpoS in Gammaproteobacteria (Chiang and Schellhorn, 2010) or sigma factor EcfG in Alphaproteobacteria (Fiebig *et al.*, 2015), but oligotrophic cells often lack any orthologous genes for entering a dormant state. As a result, oligotrophic cells have to be kept in continuous exponential cultures and do not readily come out of cryostorage, significantly increasing the effort and costs required to maintain cultures. Due to cultivation challenges like these, many bacterial taxa and subsequently their associated viruses still lack representation in culture collections (Hug *et al.*, 2016; Steen *et al.*, 2019).

1.5 *Pelagibacterales* – The SAR11 clade

The most abundant and arguably successful oligotrophic microbes belong to the alphaproteobacterial order *Pelagibacterales* (known as the SAR11 clade, named after the 11th clone from a 16S rRNA gene amplicon library prepared from a Sargasso Sea DNA sample), whose members have some of the smallest known genomes of any free living microorganisms (~1.3 megabase pairs (Mbp),

compared to 4.5 to 5.5 Mbp for *Escherichia coli*, depending on strain specific chromosomal loci and plasmids (Rode *et al.*, 1999)). SAR11 have a global population size that encompasses ~25% of all bacterioplankton on average and are found in all aquatic environments (Giovannoni, 2017). Though their exact evolutionary origin is under debate, phylogenetic reconstruction suggests that SAR11 are an ancient lineage at the root of the alphaproteobacterial phylogenetic tree together with mitochondria and *Rickettsiales*, and may directly share a marine ancestor, one branch of which then evolved a symbiosis leading to the first eukaryotes (Thrash *et al.*, 2011). Due to their huge population sizes, SAR11 have received much attention since their discovery (Giovannoni *et al.*, 1990) and initial isolation (Rappé *et al.*, 2002). However, it still took well over a decade of research to develop methods for cultivating *Pelagibacterales* in artificial media (Carini *et al.*, 2013, 2014). In recent years, DtE culturing underwent a renaissance, using information from metabolic genes identified in available genomes to design artificial seawater media that fulfil the metabolic requirements of many oligotrophic taxa, yielding hundreds of novel marine bacterial isolates, including SAR11 (Stingl *et al.*, 2007; Song *et al.*, 2009; Carini *et al.*, 2013; Henson *et al.*, 2016, 2020). As a direct result, our knowledge about the most abundant taxa such as SAR11 has improved greatly and has procured strains that span the diverse SAR11 clade. All these strains have similar phenotypes in culture with small, aerobic, chemoheterotrophic cells and doubling times of up to 48 hours (Rappé *et al.*, 2002; Giovannoni, 2017). Due to culture based experiments, SAR11 have been confirmed as a fundamental part of marine microbial communities, not only due to their abundance, but also because they are directly linked to global carbon, nitrogen, phosphate and sulfur ocean cycling (Malmstrom *et al.*, 2004; Schwalbach *et al.*, 2010; Carini *et al.*, 2015; Tsementzi *et al.*, 2016). SAR11 bacteria are predominant in surface waters, where they are responsible for about 50% of amino acid assimilation and 30% of dimethylsulfoniopropionate (DMSP), metabolising both at higher rates compared to other prokaryotes and thus effectively outcompeting them in oligotrophic conditions (Malmstrom *et al.*, 2004). Sun et al. (2011) also reported the capacity of SAR11 for demethylation of C1 carbon compounds, and thereby SAR11 population might collectively play a significant role for recycling CO_2 in the upper ocean (Sun *et al.*, 2011). SAR11

also metabolises volatile organic compounds (VOCs), a diverse group of climate active gases such as methanol, acetone and isoprene that are typically produced by phytoplankton (Halsey *et al.*, 2012, 2017). Therefore, SAR11 can transfer carbon directly from primary producers to bacterioplankton, but the influence of viral hijacking on such processes is currently unknown.

Copiotrophs often regulate metabolism on a transcriptional level via regulatory genes, but bacterial metagenomes from the Sargasso Sea revealed over three hundred different riboswitches (regulatory RNA structures), suggesting riboswitches are common in oligotrophic cell regulation (Kazanov *et al.*, 2007). Indeed, SAR11 cells have an unusually large number of riboswitches that control main functions of the cell (Meyer *et al.*, 2009). For example, SAR11 utilises intracellular glycine levels to activate glycine riboswitches that regulate its carbon metabolism (Tripp *et al.*, 2009). The advantage of riboswitches, though simpler compared to transcriptional control, is that they are more economical and respond quicker to external factors than transcriptionally controlled features do. Transcription regulatory systems in SAR11 cells are minimal (Giovannoni, Tripp, *et al.*, 2005). SAR11, like many other marine bacteria, use proteorhodopsin (Fuhrman *et al.*, 2008), though in SAR11 this typically light-dependent proton pump does not enhance light-induced growth (Giovannoni, Bibbs, *et al.*, 2005). Instead SAR11 constitutively express proteorhodopsin, which in light reduces respiration and maintains biomass during carbon starvation, whereas starved SAR11 in the dark metabolise endogenous carbon, leading to shrinking cells (Steindler *et al.*, 2011). Particularly challenging for SAR11 cultivation are its extremely slow growth rates (up to 48 hours per cell division), which also might be a result of missing regulatory features. Maximal growth rates and rRNA operon numbers have been correlated, and SAR11 only have a single rRNA operon, which may contribute to the growth rates observed in SAR11 (Giovannoni, Tripp, *et al.*, 2005; Vieira-Silva and Rocha, 2010; Giovannoni, 2017).

Due to small, streamlined genomes in SAR11, the proportion of core genes is high with 48 to 56%, yet can be separated into highly diverse lineages that are often categorized as subclades Ia-c, IIa-b, IIIa-b, IV and V (Grote *et al.*, 2012). Subclade Ia is highly conserved with 16S identity of >98% and average nucleotide identity (ANI) of >75%, and members of the subclade Ia are also

frequently used as model organisms in culture experiments, *Pelgagibacter ubique* HTCC1062 in particular is a common strain used as a reference. However, combined analyses of the internal transcribed spacer (ITS) loci and the corresponding 16S rRNA genes has revealed high micro- and macrodiversity within populations (Ngugi and Stingl, 2012) and analysis of SAR11 genetic fingerprints for spatiotemporal distribution patterns have shown that subclades can be partitioned by season, depth and marine province (Field *et al.*, 1997; Carlson *et al.*, 2009; Schattenhofer *et al.*, 2009; Giovannoni and Vergin, 2012; Salter *et al.*, 2015). Ecotypes in bacteria are thought of as genetically closely related groups with minor differences in their physiology that allow them to specialise into ecological niches (Cohan, 2006). High-resolution phylogenetic analysis highlighted that the subclades of SAR11 could be classified as ecotypic groups (Vergin *et al.*, 2013), e.g. Ia as a surface-ocean ecotype, which can be further subdivided into the cold-water surface ocean ecotype Ia.1 and the warm-water surface ocean ecotype Ia.3, which are often represented by the strains HTCC1062 and HTCC7211, respectively (Brown *et al.*, 2012, Vergin *et al.*, 2013). It is noteworthy, that the 16S rRNA gene between HTCC7211 and HTCC1062 only differs by 15 single nucleotide polymorphisms (SNPs), HTCC7211 also encodes the same set of genes except for an additional polyphosphate kinase (*ppk* and *ppx*) and a phosphonate transporter (Grote *et al.*, 2012). Using single-amino acid variants (SAAVs) Delmont et al. deduced that environmentally mediated selection drives biogeographic partitioning of SAR11 ecotypes at a strain level (Delmont *et al.*, 2019). Likely, adaptive convergence within highly conserved and streamlined SAR11 subclades allows strains to thrive in different marine habitats, despite only relatively small differences in the genotype (Salter *et al.*, 2015). Based on the study of T4-like marine cyanophages, it has been suggested that the abundance of virus populations reflects that of their hosts (Chow *et al.*, 2014). In SAR11 such closely related ecotypes might increase the proportion of viruses that can infect across multiple ecotypes, though ecotypic distribution of SAR11 virus-host systems has not yet been explored and will require cultured representatives to test host ranges.

High levels of polar endemism and bi-polar ecotypes have been reported for SAR11 (Ghiglione *et al.*, 2012; Kraemer *et al.*, 2020), thus SAR11 are likely

important in polar microbial ecosystems as well. Reports by the Intergovernmental Panel on Climate Change (IPCC) summarize climate research, which shows that the Arctic Ocean is particularly vulnerable to increasing temperatures (IPCC report 2019). This will reduce the annual sea ice extent, which based on observations from the 2007 sea ice minimum, will reduce the diversity of bacterioplankton significantly (Comeau *et al.*, 2011). However, a shift towards a seasonally ice-free Arctic Ocean might overall increase microbial productivity and respiration and thereby affect nutrient cycles (Slagstad *et al.*, 2015; Underwood *et al.*, 2019). These changing environmental pressures could cause a shift towards warm-water SAR11 ecotypes, and in turn influence the community structure of associated virioplankton. To understand and test how a warming ocean might influence ecotypic SAR11 virus-host population dynamics, it is essential to identify and isolate endemic polar strains.

1.6. *Methylophilales* – The OM43 clade

The same cultivation campaign that resulted in the isolation of SAR11 also isolated the first cultures of OM43 (Connon and Giovannoni, 2002). The OM43 clade is most closely related to the obligate methylotrophic *Methylophilaceae*, an aerobic family within the Gammaproteobacteria that relies solely on methylated carbon compounds as a nutrient source (Kalyuzhnaya *et al.*, 2006). OM43 are phenotypically similar to SAR11, with highly streamlined genomes of about 1.3 Mbp and small cell size (Giovannoni *et al.*, 2008; Jimenez-Infante *et al.*, 2016). Unlike SAR11, members of the OM43 clade make up a much smaller proportion of the microbial community, but with about 5% of coastal bacteria, they are an ecologically important part of marine ecosystems, particularly because of their role as marine methylotrophs (Morris *et al.*, 2006; Halsey *et al.*, 2012). Observations made from the Oregon coast show that OM43 abundance coincides with phytoplankton blooms (Morris *et al.*, 2006). Phytoplankton produce methylated C1 compounds and VOCs as by-products during primary production (Sinha *et al.*, 2007). Methanol is the single largest component of VOCs in the atmosphere and involved in the production of ozone in the troposphere and other climate-relevant gases. Previous work has shown that metabolism of methanol

by methylotrophic bacteria such as OM43 represents a major link between primary production and conversion of carbon back to atmospheric CO_2, making methylotrophy an important component of global carbon budgets (Halsey *et al.*, 2012, 2017; Thrash *et al.*, 2014; Beale *et al.*, 2015). Metatranscriptomics has shown that OM43-like methanol dehydrogenases (XoxF) account for up to 2.3% of all peptides in coastal marine provinces (Sowell *et al.*, 2011; Hanson *et al.*, 2014). OM43 also encodes XoxF, allowing them to utilise methanol both for conversion into biomass (assimilation) and energy production (oxidation) at a rate of 43 attomols per cell per day (Halsey *et al.*, 2012). In contrast, SAR11 uses methanol exclusively for energy production only and at a much lower rate of 1.2 attomols per cell per day (Sun *et al.*, 2011). Although OM43 is thought to specialize for C1 compounds, media enriched with complex high molecular weight (HMW) DOM selected for methylotrophic OM43 in culture experiments (Sosa *et al.*, 2015). The authors pointed out that HMW DOM contains a large fraction of methylated sugars (Quan and Repeta, 2007), and abiotic oxidation could release low molecular weight compounds that in turn serve as nutrients for methylotrophs. Indeed, quantitative transcriptomics revealed upregulation of OM43 XoxF as a response to the addition of both DOM and methanol (Gifford *et al.*, 2016). This highlights the important role of OM43 in coastal ecosystems for the oxidation of methanol and C1 compounds as well as DOM. And yet, only three reported OM43 with fully sequenced genomes have been reported from the Pacific North-West Coast (Giovannoni *et al.*, 2008), Hawaii (Huggett *et al.*, 2012) and the Red Sea (Jimenez-Infante *et al.*, 2016), as well as isolates from the LD28 group that are thought to be a freshwater variant of OM43 (Salcher *et al.*, 2015). Though likely subject to continuous updates when more strains are made available, pangenomic analysis suggested that more than half of OM43 genes are conserved between strains, similar to the highly conserved SAR11 pangenome (Jimenez-Infante *et al.*, 2016). Phylogenetic analysis of the 16S-23S ITS region further indicated that, like SAR11, OM43s can be partitioned into environmentally driven niches (Jimenez-Infante *et al.*, 2016). Though uncertain, it is thought that the element lanthanum might govern the biogeography of OM43, since it is a cofactor of their XoxF, which is fundamental for its metabolic pathways (Giovannoni *et al.*, 2008; Good *et al.*, 2019). Overall, OM43 is an ecologically

important streamlined heterotrophic taxa similar in physiology to SAR11, and is an ideal subject to test viral isolation methods for a less abundant host.

1.7 Discovery of abundant viruses for SAR11 tested paradigms of viral ecology

The innovations for isolating fastidious oligotrophic bacteria in the past two decades, and the resulting broadening of our understanding of them, have not been matched by equivalent advancements with their associated viruses. This is largely due to the inherent difficulties of isolating viruses for hosts that do not grow on solid agar plates, invalidating many established techniques of virology, such as plaque assays. As a result, no viruses could be identified for SAR11 even a decade after their isolation. This fact prompted alternative theories of 'cryptic escape' to explain the seemingly contradicting observations of high abundances of SAR11 despite slow growth rates. 'Cryptic Escape' suggested that SAR11-like oligotrophic picoplankton have either not enough biomass per cell, or do not reach sufficiently dense populations (cells/ mL) to support viral replication (Yooseph *et al.*, 2010). Another explanation was the 'defence specialism' hypothesis, which suggests that high SAR11 abundance could be caused by a lack of top-down controls, i.e. viral predation (Suttle, 2007; Våge *et al.*, 2014). Defence specialism states that SAR11 are abundant because they invest in viral defence mechanisms and thereby minimise population losses to viral predators, not by increasing competitiveness for resources against other microbes. Therefore, under defence specialism dynamics, few to no viruses of SAR11 would be expected. However, these concepts were demonstrated to be inaccurate with the isolation of HTVC010P in 2013 (Zhao *et al.*, 2013). HTVC010P is one of the first four reported phages infecting SAR11 (pelagiphages) and are now known as some of the most abundant viruses in marine viral metagenomes, even though assemblies had failed to recover it from metagenomes prior to its isolation. Though this is indisputable evidence for the existence of pelagiphages and the significant predation of viruses on SAR11, these bacteria might still allocate a large proportion of their resources towards viral defence, and, under the assumption of a cost of resistance, the existence of abundant pelagiphages does not necessarily contradict the defence specialism

hypothesis (Våge *et al.*, 2014). The King-of-the-Mountain hypothesis (Giovannoni *et al.*, 2013) was proposed in response to 'defence specialism' with large pelagiphage populations, and states that the winning bacterial strains remain dominant by passing on viral resistances, creating a positive feedback loop that rewards superior resource competition, and is supported by high genetic recombination rates in SAR11 (Vergin *et al.*, 2007). Whatever the case, the isolation of HTVC010P serves as an example of how the discovery and culturing of a single virus can change the field.

Well-studied lab model systems, e.g. *E. coli,* are known to deploy multiple strategies to defend against aggressive mobile genetic elements such as viruses. These strategies include adaptive (acquired) immunities, such as the well-characterised CRISPR-Cas mechanism (Koonin & Makarova 2013), and innate immunities, such as restriction modification proteins (Pingoud *et al.*, 2005). Other strategies such as dormancy and programmed cell death as a response to viral infection might use for example toxin-antitoxins (Engelberg-Kulka *et al.*, 2005). Screening 4961 complete prokaryotic genomes for genes involved in viral defense showed that most bacteria dedicate 1% or more of their genome to various defence mechanisms (Koonin *et al*., 2017). As SAR11 species have between 1357 and 1576 genes (Grote *et al.*, 2012), about 13 to 15 genes for viral defence would be expected. Applying bioinformatics tools (with default settings) for finding viral defence genes such as DefenseFinder v.1.0.1 (Tesson *et al.*, 2021) and the Prokaryotic Antiviral Defence Locator (PADLOC) v.1.1. (Payne *et al.*, 2021) to the three SAR11 and three OM43 host genomes, used as model systems in this work (see Chapter 2), identified no defence-associated genes in the OM43 strains. A restriction endonuclease and a DNA methyltransferase were identified in SAR11 strains HTCC7211 (PB7211_69, PB7211_1097), HTCC1062 (SAR11_RS00570, SAR11_RS00560) and H2P3α (Genes 1208-1210), which form a type II restriction-modification system that can degrade invading DNA by methylation (Pingoud *et al*., 2005). From all three screened SAR11 genomes additional dGTPase genes (hmm accession number PDLC02374) were found in HTCC7211 (PB7211_671); HTCC1062 (SAR11_RS04845) and H2P3α (gene 685) as well as OM43 strain H5P1 (gene 399). dGTPases could be used to deplete the deoxynucleotide dGTP and thereby starving the phage infection of

essential DNA building blocks (Tal *et al*., 2021). This means in a SAR11 genome of about 1.3 MBp, only about 0.2 % of the genome might be dedicated to viral defence. Furthermore, DNA methyltransferases are highly diverse, DNA methylation in the host genomes may not primarily function as phage defence but could also be involved in DNA repair or DNA replication (Russo *et al.*, 2016). Thus, the extent of phage defence in marine oligotrophs such as SAR11 and OM43 remains uncertain, and therefore does not rule out either the defence specialism or King-of-the-Mountain hypotheses. To settle the debate, experimental evidence of resource allocation in response to viral predation and further evidence of the function of possible phage defence genes is required.

The isolation of the first pelagiphages in liquid culture, whilst important in itself, also proved that the deployed DtE approach is a viable alternative to plaque assays for isolating phages of chemo-heterotrophic oligotrophs in liquid culture, which is supported by the isolation of an additional eleven pelagiphages in 2018 (Zhao *et al.*, 2018). Zhao et al. demonstrated that these types of pelagiphages could replicate by both lytic and lysogenic infection strategies, using host SAR11 tRNAs as integration sites. A further 17 pelagiphage isolates were obtained in recent years (Zhang *et al.*, 2020; Du *et al.*, 2021), which could be grouped into six phylogenetic types of podophages, hinting at high pelagiphage diversity in the oceans. With an increasing number of pelagiphage genomes that can be used as references, our capabilities to interpret metagenomic datasets have greatly improved, further revealing previously hidden genomic diversity within the marine virosphere and partially illuminating the viral “dark matter” (Mizuno *et al.*, 2013; Brum *et al.*, 2016; Zaragoza-Solas *et al.*, 2020). However, every breakthrough opens up more questions. Though pelagiphages are thought to be hugely abundant, metatranscriptomics data collected in the coastal North Atlantic over two years did not match their abundance in metagenomes, indicating that known lytic pelagiphages are less active than other viruses in that region (Alonso-Sáez *et al.*, 2018). Estimates of the microbial influence on carbon export based on metagenomics data gathered during the *Tara* Oceans expedition further suggested that phages of SAR11 lack influence on the carbon cycle despite their abundance (Guidi *et al.*, 2016). A possible explanation for these counterintuitive observations might be found in reports which show that pelagiphages assemble

particularly poorly from commonly used short-read sequencing based metagenomes (Warwick-Dugdale *et al.*, 2019). Therefore, ecological interpretations based purely on meta 'omics data could be skewed and may underestimate the true influence of pelagiphages. To resolve these intriguing questions to a satisfying degree, obtaining viral isolates in culture to measure processes such as carbon metabolism and cycling are indispensable in order to verify any ecological and biochemical models.

Life strategies are varied between phages of phototrophs and heterotrophs (Deng *et al.*, 2012). Whilst the cultivation efforts for pelagiphages have returned invaluable results, after a decade of research, viruses of streamlined heterotrophs are still difficult to attain and only scratch the surface of the marine virosphere, as viral model systems for many ecologically important microbial taxa such as OM43, SAR86 or SAR202 remain unavailable. With increasing climate concerns, it is critical to address the current paucity of model systems which will help rectify biases in our understanding of viral influences on global biogeochemical cycling. The first step needs to be the establishment of efficient high-throughput viral isolation methods targeting streamlined heterotrophs, similar to the achievements made for cyanobacteria and cyanophages. Only by having ecologically relevant model systems for both phototrophy as well as heterotrophy can we start to parameterise models of carbon flux, and understand the marine virosphere as a whole.

1.8. Summary of thesis chapters

Here, I develop a high-throughput viral isolation method, and demonstrate its efficiency and application possibilities by isolating viruses from the Western English Channel and Sargasso Sea, using SAR11 and OM43 as representative high and low abundance model host strains for bait. In **Chapter 2, '*Efficient Dilution-to-Extinction isolation of novel virus-host model systems for fastidious heterotrophic bacteria*'**, I detail the development of a highly efficient high-throughput viral isolation method based on established DtE culturing for fastidious marine heterotrophs. This workflow was designed to be low-cost, easily adaptable to other taxa, and efficient. I used two well-studied SAR11 ecotypic

strains and isolated a new strain of SAR11 and three new strains of OM43 from the Western English Channel, as model systems representing high- and low-abundance hosts. As proof of concept, I applied my method on monthly water samples from the Western English Channel, resulting in the isolation of hundreds of viral isolates. I analyse and describe a subsample of the isolate genomes, revealing the first known siphophage infecting SAR11 and the first phage infecting the OM43 clade. In this chapter I also introduce the 'Soylent Green' hypothesis, explaining the unusual growth dynamics observed in infected SAR11 cultures. Conventional viral infections using liquid cultures result in rapid lysis and cell death, however, I routinely observed that infections of SAR11 and OM43 cultures led to steady-state cell densities not dissimilar to the stationary phase associated with copiotrophic taxa, despite SAR11 lacking any genes for stationary phase metabolic reconfiguration. The Soylent Green hypothesis states that a subpopulation of the host does not fall victim to viral lysis and can sustain itself by recycling nutrients from lysed cells of the same population. The chapter concludes with a general estimation of the abundance, distribution and ecological importance of the novel phages based on read recruitment from Global Ocean Viromes (GOV2; (Gregory *et al.*, 2019)).

In Chapter 3, '***Genomic evidence for inter-class host transition between abundant streamlined heterotrophs by a novel and ubiquitous marine Methylophage***', I introduce four variant strains of Melnitz, the first OM43 myophage, isolated in both the Western English Channel and the Sargasso Sea. The chapter provides an in-depth genome analysis, revealing several unusual features, such as a curli operon and a complete virally encoded tmRNA gene (*ssrA*) controlled by a glutamine riboswitch. Curli genes have been found in MAVGs of related myophages infecting SAR11 previously (Zaragoza-Solas *et al.*, 2020), yet their function in phages remains speculative. The results presented here suggest that curli pores may have been repurposed as a functional analogue for the regulation of viral lysis. The genetic evidence also supports an inter-class host transition event between phages of OM43 and SAR11, as (1) Melnitz and pelagiphages are closely related taxonomically, (2) phylogenetic analysis of shared *ssrA* genes suggests that the gene in Melnitz was acquired from an Alphaproteobacteria, not the gammaproteobacterial OM43; and (3) alignments of

tRNA genes found in these phages suggests genetic exchange between SAR11 and OM43 virus-host systems. Combined with host range experiments on myophages infecting SAR11 and OM43, this work provides first evidence that in genomically streamlined virus-host systems, host transitioning is not restricted to phylogenetically related hosts, but instead is potentially a function of shared physical and biochemical properties.

Chapter 4, '*Dominant polar pelagiphage Skadi provides evidence for ecotypic distribution patterns of highly abundant marine viruses*', introduces the pelagiphage Skadi isolated from the Western English Channel. In this chapter I gather evidence that Skadi and HTVC010P-type viruses are a taxonomically distinct group, and propose the novel viral family *Ubiqueviridae*, following the most recent guidelines from the International Committee on Taxonomy of Viruses (ICTV) (Adriaenssens *et al.*, 2020; Turner *et al.*, 2021). Metagenomics analysis shows that Skadi is more than three times as abundant as the second most abundant phage HTVC010P –which has been shown to be the single most abundant virus in the ocean overall- (Zhao *et al.*, 2013), but only in polar regions. At lower latitudes Skadi was below detection limit in most viromes, revealing a distinct cold-water ecotypic distribution pattern. This suggests that Skadi is a specialist pelagiphage of bi-polar SAR11 ecotypes, whereas HTVC010P most likely is a generalist phage. Thus the discovery of Skadi revealed that biogeography in pelagiphages follows the biogeography of its SAR11 host and environmental constraints imposed on virus-host systems are therefore significant factors for shaping pelagiphage populations. Furthermore, I argue that the bi-polar populations of Skadi experience frequent genetic exchange via global ocean currents. The low abundance of Skadi in warm waters, i.e. low abundance of its host, thereby provides a working example supporting the viral Seed-Bank model as well as the Royal-Family model for pelagiphages.

I conclude with **Chapter 5, 'General Discussion'**, where I reflect on the experimental results and findings presented, argue the constraints of cultured marine virology in general and discuss the limitations of my work in particular. I further discuss how my work could be used for possible future virological studies, and how my research contributed to the field.

In addition to the chapters outlined above, I co-authored a review paper '**Host-hijacking and planktonic piracy: how phages command the microbial high seas**' and published another paper as first author describing the isolation of the myophage Mosig and podophage Lederberg from Bermuda (**'Draft Genome Sequences of Pelagimyophage Mosig EXVC030M and Pelagipodophage Lederberg EXVC029P, Isolated from Devil's Hole, Bermuda'**). These are presented in their published formats in the Appendices.

Chapter 2: Efficient Dilution-to-Extinction isolation of novel virus-host model systems for fastidious heterotrophic bacteria.

This chapter records an improved high-throughput virus-host systems isolation method, based on advances in cyanophage culturing and Dilution-to-Extinction isolation of fastidious marine heterotrophic bacteria. Reported here are the first phages for the OM43 clade and the first siphovirus infecting SAR11. The chapter is reformatted from my published paper:

Holger H. Buchholz, Michelle L. Michelsen, Luis M. Bolaños, Emily Browne, Michael J. Allen, Ben Temperton. 2021. Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria. *ISME J 15*, 1585-1598. https://doi.org/10.1038/s41396-020-00872-z

Author Contributions:

- *HHB conceived, designed and performed the experiments, analysed the data, prepared figures and/or tables, authored and reviewed drafts of the paper, and approved the final draft*
- *MM assisted with SAR11 stock culture maintenance*
- *LMB analyzed 16S rRNA relative contribution, prepared* Supplementary Figure 2.9 *and* Figure 2.2
- *EB performed the experiment for H2P3α temperature response*
- *MJA authored or reviewed drafts of the paper and approved the final draft*
- *BT conceived and designed the experiments, analysed the data and prepared figures and/or tables, contributed reagents, materials and analysis tools, authored or reviewed draft of the paper and approved the final draft*

2.1 Abstract

Microbes and their associated viruses are key drivers of biogeochemical processes in marine and soil biomes. While viruses of phototrophic cyanobacteria

are well-represented in model systems, challenges of isolating marine microbial heterotrophs and their viruses have hampered experimental approaches to quantify the importance of viruses in nutrient recycling. A resurgence in cultivation efforts has improved the availability of fastidious bacteria for hypothesis testing, but this has not been matched by similar efforts to cultivate their associated bacteriophages. Here, we describe a high-throughput method for isolating important virus-host systems for fastidious heterotrophic bacteria that couples advances in culturing of hosts with sequential enrichment and isolation of associated phages. Applied to six monthly samples from the Western English Channel, we first isolated one new member of the globally dominant bacterial SAR11 clade and three new members of the methylotrophic bacterial clade OM43. We used these as bait to isolate 117 new phages including the first known siphophage infecting SAR11, and the first isolated phage for OM43. Genomic analyses of 13 novel viruses revealed representatives of three new viral genera, and infection assays showed that the viruses infecting SAR11 have ecotype-specific host-ranges. Similar to the abundant human-associated phage ϕCrAss001, infection dynamics within the majority of isolates suggested either prevalent lysogeny or chronic infection, despite a lack of associated genes; or host phenotypic bistability with lysis putatively maintained within a susceptible subpopulation. Broader representation of important virus-host systems in culture collections and genomic databases will improve both our understanding of virus-host interactions, and accuracy of computational approaches to evaluate ecological patterns from metagenomic data.

2.2 Introduction

It is estimated that viral predation kills ~15% of bacterial cells in marine surface water each day (Wommack *et al.*, 2000) and is a major contributor to nutrient recycling via the viral shunt, where marine viruses make cell-bound nutrients available to the neighbouring microbial community through viral lysis of host cells (Weitz and Wilhelm, 2012; Weitz *et al.*, 2015). Viruses are key players in the modulation of carbon fluxes across the oceans (150 Gt/yr), increasing particle aggregation and sinking to depth (Jover *et al.*, 2014; Weitz *et al.*, 2015), and

accounting for 89% of the variance in carbon export from surface water to the deep ocean (Guidi *et al.*, 2016). Viruses alter host metabolism through auxiliary metabolic genes (AMGs), increasing and altering the cellular carbon intake of infected cells (Warwick-Dugdale *et al.*, 2019). Virus-host interactions also increase co-evolutionary rates of both predator and prey via Red Queen dynamics (Weitz *et al.*, 2005; Ignacio-Espinoza *et al.*, 2020). While recent metagenomic advances have provided major insight into global viral diversity and abundance (Brum *et al.*, 2016; Roux *et al.*, 2016; Gregory *et al.*, 2019; Warwick-Dugdale *et al.*, 2019), mechanistic understanding of virus-host interactions in ecologically important taxa is reliant on experimental co-culturing of model systems. In cyanobacteria, such systems have shown that viruses increase duration of photosynthetic function (Mann *et al.*, 2003) and can inhibit CO_2 fixation, providing direct evidence that viruses of abundant phototrophs play an important role in nutrient cycling and global carbon budgets (Puxty *et al.*, 2016). Furthermore, isolation of new viruses provides complete or near-complete viral genomes, with concrete evidence of known hosts. Such systems are critical to the development, ground-truthing and application of computational methods to identify and classify viral genomes in metagenomic data (e.g. VirSorter (Roux *et al.*, 2015); VirFinder (Ren *et al.*, 2017); MARVEL (Amgarten *et al.*, 2018)); quantify boundaries for viral populations (Roux *et al.*, 2017; Warwick-Dugdale *et al.*, 2019) and genera (VConTACT2 (Bin Jang *et al.*, 2019)); understand the importance of AMGs in altering nutrient flux in natural communities (Hurwitz *et al.*, 2015) and to predict host ranges of uncultured viruses in metagenomic data (e.g. WIsH (Ahlgren *et al.*, 2017; Galiez *et al.*, 2017)).

Viruses of primary producers, such as cyanophages are both well represented with model systems and well-studied in the laboratory. In contrast, virus-host model systems for similarly important and abundant marine heterotrophic bacteria are rare. Isolated viruses infecting heterotrophs are heavily biased towards those with fast-growing, copiotrophic hosts that grow readily on solid agar, enabling the use of plaque assays for viral isolation. Such systems are not representative of the vast majority of heterotrophs in nutrient-limited soil and aquatic environments, which are dominated by slow-growing, oligotrophic taxa with few regulatory mechanisms and complex auxotrophies that limit growth on

solid media (Morris *et al.*, 2002; Sun *et al.*, 2011; Becker *et al.*, 2019; Kraemer *et al.*, 2020). Advances in Dilution-to-Extinction culturing of ecologically important hosts have enabled the cultivation of many fastidious bacterial taxa that are not amenable to growth on solid media from soil (Bartelme *et al.*, 2020), marine (Connon and Giovannoni, 2002; Henson *et al.*, 2016) and freshwater environments (Henson *et al.*, 2018). Without plaque assays to facilitate isolation and purification of viral isolates, cultivation of viruses infecting fastidious taxa in liquid media is challenging, and further exacerbated by the slow growth rates and complex nutrient requirements of their hosts. Paucity of such model systems introduces significant bias in our understanding of viral influence on global carbon biogeochemical cycles. Therefore, it is important that the efforts to isolate heterotrophic bacterial taxa for experimentation and synthetic ecology is matched by efforts to isolate their associated viruses.

Here, we adapted recent advances in Dilution-to-Extinction culturing of hosts (Henson *et al.*, 2016), and protocols to isolate viruses from liquid media (Nagasaki and Bratbak, 2010), to improve efficiency of cultivating novel virus-host systems for fastidious taxa: First, we selected the ecologically significant SAR11 and OM43 heterotrophic marine clades as models for viral isolation; Second, we used inoculum concentration with tangential flow filtration (TFF) and sequential enrichment of viruses on target hosts to improve rates of viral isolation (Shkoporov *et al.*, 2018); Third, we replaced the requirement for time-intensive epifluorescent microscopy with identification of putative viral infection by comparing infected and uninfected hosts using flow cytometry, optionally to be followed by confirmation using Transmission Electron Microscopy (TEM). These clades are abundant and important to global carbon biogeochemistry (Morris *et al.*, 2006; Sun *et al.*, 2011; Halsey *et al.*, 2012; Carini *et al.*, 2014; Ramachandran and Walsh, 2015; Reintjes *et al.*, 2019), but little is known about their associated viruses. In the case of viruses infecting SAR11, two challenges limit our ability to evaluate host-virus ecology in natural communities: (1) Assembly of abundant and microdiverse genomes from viral metagenomes presents a challenge to short-read assembly methods, resulting in underrepresentation in subsequent datasets (Roux *et al.*, 2017; Warwick-Dugdale *et al.*, 2019). This was demonstrated in the successful isolation of the first known pelagiphages,

including the globally dominant HTVC010P, which, prior to its isolation, was entirely missed in marine viromes (Zhao *et al.*, 2013); (2) poor representation of viral taxa in databases limits our capacity to accurately train machine learning approaches for *in silico* host prediction (Mizuno *et al.*, 2013; Martinez-Hernandez *et al.*, 2017; Warwick-Dugdale *et al.*, 2019). This results in either a lack of host information for abundant viral contigs, or worse, incorrect assignment of hosts to viral contigs, confounding ecological interpretation of data. Some important taxa, such as the OM43 clade, which plays an important role in oxidation of volatile carbon associated with phytoplankton blooms (Morris *et al.*, 2006; Halsey *et al.*, 2012; Taubert *et al.*, 2015), lack any know viruses in culture collections or genomic databases. In addition, both SAR11 and OM43 represent model organisms for genome streamlining as a result of nutrient-limited selection (Giovannoni *et al.*, 2014). The effect of genome minimalism on viral infection dynamics is poorly understood, but critical to evaluating the impact of predator-prey dynamics on global marine carbon budgets. In this study, novel SAR11 and OM43 representatives from the Western English Channel were isolated and used as bait to isolate associated viruses. We increased the initial concentration of viruses in natural seawater samples by tangential flow filtration, followed by inoculation of cultures and one to three rounds of sequential enrichment on target hosts, and finally purification in 96-well plates. This yielded 117 viral isolates from 218 inoculated cultures from seven monthly water samples (September 2018 to July 2019). A subsample of putative viral isolates for both clades were sequenced, providing 13 novel viral genomes, including the first known siphovirus to infect SAR11 and the first known virus for OM43.

2.3 Results and Discussion

2.3.1 *Isolation of a novel SAR11 strain and three new OM43 strains from the Western English Channel to use as bait for phage isolation*

Dilution-to-Extinction culturing for host taxa using natural seawater-based medium was performed from a water sample collected in September 2017 from the Western English Channel, and yielded the first SAR11 strain (named H2P3α) and the first three OM43 strains (named C6P1, D12P1, H5P1) from this region.

The full-length 16S rRNA gene of *Pelagibacter sp.* H2P3α was 100% identical to that of the warm water SAR11 ecotype *Pelagibacter bermudensis* HTCC7211 (subclade 1a.3) and was considered to be a local variant (Schwalbach *et al.*, 2010) (Supplementary Figure 2.1A). All three novel OM43 isolates were most closely related to *Methylophilales sp*. HTCC2181, a streamlined member of the OM43 clade with a 1.3 Mbp genome, isolated from surface water of the North Eastern Pacific (Giovannoni *et al.*, 2008) (C6P1 96.17%, D12P1 96.62%, H5P1 97.79% nucleotide identity across the full 16S rRNA gene) (Supplementary Figure 2.1B). The average nucleotide identity of the 16S rRNA gene of isolates CP61, D12P1 and H5P1 to each other was ~98.46% (Supplementary Table 2.1), suggesting they are representatives of the same genus (Yarza *et al.*, 2014).

2.3.2 *An efficient, low-cost method of isolating new viruses yielded 117 new viral isolates for SAR11 and OM43 taxa*

Using the four new hosts from above and established SAR11 isolates *Pelagibacter ubique* HTCC1062 (subclade 1a.1) and HTCC7211 (subclade 1a.3), we developed an optimised viral isolation pipeline (Figure 2.1) and applied it to six monthly water samples from the Western English Channel, taken between September 2018 to April 2019 (Supplementary Table 2.2). Each month, we concentrated a natural viral community with tangential flow filtration and used it to inoculate one to two 96-well Teflon plates containing host cultures at ~10^6 cells·mL^{-1}. Plates were monitored by flow cytometry and growth of putatively infected cultures was compared to those of unamended controls over the course of ~2 weeks to account for the slow growth rates of SAR11 and OM43 (Polyphasic taxonomy of Marine bacteria; Halsey *et al.*, 2012). An additional sample was taken in July 2019 to attempt viral isolation on OM43 strains D12P1 and H5P1. Out of a total of 218 cultures amended with concentrated viral populations, 117 viruses were isolated, purified and still infective after at least three passages, with repeatable differences observed in cytograms of infected cultures and control cultures (Supplementary Figures. 2.2 – 2.5). This represents an overall isolation efficiency of 53% and an average yield of 18 viruses per environmental sample (Figure 2.2). For 90% of inoculated SAR11 cultures (94 out of 105) we observed

evidence of viral infection, and the putative viral isolate could be propagated and purified, fulfilling Koch's postulates for confirming a pathogenic agent. For OM43, 23 out of 113 (20%) inoculated cultures yielded positive infections that could be similarly propagated. All viral isolations required between one and three rounds of virus enrichment (Supplementary Table 2.3) before changes in host growth curves between infected and uninfected cultures could be observed. This suggests putatively rare viruses can be enriched within three rounds to a level at which infection can be observed on a flow cytometer.

2.3.3 *New viruses represent novel viral populations and support established ANI cut-offs for ecologically discrete viral ecotypes*

Due to the rate-limiting step of culturing sufficient biomass for extraction of viral DNA, we sub-selected 16 viral isolates based on availability in November 2018 across four different hosts (HTCC1062, HTCC7211, H2P3α and H5P1) for Illumina sequencing to >30-fold coverage (Table 2.1). Three out of 16 sequenced samples (~19%, two from host HTCC7211, one from OM43 host H5P1) failed to assemble into single viral contigs, in line with previously reported failure rates of 18-39% for phages of *Escherichia coli* and *Salmonella spp.* (Olsen *et al.*, 2020). For 11 of the remaining 13 samples (12 from SAR11 hosts and one from OM43), each individual sequence assembly was identified as a complete viral genome by VirSorter (Roux *et al.*, 2015) and 95-100% complete using CheckV (Nayfach *et al.*, 2020). All assemblies yielded a single viral contig (categories 1 or 2, >15kbp) per sample, indicating that our purification process was effective in recovering pure viral isolates. Interestingly, all viral isolate genomes were classified by VirSorter as Category 2 ("likely virus") rather than Category 1 ("most confident"), despite being complete - indicating either a lack of viral hallmark genes or a lack of enrichment of viral genes on the contigs. This finding matches our own observations for other isolated viruses infecting SAR11 (data not shown), and suggests VirSorter classification of pelagiphages is conservative. Viral isolates were named in accordance with current ICTV guidelines for prokaryotic virus taxonomy (Adriaenssens *et al.*, 2020), with names selected from Norse mythology and folklore, and contemporary culture (Table 2.1).

Table 2.1 Summary of phages isolated and sequenced in this study

Collection Number	Virus	Host	Viral Population Representative	Hypergeometric cluster	Genome	Length (bp)	G + C %	Morphotype	Accession number	Simplified phonetic spelling	Meaning and origin of the names
EXVC025P	Eistla	HTCC1062	Eistla	Cluster A	Circular	39638	32.7	Podoviridae	MT375521	ais:tla:	"stormy one", Giantess in the poetic Edda
EXVC018P	Eyrgjafa	HTCC1062	Eyrgjafa	Cluster B	Circular	38005	32.6	Podoviridae	MT375523	e:ɪrgja:fa:	"scar donor", Giantess in the poetic Edda
EXVC020P	Gjalp	HTCC1062	Eyrgjafa	Cluster B	Circular	37857	32.5	Podoviridae	MT375524	gja:lp	"Roaring one", Giantess in the poetic Edda
EXVC021P	Greip	HTCC1062	Greip	Cluster D	Circular	34916	31.5	Podoviridae	MT375525	graip	"Grasp", Giantess in the poetic Edda
EXVC014P	Ran	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375530	ran	"plundering", Goddess of the Sea
EXVC013S	Aegir	H2P3α	Kolga	Singleton	Fragment	18297	31.1	Siphoviridae	MT375519	æ:gir	"Sea", God of the Sea
EXVC012P	Jörmungand	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375528	jɔrmunˌgand	"huge monster", Giant sea serpent
EXVC016S	Kólga	H2P3α	Kolga	Singleton	Circular	48659	30.5	Siphoviridae	MT375529	kɔlga:	"Cool-wave", Daughter of Ran and Aegir
EXVC019P	Unn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375531	ʊ:n	"Wave", Daughter of Ran and Aegir
EXVC015P	Hroenn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375527	ɦrøn	"Wave", Daughter of Ran and Aegir
EXVC010P	Bylgja	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375520	bɪ:lgja:	"Billow"-wave, Daughter of Ran and Aegir
EXVC011P	Himinglæva	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375526	hi:mɪŋglæfa	"transparent-on-top"-wave, Daughter of Ran and Aegir
EXVC028S	Venkman	H5P1	Venkman	Singleton	Linear	38624	34.4	Siphoviridae	MT375522	vɛŋkmæn	Bill Murray's Character in the Ghostbusters movie

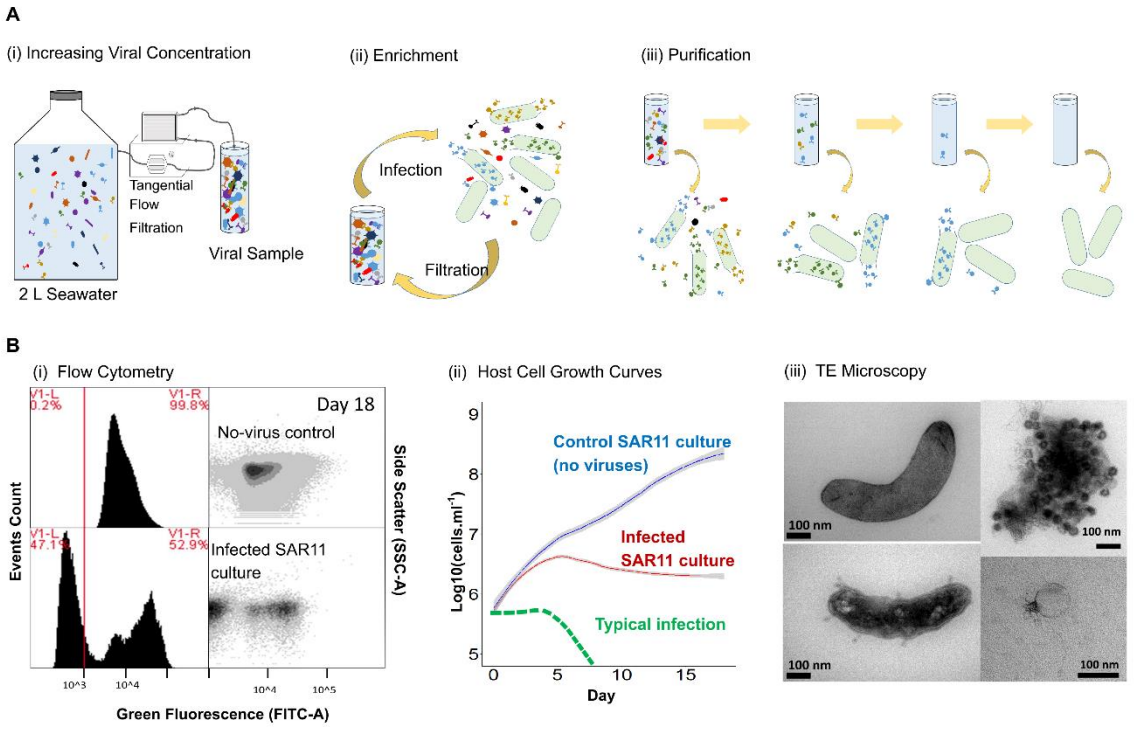


Figure 2.1 Workflow for high-throughput isolation. **(A)** i. Increasing concentration of viruses in water samples by TFF; ii. Initial infection of host cultures to enrich the sample for specific viruses; iii. Purification of viral isolates through 3 rounds of Dilution-to-Extinction; **(B)** Initial screen of viral infections using i. Flow cytometry, by comparing populations of no-virus controls and infected cultures; ii. Comparing growth curves of no-virus control culture (HTCC1062) against infected SAR11 cultures; iii. Confirming the presence of viruses in infected SAR11 cultures using TE microscopy: top left: HTCC1062 no-virus control, bottom left: infected HTCC1062, top right: aggregated cellular debris and viruses, bottom right: virus found in infected HTCC1062 culture.

Viral populations are defined as discrete ecological and evolutionary units that form non-overlapping clouds of sequence space, and previous work in cyanophage populations have shown that viral populations can be delineated into populations using an average nucleotide identity (ANI) cutoff of 95% (Deng *et al.*, 2014). Pairwise ANI was calculated between the 13 successfully sequenced viral genomes from this study and 85 other known or putative pelagiphages (Mizuno *et al.*, 2013; Våge *et al.*, 2013; Zhao *et al.*, 2018). Pairwise ANI ranged between 77.5-100%, with a discrete distribution between 96.4-100.0% (Supplementary Figure 2.6, Supplementary Table 2.4). This is in agreement with previous work in cyanophages (Deng *et al.*, 2014; Marston and Martiny, 2016) and supports the broad use of proposed boundary cutoffs to define viral populations within viral metagenomic assemblies (Roux *et al.*, 2016; Gregory *et al.*, 2019). At the

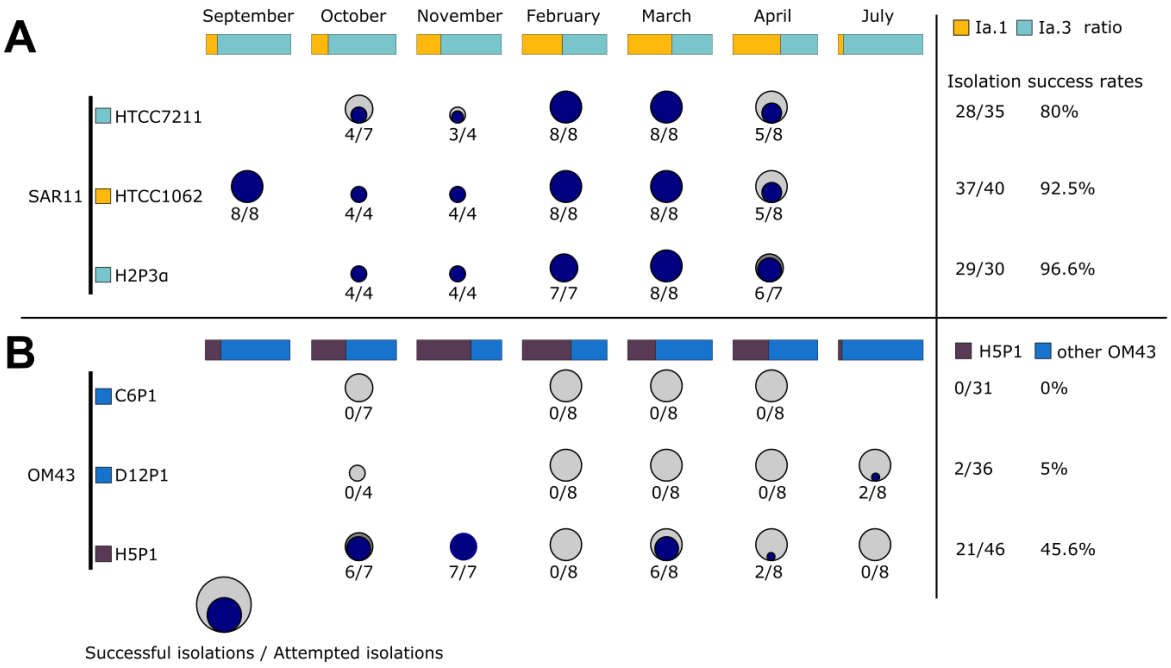


Figure 2.2 Summary of success rates for each bacterial host, as well as relative contribution to the clade community during sampling based on 16S analysis: (**A**) three SAR11 strains HTCC1062, HTCC7211 and H2P3α and (**B**) three OM43 strains C6P1, D12P1 and H5P1. This shows the water sample- target host strain combination used for infections; the numerator indicates successful isolation of a virus from that water sample (i.e. successful passage through three rounds of Dilution-to-Extinction), denominator shows the isolation attempts made for a host-water sample combination (i.e. how many cultures were treated with viral samples after multiple rounds of enrichment).

proposed ANI cutoffs of 95% over 85% length (Roux *et al.*, 2017), our 13 new viruses clustered into six viral populations, ranging from singletons to a viral population with four members (Table 2.1). Phages within the same populations were all isolated from the same environmental sample and on the same host, in agreement with their classification as discrete ecological and evolutionary units. All viruses isolated in this study formed their own exclusive viral populations, with no representatives from either known isolates (Zhao *et al.*, 2013, 2018) or fosmid-derived (Mizuno *et al.*, 2013) genomes from other studies, indicating that a high degree of viral population diversity remains to be discovered in the Western English Channel and beyond.

2.3.4 *Possible limitations and biases of the presented workflow.*

The methods described in this chapter were designed to isolate dsDNA phages for SAR11-like bacteria; other host and virus types will likely require optimisation or different approaches. Bacteriophages are often categorized based on general

arrangements of structural proteins, such as the tail and capsid proteins. Physical stress from concentration by tangential flow filtration may shear off phage tails or other fragile structural proteins, thus bias against long-tailed phages like the myophage (long contractile tail) and especially siphophage morphotypes (long non-contractile tails), thus inadvertently bias for podophages (short contractile tail). The workflow is based on, and therefore likely biased towards, available information for previously isolated pelagiphages, which are largely dsDNA podophages. Yet, for well-studied hosts, all morphotypes are typically represented in the databases, and all morphotypes are important components of the virosphere. For example, analyses of microscopy (Brum *et al.*, 2013) and viromic data (Gong *et al.*, 2018) suggest that myoviruses and siphoviruses are overall (not accounting for host types) more abundant than the podophage morphotype. Although the isolation of Kolga proves that siphoviruses can be isolated with the method described in this chapter, the overall low diversity of myophages and siphophages among viruses isolated on SAR11, previously and in this work, could indicate that known isolates do not reflect the true diversity of pelagiphage morphotypes in nature, possibly due to the biases introduced through the chosen methods. Non-tailed phages are also highly abundant in marine environments (Brum *et al.*, 2013), yet there are no known non-tailed phage for SAR11 or OM43. Considering the large SAR11 population sizes in global oceans, it seems reasonable to assume that every bacteriophage morphotype is represented among SAR11 associated viruses. But their existence remains to be demonstrated, and it is also possible that non-tailed phages infecting SAR11 do not exist. The evolutionary origin of different viral (nucleo-) capsid proteins remains unclear and may have evolved independently multiple times based on host lineages (Krupovic and Koonin 2017), yet a diverse range of viral morphotypes is known for less challenging-to-work-with host systems. For example, Kauffman *et al*. (2018) identified non-tailed dsDNA viruses with ‘double jelly roll’ capsids as a major predator of *Vibrio* species in the marine environment. Kauffman *et al.*, (2018) reported that the time until lysis was observed could be up to an order of magnitude longer for these vibriophages compared to other morphotypes. Since “typical” podophages already required about a week to cause signs of lysis in SAR11 cultures (Figure 2.1), a successful infection from a

comparable ‘double jelly roll’ non-tailed virus could have been missed due to an insufficient incubation period, resulting in a false negative isolation attempt or allow for possibly more virulent podophages to outcompete non-tailed viruses in the unnaturally dense enrichment step (Figure 2.1). Similarly, RNA phages are rarely represented in marine culture collections, and no known RNA phage for SAR11 or OM43 exist, and yet RNA phages without a lytic lifestyle similar to dsDNA phages might be common in the environment (Cai *et al.*, 2021). If lysis does not occur within the expected period of one week as mentioned above, the viral isolation attempt would have likely been dismissed following when following the protocol presented here. Other viruses with a predominantly or purely lysogenic replication strategy, as well as chronic or otherwise non-lytic infections, would therefore likely require optimisation of the detection methods and/or incubation period, as the methods used in the presented workflow rely on observing host lysis to identify successful infections.

The concentration of particles in seawater (compare Figure 2.1 A (i)) may also increase non-virus particles like particulate organic matter or high molecular weight DOM that could provide additional nutrients to cells (Sosa *et al.*, 2015). If the concentration of excess nutrients or other potentially growth inhibiting compounds crosses a certain threshold it may inhibit the growth of oligotrophs like SAR11 cells due to continuously expressed transporter proteins (Carini *et al.*, 2013, Giovannoni 2017). The resulting retarded growth curves and possible cell death could be interpreted as a false positive phage infection. Though we did not fail to obtain viral DNA from culture lysate, and therefore provide strong evidence for the presence of successful viral replication, it is important to recognise that the scalability of this method is not unlimited depending on host cell properties.

Phages that are of low abundance in the environment (I.e. the water sample used as inoculum) may also be biased against by using a Dilution-to-Extinction approach for isolation. The chance of randomly selecting a viral particle of high abundance out of the sample pool will always be higher compared to randomly selecting a viral particle of low abundance. The introduction of an enrichment step enhanced the overall isolation efficiency (Figure 2.2), but might also enhance the bias for more virulent and lytic phages as with each enrichment step the density of phage species that produce large numbers of progeny quickly, will increase

more compared to a less virulent or non-lytic virus. Thereby the fraction of the most productive phages will increase more with each enrichment step compared to the less virulent phages, and might ultimately further decrease the chance of isolation of less virulent phages through the DtE series. This could explain the repeated isolations of Skadi (Described in chapter 4). Although it is worth highlighting that the OM43 phage Melnitz (Described in chapter 3) was in fact isolated from the Sargasso Sea, where OM43 host cells are absent or below detections limits, thus assuming that the abundance of their associated phages would be equally low is reasonable. This indicates that phages of in total very low abundance can be successfully isolated with this method. But relative to all phages that are able to infect OM43 which might be present in the Sargasso Sea, Melnitz could still be more abundant than other potentially existing phages of low abundance. Thus potential isolation bias against low abundance phages relative to the viruses specific for a given host does not necessarily contradict efficient isolation of in total (relative to both host specific and non-specific) low abundance phages for the same host. Without further information about OM43 phage diversity in the Sargasso Sea it remains unclear which scenario applies here. We operated under the assumption that any isolated virus is desirable as such isolates for SAR11 and OM43 are rare or were non-existing, respectively. Our workflow indeed provides an efficient and easy to optimise basis for high-throughput viral isolation on fastidious microbes of both high and low abundances, but it is likely not universally the optimal approach for every type of virus. If e.g. only temperate viruses are sought after, optimisation will be necessary.

2.3.5 *Current pelagiphage isolates can be organised into five distinct phylogenetic clades*

To evaluate isolate diversity at higher taxonomic organisation, we picked one representative from each of our six viral populations and compared them to previous isolates and fosmid-derived phage sequences using three approaches: First, phylogenetic analyses were performed based on conserved genes in known pelagiphage isolates and closely-related taxa from viral metagenomic surveys (Gregory *et al.*, 2019; Warwick-Dugdale *et al.*, 2019); Second, raw

hypergeometric probability of shared gene content was calculated (to capture broader relationships and account for genomic mosaicism) (Lima-Mendez *et al.*, 2008); Third, genomes were organised into ICTV-recognized genera using vConTACT2 (Bin Jang *et al.*, 2019). vConTACT2 initially derives viral clusters using a hypergeometric approach, with subsequent refinement with Euclidean-distance-based hierarchical clustering to split mismatched, 'lumped' viral clusters. All three approaches were congruent - clustering on probability of shared gene content organised pelagiphage genomes into four main clusters and numerous singleton genomes (Figure 2.3). This was broadly supported by phylogenetic (Supplementary Figure 2.7) and vConTACT2 (Supplementary Figure 2.8) classification approaches. Cluster A contained 23 members (nine from fosmid-derived contigs (Mizuno *et al.*, 2013); eleven previously isolated pelagiphages (Giovannoni *et al.*, 2013; Zhao *et al.*, 2018)) and *Pelagibacter* phages Ran (EXVC014P), Bylgja (EXVC010P) and Eistla (EXVC025P) from this study. Cluster B contained two previously isolated pelagiphages, one fosmid-derived contig and *Pelagibacter* phage Eyrgjafa (EXVC018P) from this study. All viruses in Clusters A and B were assigned to a single viral genus by vConTACT2 that also contained 12 previously isolated pelagiphages (Zhao *et al.*, 2013, 2018). Cluster C only contained fosmid-derived contigs from the Mediterranean (Mizuno *et al.*, 2013), with no isolated representatives, marking it an important target for future isolation attempts. Cluster D contained eight fosmid-derived contigs, *Pelagibacter* phage HTVC010P, and *Pelagibacter* phage Greip (EXVC021P) from this study. *Pelagibacter* phage Kolga (EXVC016S) from this study and the only known Pelagimyophage HTVC008M from a previous study (Zhao *et al.*, 2013) fell outside the four main clusters. VConTACT2 split Cluster D, leaving *Pelagibacter* phages Greip and Kolga as members of two singleton clusters, suggesting they are the first cultured representatives of novel viral genera and distinct from the globally ubiquitous *Pelagibacter* phage HTVC010P.

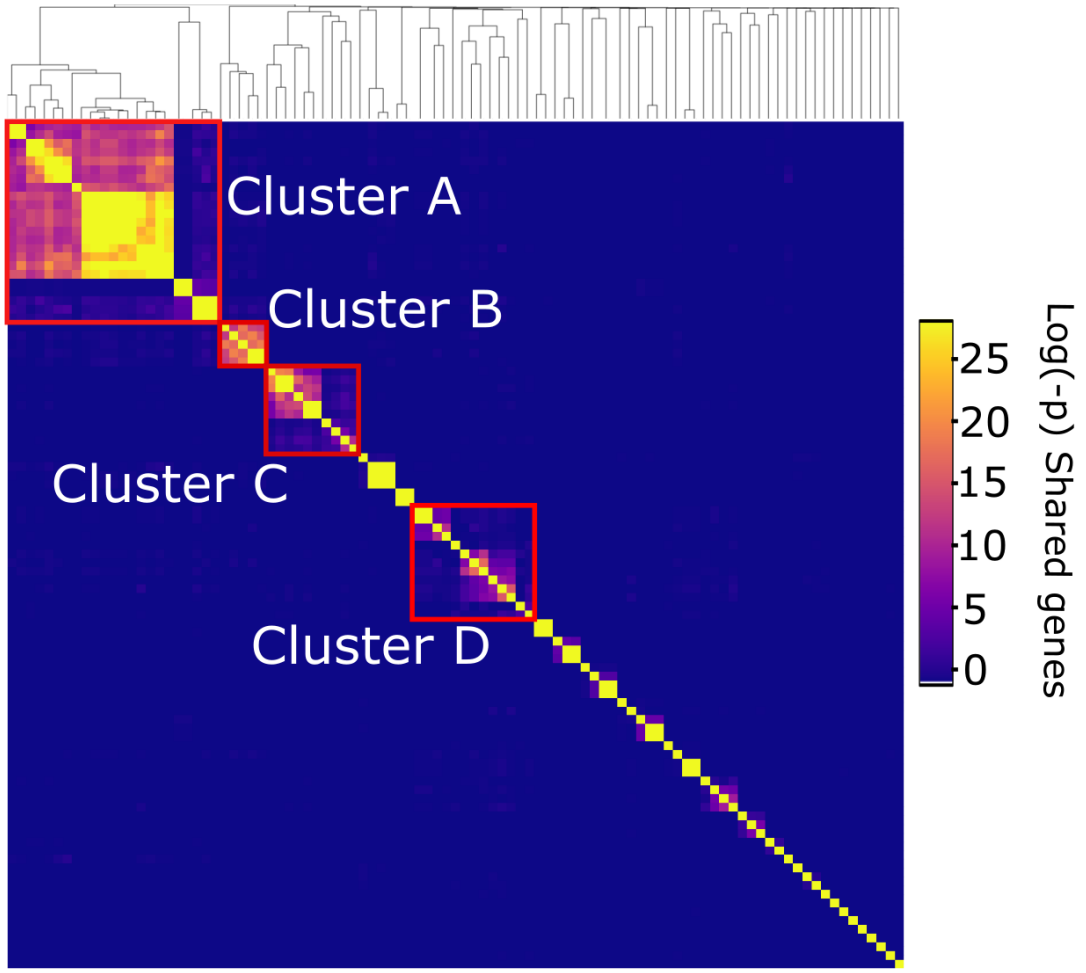


Figure 2.3 Hypergeometric testing of viral proteins. Hypergeometric testing of shared protein content in viral genomes identified four viral clusters and numerous singletons for *Pelagibacter* viruses.

2.3.6 *Novel pelagiphages are ecotype-specific and persist in the community.*

Community composition analysis using 16S rRNA genes showed that during the sampling period (September 2018 to July 2019) the SAR11 contribution relative to the total number of sequences ranged from a minimum of 28.0% in July 2019 to a maximum of 42.2% in September 2018 (Supplementary Figure 2.9). At the clade level, SAR11 composition was relatively stable over time, except for clade II roughly doubling its relative contribution from 15.1% to about 29.6% between February and April. The SAR11 community was dominated by clade I overall throughout the sampling period. Within clade I, the ratio between SAR11 subclade Ia.1 (cold-water ecotype) and Ia.3 (warm-water ecotype) showed that

the warm-water ecotype dominated from September to November as well as July (Figure 2.2A). During the coldest months in the Western English Channel (February to April) the cold-water ecotype became as abundant as the warm-water ecotype with roughly a 1:1 ratio. Overall, temperature ranged from 9.3 °C - 14.1 °C from October to April, when isolations were attempted on all strains (Supplementary Table 2.2). HTCC1062 and HTCC7211 have specific growth rates of ~0.22 and ~0.12 divisions per day at 10 °C, respectively (Polyphasic taxonomy of Marine bacteria). Our new SAR11 isolate from the Western English Channel, *Pelagibacter sp.* H2P3α, showed similar growth rates to *Pelagibacter bermudensis* HTCC7211 (Supplementary Figure 2.10), with specific growth rates of ~0.10, ~0.45 and ~0.84 divisions per day at 10, 18 and 25 °C, respectively. Therefore, measured *in situ* temperatures during our sampling period were sufficient to support slow growth of warm-water ecotypes even during winter months, potentially providing sufficient prey to support a population of warm-water ecotype specific phages. This result shows that slow growth of warm water ecotypes at *in situ* temperatures is possible, supporting the finding of the 16S community analysis that this ecotype persists throughout the year.

An alternative explanation is that isolated viruses have a broad host range encompassing both warm- and cold-water ecotypes of the SAR11 subclade Ia. We tested the host range (Table 2.2) of six pelagiphages (one of each viral population), isolated from samples in October 2018 and November 2018 (*in situ* water temperature of 14.8 °C and 14.2 °C respectively) across the three SAR11

Table 2.2 Host infectivity of viral populations isolated and sequenced in this study. H2P3α and HTCC7211 are warm-water ecotypes of *Pelagibacter spp*. subclade 1a; HTCC1062 is a cold-water ecotype of *Pelagibacter spp.* phage Bylgja is the only virus infecting both ecotypes. (Note: *Pelagibacter* phage Skadi was added from Chapter 4 for brefity)

Phage	H2P3α	HTVC1062	HTVC7211
Eistla		+	
Eyrgjafa		+	
Greip		+	
Ran	+		+
Kolga	+		+
Bylgja	+	+	+
Skadi		+	

strains (Table 2.1). *Pelagibacter* phages Eistla, Eyrgjafa and Greip all infected cold-water ecotype HTCC1062 exclusively, while *Pelagibacter* phages Ran and Kolga only infected warm-water ecotypes HTCC7211 and H2P3α. *Pelagibacter* phage Bylgia was the only virus that could infect both warm and cold-water ecotypes. Therefore, our new pelagiphages appear to be broadly ecotype-specific, confirming previous findings (Zhao *et al.*, 2018). Our results suggest overall that pelagiphages persist in the water column throughout the year in sufficient densities to be isolated by our enrichment method, despite ecotypic specificity and fluctuations in warm- and cold-water ecotype community contributions of SAR11 subclade 1a. If concentration and enrichment of viruses during isolation is sufficient to successfully isolate even low abundance phages then a comprehensive library of representative phage isolates could be generated with relatively modest sampling effort across a few locations.

2.3.7 Pelagibacter *phages Kolga and Aegir- the first siphoviruses infecting SAR11*

The 25 previously known viral isolates infecting SAR11 comprise 24 podoviruses and one myovirus (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2019). Previous cultivation efforts for viruses of SAR11 have not isolated any siphoviruses, nor are any known from viral metagenomic studies. We isolated and sequenced the two *Pelagibacter* phages Kolga and Aegir using host H2P3α as bait (Figure 2.4A-C). Transmission electron microscopy (TEM), which showed evidence of a long tail (Figure 2.4B), suggested classification as the first reported siphoviruses infecting members of the SAR11 clade. *Pelagibacter* phages Aegir and Kolga were classified as members of the same population using a boundary cutoff of 95% ANI over 85% contig length, however, *Aegir* had a length of 18,297 bp compared to 48,659 bp in Kolga, therefore we considered *Aegir* to be a partial genome of the same viral population. Kolga did not share a significant number of genes with known SAR11 podoviruses (Figure 2.3), and did not cluster with other pelagiphages using hypergeometric analysis based on shared gene content. VConTACT2 also grouped Aegir and Kolga into one cluster without any other

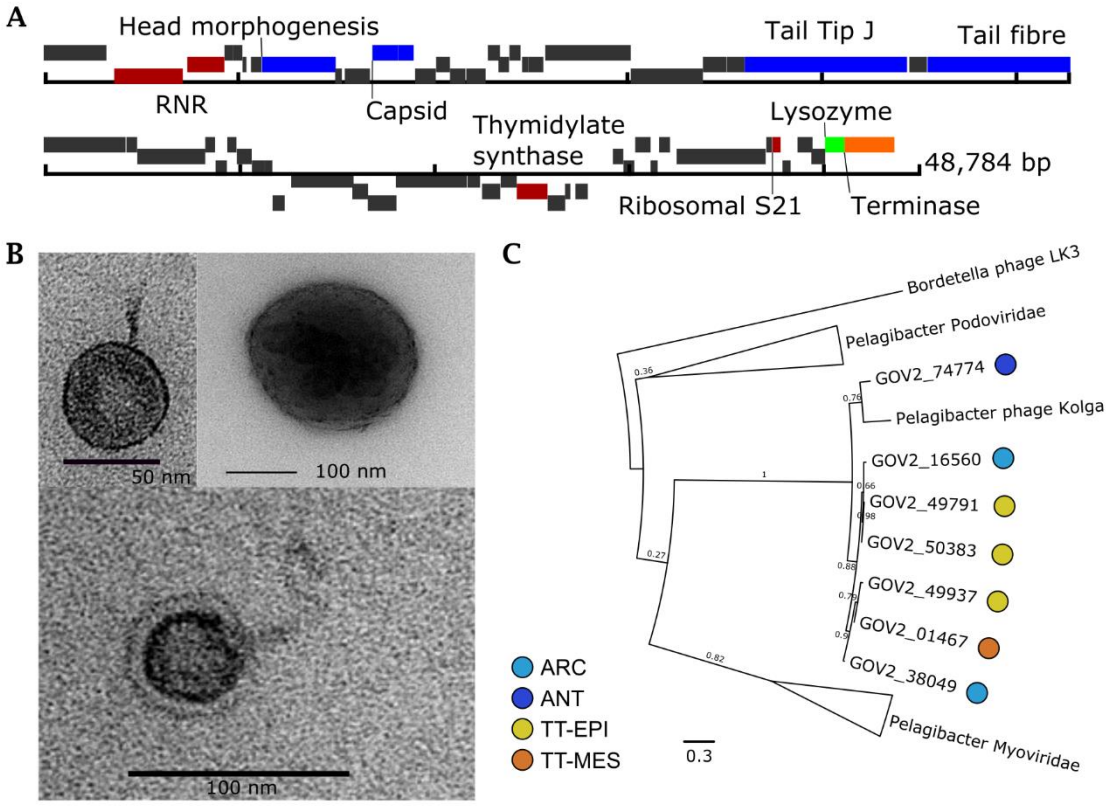


Figure 2.4 The first reported siphovirus (Kolga) infecting novel SAR11 H2P3α; **(A)** gene map of the 48,784 bp genome, which contains 80% hypothetical genes without known function; **(B)** TEMs of *Kolga* (left and right) and a H2P3α host cell infected with *Kolga* (middle). **(C)** Unrooted maximum likelihood tree of TerL genes found in Pelagiphages. The branches containing members of the *Podoviridae* and *Myoviridae* family infecting SAR11 were collapsed for clarity, the full tree is available in Supplementary Fig. 7. Sampling environment of metagenomes used to identify *Kolga*-like contigs are marked: Epipelagic (TT-EPI), mesopelagic (TT-MES), bathypelagic (BATHY), Southern Ocean/ Antarctica (ANT), Arctic Ocean (ARC).

known viruses, suggesting it represents a novel viral genus. A number of genes found in Kolga were shared with other known siphoviruses (associated with different hosts) such as the Bordetella phage LK3. Screening of contigs from the Global Ocean Virome dataset (Gregory *et al.*, 2019) identified six contigs from various ecological zones which shared a viral cluster with Kolga, but which belonged to different viral populations based on network analysis using vConTACT2. Phylogenetic analysis of *TerL* genes in *Pelagibacter* phages indicates that the closest known relatives to Kolga and Kolga-like contigs are members of the *Myoviridae* family, supporting a classification as a distinct and novel viral group (Figure 2.4C, Supplementary Figure 2.7). In Kolga 67% of encoded genes could not be functionally annotated, and out of all hypothetical

genes identified on Kolga, only three hypothetical genes were shared with SAR11 podoviruses. In contrast, on average ~90% of genes without known function identified within our novel SAR11 podoviruses are shared between different SAR11 podoviruses. Kolga possesses a tail tip J protein (Figure 2.4A), often found in phages with long non-contractile tails such as E. coli phage ƛ, where it plays a role in DNA injection during cell entry and tail assembly (Tam *et al.*, 2013). Kolga also encodes a small S21 subunit of the 30S ribosomal gene structurally similar to the ones found in *Pelagibacter* phage HTVC008M, Pelagimyophage-like contigs (Zaragoza-Solas *et al.*, 2020), and hosts HTCC7211 and H2P3α. Encoding ribosomal genes is a feature found in numerous myoviruses and siphoviruses (Mizuno *et al.*, 2019). The S21 gene is involved in translation initiation and needed for the binding of mRNA (Van Duin and Wijnands, 1981). Virally-encoded S21 genes may provide a competitive advantage for the phage as it could replace cellular S21 and assist in the translation of viral transcripts. Kolga may need its S21 gene for shifting the translational frame, as it has been shown that for some members of the Caudovirales, the production of tail components is dependent on programmed translational frameshifting (Xu *et al.*, 2004). Given the constitutive nature of gene expression in genomically streamlined bacteria (Giovannoni *et al.*, 2014), genes such as S21 may also provide the virus with a mechanism to manipulate host metabolism in the absence of typical promoters and repressors.

2.3.8 *First methylophages for marine OM43 isolated*

Isolation of novel phages for the OM43 clade yielded 23 positive infections, with efficiencies ranging from 0% (no viruses isolated on host C6P1) to 45% on H5P1 (Figure 2.2). To the best of our knowledge these are the first reported viruses infecting members of the OM43 clade. One explanation for the lower efficiency of isolation of phages infecting OM43 is simply one of lower host abundance concomitant with lower phage abundance in the viral community, reducing the likelihood of infective viruses coming into contact with susceptible and permissive cells. OM43 are closely associated with metabolism of extracellular substrates

from phytoplankton blooms (Neufeld *et al.*, 2008), but have low abundance outside of phytoplankton spring blooms (Morris *et al.*, 2006). Our water samples were not associated with high *in-situ* fluorescence (used as proxy measurement for phytoplankton), and missed the April 2019 spring bloom by about two weeks (Supplementary Figure 2.11). Based on 16S community analysis, the OM43 contribution to the bacterial community (0.7%) was lowest during sampling in April 2019 and highest in October 2018 (1.9%). Relative to all other OM43, H5P1 was the most abundant OM43 in the Western English Channel, contributing more than half of all OM43 in November (Figure 2.2B). This could explain the higher success rate of isolating phages on host H5P1 compared to C6P1 and D12P1.

Sequencing of the first OM43 phages isolate, *Methylophilales* phage Venkman (EXVC282S), returned a single genome 38,624 bp long (31.9% GC content), which was linear but complete (Supplementary Table 2.5). Venkman encodes genes (Figure 2.5A) with similar synteny and function to the siphovirus P19250A (38,562 bp) that infects freshwater *Methylophilales* LD28, which are often considered a freshwater variant of OM43 (Salcher *et al.*, 2015; Moon *et al.*, 2017). Unlike the siphovirus P19250A, TEM images indicated that OM43 phage Venkman had a short tail (Figure 2.5B) similar to podoviruses, though it is possible that tail structures were lost during grid preparation. Venkman shared a viral population with a 23kbp contig from the Western English Channel virome assembled from short-read data (WEC_HYBRID_01170), suggesting this viral type does not suffer from the issues of high abundance and microdiversity that challenge assembly of some pelagiphages. Phylogenetic analysis of concatenated *TerL* and exonuclease genes indicated that Venkman is most closely related to P19250A and other siphoviruses infecting different Proteobacteria (Figure 2.5C). However, branch support values were low, despite numerous attempts to refine the tree with different approaches (see S2.1 Supporting Methods).

2.3.9 *Global abundance of novel isolates highlights niche-specificity and low representation in existing datasets*

To evaluate relative global abundance of our new phage isolates, existing virome datasets from the Global Ocean Virome survey (Gregory *et al.*, 2019) and the

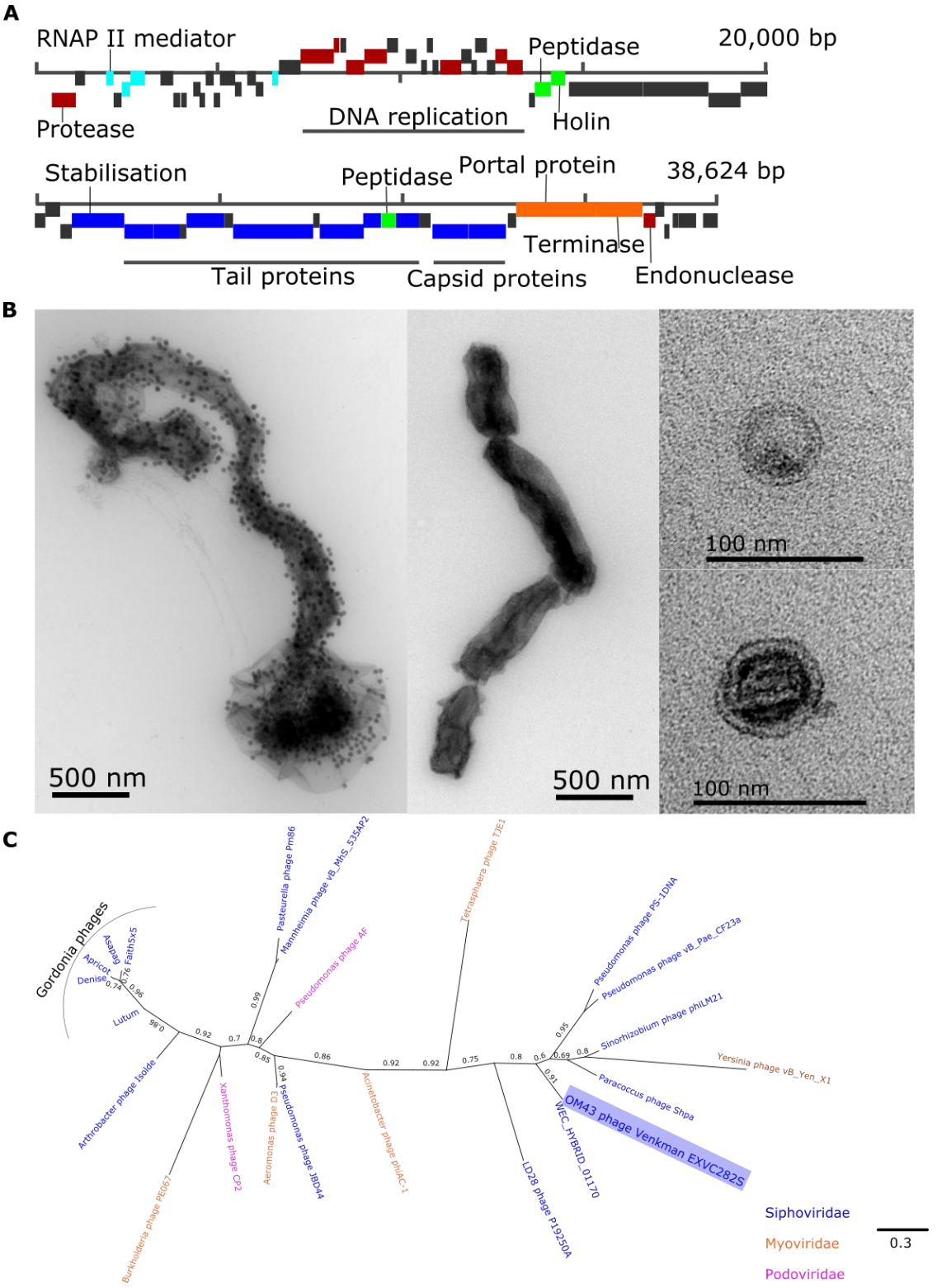


Figure 2.5 Evidence for the first reported virus Venkman infecting a member of the OM43 clade (H5P1); **(A)** Gene map displaying protein coding genes. Structural genes are shown in blue, DNA replication genes in red, lysis related genes in green, transcription genes in turquoise, packaging genes in orange, hypothetical genes are grey; **(B)** TEM images of: infected and chaining H5P1 cells (left), uninfected H5P1 chaining cells (second left), Venkman viral particles (top and bottom right); **(C)** Maximum likelihood tree (500 bootstraps) of concatenated viral *TerL* and exonuclease genes, host

families of the phages are indicated on the figure. We identified a number of common phage proteins such as a capsid protein, terminase, nucleases and tail structural proteins, and the remaining 54% of genes were hypothetical. VConTACT2 assigned OM43 phage Venkman and LD28 phage P19250A to the same genus-level cluster, therefore the OM43 phage Venkman may be a marine variant of the freshwater LD28 siphovirus P19250A.

Western English Channel (Warwick-Dugdale *et al.*, 2019) were randomly subsampled to 5 million reads and competitively recruited against genomes of viral population representatives from this study as well as previously isolated pelagiphages (see S2.1 Supporting Methods). Overall, with the exception of *Pelagibacter* phage HTVC010P, pelagiphages were poorly represented in samples from temperate-tropical mesopelagic (TT-MES) and Antarctic (ANT) ecological zones, as defined in (Gregory *et al.*, 2019) (Figure 2.6). Phages isolated in this study were neither globally ubiquitous, nor abundant in the single Western English Channel virome (with the exception of phage Ran). Phages Bylgja, Himinglaeva, Eistla and Eyrgjafa did not achieve the minimum cutoff of 40% genome coverage to be classified as present (Tithi *et al.*, 2018) in any of the viromes tested. Ran was the only phage isolated in this study with representation in at least two samples from temperate-tropical epipelagic (TT-EPI) zones, concordant with its host specificity of warm-water SAR11 ecotypes (*Pelagibacter bermudensis* HTCC7211 and H2P3α from this study (Table 2.2). Similarly, the Western English Channel virome was taken in September 2016, when waters are usually highly stratified after summer heating (Sargeant *et al.*, 2016) and warm water ecotypes of SAR11 dominate the microbial community (Figure 2.2A). Other viral populations from this study also isolated on warm-water SAR11 ecotypes (phages Kolga, Bylgja and Himinglaeva) were either absent or below limits of detection in this sample.

This coupled with the low global abundance of these viruses suggests that pelagiphage communities comprise few highly abundant taxa and a long rare tail. *Pelagibacter* phage Greip was detected in seven samples, six of which were Arctic (ARC) samples from a discrete region of the Arctic characterized by low nutrient ratios (Gregory *et al.*, 2019). In three of those samples (191_SRF, 193_SRF, 196_SRF), HTVC010P was not detected and in the other three

Figure 2.6 Global abundance of phage isolates. Metagenomic recruitment (reads recruited per kilobase of contig per million reads (RPKM)) of isolate and metagenomically-derived viral genomes against a virome from the Western English Channel and the Global Ocean Virome dataset (Warwick-Dugdale *et al.*, 2019, Gregory *et al., 2019*).

(206_SRF, 208_SRF and 209_SRF) Greip was 1.5 to 6.6-fold more abundant than HTVC010P, identifying Greip as an abundant arctic pelagiphage ecotype. The fact that Greip was isolated on the cold-water ecotype of SAR11 (*Pelagibacter ubique* HTCC1062) and does not infect either warm-water ecotype HTCC7211 or H2P3α supports the hypothesis that host niche specificity shapes the phylogeography of associated viral taxa. In the Western English Channel viral metagenomes, OM43 phage Venkman was the third most abundant (2,699 RPKM) after *Pelagibacter* phages HTVC027P (5,023 RPKM) and HTVC010P (5,026 RPKM) identifying it as an ecologically important virus in this coastal microbial community (Figure 2.6). In contrast, neither Venkman nor LD28 phage P19250A were identified in the majority of samples from GOV2, except in one sample (67_SRF) in the South Atlantic Ocean (257 RPKM against Venkman). Sample 67_SRF was also classified as a coastal biome by the authors (Gregory *et al*. 2019). These results coincide with previous reports that OM43, and presumably their associated viruses as a consequence of host abundance, are important in some coastal regions, but largely absent in open-ocean systems (Morris *et al.*, 2006; Taubert *et al.*, 2015).

2.3.10 *Unusual host-virus dynamics are prevalent in isolated phages*

Pelagibacter phages Bylgja, Eyrgjafa, Ran and Eistla all encode endonucleases and exonucleases (Supplementary Figure 2.12) and cluster by shared protein content with other pelagiphages (Cluster A and B), such as e.g. HTVC011P and HTVC025P, shown previously to integrate into host genomes (Figure 2.3) (Zhao *et al.*, 2018). We therefore predict that all our phages within Clusters A and B are temperate phages. Eyrgjafa also encodes a tRNA-Leu, that has 85% nucleotide identity over the first 34 bases of the tRNA-Leu of its host HTCC1062, suggesting a putative integration site into the host genome (Bailly-Bechet *et al.*, 2007). To date, 16 of the 29 viruses previously isolated on SAR11 strains have either been shown to be capable of lysogeny, and/or encode genes associated with a temperate infection cycle. In contrast, viruses such as the *Pelagibacter* phages Greip*,* and the abundant HTVC010P in Cluster D do not possess any genes associated with lysogeny, and would therefore be classified as exclusively lytic.

Viruses in Clusters A and B (putatively temperate phage), were of much lower abundance in the environment compared HTVC010P which was among the most abundant pelagiphages in the Western English Channel (Figure 2.6), suggesting a possible ecological difference between two groups of viruses with different infectivity strategies.

Interestingly, growth curves of hosts infected with *Pelagibacter* phage Greip and other isolated phages deviated from the expected decay in cell abundance associated with viral lysis and previously observed in isolated pelagiphages (Zhao *et al.*, 2013) (Fig.1 Bii). The first pelagiphages HTVC010P and HTVC008M were isolated from the warm waters of the Sargasso Sea, and HTVC011P and HTVC019P were isolated from the colder waters of the Oregon Coast. All four strains were propagated on the cold-water SAR11 ecotype *P. ubique* HTCC1062. In all cases, host density was reduced from ~8 $\times$ 10^6 cells·mL^{-1} at T_0 to <10^6 cells·mL^{-1} over a 60-72h period. Viruses isolated from warm waters took 17% longer than those from cold water to do so (Zhao *et al.*, 2013), suggesting that suboptimal hosts reduced the rate of infection as shown in cyanophages (Enav *et al.*, 2018). In contrast, infection dynamics of our isolates often resulted in host density of infected cultures growing to a steady-state, but at a lower cell density than uninfected cells. (Appendix 6, Supplementary Video 2.1), irrespective of cluster, population assignment or evidence of genes associated with temperate lifestyles. Out of 117 viruses isolated in this study, only 16 infections reduced host abundance below their inoculum density of 10^6 cells·mL^{-1}. In 53 infections, densities of infected cells increased to within an order of magnitude of uninfected cells (Supplementary Figure 2.13), but demonstrated clear evidence of viral infection in cytograms, TEMs and subsequent recovery of viral genomes in selected samples.

Similar patterns of infection were recently reported in the extremely abundant bacteriophage ϕCrAss001 found in the human gut, where infection of exponentially growing cells of *B. intestinalis* 919/174 did not result in complete culture lysis, but caused a delay in stationary phase onset time and final density, despite lacking genes associated with lysogeny. As with our study, the authors observed that this only occurred in liquid culture, and isolation of the virus required numerous rounds of enrichment. They postulated that the virus may

cause a successful infection in only a subset of host cells, with the remainder exhibiting alternative interactions such as pseudolysogeny or dormancy (Shkoporov *et al.*, 2018). The prevalence of similar infection dynamics in the phages isolated in this study offer two intriguing possibilities: (i) Many of the viruses isolated in this study are either not fully lytic, but fall somewhere on the continuum of persistence (Weitz *et al.*, 2019). This could be controlled by genes currently lacking a known function, with lysogeny (and associated superinfection immunity) favoured at high cell density, which would support the Piggyback-the-Winner hypothesis (Knowles *et al.*, 2016; Silveira and Rohwer, 2016); (ii) the steady-state of host and virus densities observed here are an indicator of host phenotypic bistability in these streamlined heterotrophic taxa. Viral propagation occurring in only a subset of cells could explain the requirement of multiple rounds of enrichment before sufficient viral load is reached to be able to observe lytic infection on the host population. Either strategy, or a combination of both, would provide an ecological advantage of long-term stable coexistence between viruses and hosts and offer an explanation of the paradox of stable high abundances of both predator and prey across global oceans (Kang *et al.*, 2013; Zhao *et al.*, 2013; Martinez-Hernandez *et al.*, 2018; Warwick-Dugdale *et al.*, 2019). Infection in a subset of the population could also explain the low lytic activity observed in pelagiphages *in situ*, despite high host densities (Alonso-Sáez *et al.*, 2018); and the small dynamic range and decoupled abundances of SAR11 and virioplankton in the Sargasso Sea (Parsons *et al.*, 2012). Limited lysis of subpopulations of hosts such as SAR11 and OM43 that specialize in the uptake of labile carbon enriched through viral predation (Middelboe and Lyck, 2002; Middelboe and Jorgensen, 2006) could facilitate efficient intra-population carbon recycling and explain the limited influence of SAR11 and associated viral abundances on carbon export to the deep ocean (Guidi *et al.*, 2016). We propose the moniker ‘Soylent Green Hypothesis’ for this mechanism, after the 1973 cult film in which the dead are recycled into food for the living. Further investigation leveraging our new virus-host model will provide greater insight into viral influence on ocean carbon biogeochemistry.

2.4 Conclusion

In conclusion, our method coupled Dilution-to-Extinction cultivation of hosts and associated viruses, resulting in the isolation of three new strains of OM43; a Western English Channel variant of a warm-water ecotype of SAR11; the first known methylophages for OM43; the first siphovirus infecting SAR11, as well as eleven other viruses infecting this important marine heterotrophic clade and >100 more isolates to be sequenced and explored. The described method represents an efficient and cost-effective strategy to isolate novel virus-host systems for experimental evaluation of co-evolutionary dynamics of important fastidious taxa from marine and other biomes. Coupling these methods to existing advances in host cultivation requires minimal additional effort and will provide valuable representative genomes to improve success rates of assigning putative hosts to metagenomically-derived viral contigs. Broader representation of model systems beyond cyanophages and viruses of copiotrophic, *r*-strategist hosts will reduce bias in developing methods to delineate viral population boundaries (Brum and Sullivan, 2015; Gregory *et al.*, 2016), increasing the accuracy with which we derive ecological meaning from viral metagenomic data. We therefore hope that this method will enable viruses to be included in the current resurgence of cultivation efforts to better understand the biology and ecology of phages, and the influence of the world's smallest predators on global biogeochemistry.

2.5 Methods Summary

A complete description of the materials and methods is provided in the S2.1 Supporting Methods. Four bacterial strains (*Methylophilales sp*. C6P1, D12P1 and H5P1; *Pelagibacter sp*. H2P3α) were isolated from Western English Channel station L4 seawater samples using Dilution-to-Extinction methods (Henson *et al.*, 2016). All four bacteria and two additional SAR11 strains *P. bermudensis* HTCC7211 and *P. ubique* HTCC1062 were used as bait to isolate phages from six monthly Western English Channel L4 seawater samples (50°15.00N; 4°13.00W). Briefly, water samples were concentrated for viruses using tangential flow filtration (100k Da Hydrosart membrane) and used as viral inoculum (10% v/v) in exponentially growing cultures of host bacteria in artificial seawater

medium (Carini *et al.*, 2013) in 96-well Teflon plates (Radleys, UK). Cells of the resulting lysate were filtered out (0.1 µm PVDF syringe filters) and the filtrate was used as viral inoculum in another round of isolation. This process was repeated until viral infection could be detected by flow cytometry - comparing cytograms and maximum density of infected cultures against uninfected cultures. Phages were purified by dilution-to-extinction methods (detailed protocol available here: dx.doi.org/10.17504/protocols.io.c36yrd). Phage genomes were sequenced using Illumina 2x150 PE sequencing, assembled and manually annotated as described in (Salisbury and Tsourkas, 2019). Phylogenetic classification of phages was performed on concatenated shared genes using a combination of Bayesian inference trees, maximum likelihood trees and shared-gene likelihood analyses, depending on the availability of appropriate taxa. ICTV-recognised genera based on shared gene content were assigned with VConTACT2 (Bin Jang *et al.*, 2019). The relative abundance of novel phages in the Western English Channel (Warwick-Dugdale *et al.*, 2019) and Global Ocean Viromes (Gregory *et al.*, 2019) was estimated in RPKM by competitive read recruitment of five million randomly subsampled reads against pelagiphage and methylophage genomes from this study and others (Zhao *et al.*, 2013, 2018; Moon *et al.*, 2017; Zhang *et al.*, 2020).

2.6 Data availability and Acknowledgements

All reads can be found in the SRA database under BioProject number PRJNA625644 as BioSamples SAMN14604128 - SAMN14604140. Annotated phage genomes are deposited as GenBank submissions under accession numbers MT375519 - MT375531. Sequences used for phylogenetic analysis are deposited under https://github.com/ViralPirates/viral-dte.

2.6.1 *Acknowledgements*

We would like to thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing the TE microscopy and imaging. We would also like to thank the crew of the *R/V Plymouth Quest* and our collaborators at PML for collecting water samples, and the driver Magic for delivering water

samples from Plymouth to Exeter. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk) which is supported by the BBSRC (grant number BB/L024209/1). This project used equipment funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). Bioinformatic analyses were conducted using the high-performance computing, ISCA, provided by the University of Exeter. The Western Channel Observatory provided environmental data from station L4 and is funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science, grant number NE/R015953/1

2.6.2 *Funding information*

The efforts of Holger Buchholz in this work were funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training program. Michelle Michelsen and Ben Temperton were funded by NERC (NE/R010935/1) and by the Simons Foundation BIOS-SCOPE program.

Chapter 2: Supplementary Information

S2.1 Supporting Methods

S2.1.1 *Water sampling*

A total of 20 L of seawater was collected in rosette-mounted Niskin bottles at a depth of 5m from the Western Channel Observatory (WCO; http://www.westernchannelobservatory. org.uk/) coastal station 'L4' (50°15.00N; 4°13.00W) on the following dates: 2018-09-24, 2018-10-17, 2018-11-05, 2019-02-11, 2019-03-11, 2019-04-01, 2019-07-22 (for details see Supplementary Table 2). Seawater was transferred immediately to a clean 2 L acid-washed polycarbonate (PC) Nalgene bottles (Thermo Fisher Scientific, Waltham, USA) and placed in a cooler box (Igloo, Katy, USA) at ambient temperature. Upon return to shore, water was transported to the University of Exeter for immediate processing (Two hours maximum duration from collection to processing).

S2.1.2 *Isolation of SAR11 strain H2P3α and OM43 strains*

1 L seawater was collected as described above in September 2017, filtered through a 142 mm Merck Millipore pore-size 0.2 µm PC filter and autoclaved in 2 L Duran bottles (DWK Life Sciences GmbH, Mainz, Germany) with the lid tightly shut to maintain carbon chemistry (Henson *et al.*, 2016). Upon cooling, nutrients were added to make natural seawater medium (NSW): 1 mM NH_4Cl, 10 µM K_2HPO_4, 1 µM $FeCl_3$, 25 µM glycine, 25 µM methionine, 100 µM pyruvate (Carini *et al.*, 2013). 1 mL of medium was placed into each 2 mL well in three sterile 96-well acid washed Teflon plates (Radleys, UK). Surface water from station L4 was filtered through a 142 mm Merck Millipore 0.2 µm PC filter to remove larger plankton and bacteria and enrich for smaller bacteria. A 200 µL aliquot was stained with SybrGreen and then quantified on a C6 Accuri flow cytometer. The remaining retained fraction was diluted in NSW to a density of ~ 1 cell·μL^{-1}. 1 µL of diluted inoculum was added to each well, with eight wells left blank as negative controls and eight wells with undiluted low-nucleic acid community fraction as positive controls. Plates were covered and incubated in the dark at 15 °C for three

months and checked monthly for positive growth by flow cytometry as described above. Positive wells (>10^6 cells·mL^{-1}) were transferred into 50 mL cultures in NSW medium and passaged three times once they achieved >10^6 cells·mL^{-1}). To verify culture purity and to taxonomically classify new isolates, the 16S rRNA gene was amplified from 10 µL of exponentially growing culture, heated to 95°C for 10 minutes, followed by PCR (25 cycles, 30 s @ 98 °C; 45 s @ 59 °C; 45 s @ 72 °C; final extension of 90 s @ 72 °C) using commercially available 27F and 1492R primers by Eurofins Genomics. Amplicon DNA was purified for Sanger sequencing (Eurofins) using a QIAquick PCR purification kit (Qiagen) following the manufacturer's standard protocol. 16S rRNA sequences were compared to known bacterial sequences in the SILVA SSU database with SINA (Pruesse *et al.*, 2012). Basic phylogenetic analysis was performed on these 16S genes. Genes were aligned using T-coffee (Notredame *et al.*, 2000), curated using Gblocks (Talavera and Castresana, 2007) and maximum likelihood trees for both alignments were created using PhyML with standard settings and 500 bootstraps (Guindon *et al.*, 2005). The trees were visualised using FigTree (v1.4.4, available https://github.com/rambaut/figtree).

S2.1.3 *Host cultivation*

SAR11 strains HTCC1062 and HTCC7211 were kindly provided by the Giovannoni lab (Oregon State University, USA) and were kept in continuous culture in 50 mL polycarbonate flasks at 15 °C on artificial seawater medium ASM1 (Carini *et al.*, 2014) throughout the duration of the study. Following isolation, cultures of OM43 and H2P3α were maintained on ASM1; medium for OM43 was further amended with 100 mM MeOH and 32 µL of commercially available MEM amino acid solution per 1.6 L of medium (Sigma, M5550), to improve cellular yields and reduce occurrences of chaining cells (observed through EM, Fig. 5, and a phenotype of alanine starvation in SAR11 (Carini *et al.*, 2013)), respectively.

S2.1.4 *Viral concentration and inoculation*

Monthly samples of 2 L of surface water were collected from station L4 as described above and sequentially filtered through a 142 mm Whatman GF/D filter (2.7-µm pore size) and a 142 mm 0.2-µm pore polycarbonate filter (Merck Millipore) using a peristaltic pump, to remove larger sized bacterioplankton. The filtrate was concentrated to 50 mL (40-fold concentration) using a 50R VivaFlow tangential flow filtration unit (Sartorius Lab instruments, Goettingen, Germany). The concentrate was filtered through a 0.1-µm pore PVDF membrane syringe-filter to remove any residual small cells and used as inoculum for viral isolation. 1 mL of exponentially growing bacterial host in ASM1 medium (amended with nutrients for OM43 strains as described for host cultivation above) was placed into each 2 mL well of a 96-well, acid washed Teflon plate. Eight wells were used as blank medium controls (no cell controls). 100 µL of viral inoculum was added to each well in the 96-well plate, with eight wells amended with 10% (v/v) medium instead of viral inoculum to serve as no-virus controls. The plates were incubated for ~2 weeks until no-virus controls reached maximum cell density (~2 weeks for strains HTCC1062, HTCCC7211 and H2P3α; about 1 week for OM43 strains H5P1, C6P1, D12P1). Growth was monitored at the end of the incubation period by flow cytometry; incubation periods were determined by the average growth times and data of the strains used for infections. Cytograms of virus-amended wells and no-virus controls were compared to identify infections as described below. A detailed protocol of the viral isolation process is available on www.protocols.io (DOI: dx.doi.org/10.17504/protocols.io.c36yrd). Once the no-virus controls reached maximum cell density, inoculated wells that did not show signs of infection were pooled (according to host type), filtered through a 0.1-µm pore PVDF membrane syringe (Durapore) filter to remove the cells and then used as fresh inoculum in a new 96-well plate of target host. This was repeated up to three times to amplify host-specific viral particles to a density whereby signs of viral infectivity could be observed on the flow cytometer.

S2.1.5 *Identification of positive infection via flow cytometry*

We observed positive infections in cells via flow cytometry by comparing cytograms of wells inoculated with viruses against no-virus controls. Cytograms of wells containing infected cultures were identifiable by (1) a population shift of up to a 10-fold increase in green fluorescence compared to uninfected cultures; (2) an increase in the 'noise' fraction (presumably from cellular debris following lysis as well as viral particles) over time (Appendix 6: Supplementary Video 2.1); and (3) a reduced maximum cell density (ranging from 10 to 1000-fold depending on initial cell density and viral inoculum concentration) in infected cultures compared to no-virus controls (Figure 2.1B). The presence of viral-like particles in wells where a shift in green fluorescence, increased noise, and reduced maximum cell density were observed, were verified by TEM and reinfection of new cultures in fresh medium.

S2.1.6 *Viral purification*

Following identification of a positive infection within a well, the contents of the well were filtered through a 0.1 µm pore-size PVDF syringe filter (Durapore) and used as inoculum in three rounds of viral purification. Briefly, a 96-well Teflon plate was inoculated with bacterial hosts as described previously. A 10-fold serial dilution series (from 10^{0}-10^{-10}) of the viral inoculum in ASM1 was added to a single row of the plate at 10 % v/v, with one well per row as no-virus controls (total of eight no-virus controls/plate). Plates were incubated at 18 °C for ~2 weeks until the no-virus controls reached maximum cell density and then screened for signs of viral infection using flow cytometry as described above. For each dilution series, the well amended with the lowest number of viruses that showed positive signs of infection was identified, and used as the inoculum in another round of viral purification (Nagasaki and Bratbak, 2010). Using this format, we were able to purify eight to twelve viral isolates simultaneously per plate.

S2.1.7 *Viral isolation costs and handling time*

For each sample (processed environmental water sample), all steps of the initial viral isolation process (counting one to two plates for a total of three times)

required ~6 hours run time on a flow cytometer. For the subsequent three rounds of purification, one plate was required for each host-sample combination (another ~6 hours of cytometer run time in total for one host-sample combination). Initial isolation of viruses from environmental water samples and three rounds of purification took ~10 weeks of incubation time in total. Between all steps, our protocol required ~7 hours of handling time per sample. Following three rounds of viral purification, generating sufficiently high viral titres to extract enough DNA for sequencing (approximately two weeks incubation time, and roughly four hours handling time over two days) was the rate-limiting step, and so required subselection of available viruses for sequencing. Future advances in DNA extraction efficiency, reducing DNA input requirements for sequencing and/or automation of viral DNA extraction will enable all isolated viruses to be sequenced. We estimate the cost of isolating a single virus is ~£20 for cultivation, e.g. flow cytometry and DNA extraction consumables, and ~£50-100 for genome sequencing at 30-fold coverage required for successful assembly, giving a total cost of £70-120 per sequenced viral isolate, not including the costs for person-hours. Thus, our protocol provides a high-throughput and scalable approach to viral isolation.

S2.1.8 *Transmission Electron Microscopy of viral isolates*

For ultrastructural analysis, virus particles were transferred onto pioloform-coated electron microscopy (EM) copper grids (Agar Scientific, Standsted, UK) by floating the grids on droplets of virus-containing suspension for 3 min. Following a series of four washes on droplets of deionized water, the bound virus particles were contrasted with 2 % (w/v) uranyl acetate in 2 % (w/v) methyl cellulose (mixed 1:9) on ice for 8 min and the grids then air-dried on a wire loop after carefully removing excess staining solution with a filter. Dried grids were inspected with a JEOL JEM 1400 transmission electron microscope operated at 120 kV and images taken with a digital camera (ES 1000W CCD, Gatan, Abingdon, UK).

S2.1.9 *Host ranges of viral isolates*.

To test the infectivity of phage isolates against different hosts, 2 mL 96-well Teflon plates were prepared with 1 mL of exponentially growing bacterial host (HTCC1062, HTCC7211, H2P3α) in ASM1 medium. As medium control, ASM1 without any bacteria was added to eight wells. 100 µL of phage inoculum was added to eight wells per host type. As no-virus control, for each host further eight wells were left without any viruses added to them. The plate was incubated for ~2 weeks and analysed for signs of infection using flow cytometry as described above. A phage was classified as unable to infect a certain host if none of the eight amended replicates showed signs of infection.

S2.1.10 *DNA preparation and sequencing of viral isolates*

50 mL ASM1 (amended with 100 mM MeOH and amino acid solution for OM43 as described above) in 250 mL acid-washed, polycarbonate flasks were inoculated with host cultures to 10^6 cells·mL^{-1}. The cultures were infected with 10% v/v viruses in ASM1 medium and incubated until after host cell lysis was detected using flow cytometry. The cultures were transferred to 50 mL Falcon tubes and the cellular fraction was removed by centrifugation (GSA rotor, Thermo Scientific 75007588) at 8,500 rpm/10,015 × g for 120 minutes. The supernatant was subsequently filtered through pore-size 0.1 µm PVDF syringe filter membranes to remove any remaining smaller cellular debris. Phages in the filtrate were precipitated using a modified version of an established PEG8000/NaCl DNA isolation method (Solonenko, 2016). Briefly, the filtered phage lysate was transferred into 50 mL Falcon tubes with pre-weighted 5 g PEG8000 and 3.3 g NaCl, shaken until both dissolved and incubated on ice overnight. The phage particles were then pelleted by centrifugation at 8,500 rpm/10,015 × g for 90 min at 4 °C. Supernatant was discarded and phage particles re-suspended by rinsing Falcon tubes twice with 1 mL SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-Cl). DNA was extracted using the Wizard DNA Clean-Up system (Promega) following manufacturer's instructions, but instead of TE we used pre-heated 60 °C nuclease free water to elute DNA. DNA was sequenced by MicrobesNG (Birmingham, UK) with Illumina paired-end HiSeq 2500. Default settings were used for all programs listed below, unless specified otherwise. All

reads were trimmed, quality-controlled and error-corrected using bbmap and tadpole (Bushnell *et al.*, 2017). Contigs were assembled using SPAdes v3.13 (Bankevich *et al.*, 2012) and evaluated with QUAST (Gurevich *et al.*, 2013). The trimmed reads were mapped back against the contigs for scaffolding using bowtie2 (Langmead and Salzberg, 2012), BamM (alignment 0.9, identity 0.95) (available at https://github.com/minillinim/BamM) and samtools (Li *et al.*, 2009). Viral contigs were confirmed with VirSorter (categories 1 or 2, >15kbp) and only accepted if the representative contig showed mean depth coverage an order of magnitude greater than the next highest recruiting contig (to filter out any cellular DNA carryover). Gene calls returned by VirSorter were imported into DNA Master for manual curation (Salisbury and Tsourkas, 2019). Additional gene calls were made using GenMark (Borodovsky and McIninch, 1993), GenMarkS (Besemer *et al.*, 2001), GenMarkS2 (Lomsadze *et al.*, 2018), GenMark.hmm (Zhu *et al.*, 2010), GenMark.heuristic (Besemer and Borodovsky, 1999), Glimmer v.3.02 (Delcher *et al.*, 2007), and Prodigal v.2.6.3 (Hyatt *et al.*, 2010). All gene calls were listed and compared using a scoring system which evaluates gene length, gene overlap and coding potential of ORFs (Salisbury and Tsourkas, 2019). ORFs were then annotated using BLASTp against the NCBI's non-redundant protein sequences (Pruitt *et al.*, 2007), Phmmer (Potter *et al.*, 2018) against UniProts UniProtKB and uniprotrefprot (UniProt Consortium, 2019), Swissprot (Bairoch and Apweiler, 2000) as well as InterProScan (Jones *et al.*, 2014) and Pfam (Finn *et al.*, 2014). Virally-encoded tRNAs were identified with tRNAScan-SE v2.0 using settings `tRNAscan-SE -qQ --detail -o# -m# -f# -l# -c tRNAscan-SE.conf -s# -B` (Lowe and Chan, 2016) to identify putative tRNA genes. Matching tRNA genes (e.g. viral and host tRNA-Leu) were then aligned using BLASTn (Altschul *et al.*, 1990). All viral genomes were organised into viral clusters (ICTV-recognised genera) based on shared gene content with VConTACT2 (Bin Jang *et al.*, 2019), and into populations using a boundary cutoff of 95% ANI over 85% contig length (Roux *et al.*, 2017).

S2.1.11 *Evaluation of gene-sharing networks with hypergeometric testing*

For the hypergeometric testing we gathered all annotated genomes from this study, other known pelagiphage isolates (Zhao *et al.*, 2013, 2018) and we arbitrarily picked one representative contig from each cluster of complete phage genomes that were retrieved from a metagenomic Mediterranean deep chlorophyll maximum sample, that the authors speculated to be infective towards 1A SAR11s and included them in the phylogenetic analysis (Mizuno *et al.*, 2013). The contigs were imported into anvio v6.1 (Eren *et al.*, 2015) where open reading frames were identified internally using Prodigal v2.6.3. The contigs were then used to run a pangenome analysis using the --use-ncbi-blast option. The binning function was then used to export the protein clusters, from which a matrix of the sum of shared proteins between all phages was created. The hypergeom function in the SciPy.stats python package was used to calculate the probability of sharing a protein between phages, based on which we created an average correlation cluster map.

S2.1.12 *Viral phylogenetic analysis*

First we identified additional contigs to include in the phylogenetic analysis. The Global Ocean Virome (GOV2) (Gregory *et al.*, 2019) and a virome from the Western English Channel (Warwick-Dugdale *et al.*, 2018) were screened for contigs that belonged to the same population (95% ANI over 80% length using ClusterGenomes.pl - https://github.com/simroux/ClusterGenomes) or the same genus (clustered by VConTACT2 (Bolduc *et al.*, 2017) into the same viral cluster, using default parameters) as isolate genomes. Additionally, all available genomes from Pelagibacter phage isolates were included, as well as selected fosmid derived contigs from Mediterranean metagenomes and 26 sequences from metagenomic mining that were identified as putative Pelagimyophages (Mizuno *et al.*, 2013; Zhao *et al.*, 2013, 2018; Zaragoza-Solas *et al.*, 2020; Zhang *et al.*, 2020). All sequences were subjected to gene calling using Prodigal v2.6.3. (default parameters) and screened for *TerL* genes by aligning all proteins against known *TerL* genes in annotated genomes (belonging to clusters A, B and D, as well as siphovirus Kolga) from this study using Protein BLAST (default parameter)

(Altschul *et al.*, 1990; Hyatt *et al.*, 2010). The *TerL* gene of *Bordetella* phage LK3 was added as an outgroup. The *TerL* genes were aligned within the Phylogeny.fr online service (Dereeper *et al.*, 2008) opting for MUSCLE alignment (Edgar, 2004) and Gblocks curation (Talavera and Castresana, 2007) with default settings. A maximum-likelihood tree was calculated with PhyML (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006), visualised using FigTree (v.1.4.4, available at http://tree.bio.ed.ac.uk/software/figtree/) and edited in inkscape (www.inkscape.org) for aesthetics.

We only identified a single additional contig from metagenomes using the approach described above that was similar to OM43 phage Venkman. Only one other known phage for a related host (LD28 phage P19250A (Moon *et al.*, 2017)) was available. These three contigs as well as all 14,883 viral contigs publicly available in January 2020 on http://millardlab.org/ were gene called using Prodigal v2.6.3.(Hyatt *et al.*, 2010). All called genes were compared to vFams (Skewes-Cox *et al.*, 2014) using hmmsearch (Eddy, 1998) and sorted by E-value for best hits. The best matching gene sequences that were identified as similar to shared vFams found in *Venkman* and P19250A were important to the Phylogeny.fr online service (Dereeper *et al.*, 2008). Sequences of were aligned individually using T-coffee (Notredame *et al.*, 2000), curated using Gblocks (Talavera and Castresana, 2007) and concatenated manually. Maximum likelihood trees for both alignments were created using PhyML with standard settings and 500 bootstraps (Guindon *et al.*, 2005; Anisimova and Gascuel, 2006). The trees were visualised with FigTree (v.1.4.4) and edited in inkscape (www.inkscape.org) for aesthetics. All sequences used for phylogenetic analysis as well as raw trees in the form of newick files are available at https://github.com/ViralPirates/viral-dte.

S2.1.13 *Calculation of Average Nucleotide Identity*

All-vs-all average nucleotide identity (ANI) for our new isolates and existing pelagiphages using nucmer v.4.1 with the --nooptimize flag, followed by --show-coords to convert the delta file into tabular format. ANI was calculated by taking the mean of all matching regions between two isolates.

S2.1.14 *Viral metagenomic recruitment*

For assessing relative abundances of phage contigs in global datasets, a virome from the Western English Channel and all samples of the Global Ocean Virome dataset (GOV2) were used for recruitment (Warwick-Dugdale *et al.*, 2018; Gregory *et al.*, 2019). Metagenomic reads were downloaded from the European Nucleotide Archive ERR2625613 and subsampled to five million reads using bbmap's reformat.sh command (available at https://sourceforge.net/projects/bbmap/). A bowtie2 (Langmead and Salzberg, 2012) index of dereplicated contigs was created from 25 known pelagiphage genomes from isolates (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020), LD28 phage P19250A (Moon *et al.*, 2017) and from this study six viral population representative genomes isolated on SAR11 hosts and one Methylophilales phage *Venkman* isolated on OM43 host. Reads were mapped against the genomes for each metagenome sample using bowtie2 using these commands: bowtie2 --seed 42 --non-deterministic. To calculate coverage and Reads Per Kilobase of contig per Million reads (RPKM) we used coverm [https://github.com/wwood/CoverM], with the following commands: coverm contig --bam-files *.bam --min-read-percent-identity 0.9 --methods rpkm --min-covered-fraction 0.4. To minimize false positive detection rates (Roux *et al.*, 2017), contigs that did not meet the 40% minimum genome coverage in a given sample were given an RPKM value of 0.

S2.1.15 *Evaluation of isolation efficiency as a function of matching host and virus geography.*

In total, 105 attempts were made for isolating viruses for SAR11 across the three host strains (HTCC7211: 35, HTCC1062: 40; H2P3α: 30). For each strain, the number of successful attempts were recorded in a vector, v_s, of length n where n is the number of attempts made to isolate a virus, such that v_i=1 if a successful attempt was made and v_i=0 if an unsuccessful attempt was made. To test whether observed rates of success for H2P3α were significantly elevated compared to those for HTCC1062 and HTCC7211, the observed rates were evaluated against the null hypothesis that host strain did not make a difference to isolation success.

Briefly, for each pair (H2P3α vs HTCC1062; H2P3α vs HTCC7211), the two vectors v_{s1} *and* v_{s2} were concatenated, and then randomly subsampled 999 times with replacement to a size of the smallest value of *n* between the vectors. The number of successes in each subsample were counted and the distribution of values across the 999 bootstraps was recorded. *P*-values were calculated by evaluating how many times within the 999 bootstraps a value as extreme as the observed value had been recorded.

S2.1.16 *Evaluation of influence of* in situ *water temperature as a predictor of isolation success*

A generalized linear model with a quasi-binomial error distribution was used to evaluate whether host ecotype, temperature or an associated interaction term was significantly correlated with rate of isolation success in SAR11, using base R commands:

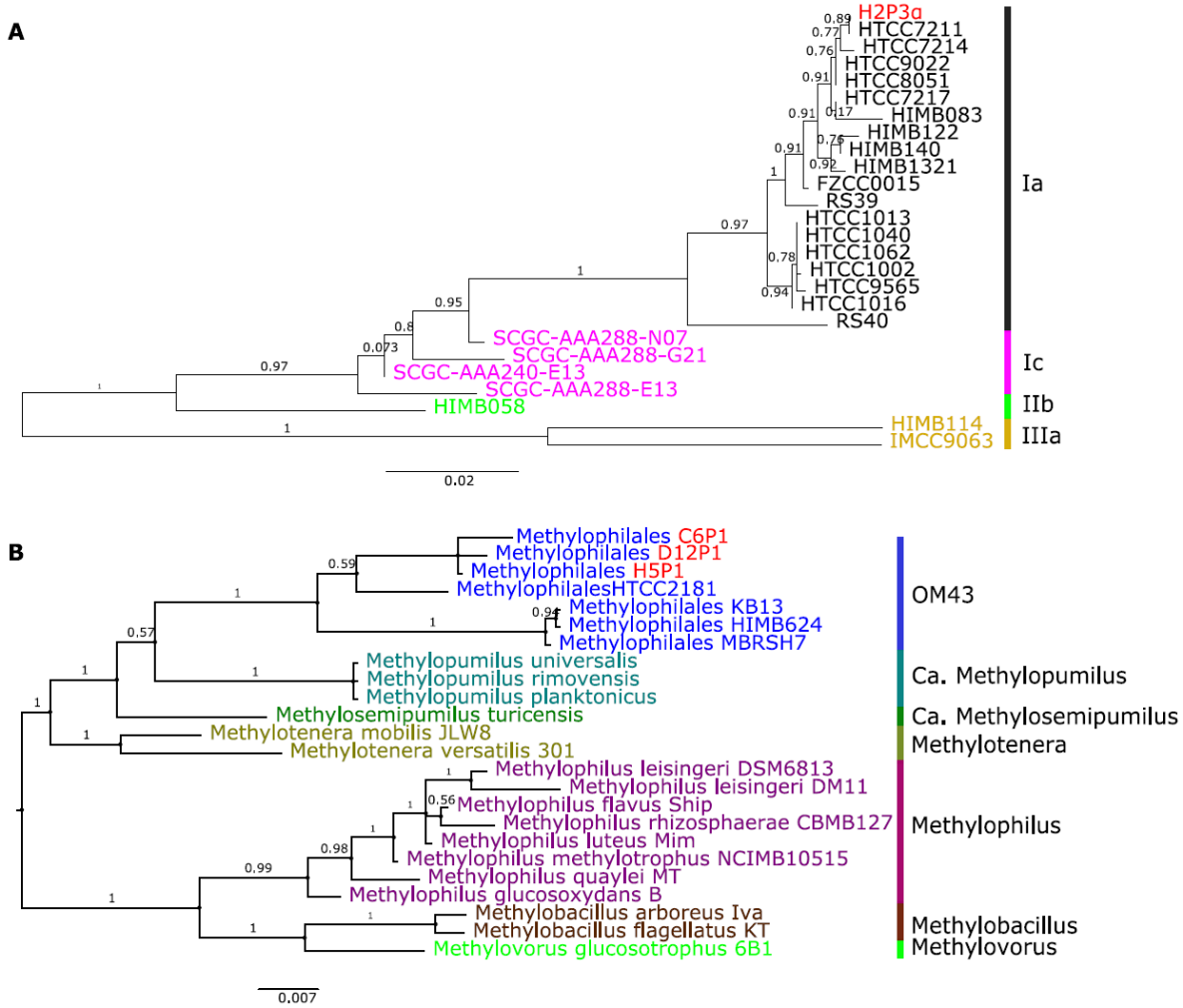
```
fit<-glm(rates~ecotype*temperature,
data=success_rates,family=quasibinomial())
```

S2.1.17 *16S rRNA high-throughput amplicon sequencing and analysis*

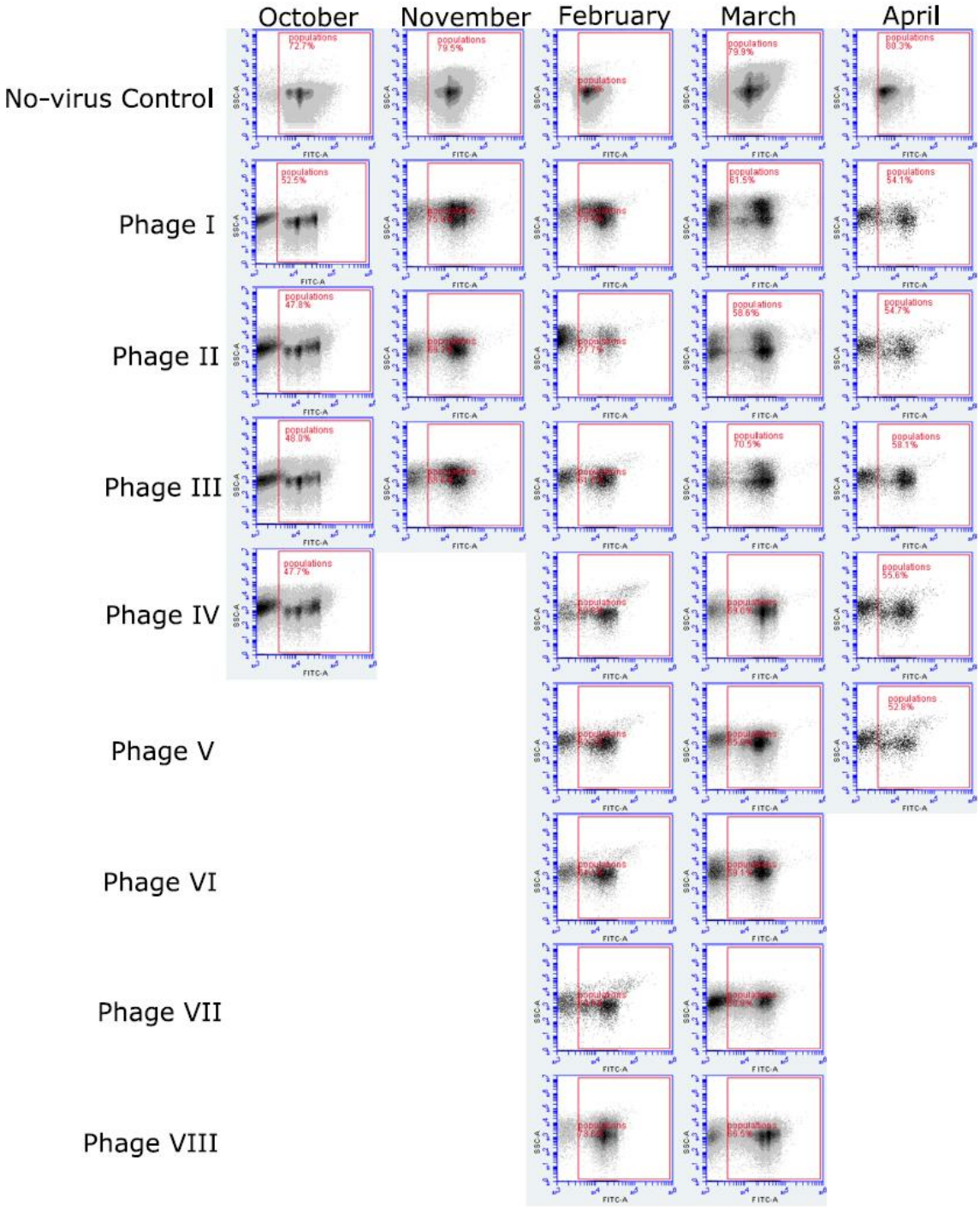
Genomic DNA was extracted using Qiagen DNeasy PowerWater Kit (REF 14900-50-NF) from biomass retained on 0.2 µm PC filters following the manufacturer's protocol with minor modifications to increase the yield. Step 7 was modified from a 5 minute to 10 minute vortex bead beat. Step 21 was changed to have a 2 minute incubation with EB warmed to 55C. Amplification of the hypervariable V4 region of the 16S rRNA gene was performed using the 515fB (5'-GTGYCAGCMGCCGCGGTAA-3') and 806rB (5'-GGACTACNVGGGTWTCTAAT-3') primers. Libraries for each reaction were done attaching dual indices and Illumina adapters with the NexTera XT Index Kit (Illumina Inc.) Purified libraries were pooled and sequenced in a MiSeq platform. Primer sequences from paired-end fastq files were trimmed using CutAdapt (Martin, 2011). Trimmed sequences were quality filtered, dereplicated and merged with DADA2 R package (version 1.8 (Callahan *et al.*, 2016)). Taxonomic assignment was performed using the “assignTaxonomy” command within dada2

pipeline and silva_nr_v132 database (Quast *et al.*, 2013) as training set. SAR11 ASVs were extracted for further characterization using Phyloassigner v6.166 (Vergin *et al.*, 2013) and oligotyping (Eren *et al.*, 2014). SAR11 phylogenetic database SAR11_Phy_DB used in this study is available at https://www.github.com/lbolanos32/NAAMES_2020. Relative contribution barplots were done in R using ggplot2 (Wickham, 2016) and edited in inkscape (www.inkscape.org) for aesthetics.

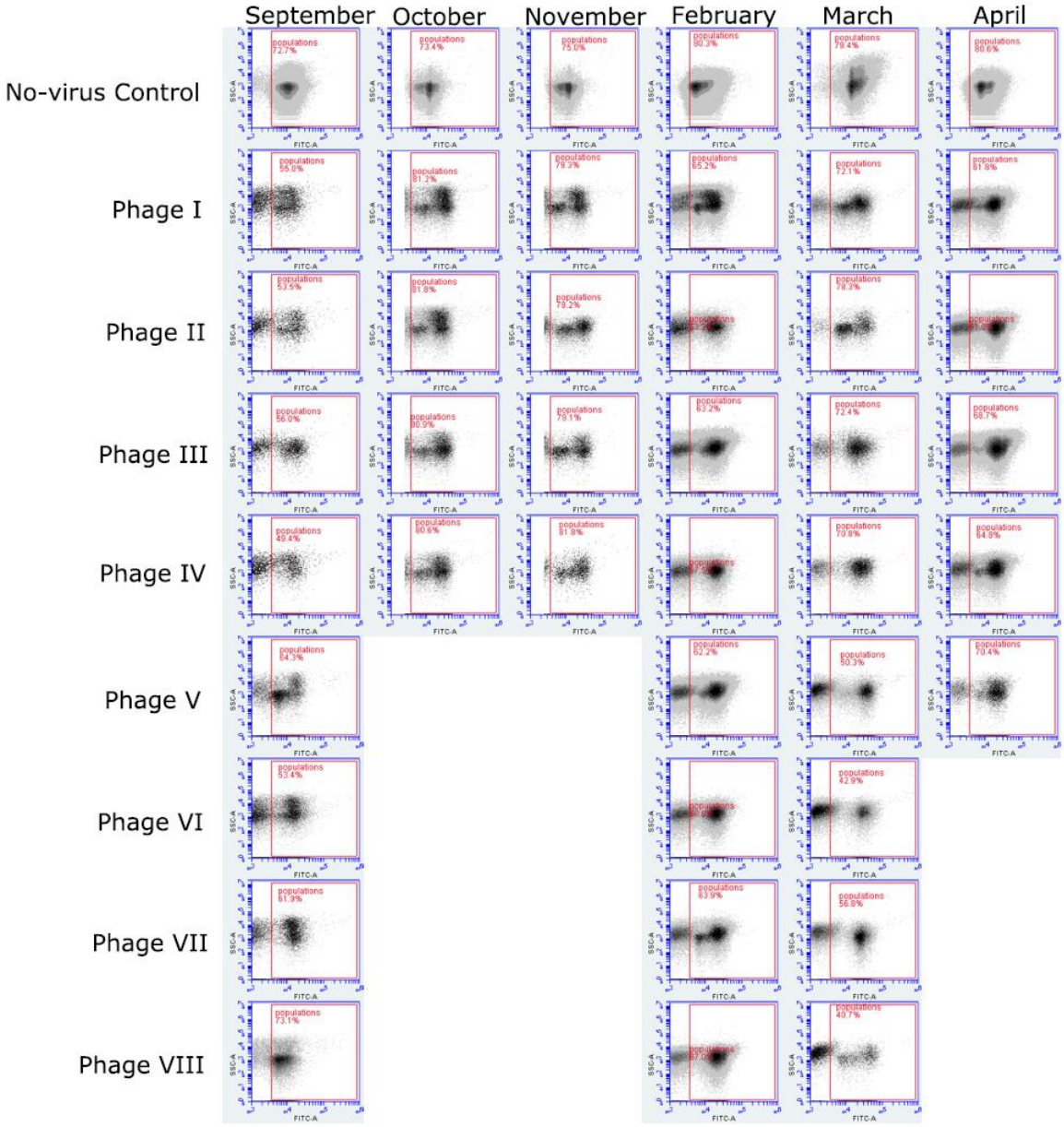
S2.2 Supporting Materials



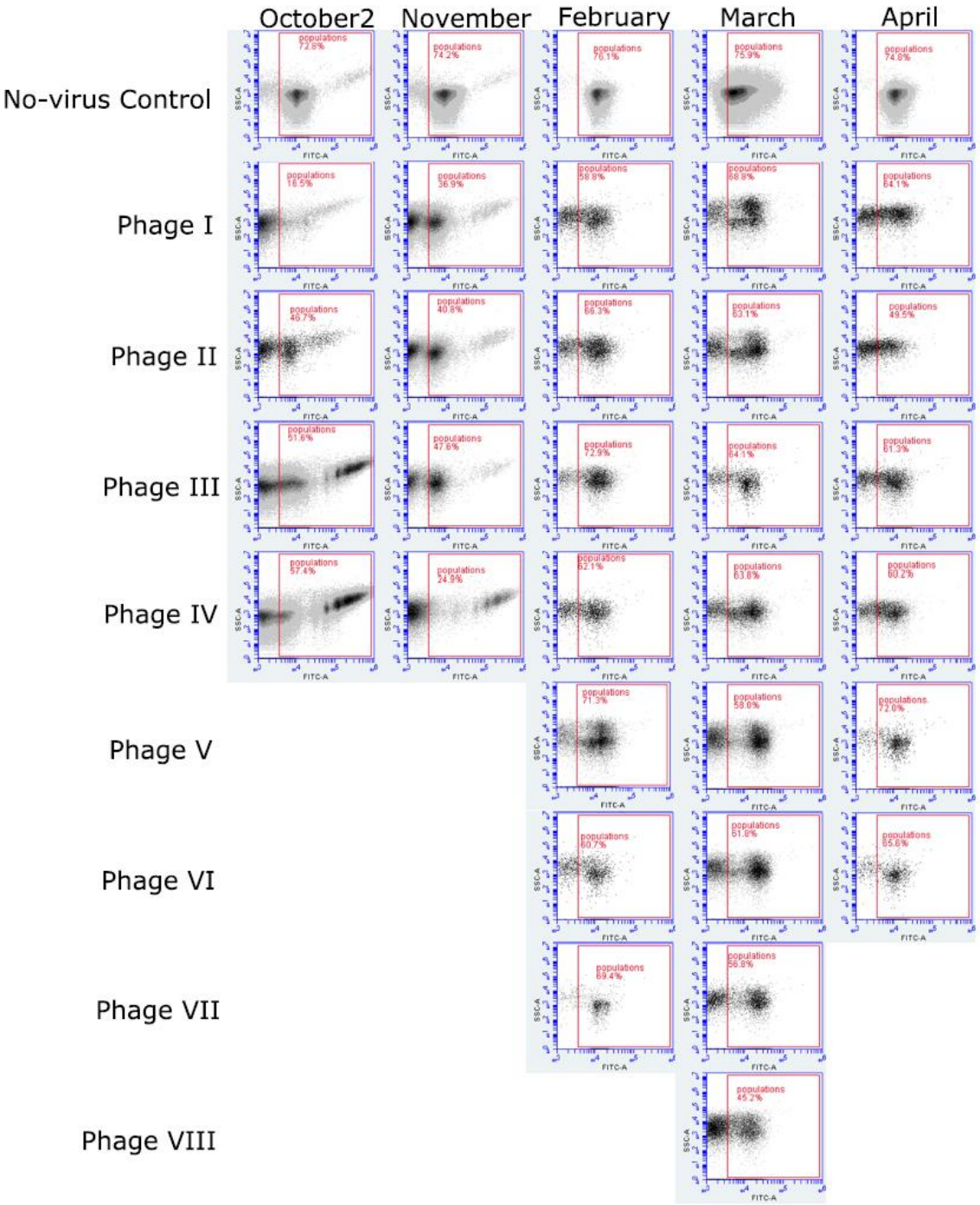
Supplementary Figure 2.1 Maximum likelihood trees (100,000 generations, sampled every 10 generations, first 5000 trees burnin) of 16S genes of **(A)** A range of SAR11 strains and our isolate H2P3α, rooted with the IIIa; **(B)** OM43 and other related members of Methylophilaceae, rooted with the *Methylophilus* branch



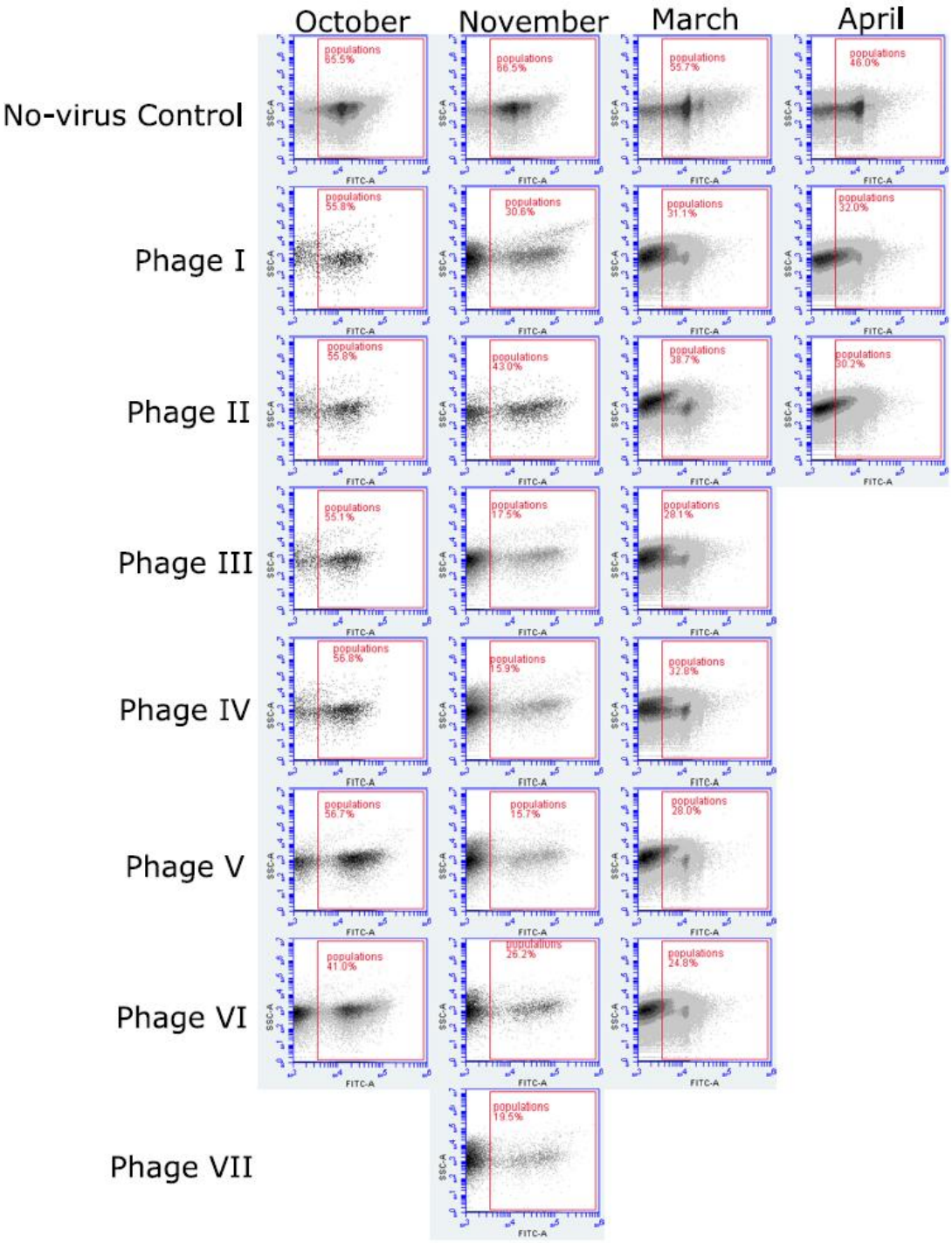
Supplementary Figure 2.2 Cytograms of infected vs uninfected SAR11 warm-water ecotype HTCC7211 cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 72.7% - 84.3% of total events, whereas infected cultures range from 27.7% - 78.6%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to spot infections in cultures.



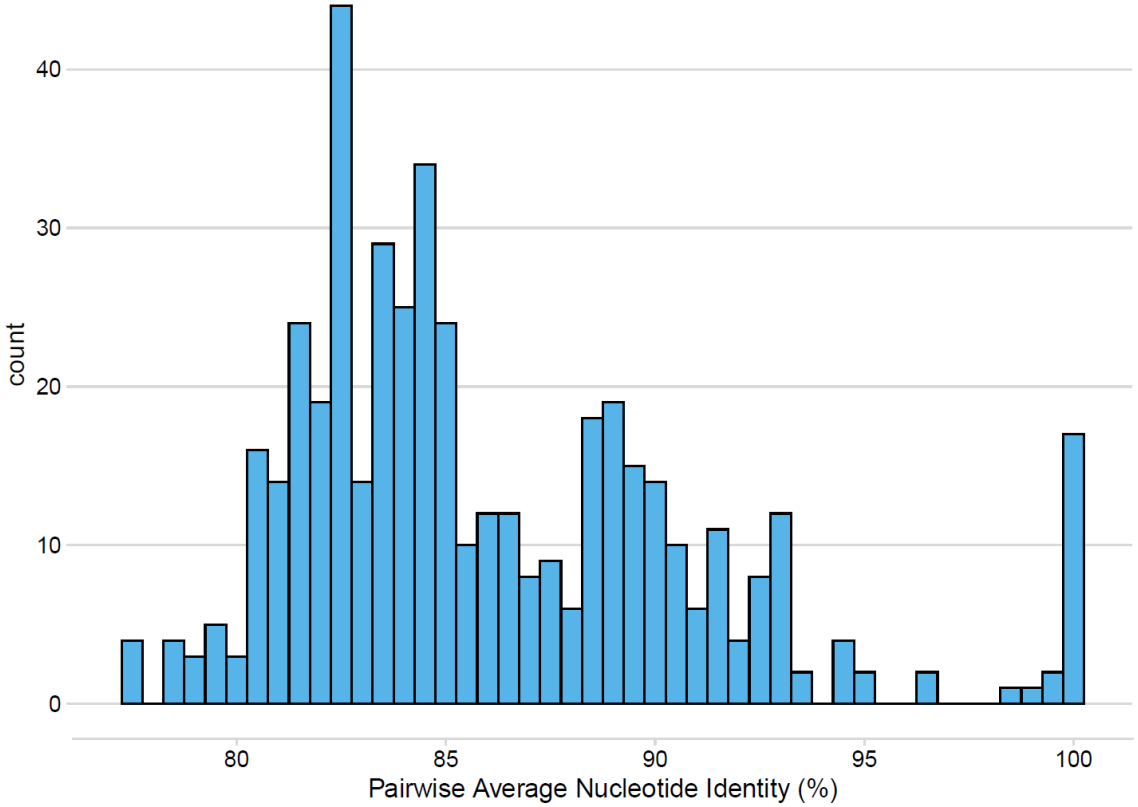
Supplementary Figure 2.3 Cytograms of infected vs uninfected SAR11 cold-water ecotype HTCC1062 cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 72.7% - 80.5% of total events, whereas infected cultures range from 40.7% - 81.8%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to spot infections in cultures.



Supplementary Figure 2.4 Cytograms of infected vs uninfected new SAR11 isolate H2P3α cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 72.8% - 76.1% of total events, whereas infected cultures range from 16.5% - 72.0%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to spot infections in cultures.



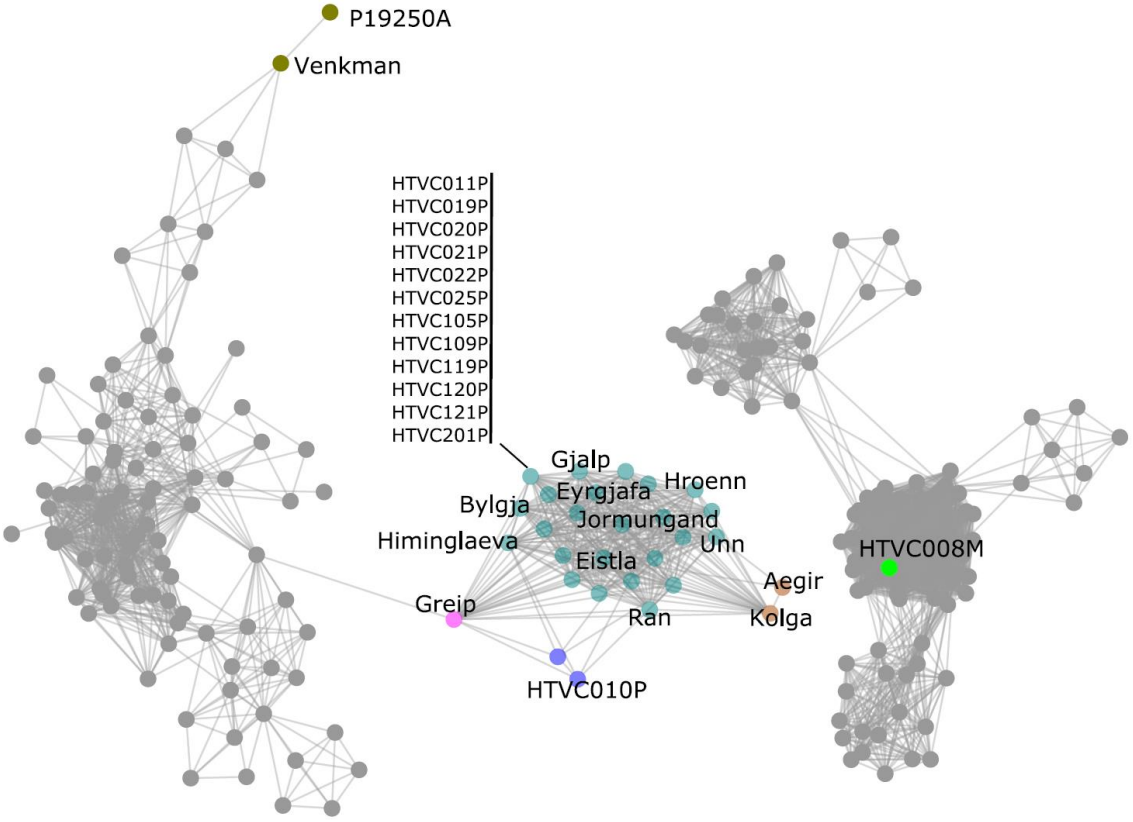
Supplementary Figure 2.5 Cytograms of infected vs uninfected new OM43 isolate H5P1 cultures to isolate phages. Events within the acquisition range ("populations") of no-virus controls range from 46.0% - 66.5% of total events, whereas infected cultures range from 15.7% - 56.8%. Note the increase in low fluorescence (FITC-A) events for phage cultures, which can be used to spot infections in cultures. There was no clear discrimination between noise and the population peaks of the cytograms for infected cultures in March and April, therefore calculated percentages are likely influenced by application of gating parameters defined in other samples.



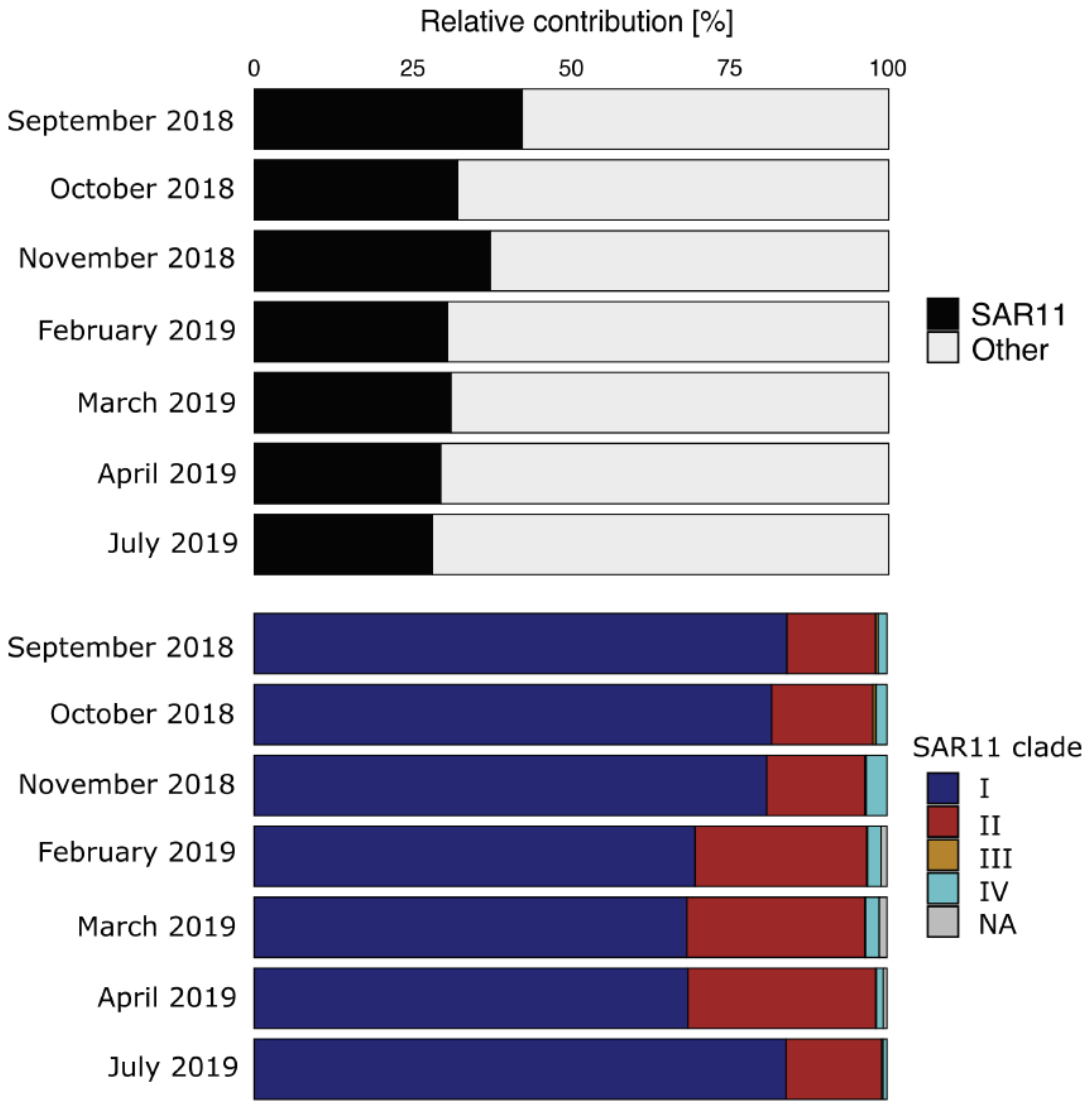
Supplementary Figure 2.6 Count of pairwise average nucleotide identity of all phages included in chapter 2


Bordetella phage LK3
GOV2_43335
GOV2_34415
GOV2_62974
GOV2_04394
GOV2_50999
uvMED-CGR-U-MedDCM-OCT-S46-C34
uvMED-CGR-U-MedDCM-OCT-S46-C10
uvMED-CGR-C54-MedDCM-OCT-S36-C37
uvMED-CGR-C54A-MedDCM-OCT-S39-C35
GOV2_55800
GOV2_09788
GOV2_00272
GOV2_74878
GOV2_06626
GOV2_03201
GOV2_53856
Pelagibacter phage HTVC010P
GOV2_60859
GOV2_64724
uvMED-CGR-C26B-MedDCM-OCT-S35-C55
GOV2_74674
GOV2_08941
GOV2_51756
GOV2_51906
GOV2_57722
Pelagibacter phage Lederberg
GOV2_65848
uvMED-CGR-C23-MedDCM-OCT-S37-C51
Pelagibacter phage Greip
uvMED-CGR-C23B-MedDCM-OCT-S39-C49
uvMED-CGR-C23A-MedDCM-OCT-S43-C49
Pelagibacter phage HTVC019P
Pelagibacter phage Ran
Pelagibacter phage Jormungand
Pelagibacter phage HTVC022P
Pelagibacter phage HTVC201P
uvMED-CGR-C99-MedDCM-OCT-S44-C25
uvMED-GF-C93-MedDCM-OCT-S38-C9
Pelagibacter phage HTVC021P
uvMED-GF-C93-MedDCM-OCT-S32-C16
Pelagibacter phage HTVC105P
Pelagibacter phage HTVC121P
GOV2_10162
GOV2_03142
GOV2_54999
GOV2_01083
Pelagibacter phage HTVC109P
GOV2_61260
Pelagibacter phage HTVC031P
Pelagibacter phage Eistla
GOV2_56275
GOV2_41538
GOV2_34604
GOV2_01612
GOV2_67468
GOV2_75893
GOV2_41506
GOV2_12033
uvMED-CGR-C97-MedDCM-OCT-S42-C7
uvMED-CGF-C97-MedDCM-OCT-S27-C11
GOV2_57174
GOV2_10184
uvMED-CGR-U-MedDCM-OCT-S35-C6
Pelagibacter phage HTVC120P
Pelagibacter phage Bylgja
Pelagibacter phage Himinglaeva
Pelagibacter phage Hroenn
Pelagibacter phage Unn
GOV2_65768
GOV2_47297
uvMED-GF-U-MedDCM-OCT-S32-C31
Pelagibacter phage HTVC119P
Pelagibacter phage HTVC200P
uvMED-CGR-C62A-MedDCM-OCT-S28-C10
GOV2_03365
GOV2_11471
GOV2_44654
Pelagibacter phage HTVC011P
GOV2_64558
GOV2_61346
GOV2_49634
uvMED-GF-U-MedDCM-OCT-S28-C24
GOV2_48181
Pelagibacter phage HTVC025P
Pelagibacter phage Eyrgjafa
Pelagibacter phage Gjalp
GOV2_35636
GOV2_74774
Pelagibacter phage Kolga
GOV2_16560
GOV2_49791
GOV2_50383
GOV2_49937
GOV2_01467
GOV2_38049
PMP-MAVG-15
PMP-MAVG-13
PMP-MAVG-21
PMP-MAVG-4
Pelagibacter phage Mosig
PMP-MAVG-25
PMP-MAVG-12
PMP-MAVG-19
PMP-MAVG-3
PMP-MAVG-26
PMP-MAVG-2
PMP-MAVG-14
PMP-MAVG-16
PMP-MAVG-8
Pelagibacter phage HTVC008M
PMP-MAVG-18
PMP-MAVG-6
PMP-MAVG-7
PMP-MAVG-5
PMP-MAVG-11
PMP-MAVG-23
PMP-MAVG-9
PMP-MAVG-24
PMP-MAVG-22
PMP-MAVG-10
PMP-MAVG-1
PMP-MAVG-17
Podoviridae Cluster D
Podoviridae Cluster A
Podoviridae Cluster B
Siphoviridae
Myoviridae
BATHY
ARC
ANT
TT-EPI
TT-MES
Outlier
0.75 - 1
0.5 - 0.75
0 - 0.5
0.3

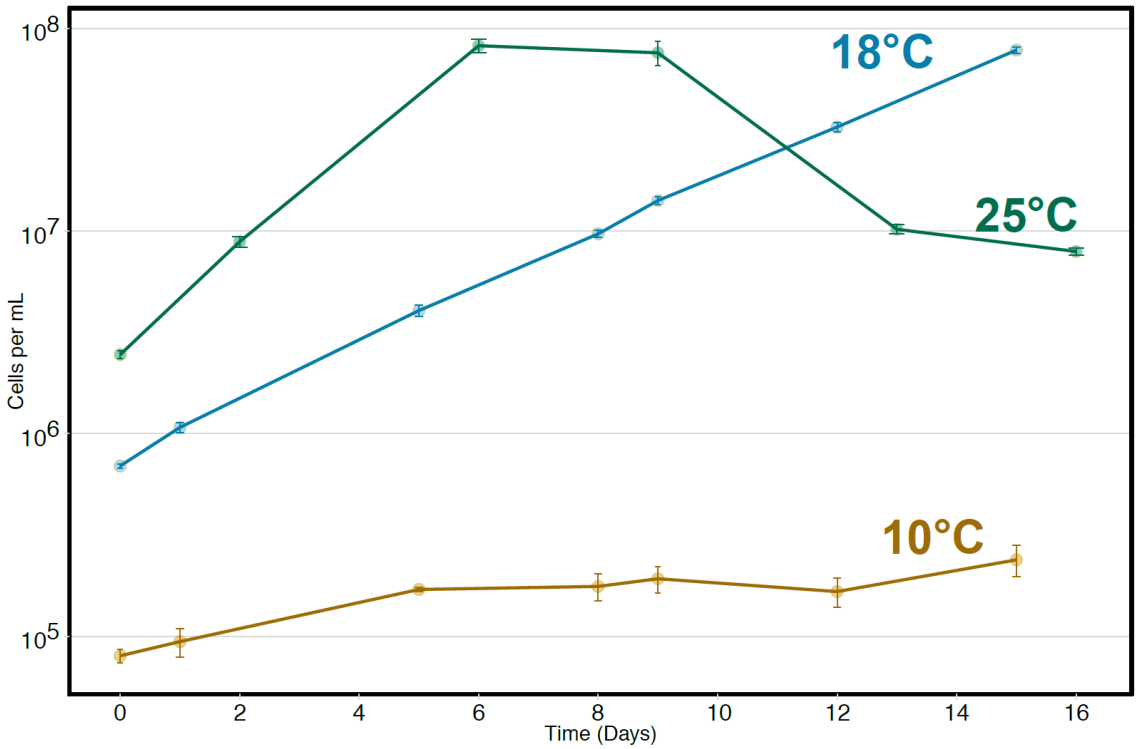
Supplementary Figure 2.7 Unrooted maximum-likelihood tree of TerL genes found in *Pelagibacter* phages, using *Bordetella* phage LK3 as outgroup. Distinct branches were highlighted and labelled according to their respective hypergeometric cluster and viral family. Sampling environment of metagenomes used to identify pelagiphage-like contigs are marked: Epipelagic (TT-EPI), mesopelagic (TT-MES), bathypelagic (BATHY), Southern Ocean/ Antarctica (ANT), Arctic Ocean (ARC).



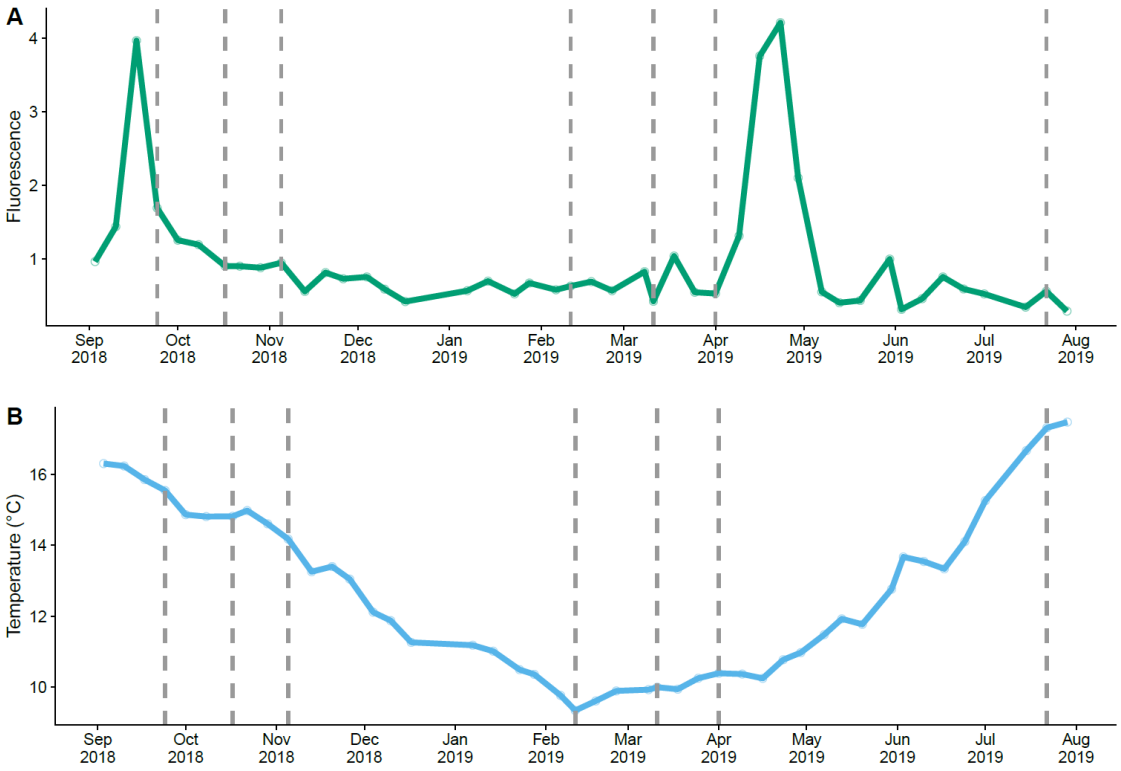
Supplementary Figure 2.8 Shared protein cluster network content of known phage isolates (Zhao *et al.*, 2013, 2018) and our viral isolates against the NCBI Bacterial and Archaeal Viral RefSeq V85 with ICTV + NCBI taxonomy database (VConTACT2, default settings); different colours represent the ICTV-recognised genera assignments.



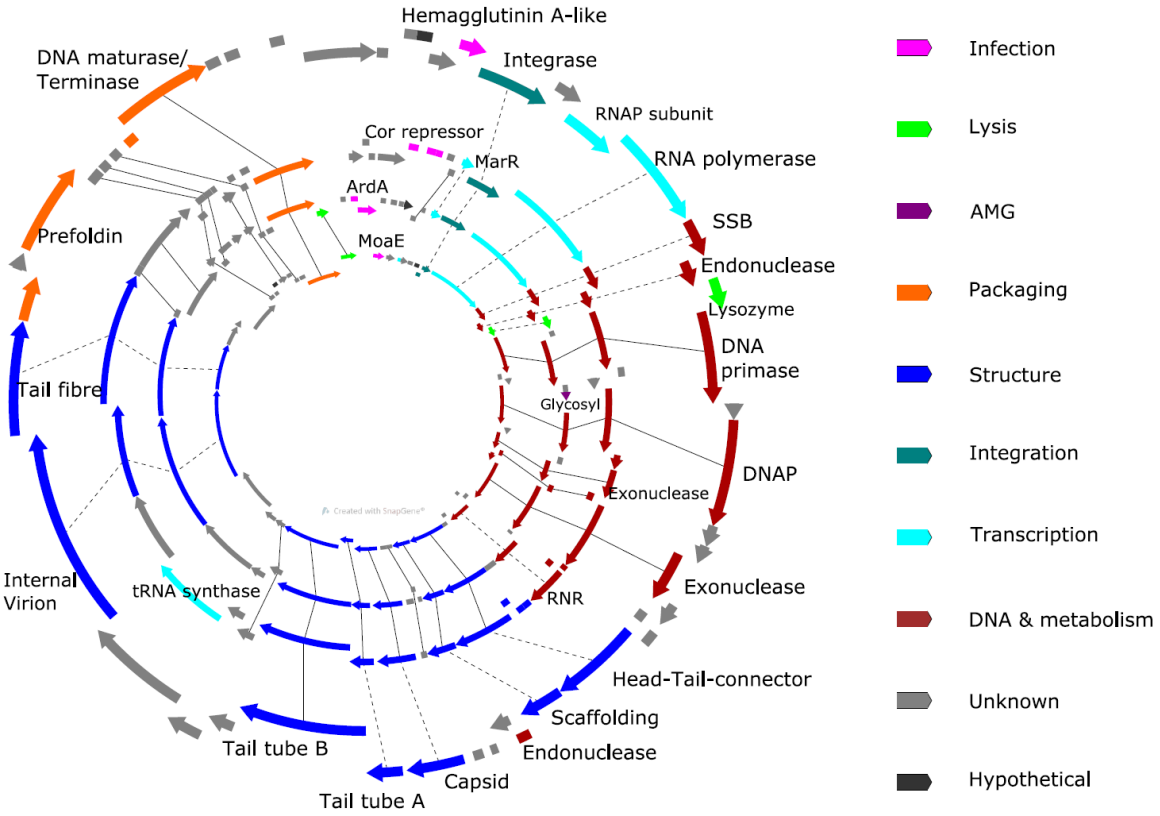
Supplementary Figure 2.9 SAR11 16S rRNA gene relative contribution and clade composition of the collected surface water. (A) Relative contribution of SAR11 amplicon sequence variants to the total 16S rRNA dataset for each time sampled. (B) Clade composition of the SAR11 fraction. The stacked barplots depict the relative contribution of each of the four clades to the total SAR11 fraction. NA: Not assigned to any of the four clades.



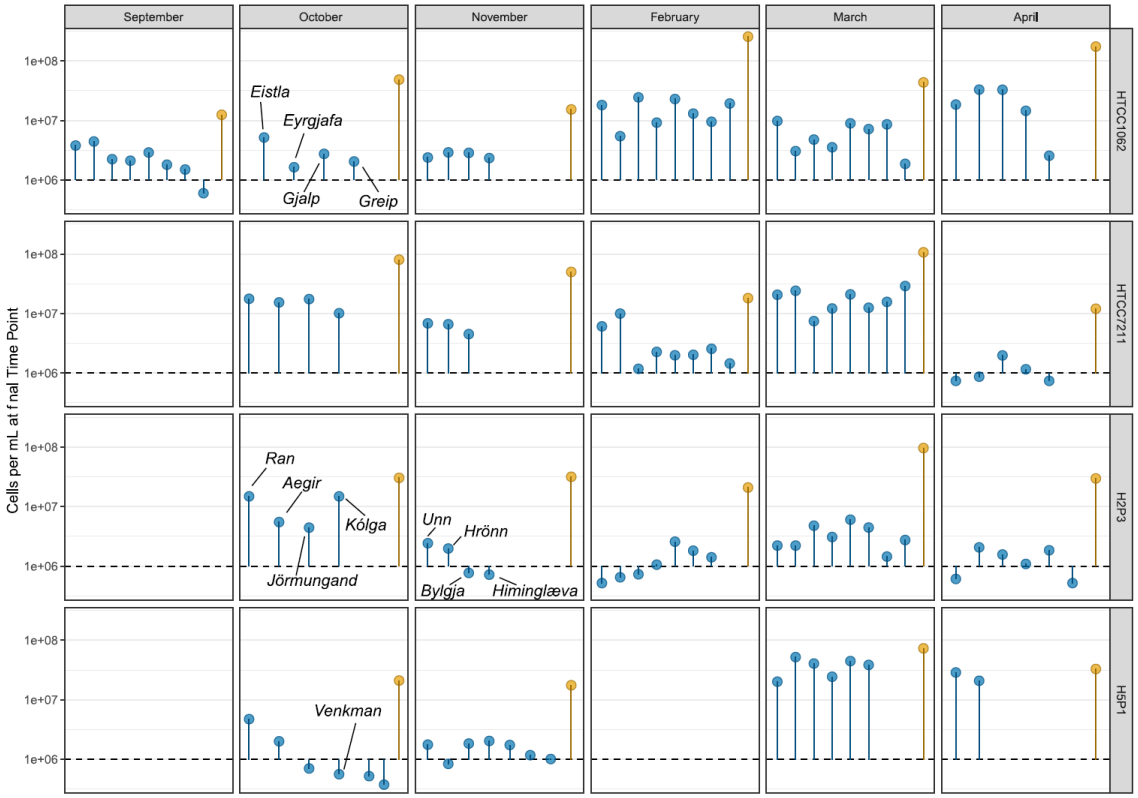
Supplementary Figure 2.10 Average growth curves of H2P3α, based on 8 replicates in 125 mL flask.



Supplementary Figure 2.11 (A) Fluorescence and **(B)** Temperature at Western English Channel sampling station L4 (50°15.00N; 4°13.00W) over the course of the study. Grey dashed lines represent samples for viral isolation.



Supplementary Figure 2.12 Circular gene map of pelagiphages in Clusters A and B arranged inner to outer: Eistla (EXEVC025P), Ran (EXEVC014P), Bylgja (EXEVC010P) and Eyrgjafa (EXEVC018P). Solid lines indicate orthologous genes, dashed lines indicate genes of similar function.



Supplementary Figure 2.13 Final cell densities of infected (blue) vs. uninfected (orange) hosts for 115 of the 117 viruses isolated in this study (two viruses isolated on OM43 D12P1 in July omitted). Dashed line at 10^6 cells per mL represents host density at T_0. Points below this line indicate viruses reduced host density below the inoculum density. Named points refer to sequenced genomes.

Supplementary Table 2.1 Average nucleotide identity of the 16S rRNA gene of OM43 isolates to each other.

	C6P1	D12P1	H5P1	HTCC2181	KB13	HIMB624	MBRSH7
C6P1		98.24	98.31	96.17	95.00	94.10	95.06
D12P1	98.24		98.83	96.62	95.39	94.50	95.46
H5P1	98.31	98.83		97.79	96.55	95.80	96.62
HTCC2181	96.17	96.62	97.79		96.49	95.40	96.55
KB13	95.00	95.39	96.55	96.49		99.70	99.93
HIMB624	94.10	94.50	95.80	95.40	99.70		99.60
MBRSH7	95.06	95.46	96.62	96.55	99.93	99.60	

Supplementary Table 2.2 General information for water samples, environmental parameters provided by the Western Channel Observatory

Collection Time	Collection Date	Temperature (°C)	Fluorescence	Depth (m)	Density	Salinity (psu)	CStarTr0	Par	Oxy
10:21:14	Sept 24 2018	15.546	1.6909	5	1025.87	34.9962	85.2961	0.4093	239.931
09:20:37	Oct 17 2018	14.8181	0.9033	5	1025.709	34.5798	69.7736	2.1833	236.302
10:27:16	Nov 05 2018	14.1766	0.9475	5	1026.269	35.126	88.7715	0.6285	240.449
10:17:04	Feb 11 2019	9.3449	0.6315	5	1026.494	34.2436	79.7166	1.6396	269
10:13:55	Mar 11 2019	10.001	0.5425	5	1026.989	35.0152	89.5834	0.6723	267.411
09:21:13	Apr 01 2019	10.394	0.5298	5	1026.958	35.0672	89.4219	0.4996	270.531
09:20:27	Jul 22 2019	17.3155	0.5612	5	1025.634	35.2245	94.7492	0.1291	237.998
09:09:43	Aug 12 2019	16.9301	0.5437	5	1025.753	35.2596	92.0995	0.2526	242.956

Supplementary Table 2.5 List of general data and descriptions of sequenced viral isolates.

Viral Collection number	Isolate name	Species Representative	viral group (vContact2 cluster)	Hypergeometric cluster	Reference	Genome	% completeness (checkV)	Genome (bp)	% GC	Virus_Family	Host strain	Host clade	Enrichment rounds required	BioSample	BioProject	Accession	SRA
EXVC025P	Eistla	Eistla	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	95.37	39638	32.7	Podoviridae	HTCC1062	SAR11 1a	3	SAMN1460	PRJNA6256	MT375521	SRR11559273
EXVC018P	Eyrgjafa	Eyrgjafa	A	Cluster B	This Study, Zhao 2013, Zhao 2018	Circular	100	38005	32.6	Podoviridae	HTCC1062	SAR11 1a	3	SAMN1460	PRJNA6256	MT375523	SRR11559269
EXVC020P	Gjalp	Eyrgjafa	A	Cluster B	This Study, Zhao 2013, Zhao 2018	Circular	99.77	37857	32.5	Podoviridae	HTCC1062	SAR11 1a	3	SAMN1460	PRJNA6256	MT375524	SRR11559268
EXVC021P	Greip	Greip	B	Cluster D	-	Circular	101	34916	31.5	Podoviridae	HTCC1062	SAR11 1a	3	SAMN1460	PRJNA6256	MT375525	SRR11559267
EXVC014P	Ran	Ran	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	99.62	41529	34.1	Podoviridae	H2P3α	SAR11 1a	3	SAMN1460	PRJNA6256	MT375530	SRR11559266
EXVC013S	Aegir	Kolga	C	Singleton	-	Fragment	38.85	18297	31.1	Siphovirida	H2P3α	SAR11 1a	3	SAMN1460	PRJNA6256	MT375519	SRR11559265
EXVC012P	Jörmunga	Ran	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	99.62	41529	34.1	Podoviridae	H2P3α	SAR11 1a	3	SAMN1460	PRJNA6256	MT375528	SRR11559264
EXVC016S	Kólga	Kolga	C	Singleton	-	Circular	103.24	48659	30.5	Siphovirida	H2P3α	SAR11 1a	3	SAMN1460	PRJNA6256	MT375529	SRR11559263
EXVC019P	Unn	Bylgja	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	98.1	41069	33.5	Podoviridae	H2P3α	SAR11 1a	2	SAMN1460	PRJNA6256	MT375531	SRR11559262
EXVC015P	Hroenn	Bylgja	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	98.13	41069	33.5	Podoviridae	H2P3α	SAR11 1a	2	SAMN1460	PRJNA6256	MT375527	SRR11559272
EXVC010P	Bylgja	Bylgja	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	98.1	41069	33.5	Podoviridae	H2P3α	SAR11 1a	2	SAMN1460	PRJNA6256	MT375520	SRR11559271
EXVC011P	Himinglæv	Bylgja	A	Cluster A	This Study, Zhao 2013, Zhao 2018	Circular	98.1	41069	33.5	Podoviridae	H2P3α	SAR11 1a	2	SAMN1460	PRJNA6256	MT375526	SRR11559270
EXVC028S	Venkman	Venkman	D	Singleton	Moon 2017	Linear	103.86	38624	34.4	Siphovirida	H5P1	OM43	3	SAMN1460	PRJNA6256	MT375522	SRR11559274

Supplementary Table 2.3 Rounds of enrichment required for each water sample vs. target host combination.

Water Sample	Targeted host	Rounds of Enrichment
September-1	1062	2
October-1	7211	3
October-1	1062	3
October-1	H2P3alpha	3
October-1	H5P1	3
November	7211	2
November	1062	2
November	H2P3alpha	2
November	H5P1	2
February	7211	1
February	1062	1
February	H2P3alpha	1
February	H5P1	2
March	7211	1
March	1062	1
March	H2P3alpha	1
March	H5P1	2
April	7211	1
April	1062	1
April	H2P3alpha	1
April	H5P1	2
July	D12P1	2

Supplementary Table 2.4 List of ANI pairings for viruses from (Zhao *et al.*, 2013, 2018) and this study. For brevity only phages from this study are shown and uvMED MAVGs were excluded (Mizuno *et al.*, 2013). Full list is available at: Appendix 8

reference	query	ANI
Aegir	Kolga	99.74
Bylgja	Eistla	80.73
Bylgja	Himinglaeva	99.99
Bylgja	Hroenn	99.995
Bylgja	HTVC019P	81.32
Bylgja	HTVC022P	81.68
Bylgja	HTVC031P	80.7
Bylgja	HTVC119P	87.525
Bylgja	HTVC120P	89.447273
Bylgja	HTVC200P	81.15
Bylgja	Jormunganc	82.26
Bylgja	Kolga	91.55
Bylgja	Ran	82.26
Bylgja	Unn	99.995
Eistla	Bylgja	80.73
Eistla	Eyrgjafa	93.24
Eistla	Gjalp	93.24
Eistla	Greip	95.175
Eistla	Himinglaeva	80.73
Eistla	Hroenn	80.73
Eistla	HTVC011P	90.79
Eistla	HTVC019P	84.745
Eistla	HTVC021P	82.385
Eistla	HTVC022P	83.302
Eistla	HTVC025P	84.38
Eistla	HTVC031P	85.74125
Eistla	HTVC105P	82.49
Eistla	HTVC109P	80.56
Eistla	HTVC119P	83.115
Eistla	HTVC120P	80.02
Eistla	HTVC121P	82.426667
Eistla	HTVC201P	82.9875
Eistla	Jormunganc	84.8
Eistla	Ran	84.8
Eistla	Unn	80.73
Eyrgjafa	Eistla	93.1
Eyrgjafa	Gjalp	99.996667
Eyrgjafa	HTVC011P	88.623333
Eyrgjafa	HTVC019P	92.27
Eyrgjafa	HTVC022P	92.73
Eyrgjafa	HTVC025P	81.946667
Eyrgjafa	HTVC031P	87.82
Eyrgjafa	HTVC119P	89.19
Eyrgjafa	HTVC201P	92.755
Gjalp	Eistla	93.1
Gjalp	Eyrgjafa	99.996667
Gjalp	HTVC011P	88.623333
Gjalp	HTVC019P	92.27
Gjalp	HTVC022P	92.73
Gjalp	HTVC025P	81.946667
Gjalp	HTVC031P	87.82
Gjalp	HTVC119P	89.19
Gjalp	HTVC201P	92.755
Greip	Eistla	95.175
Greip	HTVC010P	82.33
Greip	HTVC011P	90.58
Greip	HTVC022P	84.44
Greip	HTVC031P	88.35
Greip	HTVC201P	88.83
Himinglaeva	Bylgja	99.99
Himinglaeva	Eistla	80.73
Himinglaeva	Hroenn	99.995
Himinglaeva	HTVC019P	81.32
Himinglaeva	HTVC022P	81.68
Himinglaeva	HTVC031P	80.7
Himinglaeva	HTVC119P	87.525
Himinglaeva	HTVC120P	89.447273
Himinglaeva	HTVC200P	81.15
Himinglaeva	Jormunganc	82.26
Himinglaeva	Kolga	91.55
Himinglaeva	Ran	82.26
Himinglaeva	Unn	99.995
Hroenn	Bylgja	99.995
Hroenn	Eistla	80.73
Hroenn	Himinglaeva	99.995
Hroenn	HTVC019P	81.32
Hroenn	HTVC022P	81.68
Hroenn	HTVC031P	80.68
Hroenn	HTVC119P	87.525
Hroenn	HTVC120P	89.146
Hroenn	HTVC200P	81.15
Hroenn	Jormunganc	82.293333
Hroenn	Kolga	91.55
Hroenn	Ran	82.293333
Hroenn	Unn	99.985
Jormunganc	Bylgja	82.26
Jormunganc	Eistla	84.8
Jormunganc	Himinglaeva	82.26
Jormunganc	Hroenn	82.26
Jormunganc	HTVC019P	89.545455
Jormunganc	HTVC021P	83.305455
Jormunganc	HTVC022P	89.695
Jormunganc	HTVC031P	84.116
Jormunganc	HTVC105P	82.575
Jormunganc	HTVC119P	84.7
Jormunganc	HTVC120P	88.71
Jormunganc	HTVC121P	83.717273
Jormunganc	HTVC201P	86.2675
Jormunganc	Kolga	93.04
Jormunganc	Ran	99.995
Jormunganc	Unn	82.26
Kolga	Aegir	99.725
Kolga	Bylgja	91.55
Kolga	Himinglaeva	91.55
Kolga	Hroenn	91.55
Kolga	HTVC119P	89.55
Kolga	HTVC120P	90.6125
Kolga	Jormunganc	93.04
Kolga	Ran	93.04
Kolga	Unn	91.55
Ran	Bylgja	82.293333
Ran	Eistla	84.8
Ran	Himinglaeva	82.293333
Ran	Hroenn	82.293333
Ran	HTVC019P	89.389
Ran	HTVC021P	83.128182
Ran	HTVC022P	89.395714
Ran	HTVC031P	84.116
Ran	HTVC105P	82.395
Ran	HTVC119P	84.7
Ran	HTVC120P	88.71
Ran	HTVC121P	83.717273
Ran	HTVC201P	86.038182
Ran	Jormunganc	99.995
Ran	Kolga	93.04
Ran	Unn	82.293333
Unn	Bylgja	99.995
Unn	Eistla	80.73
Unn	Himinglaeva	99.995
Unn	Hroenn	99.985
Unn	HTVC019P	81.32
Unn	HTVC022P	81.68
Unn	HTVC031P	80.7
Unn	HTVC119P	87.525
Unn	HTVC120P	89.146
Unn	HTVC200P	81.15
Unn	Jormunganc	82.26
Unn	Kolga	91.55
Unn	Ran	82.26

Chapter 3: Genomic evidence for inter-class host transition between abundant streamlined heterotrophs by a novel and ubiquitous marine Methylophage.

This chapter applies the High-throughput viral isolation methods and reports the first myophage of the OM43 clade, isolated from the Western Channel and Sargasso Sea. It provides genomic evidence of host transitioning coupled with host range experiments. It was reformatted from the manuscript submitted for publication (in review by the ASM journal) and has been reviewed by my co-authors. A pre-print of the paper is available at:

Holger H. Buchholz, Luis M. Bolaños, Ashley G. Bell, Michelle L. Michelsen, Michael J. Allen, Ben Temperton. 2021. *bioRxiv*; doi:10.1101/2021.08.24.457595

Authors Contributions:

- *HHB conceived, designed and performed the experiments, analysed the data, prepared figures and tables, authored and reviewed drafts of the paper, and approved the final draft*
- *LMB analysed the data*
- *AGB analysed the tRNA alignment data*
- *MLM performed the experiments*
- *MJA authored or reviewed drafts of the paper*
- *BT performed structural protein analysis and prepared protein 3D models, conceived and designed the experiments, analysed the data and prepared figures and tables, contributed reagents, materials and analysis tools, authored or reviewed drafts of the paper and approved the final draft*

3.1 Abstract

The methylotrophic OM43 clade are Gammaproteobacteria that comprise some of the smallest free-living cells known and have highly streamlined genomes. OM43 represents an important microbial link 0between marine primary production and remineralisation of carbon back to the atmosphere. Bacteriophages shape microbial communities and are major drivers of microbial mortality and global marine biogeochemistry. Recent cultivation efforts have

brought the first viruses infecting members of the OM43 clade into culture. Here we characterize a novel myophage infecting OM43 called Melnitz. Melnitz was isolated independently on three separate occasions (with isolates sharing >99.95% average nucleotide identity) from water samples from a subtropical ocean gyre (Sargasso Sea) and temperate coastal (Western English Channel) systems. Metagenomic recruitment from global ocean viromes confirmed that Melnitz is globally ubiquitous, congruent with patterns of host abundance. Bacteria with streamlined genomes such as OM43 and the globally dominant SAR11 clade use riboswitches as an efficient method to regulate metabolism. Melnitz encodes a two-piece tmRNA (*ssrA*), controlled by a glutamine riboswitch, providing evidence that riboswitch use also occurs for regulation during phage infection of streamlined heterotrophs. Virally encoded tRNAs and *ssrA* found in Melnitz were phylogenetically more closely related to those found within the alphaproteobacterial SAR11 clade and their associated myophages than those within their gammaproteobacterial hosts. This suggests the possibility of an ancestral inter-class host transition event between SAR11 and OM43. Melnitz and a related myophage that infects SAR11 were unable to infect hosts of the SAR11 and OM43, respectively, suggesting host transition rather than a broadening of host range.

3.1.1 Importance

Isolation and cultivation of viruses is the foundation on which the mechanistic understanding of virus-host interactions and ground-truthing is based. This study isolated and characterised the first myophage known to infect the OM43 clade, expanding our knowledge of this understudied group of microbes. The near-identical genomes of four strains of Melnitz isolated from different marine provinces and global abundance estimations from metagenomic data suggest that this viral population is globally ubiquitous. Genome analysis revealed several unusual features in Melnitz and related genomes recovered from viromes, such as a curli operon and virally encoded tmRNA controlled by a glutamine riboswitch, neither of which are found in the host. Further phylogenetic analysis of shared genes indicates that this group of viruses infecting the gammaproteobacterial OM43 shares a recent common ancestor with viruses infecting the abundant alphaproteobacterial SAR11 clade. Host ranges are affected by compatible cell

surface receptors, successful circumvention of superinfection exclusion systems and the presence of required accessory proteins, which typically limits phages to singular narrow groups of closely related bacterial hosts. This study provides intriguing evidence that for streamlined heterotrophic bacteria, virus-host transitioning is not necessarily restricted to phylogenetically related hosts, but is a function of shared physical and biochemical properties of the cell.

3.2 Introduction

Bacteriophages are the most abundant and diverse biological entities in the oceans and are, on average, an order of magnitude more abundant than their bacterial hosts in surface water (Suttle, 2005; Brussaard *et al.*, 2008). Viral predation kills a large proportion of bacterial cells in marine surface waters each day (Wommack and Colwell, 2000) and is a main contributor to nutrient recycling by releasing cell-bound organic compounds into the environment (Weitz and Wilhelm, 2012; Weitz *et al.*, 2015). Viral infection can also alter host metabolism through metabolic hijacking (Breitbart *et al.*, 2018; Warwick-Dugdale *et al.*, 2019), which has been shown to reprogram resource acquisition and central carbon and energy metabolism (Howard-Varona *et al.*, 2020), influencing oceanic nutrient cycles. The selective pressure of the predator-prey relationship of bacteria and phages is also a main driver of microbial evolution (Martiny *et al.*, 2014), where the constant arms race requires phages to evolve and improve strategies to overcome host defence mechanisms (Avrani *et al.*, 2012). Recent advances in culture-independent sequencing technology such as single-cell genomics and metagenomics have expanded our understanding of the enormous diversity of marine viruses (Brum *et al.*, 2015; Martinez-Hernandez *et al.*, 2017, 2019; Gregory *et al.*, 2019). However, many of these sequences lack representation in viral culture collections, limiting experimental determination of viral characteristics such as host-range and infection dynamics. A resurgence in bacterial cultivation efforts and improved viral isolation methods has led to the discovery of many new phages infecting abundant but fastidious marine bacteria such as the abundant and ubiquitous SAR11 clade. Combining genomes from viral cultures with metagenomics identified these viruses to be some of the most abundant on Earth across all marine ecosystems (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020). Yet,

many more virus-host systems occupying a range of important ecological niches such as methylotrophy remain poorly understood.

Members of the OM43 clade are small, genomically streamlined (genomes ~1.3 Mbp) Type I methylotrophs of the class Gammaproteobacteria (Giovannoni *et al.*, 2008). The catabolism of methanol and other volatile organic compounds (VOCs) is an important link between primary production and remineralisation of carbon back to atmospheric CO_2 (Halsey *et al.*, 2012; Thrash *et al.*, 2014; Beale *et al.*, 2015). Epipelagic peak abundance of OM43 coincides with phytoplankton blooms that provide by-products which are the main carbon source for OM43 (Halsey *et al.*, 2012). OM43 are particularly abundant in coastal ecosystems where they comprise up to 5% of the microbial community (Rappé *et al.*, 1997). Members of the OM43 clade are challenging to grow in the laboratory: increased levels of auxotrophy and largely constitutive metabolism renders them sensitive to media composition (Giovannoni *et al.*, 2008). As a result, only two OM43 phages have been reported, therefore the influence of viral predation on OM43 is virtually unexplored. The isolation of Venkman, the first viruses infecting OM43 (Figure 2.5), from the coastal Western English Channel (WEC), and MEP301 from the Bohai Sea were both reported in 2021 (Yang *et al.*, 2021). Venkman was the third most abundant phage in the WEC sample, indicating that phages of methylotrophs are a major component of this coastal ecosystem (Figure 2.6). In contrast, recruiting reads from global ocean viromes against MEP301 and Venkman, indicated that their relative abundance was below detection limit in most lower-latitude pelagic viromes (Yang *et al.*, 2021). Thus, phages infecting OM43 were thought to be predominantly found at higher latitudes in regions of high primary productivity.

Here we report the isolation and genomic analysis of Melnitz, representing a novel population of myophages infecting OM43. Four representatives of this virus that shared >99.5% average nucleotide identity were isolated independently on three separate occasions from the temperate coastal WEC and from the Sargasso Sea located within the North Atlantic subtropical gyre. This indicates that despite results indicating low relative abundance at low latitudes, Melnitz was sufficiently abundant to be isolated through enrichment techniques. The genomic similarity between independent isolates suggests a cosmopolitan global distribution, which was supported by metagenomic read recruitment from Global

Ocean Viromes (GOV2, (Gregory *et al.*, 2016). Genome analysis of Melnitz revealed a two-piece tmRNA gene (*ssrA*), controlled by a glutamine riboswitch. Riboswitch control of regulation is a feature of streamlined organisms such as OM43 and SAR11 (Giovannoni *et al.*, 2014), and here we show that it is also a feature of their associated viruses. Like previously reported SAR11 myophages (Appendix 2, Zhao *et al.*, 2013), Melnitz also encoded the portal proteins of a curli operon that is absent in the host. Structural analysis suggests a putative reconfiguration allows this phage-encoded protein to serve as a novel gated secretin or pinholin, with gene synteny indicating a role in timing the release of viral progeny. Phylogenetic analysis revealed that both the *ssrA* gene and tRNA genes encoded by Melnitz were more closely related to those found within the alphaproteobacterial SAR11 host or its associated viruses, than those of its own gammaproteobacterial host. These findings point towards a recent shared ancestor indicative of host transitioning between OM43 and SAR11.

3.3. Results and Discussion

3.3.1 *Phage Melnitz infecting* Methylophilales sp. H5P1 *shares viral clusters with* Pelagibacter *phages.*

Two bacteriophages were isolated on the OM43 *Methylophilales* strain H5P1 from two environmental water samples taken from the WEC previously in April and July 2019 (Table S2.5, Buchholz *et al.*, 2021). Two more phages were obtained from an additional water sample taken at the Bermuda Atlantic Time Series (BATS) station in the Sargasso Sea (Table 3.1). Phages were considered successfully purified after three rounds of DtE isolation, sequenced from cultures and assembled into single circular contigs. CheckV comparison to publicly available metagenomes and phage genomes suggested that the viral contigs were complete, circularly permuted genomes without terminal repeats. All four phage genomes shared 99.95-100% average nucleotide identity across their full genomes therefore all four phages should be regarded as the same viral species (Turner *et al.*, 2021). We named this species “Melnitz" after a character in the popular Ghostbusters franchise, continuing the theme of another phage previously isolated on OM43 (Venkman) (Buchholz, *et al.*, 2021). For clarity, where individual phages within the population (Melnitz) are specified, they will

subsequently be referred to with a numerical suffix (e.g. Melnitz-1). General features of the four phages are summarised in Table 3.1. Shared gene network analysis (VConTACT2) using assembled contigs from GOV2 (Gregory *et al.*, 2019) and RefSeq (V88 with ICTV and NCBI taxonomy) viruses assigned Melnitz to the same cluster as the *Pelagibacter* myophages HTVC008M and Mosig (Zhao *et al.*, 2013; Appendix 2), suggesting they belong to the same family (Figure 3.1). The four Melnitz genomes were shared across two clusters, the first contained an additional 15 viral contigs from nine different metagenomes. The second cluster contained three virome contigs shared between the two clusters, as well as 25 pelagimyophage genomes assembled from metagenomes (PMP-MAVGs) (Zaragoza-Solas *et al.*, 2020). *Pelagibacter* myophage isolates HTVC008M and Mosig were also placed in this second cluster. A total of 66 contigs clustering with Melnitz isolates were identified from GOV2 and WEC viromes (Gregory *et al.*, 2019; Warwick-Dugdale *et al.*, 2019). Phylogenetic analysis showed that only two environmental contigs consistently shared a branch with Melnitz, based on single shared genes encoding: tail sheath protein (Supplementary Figure 3.2), terminase large subunit (Supplementary Figure 3.3) and scaffolding proteins (Supplementary Figure 3.4), as well as four concatenated structural genes (Figure 3.2). All phages within this viral group were either myophage isolates known to infect streamlined heterotrophs, or were previously predicted to do so based on phylogenetic similarity (Zaragoza-Solas *et al.*, 2020). The sequence data and phylogenetic evidence thus suggests that Melnitz phages are to be considered myophages. Unfortunately, conformational evidence from transmission electron microscopy (TEM) was flawed. The images showed straight contractile tails, which are likely already contracted since no tail plates or tail fibres were visible, indicative of myophage morphology. But TEM image quality was consistently poor and/or displayed signs of preparation artifacts. Furthermore, different capsid sizes could be seen in some images (Supplementary Figure 3.1). This may be due to the image aspect, viewing the elongated capsids common in T4-like myophages from two different angles. We found no evidence in the sequencing data of another virus in the cultures, but it is possible that a contaminating phage was present and did not return sequences for unknown reasons. Even though TEM evidence is of poor quality and does not allow for a clear identification based on morphology, the Melnitz genomes

were closely related to both confirmed HTVC008M (Zhao *et al.*, 2013) and putative pelagimyophages (Zaragoza-Solas *et al.*, 2020), suggesting Melnitz to be a putative myophage.

3.3.2 *Metagenomic analysis shows cosmopolitan nature of Melnitz-like phages*

To establish the distribution patterns of Melnitz, GOV2 virome reads were mapped against phage genomes (≥95% nucleotide identity over ≥90% read length; genome coverage of <40% was required to register a phage as 'present' to avoid false positives (Roux *et al.*, 2017)) (Supplementary Figure 3.5). Relative abundance of phages was calculated based on the number of reads mapped to a contig, normalised by contig length and sequencing depth (mapped reads per kilobase pair of metagenome; RPKM). Linear regression of relative abundance of all three phages known to infect OM43 (Melnitz, Venkman and MEP301) as a function of temperature showed a significant and negative relationship (Melnitz: $p = 4.69 \times 10^{-5}$, $R^2 = 0.199$; Venkman: $p < 2.20 \times 10^{-16}$, $R^2 = 0.435$; MEP301: $p = 1.85 \times 10^{-7}$, $R^2 = 0.198$). Relative abundance was positively correlated with absolute latitude (Melnitz: $p = 1.78 \times 10^{-5}$, $R^2 = 0.094$; Venkman: $p < 2.20 \times 10^{-16}$, $R^2 = 0.545$; MEP301: $p = 4.67 \times 10^{-9}$, $R^2 = 0.247$). This suggests that these viruses are most abundant in cold, high-latitude waters. In contrast, relative abundances of Melnitz-related *Pelagibacter* phages HTVC008M and Mosig were not correlated to temperature (Mosig: $p = 0.094$, $R^2 = 0.047$; HTVC008M: $p = 0.217$, $R^2 = 0.004$) or latitude (Mosig: $p = 0.028$, $R^2 = 0.031$; HTVC008M: $p = 0.077$, $R^2 = 0.018$). OM43 phages MEP301 and Venkman were classed as 'present' (>40% genome coverage based on (Roux *et al.*, 2017) in 39.1% and 53.9% of 131 GOV2 viromes, respectively. Melnitz showed greater global ubiquity, being classed as 'present' in 78.8% of GOV2 viromes. Whilst there are no GOV2 samples from the Sargasso Sea, this ubiquity may explain how we were able to isolate Melnitz from both the WEC and the Sargasso Sea.

Table 3.1 General features of OM43 strain H5P1 and four Melnitz phages isolated in this study.

Phage	Culture ID	Host	Phage group	Morpho-type	Genome size (bp)	G + C %	ORFs	tRNAs	Source	Date	Latitude	Longitude
Melnitz	EXVC044M	H5P1	Melnitz	Myovirus	141,548	37.6	224	4	BATS	01-Jun-19	N31°40'	W64°10'
Melnitz-1	EXVC043M	H5P1	Melnitz	Myovirus	141,552	37.6	226	4	BATS	01-Jun-19	N31°40'	W64°10'
Melnitz-2	EXVC040M	H5P1	Melnitz	Myovirus	141,548	37.6	224	4	WEC	22-Jul-19	N50°15'	W04°13'
Melnitz-3	EXVC039M	H5P1	Melnitz	Myovirus	141,372	37.6	224	4	WEC	01-Apr-19	N50°15'	W04°13'
		H5P1	*Methylo-philales* sp.	Gamma-proteo-bacteria	1,336,408	34.4	1382	38	WEC	Sep-18	N50°15'	W04°13'

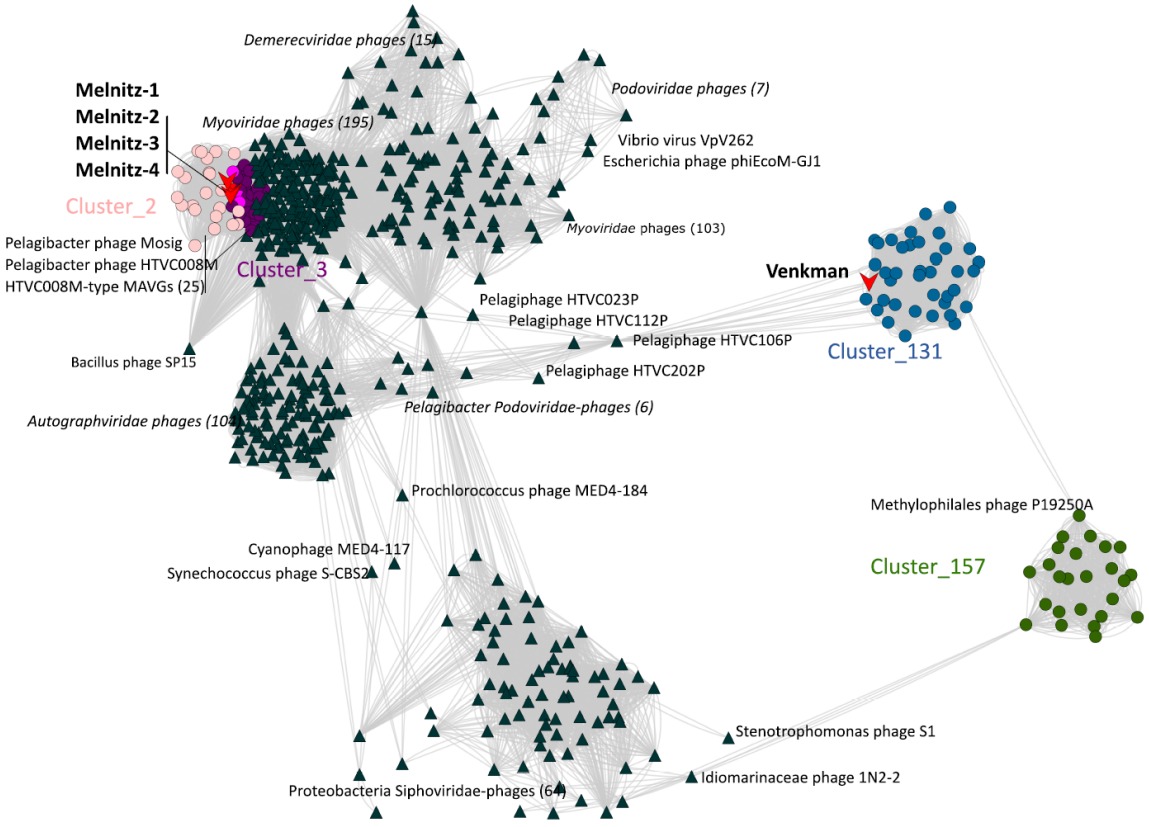


Figure 3.1. **Viral shared gene-content network of OM43 phages**, related bacteriophages from NCBI and related sequences from the Global Ocean Virome (GOV2.0). Nodes represent viral genomes; edges represent the similarity between phages based on shared gene content. NCBI reference genomes more than two neighbouring edges removed were excluded for clarity. Phage isolates are indicated with red arrows. Coloured circles represent genomes and virome contigs within the same cluster as OM43 phage isolate genomes. Nodes shared between Cluster_2 (light pink) and Cluster_3 (purple) are highlighted in hot pink.

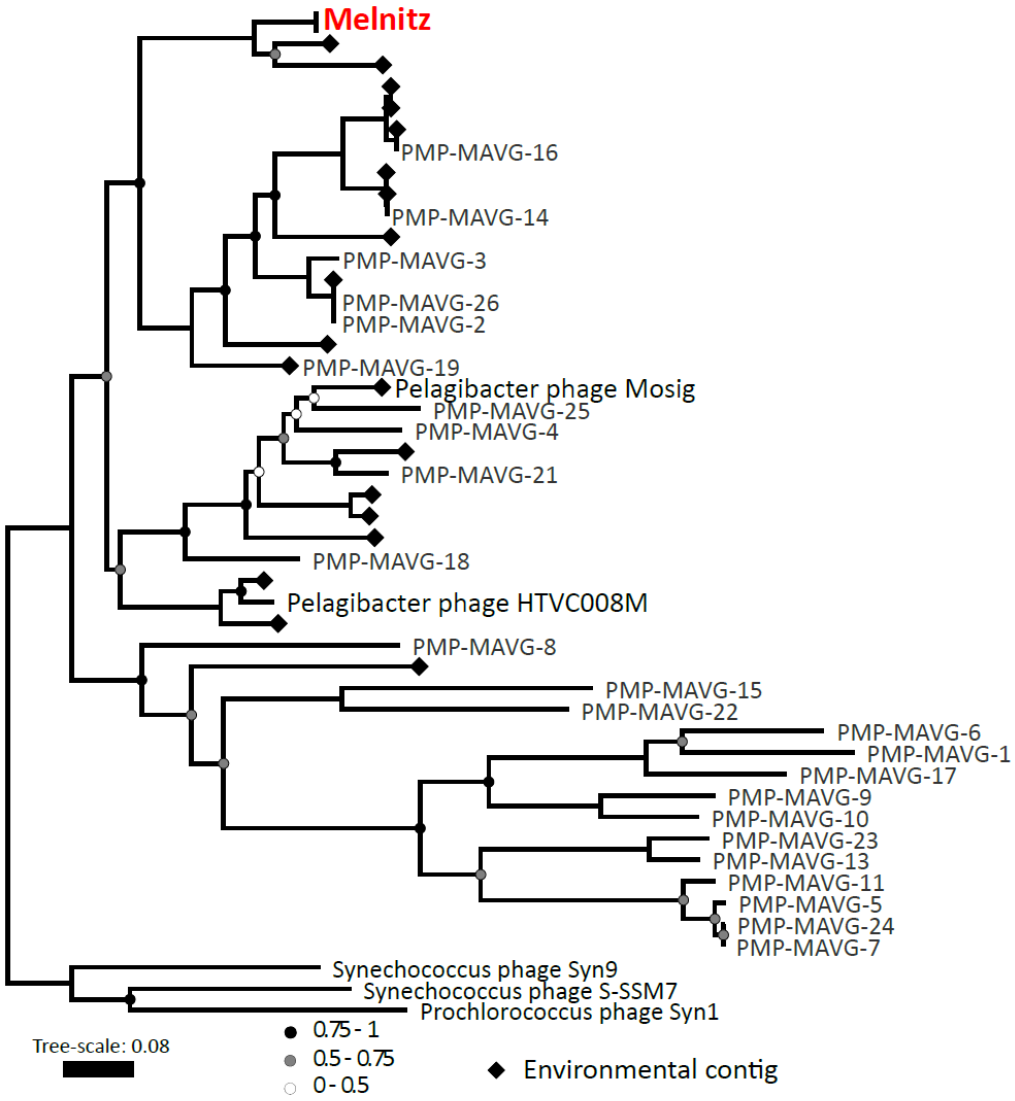


Figure 3.2. **Phylogenetic tree of metagenomic contigs and marine Melnitz-type myophages**. Neighbour-joining tree (500 bootstraps) based on four individually aligned and concatenated structural genes (capsid assembly, major capsid, sheath subtilisin, terminase large subunit) of genomes and contigs that were clustered with OM43 phage Melnitz, with the exception of phages infecting *Synechococcus* spp. that were used to root the tree. All four Melnitz-like isolates were included but the branch was collapsed for clarity. Branch support values of 1 are not shown. Leaves without labels indicate contigs from the Global Ocean Virome (GOV2) dataset (Gregory *et al.*, 2019), which were omitted for clarity.

It has been demonstrated that the marine biosphere maintains persistent bacterial Seed-Banks (Gibbons *et al.*, 2013), meaning there is a high probability that any given marine bacteria can be found in any marine ecosystem, albeit in extremely low abundance, awaiting favourable conditions for growth. Similarly, many viruses infecting globally distributed bacteria such as SAR11 are found in viromes from all oceans (Zhang *et al.*, 2020). Studies using “viral tagging” on cyanobacteria and their associated cyanophages suggest that at single sites,

viral populations in metagenomes comprise non-overlapping viral sequence space resembling “clouds”, which has benchmarked metagenomic analyses and subsequent interpretations (Deng *et al.*, 2012, 2014). In that model, single strains within a viral population encode genomic variations that enable the overall population to adapt more rapidly to environmental selection pressures. In the case of Melnitz, the isolation of the same virus with up to 100% nucleotide identity across the full genome on three separate occasions in the Western English Channel and BATS station in the Sargasso Sea (~5000 km distance between sites) suggests either low population-level variance, or that the isolation conditions used favour this strain. Nonetheless, the presence of cultivable Melnitz populations at both sites supports the virus Seed-Bank hypothesis where viral populations are conserved and persistent in the environment, being passively transported across oceans via global currents until favourable conditions select them for propagation (Breitbart and Rohwer, 2005; Brum *et al.*, 2016). The ‘environmental selection’ in the case of Melnitz would likely be the enrichment culturing used for isolation, providing enough suitable hosts and nutrients for viral replication. The possibility remains that the Melnitz population at the BATS site is maintained by a resident “seed” population of OM43, though OM43 is seldom reported in microbial communities at BATS.

3.3.3 Detecting ssDNA and RNA phage with marine viral metagenomics is challenging

Beside the different genome arrangement, most single-stranded DNA phages have on average smaller genome sizes (4.4 to 12.4 kb) compared to the better studied tailed double-stranded DNA phages (Up to 500 kb, reviewed in Székely & Breitbart 2016). Though not exclusive, ssDNA phages are also recognised for their lysogenic capabilities, which, for example in filamentous ssDNA phages, can cause chronic infections in bacteria where phage progeny is produced and released without cell death (Krupovic & Forterre 2015). Similar to dsDNA phages, ssDNA phages are cosmopolitan in global oceans, though species tend to have a narrower, regional distribution compared to e.g. ubiquitous pelagiphages, and there are no known ssDNA viruses of SAR11 or OM43 (Labonte & Suttle 2013, Labonte *et al.,* 2015). Viromes can be extremely useful tools to identify both single-stranded and double-stranded DNA phages, for example, Roux *et al.*,

(2019) have discovered 10,295 novel ssDNA sequences belonging to filamentous inoviruses. This demonstrated that Inoviruses are a significant component of the marine virosphere that is often overlooked. However, preparation of viromes, which are often used to study the distribution ranges, are subject to biases due to the differences in used methods (Kleiner *et al*., 2015). Though in recent years virome datasets have greatly enhanced the quantity of known viral sequences (Gregory *et al.*, 2019), they are usually designed for the study of dsDNA viruses and our knowledge about marine phages remains heavily biased towards double-stranded DNA viruses. A major challenge is that many available viromes introduce significant biases, causing a distorted view on ssDNA phages. This is because depending on the protocols used to produce viromes (Vega-Thurber *et al.*, 2009, Székely & Breitbart 2016), several common procedural steps are known to affect ssDNA differently compared to dsDNA, impacting for example the (1) Virus concentration, (2) Sample preparation, (3) Virion purification, (4) DNA amplification, and (5) library preparation:

(1) The initial concentration of viral particles in the sample impacts the recovery rate of different phage types, for example small circular ssDNA phages were overrepresented in waste water samples when concentrated, whereas large dsDNA genomes were increasingly recovered when the sample was diluted prior to preparing the virome (Brinkman *et al.*, 2018). (2) The make of DNA extraction kits or methods used to obtain genetic material can also distort the recovery rates for different phages, for example bead-beating kits used to lyse cells removed ssDNA phages entirely (Trubl *et al*., 2019). (3) Density gradient centrifugation using cesium chloride is a well-established purification approach to separate free DNA and viral DNA (Cremonese et al. 1969). Different phage groups have different buoyant densities depending on virion and genome size and composition; ssDNA phages are on average less dense compared to dsDNA phages (reviewed by Székely & Breitbart 2016). Density gradient centrifugation can therefore bias results for or against low-density ssDNA virions, depending on what phase densities are recovered. Whilst CeCl purification is commonly used in environments where heavy DNA contamination has to be expected, such as coral (Vega-Thurber *et al.,* 2009, Quistad *et al.,* 2016) or human associated viromes (Pride *et al.*, 2012), this approach is not often used in pelagic viromes, where DNases and RNases are frequently used (e.g. Gregory *et al.*, 2019), and

therefore less of a concern. Nonetheless, it is important to consider the limitations of different methods, for example for the comparison and re-interpretation of historic data. (4) Environmental samples with relatively low concentration of phage, such as oligotrophic water, often requires DNA amplification, such as multiple displacement amplification or linker amplification (Vega-Thurber *et al*., 2009). Linker amplification uses the ligation of dsDNA adapters and sheared dsDNA fragments, which is inefficient with ssDNA strands and is therefore likely to cause distorted representation of ssDNA phages within the virome (Duhaime *et al*., 2012). Multiple displacement amplification on the other hand is based on the polymerization of debranched DNA strands, and is known to disproportionately amplify circular ssDNA, thus leads to overestimations of the relative ssDNA phages abundance in the virome (Kim & Bae 2011). (5) Finally, library preparation can introduce further biases when using indexes, as they are commonly designed to tag onto dsDNA, not ssDNA, and therefore will not work properly with ssDNA phages (Solonenko *et al.*, 2013). Overall, investigating ssDNA phages with the help of currently available viromes is challenging. Whilst it is possible to detect ssDNA phages in some viromes, methodological inconsistencies between different studies can lead to over- or, more often, underestimates of ssDNA phages. This greatly impedes comparisons between viromes. It is important to recognise that conclusions drawn from most available virome data should be considered to encompass only the dsDNA phage fraction of the viral world, but likely does not accurately represent the true diversity of ssDNA phages.

Similar to DNA viruses, RNA viruses can have single- (positive or negative) strand-RNA or double-stranded genomes (Loeb & Zinder 1961, Vidaver *et al.,* 1973). Retroviruses have RNA genomes but use viral reverse transcriptases to create DNA that can integrate into the host genome. The few examples for ss/dsRNA phages that are available are non-tailed virions with small genomes that are roughly between 5 and 30 kb in size (Dion *et al.*, 2020). They are known to infect all branches of life, and reverse-transcribed whole-genome shotgun sequencing revealed large and diverse reservoirs of mostly unknown RNA viruses in the ocean (Culley *et al.,* 2006). However, most studies of RNA viruses are on eukaryotic pathogens, and these RNA viruses can be indeed highly abundant, possibly rivalling the abundance of dsDNA phages (Steward *et al*.,

2013). Yet, the first reported marine RNA bacteriophage, isolated on *Pseudomonas* in 1971, remained the only known marine RNA phage for decades (Hidaka 1971, Lang *et al.*, 2009), e.g. the first dsRNA phage for *Pseudomonas* was isolated in 2016 from wastewater (Thus not marine, Yang *et al.*, 2016). Unsurprisingly, as is the case for ssDNA viruses, no single- or double-stranded RNA phages infecting SAR11 or OM43 are known and it is clear that marine RNA phages have been neglected for decades, which might be due to methodological limitations. For instance, a known isolation method requires host cultures to be grown with 5-fluorodeoxyuridine to limit DNA synthesis, enriching for RNA phages over DNA phages (Davern 1963), but floxuridines are also known as antibacterial agents (Wambaugh *et al.*, 2017), which creates the obvious challenge of keeping host cultures alive for viral replication.

Using DNA viromes for the study of RNA phages is also challenging. RNA sequencing often requires additional steps to reverse transcribe RNA into cDNA (Culley *et al.*, 2006), but the fragility of RNA genomic material complicates RNA inclusive virome preparations, and the vast majority of protocols are likely to bias against RNA phages (Grasis 2018). For the preparation of marine metagenomes it is often considered desirable to remove free DNA and RNA from microbes or eukaryotes in the water sample, because contamination from their large genomes can skew the results in disfavour of relatively small viral genetic material (Vega-Thurber *et al.*, 2009). Added RNases and DNases can reduce contamination but can also degrade RNA viral particles (Acheson & Tamm 1969, Vega-Thurber *et al*., 2009), and therefore complicates the search for RNA phages in many of the publically available viromes.

Even though RNA phages have been largely ignored in marine virology in the past, the rapid advancements in genomics technologies have nonetheless led to greatly improved knowledge about marine RNA viruses and phages in recent years. In 2016, Krishnamurthy *et al.* identified 122 partial RNA phage genomes from public transcriptomes and RNA inclusive metagenomes, drastically increasing available genomes from the 16 complete ssRNA and dsRNA phage genomes that were deposited in NCBI before then. Another 1015 near-complete ssRNA phage genomes as well as 15,611 incomplete RNA phage sequences were identified in publically available meta-transcriptomes, expanding the number of available genomes by nearly an order of magnitude (Callanan *et al*.,

2020). Using complementary DNA viromes, Wolf *et al.* (2020) identified another 4,593 distinct RNA viruses, nearly doubling the number of known RNA viruses. Compiling RNA sequences from reverse-transcribed whole-genome shotgun libraries, (meta)-transcriptomes and other sources of marine origin allowed for the identification of 44,779 contigs encoding RNA-directed RNA polymerase, which is cited as a gene marker for the RNA virus kingdom *Orthornavirae* (Zayed *et al.*, 2022), which includes the two RNA phage families *Cystoviridae* and *Leviviridae* (reviewed by Callanan *et al.*, 2018, though *Liviviridae* may encompass two distinct family-level lineages (Wolf *et al*., 2018) and was proposed to be re-organised and renamed *Fiersviridae*). Zayed *et al.* (2022) proposed five new orthonaviran phyla, doubling the number to 10, and, together with the vast number of novel sequences, they create the first substantial framework to systematically study and integrate RNA viruses into marine ecological models. The rapid and impressive pace of discoveries over the past years will undoubtedly lead to more fascinating insights about RNA virus diversity and evolution, but until then the marine RNA phage ecology remains virtually unexplored. The foundations are in place now with improved viromes and metatranscriptomes, it is only a matter of time until this part of the marine "viral dark matter" will be illuminated. And, as has been the case for SAR11 and pelagiphages, advancing our understanding through (meta)-genomics will likely enable improved study efforts and allow for in depth exploration of RNA phage ecology.

3.3.4 *Genomic characterization of Melnitz-like* Methylophilales *phages.*

Genomes of the four Melnitz-like myophage isolates were between 141,372 - 141,552 bp in length, all with G+C content of 37.6 % (Table 3.1), similar to that of its host (34.4%). For each, 224 open reading frames (ORFs) were predicted, except Melnitz-2, which had two additional small ORFs of unknown function (226 ORFs total). Four additional tRNA sequences for all four genomes were identified, and functional annotation of ORFs suggest that all Melnitz strains encoded the same set of genes. Out of 224 ORFs, 143 (~63%) had unknown function (Figure 3.3A). The tail-assembly associated region in the Melnitz-4 genome had nine genes with altered length compared to the equivalents in the other three phages,

but where annotation was possible, the genes were predicted to have the same function. Predicted protein structures (PHYRE2) had low confidence and did not allow for a meaningful structural comparison (data not shown). Though structural variations in tail and receptor genes are often considered to be important factors for defining strain-level host ranges (Mahichi *et al.*, 2009; Casey *et al.*, 2015), all four Melnitz variants had identical host ranges when screened against a panel of other OM43 isolates from the WEC (*Methylophilales* spp. C6P1, D12P1 and H5P1; Supplementary Figure 2.1B) (Supplementary Figure 3.11). Melnitz possessed a set of structural genes typically associated with *E. coli* T4-type myophages, including T4-like baseplate, tail tubes, base plate wedges, tail fibres, virus neck, tail sheath stabilization, prohead core and capsid proteins. Melnitz encodes orthologs of the auxiliary metabolic genes (AMGs) *mazG* and *phoH*, which are involved in cellular phosphate starvation induced stress responses and are a common feature of phages from P-limited marine environments (Morris *et al.*, 2006; Bryan *et al.*, 2008; Sullivan *et al.*, 2010). Additionally, *hsp20* was identified, which together with *mazG* and *phoH* are considered part of the core genes in T4-like cyanomyophages (Sullivan *et al.*, 2010; Rihtman *et al.*, 2019). Both Melnitz and its host, *Methylophilales* sp. H5P1, encode for a type II DNA methylase; DNA methylase targets foreign DNA for cleavage by the restriction endonucleases (Murphy *et al.*, 2013). In T-even phages, DAM methylase (similar

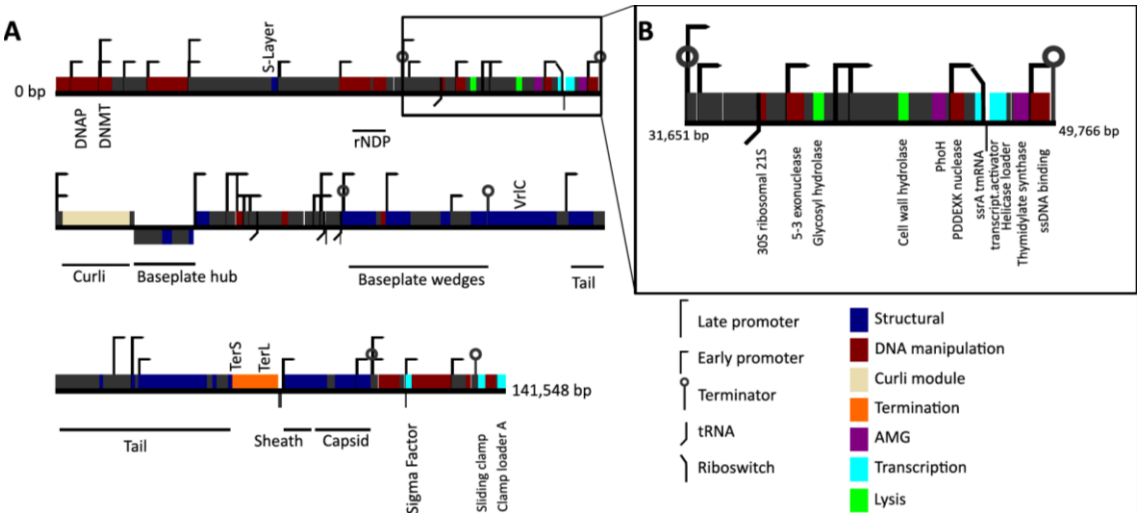


Figure 3.3. **Gene map showing identified genomic features of the OM43 phage Melnitz**. **A:** Gene map of the 141,548 bp genome of Melnitz contains 143 hypothetical ORFs (62 %) without known function (indicated dark grey); **B:** Section of the Melnitz genome between two terminator sequences that contains the *ssrA* gene, glutamine riboswitch and transcription co-activator gene for transcription-translation regulation.

to type II DNA methylase) protects phage DNA from restriction endonucleases through competitive inhibition (Kossykh *et al.*, 1995), thus DNA methylase in Melnitz could also be involved in protection from host restriction endonucleases during infection.

For DNA replication and metabolism, Melnitz encodes *nrdA/nrdE* and *nrdB/nrdF* genes that together form active ribonucleotide reductases for catalysing the synthesis of deoxyribonucleoside triphosphates (dNTPs) required for DNA synthesis (Jordan *et al.*, 1994). Additional DNA replication and manipulation associated genes were *polB*, *recA*, *dnaB,* as well as two DNA primases and a 2OG-Fe oxygenase. A *rpsU* gene (encoding the bacterial 30S ribosomal subunit 21S) was found downstream adjacent to a tRNA-Arg (tct). Virally encoded S21 genes were previously identified in the *Pelagibacter* phages HTVC008M (Zhao *et al.*, 2013) and Mosig (Appendix 2) as well as the OM43 phage Venkman (Figure 2.5A). It is thought that the 21S subunit in phages might be required for initiating polypeptide synthesis and mediation of mRNA binding (Mizuno *et al.*, 2019); the proximate tRNA sequence may support a translational and protein synthesis role of the viral ribosomal gene. Within the same region, Melnitz also encodes for a surface layer-type (S-layer) protein, previously associated with a strong protection mechanism against superinfection in bacteriophages of *Bacillus spp*. and *Pseudomonas aeruginosa* (Lewis and Yousten, 1988; Plaut *et al.*, 2014; Bondy-Denomy *et al.*, 2016). We postulate that phage-encoded S-layer proteins in Melnitz may be used to alter host cell surface receptors and thereby protect against superinfection from viral competitors.

Temporal regulation of T4-like phages primarily occurs at the transcriptional level and requires concomitant DNA replication and promoter/terminator sequences to organise into early, middle and late stages (Yang *et al.*, 2014). In Melnitz, a total of 46 promoters (31 early and 15 late) and six terminator sequences were identified – considerably fewer than the >124 promoter sequences found in T4 (Miller *et al.*, 2003; Hinton, 2010). Like other marine T4-like phages, such as those infecting streamlined cyanobacteria (Doron *et al.*, 2016), Melnitz lacked identifiable middle promotors. Reduced numbers of regulatory elements are a common feature of genomically streamlined bacteria such as SAR11 and OM43 (Giovannoni *et al.*, 2005). Therefore, we propose that temporal regulation of Melnitz is similarly simplified as a result of host streamlining. An alternative

hypothesis is that additional (middle-) promoter sequences in Melnitz are too divergent from known sequences for detection.

Like other T4-like phages, Melnitz did not encode for RNA polymerase, relying on host transcription machinery after infection for synthesis of phage proteins. Melnitz possessed the essential T4-like transcriptional genes: σ^{70}-like; gp45-like sliding clamp C terminal; clamp loader-A subunit; transcription coactivator; and a translational regulator protein. In T4, the DNA sliding clamp subunit of a DNA polymerase holoenzyme coordinates genome replication in late-stage infection, forming an initiation complex with a sigma-factor protein and host RNA polymerase (Twist *et al.*, 2011). To activate it, the sliding clamp is loaded by a clamp-loader DNA polymerase protein complex, allowing the complex to move along DNA strands (Nechaev *et al.*, 2004; Geiduschek and Kassavetis, 2010). As the same genes were found in Melnitz it is likely using them to form a T4-like protein complex for transcription and genome replication.

3.3.5 *Marine Melnitz-like phages may use glutamine riboswitches to regulate genome expression*

ssrA encodes for a tmRNA that in bacterial *trans*-translation together with *smpB* and ribosomal protein S1 is important to release ribosomes that have stalled during protein biosynthesis. These proteins are subsequently tagged and degraded (Withey and Friedman, 2003). In Melnitz, the *ssrA* gene encoding two-piece tmRNA was situated on the same operon and directly upstream of the transcription co-activator (Figure 3.3B). Host-encoded *ssrA* can be used as sites of integration for prophages in deep-sea *Sherwanella* isolates (Plaut *et al.*, 2014) and fragments of *ssrA* have previously been identified in prophage genomes (Bondy-Denomy *et al.*, 2016), most likely as a result of imprecise prophage excision that transfers host genetic material to the excising phage. Though we did not find integrases or other evidence that Melnitz is able to integrate into host genomes, a speculative temperate ancestral strain may explain the presence of a complete *ssrA* gene in Melnitz. A possible role for a complete viral tmRNA is to either help maintain the hijacked bacterial machinery, or the phage might use tmRNA to selectively tag and degrade host proteins to recycle amino acids for viral protein synthesis. We evaluated the frequency that *ssrA* genes were found

in 18,146 publicly available genomes in the phage genome database at millardlab.org (last updated on January 21, 2021), which includes all RefSeq genomes. Only 402 phages (2.3% of all available genomes) encoded 133 unique tmRNA genetic structures. Of these, 53 were suggested to be two-piece tmRNA by ARAGORN (Laslett and Canback, 2004), of which 50 were encoded by phages isolated from marine or aquatic samples. In contrast, only 2.5% of non-permuted tmRNAs were marine, suggesting that viral permuted tmRNA is a feature more prevalent in marine and aquatic phages compared to phages from other environments. Of the 47 contigs (27 were confirmed complete) that were clustered with Melnitz based on shared gene content (Figure 3.1), 17 had complete *ssrA* genes, indicating that *ssrA* encoded tmRNA is a common, but not defining feature of this viral group.

The 3' domain of the *ssrA* gene in Melnitz encodes a predicted glutamine riboswitch that resembles *glnA* RNA motifs found in cyanobacteria and other marine bacteria (Weinberg *et al.*, 2010). Riboswitches are a common regulatory mechanism in streamlined marine bacteria due to their low metabolic maintenance cost compared to protein-encoded promotors and repressors (Giovannoni *et al.*, 2005). Like tmRNAs, riboswitches are a rare feature in phages. A phage-encoded riboswitch putatively controlling regulation of *psbA* was previously identified in a cyanophage (Geiduschek and Kassavetis, 2010). Of the 133 isolate phage genomes encoding tmRNAs only two phages other than Melnitz possessed a riboswitch (both glutamine): *Pelagibacter* phage Mosig and *Prochlorococcus* phage AG-345-P14 (Berube *et al.*, 2018). Additional riboswitches were identified in metagenomic assembled virome contigs. In ten metagenomically assembled viral genomes of pelagimyophages thought to infect SAR11 (PMP-MAVGs) and five GOV2 contigs related to Melnitz, only one of these contigs had both tmRNA and a riboswitch; eleven PMP-MAVGs encoded neither. This suggests that: (1) glutamine riboswitches are a common but not defining feature in Melnitz-like marine phages; (2) phages of that group use riboswitches or tmRNA, but rarely both, with the only two identified examples occurring in isolated phages, not metagenomically derived genomes. Curiously, the bacterial OM43 host of Melnitz (H5P1) does not encode a glutamine riboswitch - only one cobalamin riboswitch was found located upstream of the Vitamin B12 transporter gene *btuB*. Other members of the *Methylophilaceae*

(OM43 strains HTCC2181, KB13, MBRSH7 and *Methylopumilus planktonicus, M. rimovensis* and *M. turicensis*) also lack glutamine riboswitches. In cyanobacteria, the glutamine riboswitches were previously found to regulate the glutamine synthase *glnA* and are strongly associated with nitrogen limitation (Klähn *et al.*, 2018). Phages infecting *Synechococcus* have been shown to use extracellular nitrogen for phage protein synthesis (Waldbauer *et al.*, 2019), which might indicate a similar role for phage-encoded glutamine riboswitches in Melnitz. However, neither OM43 nor Melnitz encode the glutamine synthase *glnA* or homologous genes required for glutamine synthesis. This suggests that Melnitz-like phages use viral riboswitches to regulate their own genes rather than hijacking the cellular machinery.

3.3.6 *Phage encoded curli operons may be involved in regulating cell lysis*

Curli genes are typically associated with the bacterial production of amyloid fibres that are part of biofilm formation, where the CsgGF pore spans the outer membrane as part of a Type VIII secretion system, allowing for the secretion of CsgA and CsgB that assemble into extracellular amyloid fibres (Barnhart and Chapman, 2006). In complete curli modules, CsgG and CsgF form an 18-mer heterodimer comprising nine subunits with 1:1 stoichiometry between CsgF and

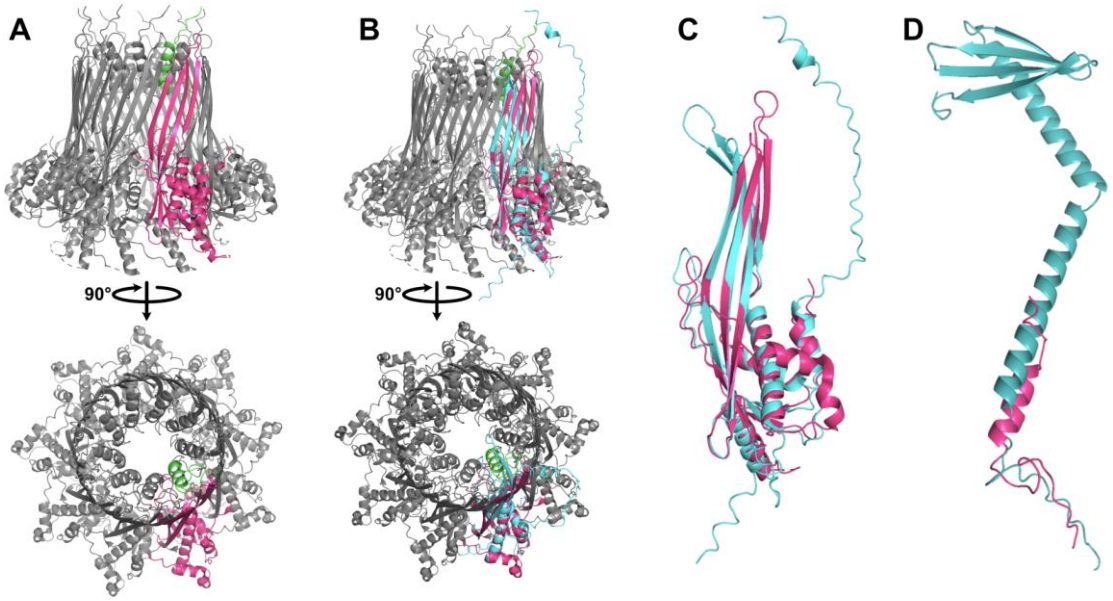


Figure 3.4. **Structural prediction of CsgGF complex encoded by Melnitz**. **A:** Predicted structure of CsgGF complex in *E. coli* (PDB model 6l7a) comprises a hetero-18-mer with 1:1 stoichiometry of CsgG (pink) and CsgF (green), forming a pore in the outer membrane with two constrictions - one provided by CsgG at the base of the barrel

and one provided by CsgF at the neck of the barrel. **B:** Structural prediction of Melnitz encoded CsgG using AlphaFold2 (teal) showed structural conservation with CsgG from *E. coli* in the periplasmic α-helices and β-barrel structure. **C:** Expanded view of the structural alignment of Melnitz CsgG with *E. coli* CsgG shows a putative narrowing of the pore at the top of the barrel, matching the pore diameter at the top of the barrel in CsgGF in E. coli. **D:** Alignment of predicted structure of Melnitz-encoded CsgF (teal) to that of CsgF in *E. coli* (pink) showed low structural similarity, with Melnitz-encoded CsgF comprising two alpha-helices and a beta sheet. Alignment indicated that the additional structures of Melnitz-encoded CsgF extend out of the CsgG pore, with unknown function.

CsgG. The structure of the pore is dictated by CsgG, which forms a channel ~12.9Å in diameter. CsgF forms a secondary channel ~14.8Å in diameter at the neck of the beta barrel (Figure 3.4A, Supplementary Figure 3.6 A and B) and assists in excretion of the amyloid fibre. Like previously described pelagimyophages (Zaragoza-Solas *et al.*, 2020), Melnitz possesses *csgF* and *csgG,* but lacks the genes for amyloid fibre production - *csgA* and *csgB* (Figure 3.3). Zaragoza-Solas et al. speculated that phage-encoded curli pores may allow for the uptake of macromolecules, or together with unidentified homologues of missing curli genes form a complete, functional curli operon producing amyloid fibres for “sibling capture” of proximate host cells (Zaragoza-Solas *et al.*, 2020). However, similar to results in pelagimyophages, we did not find evidence for a complete phage encoded curli biogenesis pathway in Melnitz, nor does the bacterial H5P1 host encode for any curli associated genes, suggesting that these genes were acquired by an ancestral strain of Melnitz and pelagimyophages before host transition to OM43 and SAR11, respectively.

Structural homology modelling of the Melnitz-encoded CsgGF portal with SwissModel (Waterhouse *et al.*, 2018) yielded low structural similarity to known CsgGF structures (GMQE <0.5 with Q-MEAN scores < -3 and z-scores >>2 outside normalised Q-mean score distributions from a non-redundant set of structures from PDB). Structural similarity was greatest in regions 238-258 of CsgG where Q-MEAN scores were ~0.7, with conserved regions in the periplasmic-facing α-helices of CsgG. Melnitz-encoded CsgF showed poor structural similarity to any models available within Swiss-Model. Therefore, we used AlphaFold2 (Jumper *et al.*, 2021) to predict the structure of phage-encoded CsgG and CsgF *de novo* and compared predicted structures to known CsgGF structures. The predicted structure of Melnitz-encoded CsgG showed close structural similarity to known CsgG, with two notable exceptions: (1) a narrowing

of the barrel at the neck; (2) an extension of the α-helix outside the barrel (Figure 3.4B and C). In contrast there was little structural similarity between Melnitz-encoded CsgF and the structure of CsgF in *E. coli*, barring a shared α-helix domain. Phage-encoded CsgF comprised two long α-helices ending in a β-sheet (Figure 3.4D). Similar modelling of CsgG and CsgF encoded by pelagiphage HTVC008M also contained these features (Supplementary Figure 3.7 A and B, respectively). Assuming the pore orientation in an infected OM43 cell is the same as that of CsgG in *E. coli*, the additional domains on phage-encoded CsgF would either place the terminal β-sheet inside the barrel of CsgG (unlikely due to steric hindrance) or the extended structure would point outwards into the extracellular milieu, in an unusual conformation with unknown function. It is more likely that in this configuration of CsgG, phage-encoded CsgG and CsgF have diverged and evolved independently. Therefore they may no longer form a heterodimer, with CsgG retaining the function as a pore, and CsgF evolving independently to provide an alternative, unknown function. Indeed, the narrowing of the barrel neck in Melnitz-encoded CsgG creates a channel 14.7Å in diameter, matching that provided by CsgF in *E. coli* (Supplementary figure 3.6 B and C). One possibility is that phage-encoded CsgG is functionally analogous to pinholins, used by phage λ to regulate cell lysis. Pinholins form channels ~15Å in diameter in the inner membrane, resulting in rapid membrane depolarization and subsequent activation of membrane-bound lysins (Pang *et al.*, 2009).

An alternative hypothesis is that in Melnitz, CsgGF forms a heterodimer in the outer membrane similar to CsgGF in *E. coli*, but that the structure is inverted so that the extended α-helices of CsgF point into the periplasm. In this conformation, the extension from Melnitz-encoded CsgG (Figure 3.4C) can interact to stabilise the elongated hinged α-helices of CsgF. The C-terminal β-sheets of CsgF could then form a second channel beneath the CsgG barrel (Supplementary Figure 3.8 A). In this conformation, CsgGF is structurally similar to a secretin, a large protein superfamily used for macromolecule transport across the outer membrane such as DNA for natural competence (e.g. PilQ within the type IV secretion system in *Vibrio cholerae*) or extrusion of filamentous phages during chronic infection (e.g. pIV in bacteriophages Ff) (Conners *et al.*, 2021). Like CsgGF in an inverted conformation, PilQ comprises a β-barrel, with a secondary pore below, attached by two hinged α-helices. The inner surface of PilQ is negatively charged to repel

the negatively charged backbone of DNA and assist in transportation across the membrane. In contrast, the inner surface of the CsgGF channel is positively charged and narrower (14.7Å compared to 20.6Å in PilQ), making a role in DNA transport across the membrane unlikely. While the function of phage-encoded CsgGF is not yet clear, the additional domains of CsgF and the lack of other curli genes within either the phage or the host suggest that it is a novel pore structure whose function has diverged from ancestral CsgGF and would be a worthy target for future structural resolution.

Whether the CsgGF complex in Melnitz acts as a pinholin or a secretin, gene synteny supports a putative role in regulation of the timing of cell lysis. First, genes *csgGF* are located immediately downstream of thymidylate synthase *thyX* and *ssrA* and transcription co-activator genes (Figure 3.3). Overexpression of thymidylate synthase in the D29 phage infecting *Mycobacteria tuberculosis* results in delayed lysis and higher phage yields (Ghosh *et al.*, 2020). In phage T4, postponing lysis is used to delay the release of viral progeny until conditions are favourable, thereby maximising virion production before viral release and successful replication after (Catalão *et al.*, 2013). In the T4-like Melnitz, *thyX* is therefore likely to be involved in postponing lysis as well. Lytic control may putatively be related to the upstream glutamine riboswitch. H5P1 lacks glutamine synthesis pathways, but has a complete peptidoglycan biosynthesis pathway necessary to produce peptidoglycans from glutamine. Therefore, both the H5P1 host and Melnitz phage are restricted to using glutamine for protein and peptidoglycan synthesis only. In phage λ, the depletion of peptidoglycan precursors can trigger lysis through activation of a spanin complex (Rajaure *et al.*, 2015). We speculate that during the course of the infection cycle, glutamine levels are kept low through incorporation into viral proteins. When protein synthesis is complete, intracellular glutamine levels increase, activating the glutamine riboswitch. This in turn could trigger an opening of the curli portal, which may result in rapid depolarization of the membrane, triggering activation of lysins and rapid cell lysis.

Structural homology modelling (SwissModel) of a Melnitz-encoded enzyme (gp67), putatively annotated as a glycosyl hydrolase, revealed structural similarity (38% sequence identity at 98% coverage, with a global model quality estimate (GMQE) of 0.82) to autolysin SagA encoded by *Brucella abortus* (PDB model

7DNP). SagA acts to generate localised gaps in the peptidoglycan layer for assembly of type IV secretion systems (Del Giudice *et al.*, 2013). Therefore, it is likely that gp67 and CsgGF work in concert in Melnitz. Peptides involved in binding peptidoglycan at the active site (Glu17, Asp26, Thr31) were conserved between Melnitz gp67 and SagA (Supplementary Figure 3.9) (Hyun *et al.*, 2021). Outside of this active site, structural similarity to T4 endolysin (PDB model 256l.1) was low (13% identity over 51% coverage; GMQE: 0.18). Melnitz lacked any other lysin-like genes. Like the T4-encoded endolysin *e* (Miller *et al.*, 2003), gp67 is under control of an early transcription promotor (Figure 3.3B). We therefore propose that the gp67 gene in Melnitz potentially serves two functions: (1) as an endolysin for degradation of cellular peptidoglycans during lysis; or (2) as an autolysin to enable assembly of the CsgGF or CsgG-only portal protein. Phages infecting *Streptococcus pneumoniae* upregulate host-encoded autolysins alongside phage-encoded lysins during late-stage infection to accelerate cell lysis (Frias *et al.*, 2009). The close structural match between gp67 and SagA suggests that an ancestor of Melnitz acquired a host-encoded autolysin as an alternative lysin during its evolutionary history. Whether this autolysin-derived phage lysin could be activated by membrane depolarization through CsgG is unknown.

3.3.7 *Melnitz-encoded genes suggest a possible host transition event from SAR11 to OM43*

Phage host ranges are largely determined by interactions between host receptor proteins and phage structural proteins such as tail fibres that enable the phage to adsorb and inject its genetic material. Mutation in phage proteins, either through point mutations or recombination during co-infection can result in host range expansion or transition (Burrowes *et al.*, 2019). Host range expansion or transition within species boundaries is more common, but rare transition events between hosts from different genera can occur through mutations in tail fibres (Tétart *et al.*, 1996). Here, two separate lines of evidence converge to suggest that Melnitz underwent a recent host transition from SAR11 to OM43. First, *ssrA* encoded by Melnitz was identified to be more closely related to homologs within the alphaproteobacterial lineage. Two-piece circularly permuted tmRNA is common in three major lineages: Alphaproteobacteria, Betaproteobacteria (now

part of Gammaproteobacteria (Parks *et al.*, 2018)) and Cyanobacteria (Keiler *et al.*, 2000; Sharkady and Williams, 2004). The phylogenetic evidence for alphaproteobacterial *ssrA* in Melnitz rather than *ssrA* that matches the gammaproteobacterial lineage of its OM43 host suggests that this gene was acquired from an Alphaproteobacteria (Figure 3.5). Second, two of the four tRNAs encoded by Melnitz were more closely related to tRNAs encoded by pelagimyophages than those of their host. Melnitz encodes two versions of tRNA-Arg(TCT) – one most closely related to that of its host H5P1, and one closely related to *Pelagibacter* phage Mosig. Phage-encoded tRNAs enable large phages to sustain translation as the host machinery is degraded to fuel phage synthesis (Yang *et al.*, 2021). Phage-encoded tRNAs also enable phages to optimise protein synthesis in a host with different codon usage and thus serve as both a marker of increased host range and evolution through different hosts (Santos *et al.*, 2011). Indeed, tRNAs are often used for computational host-prediction of phages due to high sequence conservation of the gene between host and viral forms (Edwards *et al.*, 2016; Coutinho *et al.*, 2021). We postulated

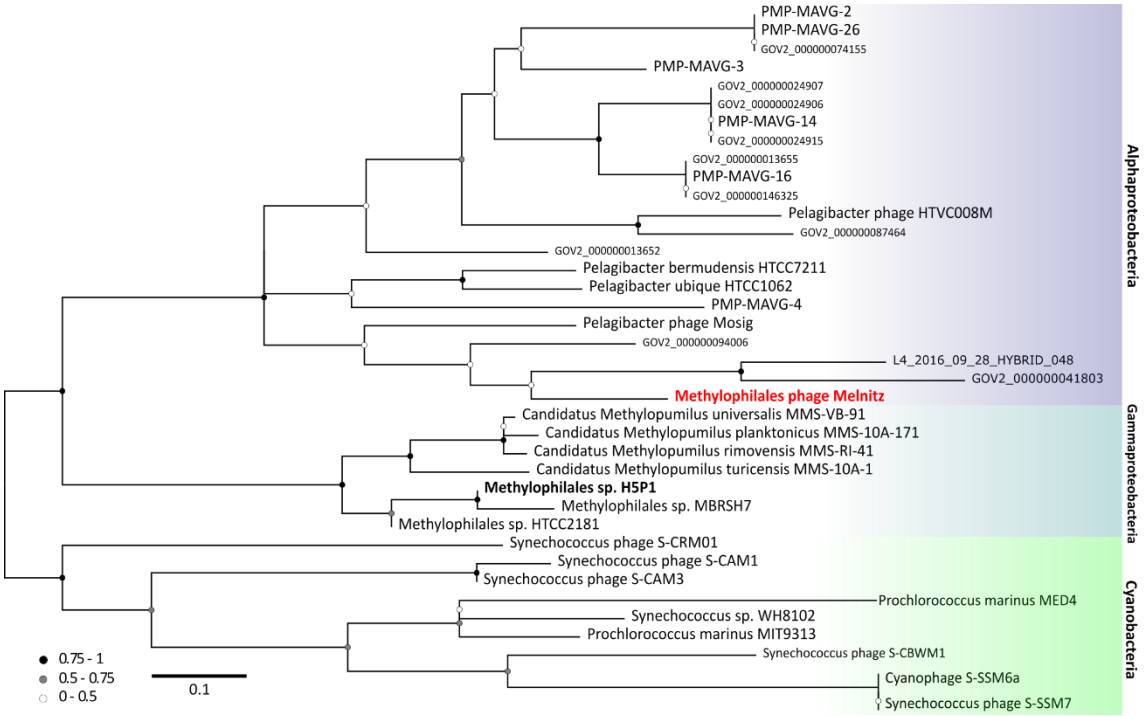


Figure 3.5. **Phylogeny of tmRNA genes in major marine lineages**. Neighbour-joining tree (100 bootstraps) of tmRNA genes found in marine phages and host lineages (not exhaustive) suggests that three the three known major lineages between Cyanobacteria, Gammaproteobacteria and Alphaproteobacteria are shared with their associated phages, except for OM43 phage Melnitz (infecting H5P1 on the Gammaproteobacterial branch) which has a tmRNA gene more closely related to genes found in Alphaproteobacteria and their phages.

that the host-range of Melnitz would be evident in the four tRNAs found within its genome and reflect potential hosts in the OM43 clade. Alternatively, if a recent host transition occurred, as *ssrA* phylogeny suggests, tRNAs would be similar to those found in the SAR11 bacteria and viruses. A search for tRNA genes in the genomes of isolated *Pelagibacter* phages and PMP-MAVGs (47 genomes) and isolated phages infecting OM43 (three genomes) identified tRNAs in seven *Pelagibacter* phage genomes. Melnitz was the only phage known to infect OM43 that encoded tRNAs, with four tRNAs in total (Figure 3.6). Sequences encoding tRNAs were aligned using all-vs-all BLASTN with a minimum expect-value of 1×10^{-5}. Two out of four tRNAs in Melnitz were tRNA-Arg(TCT), the first aligning most closely with its H5P1 host. The second tRNA-Arg best aligned with the tRNA gene found in the *Pelagibacter* phage Mosig. The alignment of that particular tRNAIts alignment in both phages to bacterial tRNA matched OM43 strains H5P1 and HTCC2181, respectively (>90% identity). The third tRNA-Leu found in Melnitz aligned with PMP-MAVG-17, previously classified as *Pelagibacter* phage (Zaragoza-Solas *et al.*, 2020), but only had 45% nucleotide identity. The fourth tRNA-Trp (CCA) found in Melnitz did not align (e-values > 1×10^{-5}) with any tRNA found in OM43, SAR11 or any of their respective phages. Using tRNA as host

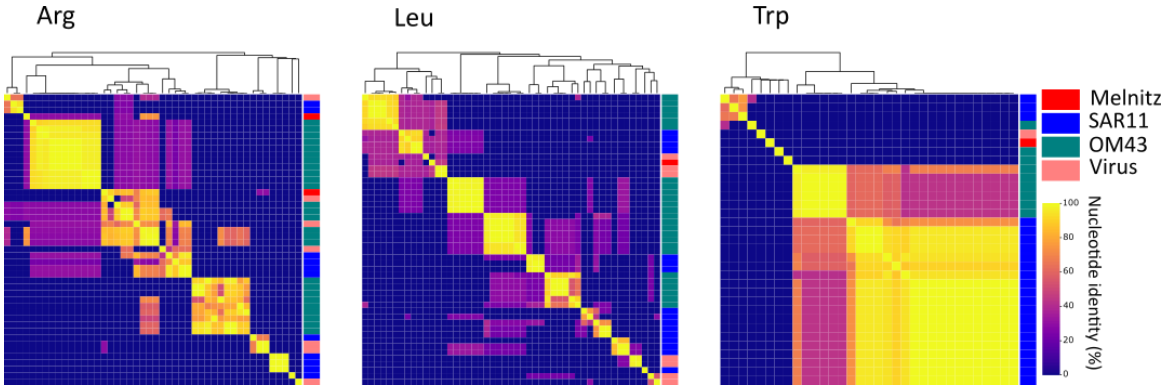


Figure 3.6. **Alignment of tRNA genes found in Melnitz, SAR11 and OM43 lineages**. Heatmaps and dendrograms (Euclidian similarity matrices) based on similarity between alignments for arginine (Arg), leucine (Leu) and tryptophan (Trp) tRNA genes found in OM43 phage Melnitz, OM43 and SAR11 as well as tRNA derived from isolated SAR11 and OM43 phages. List of information for individual alignments available in Appendix 7.

indication in Melnitz therefore reflects its OM43 host, but also indicates genetic exchange with SAR11 virus-host systems. Furthermore, the related *Pelagibacter* phage Mosig possessed tRNAs matching OM43 and Melnitz as well as a second tRNA aligning with its SAR11 host (93%). This may suggest either a relatively recent genetic exchange between OM43 and SAR11 virus-host systems, and/or

a surprisingly broad host-range for *Pelagibacter* phage Mosig and *Methylophilales* phage Melnitz. To assess the breadth of host ranges, we challenged the SAR11 strains HTCC7211 and HTCC1062 and OM43 strains H5P1, D12P1 and C6P1 with the phages Melnitz and Mosig, but found no evidence that either phage is able to replicate in cells beyond host class boundaries as observed via cell lysis (Supplementary Figure 3.10). Melnitz only infected D12P1 and H5P1, but not C6P1 or any SAR11, whereas Mosig only caused lysis in the *Pelagibacter ubique* HTCC1062. Though the limited number of available bacterial strains could have missed potential permissive hosts in a different subclade, there is no evidence that non-canonical matching tRNA between SAR11/OM43 and their phages is due to a speculative broad host range in these systems, and instead supports the phylogenetic evidence we presented for a recent host transition event.

Phage host-range expansion and transition within closely related (i.e. between strains) is a common feature of phage evolution, and has been shown to occur *in vitro* via homologous recombination between co-infecting phages (Riede, 1986; Burrowes *et al.*, 2019). In contrast, expansion and transition between distantly related taxa is rare but has been previously observed *in vitro*. Naturally occurring, low-abundance mutants of T4 transitioned from *E. coli* to *Yersinia pseudotuberculosis* through modification of tail fibre tips in gene 37, yielding variants that could infect both hosts and others that could only infect *Y. pseudotuberculosis* (Tétart *et al.*, 1996). To our knowledge, host transition between bacteria belonging to different classes, as proposed here, has not previously been observed, although host-prediction of viral contigs from metagenomes has identified rare phages (115 out of 3,687) possibly able to infect multiple classes (de Jonge *et al.*, 2019). We propose that two properties of SAR11 and OM43 increase the likelihood of such an event in natural communities: (1) Both SAR11 hosts and their associated phages possess extraordinarily large, globally ubiquitous effective population sizes, making even extreme rare events likely. This study has shown that phages infecting OM43 can also be abundant at higher latitudes, further increasing the likelihood of rare strain variants to occur; (2) Both OM43 and SAR11 hosts share highly streamlined genomes with elevated levels of auxotrophy, minimum regulation and similar

G+C content, shaped by selection pressure to maximise replication on minimum resources in nutrient-limited marine environments (Giovannoni *et al.*, 2008; Giovannoni, 2017). Both fastidious hosts can be cultured on identical minimal medium, as long as additional methanol is provided as a carbon source for the methylotrophic OM43 (see Section S2.1.3). Therefore, the phenotypic difference between OM43 and SAR11 might be smaller than suggested by their taxonomic classification. Given viral selection occurs at the phenotypic level, we propose the possibility of a rare event where a mutant of a T4-like phage infecting SAR11 was able to successfully adsorb and inject its genome into an OM43 host, possibly co-infected with a methylophage that enabled a host-transition event through homologous recombination. Once a transition event occurred, the phage likely rapidly evolved specialism on the new host, losing the ability to infect the original host, as demonstrated in host-range experiments. Compared to copiotrophic, *R*-strategist microbial taxa such as *E. coli* and *Vibrio* spp., very little is known about the processes governing viral host-range and co-evolution in *K*-strategist bacteria such as the genomically streamlined taxa that dominate oligotrophic oceans. The early evidence shown here may suggest that phages can transition to new hosts from distantly related taxa in natural communities. Such events would explain in part the diverse host ranges predicted amongst viruses within gene-sharing clusters (Bolduc *et al.*, 2017; Bin Jang *et al.*, 2019).

3.4 Conclusion

In this study, we provide evidence that supports a putative inter-class host transition event between two important clades of streamlined marine heterotrophs and expand our knowledge about the dynamics and characteristics in genomically streamlined heterotrophic virus-host systems. We isolated four near-identical strains of the new myophage Melnitz from subtropical and temperate marine provinces infecting the important methylotrophic OM43 clade, which we showed to be closely related to myophages infecting the abundant SAR11 clade. The analysis of metagenomic datasets provides evidence that this phage group is ubiquitous in global oceans despite relatively low overall abundance, supporting the viral Seed-Bank hypothesis. Our genomic analysis of

Melnitz revealed an incomplete curli module similar to reported curli pores in pelagimyophages, representing a rare and intriguing protein dimer that is absent in their respective host clades. We propose that these virally encoded curli pores may have been repurposed as a functional analogue for the regulation of viral lysis. We also identified an *ssrA* gene encoding for a complete viral tmRNA protein controlled by a glutamine riboswitch, showing that virus-host interactions can be regulated through riboswitches reflecting the extensive use of riboswitches in streamlined marine heterotrophs. Further phylogenetic analysis showed that the *ssrA* gene is related to the alphaproteobacterial SAR11 lineage, not the gammaproteobacterial OM43 lineage, providing evidence for host transition events in natural marine microbial communities, which was supported by the alignment of viral and bacterial tRNA genes of both lineages. These findings support the conclusion that in heterotrophic streamlined virus-host systems evolution of viral diversity is likely to be driven by host transition and expansion between closely related phages infecting hosts across broad taxonomic groups, likely increasing mosaicism and genetic exchange.

3.5 Materials & Methods

3.5.1 *OM43 strain, media, and growth conditions*

The OM43 strains *Methylophilales* H5P1 and D12P1 were isolated previously from surface water from the Western English Channel (S2.1.1). Continuous cultures were grown using artificial seawater-based artificial seawater medium (ASM1) (Carini *et al.*, 2013) amended with 1 mM NH_4Cl, 10 µM KH_2PO_4, 1 µM $FeCl_3$, 100 µM pyruvate, 25 µM glycine, 25 µM methionine as well as 1 nM each of 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP), pantothenate, biotin, pyrroloquinoline quinone (PQQ) and B12. Additional 1 mM methanol and 5 µL of an amino acid mix (MEM Amino Acids (50x) solution, Sigma-Aldrich) were added per 100 mL medium. Bacterial stocks were grown in 50 mL acid-washed polycarbonate flasks at 15 °C without shaking and 12-hour light-dark cycles.

3.5.2 *Water sources and phage isolation*

Water samples were collected from two different stations: Station L4 in the Western English Channel (WEC; 50°15'N; 4°13'W) and the Bermuda Atlantic Time Series sampling station (BATS; 31°50'N; 64°10'W) in the Sargasso Sea. For each sample we used Niskin bottles mounted on a CTD rosette to collect 2 L of seawater at 5m depth (Table 3.1) into acid-washed, sterile, polycarbonate bottles. To obtain a cell-free fraction, water samples were filtered sequentially through a 142 mm Whatman GF/D filter (2.7 µm pore size), a 142 mm 0.2 µm pore polycarbonate filter (Merck Millipore) and a 142 mm 0.1 µm pore polycarbonate filter (Merck Millipore) using a peristaltic pump. The cell-free viral fraction was concentrated to about 50 mL with tangential flow filtration (50R VivaFlow, Sartorius Lab instruments, Goettingen, Germany) and used as inoculum in a multi-step viral enrichment experiment followed by Dilution-to-Extinction purification as described (Figure 2.1). Briefly, viral inoculum (10% v/v) was added to 96-well Teflon plates (Radleys, UK) with exponentially growing host cultures. After a 1 to 2 weeks incubation period, cells and cellular debris were removed with 0.1 µm syringe PVDF filters. The filtrate was added as viral inoculum to another 96-well Teflon plate with exponentially growing host culture. This process was repeated until viral infection was detected by observing cell death using flow cytometry. Phages were purified by Dilution-to-Extinction methods (A detailed step-by-step protocol is available here: 10.17504/protocols.io.c36yrd).

3.5.3 *Assessment of viral host ranges*

An acid-washed (10% hydrochloric acid) 48-well Teflon plate was prepared with 2 mL per well of ASM1 amended with 1 mM methanol and 5 µL of an amino acid mix (MEM Amino Acids (50x) solution, Sigma-Aldrich). Wells were inoculated to $1x10^{6}$ cells·mL^{-1} with *Pelagibacter* hosts HTCC1062 or HTCC7211, or OM43 hosts *Methylophilales spp*. C6P1, D12P1 and H5P1 cultures in replicates of three for each of the Mosig and Melnitz phages, plus one set of no virus controls for each host. The wells marked for viral infections were infected with 200 µL of viral culture. The plate was then incubated at 15 °C and monitored daily using flow cytometry for ~2 weeks. Successful infections were identified observing cell lysis in virus amended wells compared to no-virus controls.

3.5.4 *Phage DNA preparation, genome sequencing and annotation*

For each viral isolate, 50 mL OM43 host cultures were grown in 250 mL acid-washed, polycarbonate flasks and infected at a cell density of about 1-5x 10^6 cells·mL^{-1} with 10% v/v viral inoculum. Infected cultures were incubated at 15 °C for 7 to 14 days after which the lysate was transferred to 50 mL Falcon tubes. Larger cellular debris was removed by centrifugation (GSA rotor, Thermo Scientific 75007588) at 8,500 rpm/ 10,015 x g for one hour. Supernatant was filtered through pore-size 0.1 µm PVDF syringe filter membranes to remove any remaining smaller cellular debris. Phage particles were precipitated using a PEG8000/NaCl flocculation approach (Solonenko, 2016). Briefly, 50 mL lysate was amended with 5g PEG8000 and 3.3 g NaCl (Sigma), shaken until dissolved and incubated on ice overnight. Precipitated phages were pelleted by centrifugation at 8,500 rpm / 10,015 x g for one hour at 4 °C. After discarding the supernatant, phage particles were resuspended by rinsing the bottom of the tubes twice with 1 mL SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-Cl). DNA from the suspended phages was extracted with the Wizard DNA Clean-Up system (Promega) following manufacturer's instructions, using pre-heated 60 °C PCR-grade nuclease free water for elution.

Nextera XT DNA libraries were prepared and sequenced by the Exeter Sequencing Service (Illumina paired end [2 x 250 bp], NovaSeq S Prime [SP], targeting 30-fold coverage). Reads were quality controlled, trimmed and error corrected with the tadpole function (default settings) within BBMap v38.22 ((Bushnell *et al.*, 2017) available at https://sourceforge.net/projects/bbmap/). Contigs were assembled using SPAdes v.3.13 with default settings (Bankevich *et al.*, 2012). Viral contigs were confirmed with VirSorter v1.05 (Roux *et al.*, 2015) (categories 1 or 2, > 15kbp). Quality and completeness of contigs as well as terminal repeats were evaluated using CheckV v0.4.0 with standard settings (Nayfach *et al.*, 2020). Genes were called with phanotate v2019.08.09 (McNair *et al.*, 2019) and imported into DNA Master for manual curation (Salisbury and Tsourkas, 2019) using additional gene calls made by GenMarkS2 (Lomsadze *et al.*, 2018), GenMark.heuristic (Besemer and Borodovsky, 1999), Prodigal v2.6.3 (Hyatt *et al.*, 2010) and Prokka v1.14.6 (Seemann, 2014). ORFs were functionally

annotated using BLASTp against NCBI's non-redundant protein sequences (Pruitt *et al.*, 2007), phmmer v2.41.1 against Pfam (Finn *et al.*, 2014) and SWISS PROT (Bairoch and Apweiler, 2000). All genes called were listed and compared using a scoring system evaluating length and overlap of ORFs as well as quality of annotation (Salisbury and Tsourkas, 2019); tRNA and tmRNA were identified with tRNAScan-SE v2.0 (Lowe and Chan, 2016) and ARAGORN v1.2.38 (Laslett and Canback, 2004). Genomes were scanned for riboswitches using the web application of Riboswitch Scanner (Singh *et al.*, 2009; Mukherjee and Sengupta, 2016). FindTerm (energy score < -11) and BPROM (LDF > 2.75) from the Fgenesb_annotator pipeline (Solovyev and Salamov, 2011) were used to predict promoter and terminator sequences, using default parameters. The σ^{70} promoters predicted this way were considered early promoters. Known T4-like late promoter sequence 5'-TATAAAT-3' (Miller *et al.*, 2003; Geiduschek and Kassavetis, 2010) and middle-promoter MotA box (TGCTTtA) dependent middle promoters were used as query for a BLASTN search over the whole genome. Promoters and/or terminators were excluded if they were not intergenic or not within 10 bp overlap of the start/end of ORFs.

3.5.5 *Host DNA preparation, genome sequencing and annotation*

Host cultures were grown in 50 mL ASM1 medium amended with 1mM methanol in 250 mL polycarbonate flasks. Upon reaching maximum cell density, genomic DNA was extracted using Qiagen DNAeasy PowerWater Kit (REF 14900-50-NF) from biomass retained on 0.1 µm PC filters following the manufacturer's protocol with minor modifications to increase the yield. The bead beating step was lengthened from 5 minutes to 10 minutes. DNA elution was performed with a 2 minute incubation with elution buffer warmed to 55 °C. Nextera DNA libraries were prepared and sequenced by MicrobesNG (Birmingham, UK) for Illumina short read sequencing on the HiSeq2500 (Illumina paired end [2 x 250 bp], targeting 30-fold coverage). Additionally, long read sequencing was prepared using a MinION flow cell. Reads were quality controlled, trimmed and error corrected with the tadpole function (default settings) within BBMap v38.22 ((Bushnell *et al.*, 2017) available at https://sourceforge.net/projects/bbmap/). Contigs were assembled using SPAdes v.3.13 with default settings (Bankevich

et al., 2012). Gene calls were made with PROKKA v1.14.6 (Seemann, 2014) and submitted to BlastKOALA (Kanehisa *et al.*, 2017) for further annotation and prediction of KEGG pathways using the *Methylophilales* strain HTCC2181 as reference.

3.5.6 *Phylogeny and Network analysis*

All contigs from the Global Ocean Virome (GOV2) dataset and a WEC virome (Gregory *et al.*, 2019; Warwick-Dugdale *et al.*, 2019), *Methylophilales* phage Venkman (Figure 2.5) and LD28 phage P19250A (Moon *et al.*, 2017), isolated *Pelagibacter* phages (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Buchholz *et al.*, 2021a; Buchholz *et al.*, 2021b) and *Pelagibacter*-like MAGs (Zaragoza-Solas *et al.*, 2020) were screened for contigs that share a viral population with the genomes of OM43 phages (95% ANI over 80% length) using ClusterGenomes.pl v5.1 (https://github.com/simroux/ClusterGenomes). Genes of all contigs were identified by Prodigal v2.6.3 (Hyatt *et al.*, 2010) and imported into the Cyverse Discovery Environment 2.0 (available at https://de.cyverse.orge/), where vContact-Gene2Genome 1.1.0 was used to prepare protein sequences before protein clustering using VConTACT2 (v.0.9.8) with default settings (Bolduc *et al.*, 2017) to assess relatedness via shared gene networks. The gene sharing network was visualised using Cytoscape v3.7.1 (Shannon *et al.*, 2003). For the single-gene based phylogeny, genes of contigs that fell into the same viral protein clusters as the isolated OM43 phages from this study were aligned to selected genes in annotated OM43 phage genomes, and other respective genes of interest, with BLASTp (default parameters) (Altschul *et al.*, 1990). Genes were aligned within the Phylogeny.fr online server (Dereeper *et al.*, 2008) opting for MUSCLE alignment (Edgar, 2004) and built-in curation function (Talavera and Castresana, 2007) with default settings, removing positions with gaps for calculating phylogenetic trees. Maximum-likelihood trees were calculated with PhyML (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006) using 100 bootstraps. The trees were visualised using FigTree (v1.4.4 available at http://tree.bio.ed.ac.uk/software/figtree/). Figures were edited with Inkscape (www.inkscape.org) for aesthetics.

3.5.7 *Metagenomic reads recruitment*

Marine virome datasets were used to assess the relative abundance of phage contigs including a single virome from the Western English Channel and 131 samples from the Global Ocean Virome dataset (GOV2) (Gregory *et al.*, 2019; Warwick-Dugdale *et al.*, 2019). Metagenomic reads were subsampled to 5 million reads using the reformat.sh command within the bbmap suite. Bowtie2 (Langmead and Salzberg, 2012) indexes of de-replicated contigs were created for all known pelagiphage isolate genome (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Buchholz *et al.*, 2021; Buchholz *et al.*, 2021), the LD28 phage P19250A (Moon *et al.*, 2017), a selection of cyanophages, a selection of abundant *Roseobacter* phages (Zhang *et al.*, 2019), *Enterobacteria* phages T4 and T7 (as negative controls) as well as one genome from the viral population isolated in this study (Melnitz). Metagenomic reads from each virome were mapped against all contigs with bowtie2 (bowtie2 --seed 42 --non-deterministic). To calculate coverage and Reads Per Kilobase of contig per Million reads (RPKM) we used coverm (available at https://github.com/wwood/CoverM), with the following commands: “coverm contig --bam-files *.bam --min-read-percent-identity 0.9 --methods rpkm --min-covered-fraction 0.4”.

3.5.8 *Search and alignment of tRNA*

tRNAs were identified from bacteria and virus genomes using ARAGORN v.1.2.38 (Laslett and Canback, 2004) using the ‘-t –gcbact –c –d –fons’ flags and tRNAscan-SE v2.0.7 using flags ‘-B –fasta’ (Lowe and Chan, 2016). tRNAs were deduplicated using seqkit v0.15.0 (Shen *et al.*, 2016) with ‘rmdup –by-seq –ignore-case’. A BLAST database of tRNA genes was made in BLAST v2.5.0+ (Altschul *et al.*, 1990) and used for sequence alignment with BLASTN with the flags `-outfmt '6 std qlen slen' -evalue 1e-05 -task blastn-short`. Percentage identity was calculated by dividing alignment-length by query-length times alignment-percentage.

3.5.9 *Structural analysis of a putative endolysin.*

The predicted amino acid sequence of gene product 67 (gp67) was used as a query to identify putative structures on the SwissModel server (Waterhouse *et al.*, 2018) using BLAST and HHBits. Putative models were downselected based on suitable quaternary structure properties and GQME>0.7. Autolysin SagA from *Brucella abortus* (PDB model 7dnp.1) was selected as the best-hit and used for subsequent modelling of the structure of gp67. Models and associated figures were visualised in PyMOL v. 2.5.1 (https://pymol.org/2/).

3.5.10 *Structural analysis of CsgGF*

The predicted amino acid sequences of CsgG and CsgF from Melnitz were used as a query to identify putative structures on the SwissModel server using BLAST and HHBits. The best-hit was determined by predicted quaternary structure properties. Global GQME scores were <0.5 and Q-MEAN scores identified the best predicted structures as unreliable (ranging from -3.28 to -5.11), although localised regions had Q-MEAN scores > 0.7. Therefore, to improve structural predictions, amino acid sequences for CsgG and CsgF were independently run through AlphaFold2 using the available Colab web interface (https://colab.research.google.com/drive/1LVPSOf4L502F21RWBmYJJYYLDlOU2NTL). Structures were determined both with and without post-prediction relaxation (use_amber) and use of MMSeqs2 templates (use_templates). As no noticeable differences were observed between these runs, we selected the top scoring unrelaxed model for downstream comparison to known CsgG and CsgF structures. Predicted structures from AlphaFold2 were downloaded, visualised and aligned to *E. coli* CsgGF (7NDP) in PyMOL v. 2.5.1. CsgGF from pelagimyophage HTVC008M was similarly analysed with AlphaFold2 to confirm structural similarity to that of Melnitz. Structural prediction of the CsgGF heterodimer in Melnitz was assumed to conform to the 18-mer structure of CsgGF in *E. coli.* Scripts for generating structures in PyMOL are available here https://github.com/HBuchholz/Genomic-evidence-for-inter-class-host-transition-between-streamlined-heterotrophs. Electrostatic potential of protein surfaces was calculated and visualised using the APBS Electrostatics plugin available within PyMOL.

3.5.11 *Transmission Electron Microscopy*

For ultrastructural analysis, bacterial cells and/or phages were adhered onto pioloform-coated 100 mesh copper EM grids (Agar Scientific, Standsted, UK) by floating grids on sample droplets placed on parafilm for 3 min. After 3 x 5 min washes in droplets of deionized water, structures were contrasted on droplets of 2% (w/v) uranyl acetate in 2% (w/v) methyl cellulose (ratio 1:9) on ice for 10 min, the grids picked up in a wire loop and excess contrasting medium removed using filter paper. The grids were then air dried, removed from the wire loop and imaged using a JEOL JEM 1400 transmission electron microscope operated at 120 kV with a digital camera (Gatan, ES1000W, Abingdon, UK).

3.6 Data availability and acknowledgements

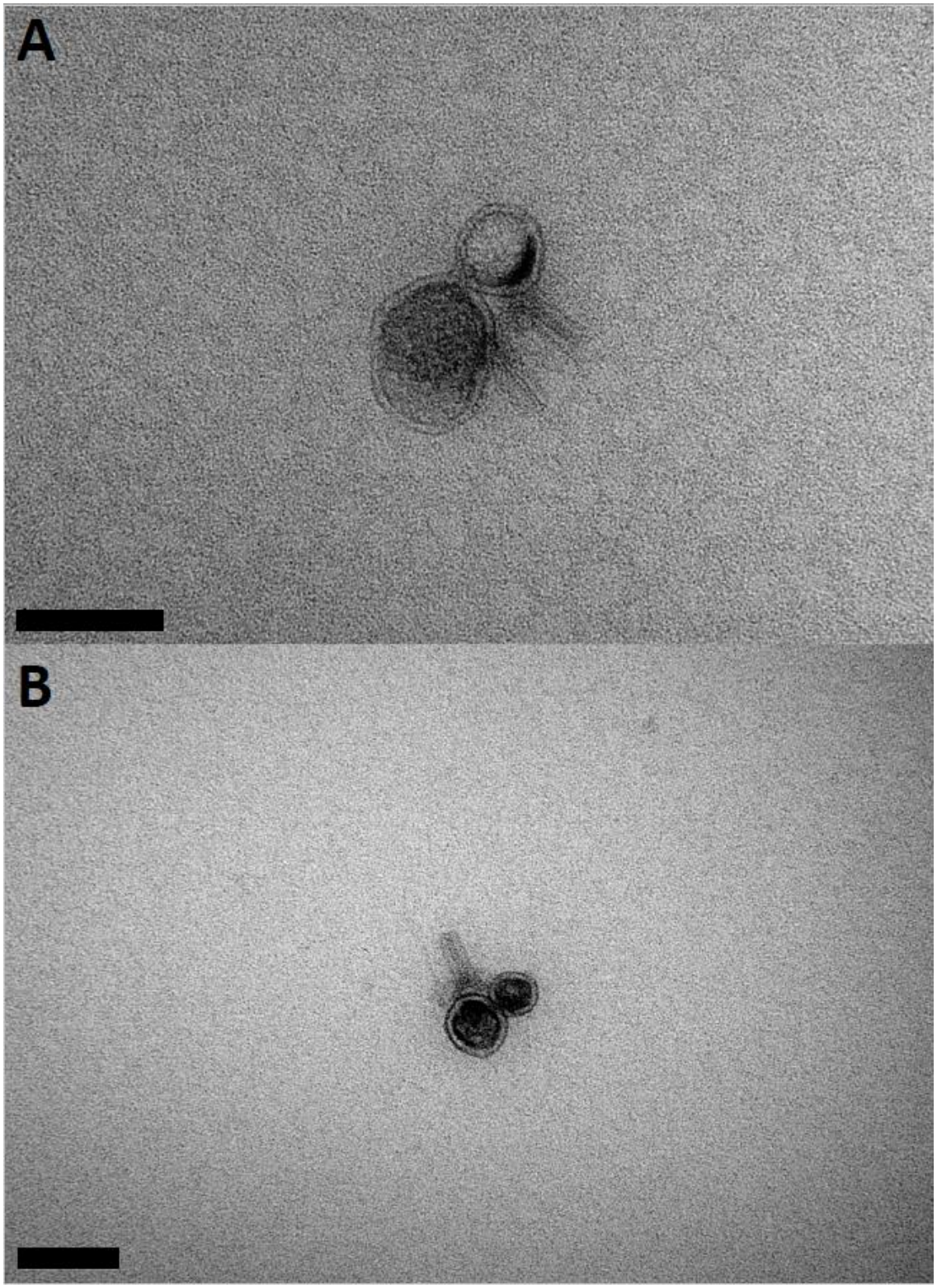
All four Melnitz-like genome were deposited as GenBank entry on NCBI under accession number MZ577095-MZ577098 of BioProject PRJNA625644; the reference genome used for the analysis was deposited under MZ577097. Sequencing data for all phages sequenced in this study can be found on the SRA databank under Accession numbers SAMN18926670-SAMN18926674. Reads for Methylophilales bacterial host H5P1 are available under SAMN20856461.

3.6.1 *Acknowledgements*

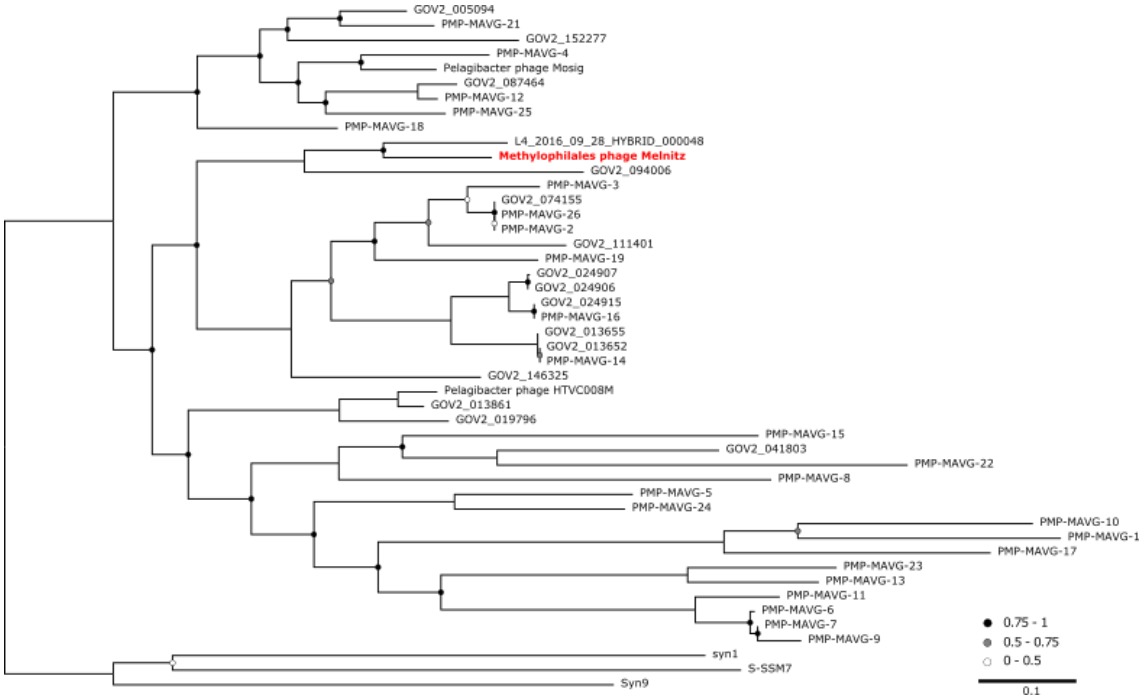
We would like to thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing the TE microscopy and imaging. We would also like to thank the crew of the R/V Plymouth Quest and our collaborators at Plymouth Marine Laboratory for collecting water samples, and the driver Magic for delivering water samples from Plymouth to Exeter. We would also like to thank the crew of the R/V Atlantic Explorer and our collaborators at the Bermuda Institute of Ocean Sciences. We acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. Genome sequencing was provided by the Exeter Sequencing Service. This project utilised equipment funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi User Equipment Award (WT101650MA)

and BBSRC LOLA award (BB/K003240/1). The efforts of Holger H. Buchholz in this work were funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training program. LMB, MLM and BT were funded by NERC (NE/R010935/1) and by the Simons Foundation BIOS-SCOPE program.

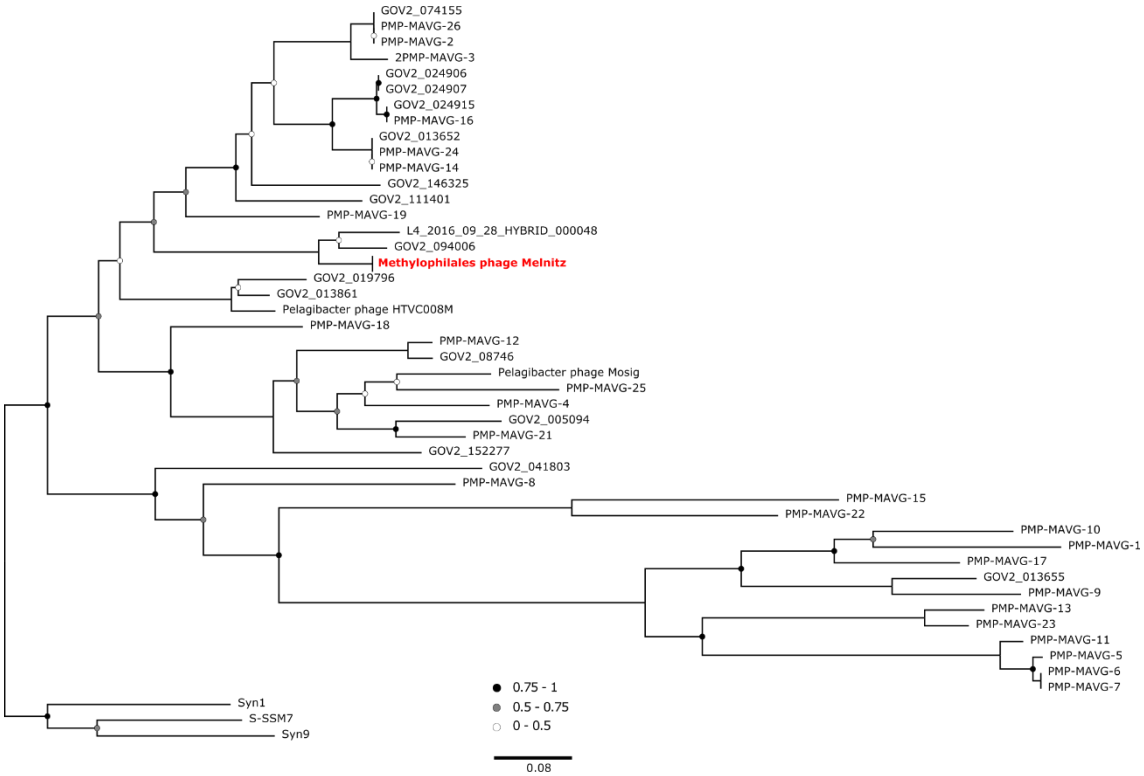
Chapter 3: Supplementary Figures



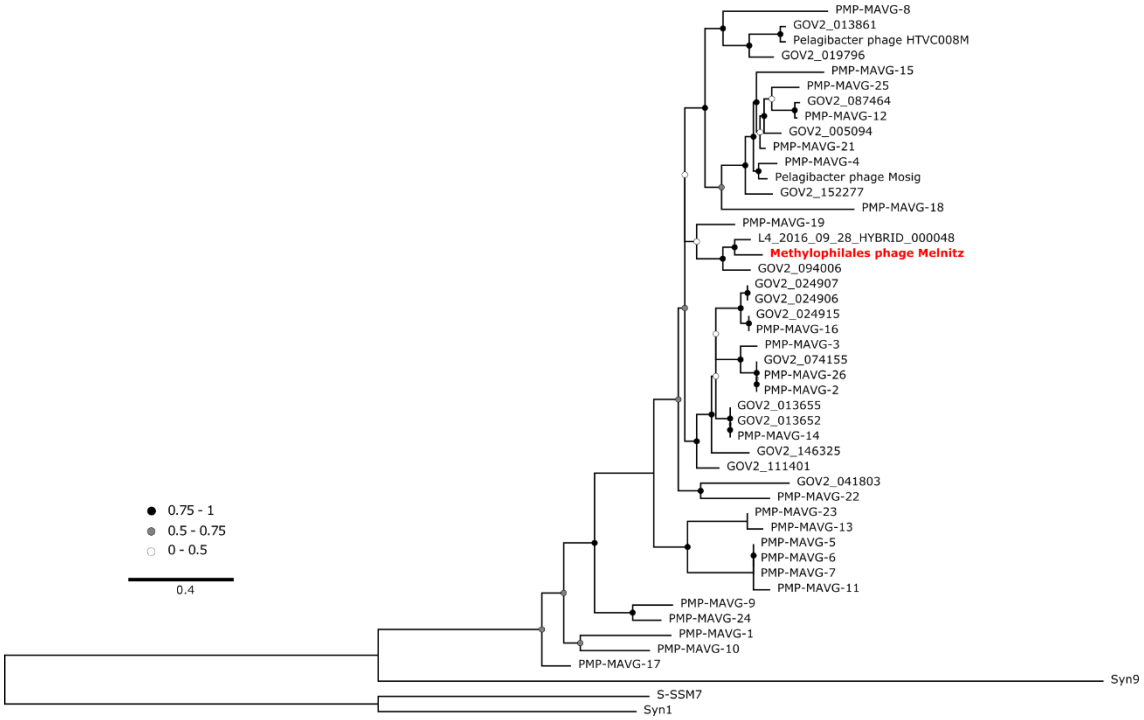
Supplementary Figure 3.1 TEM images of phage Melnitz virions. Bar indicates 100 nm.



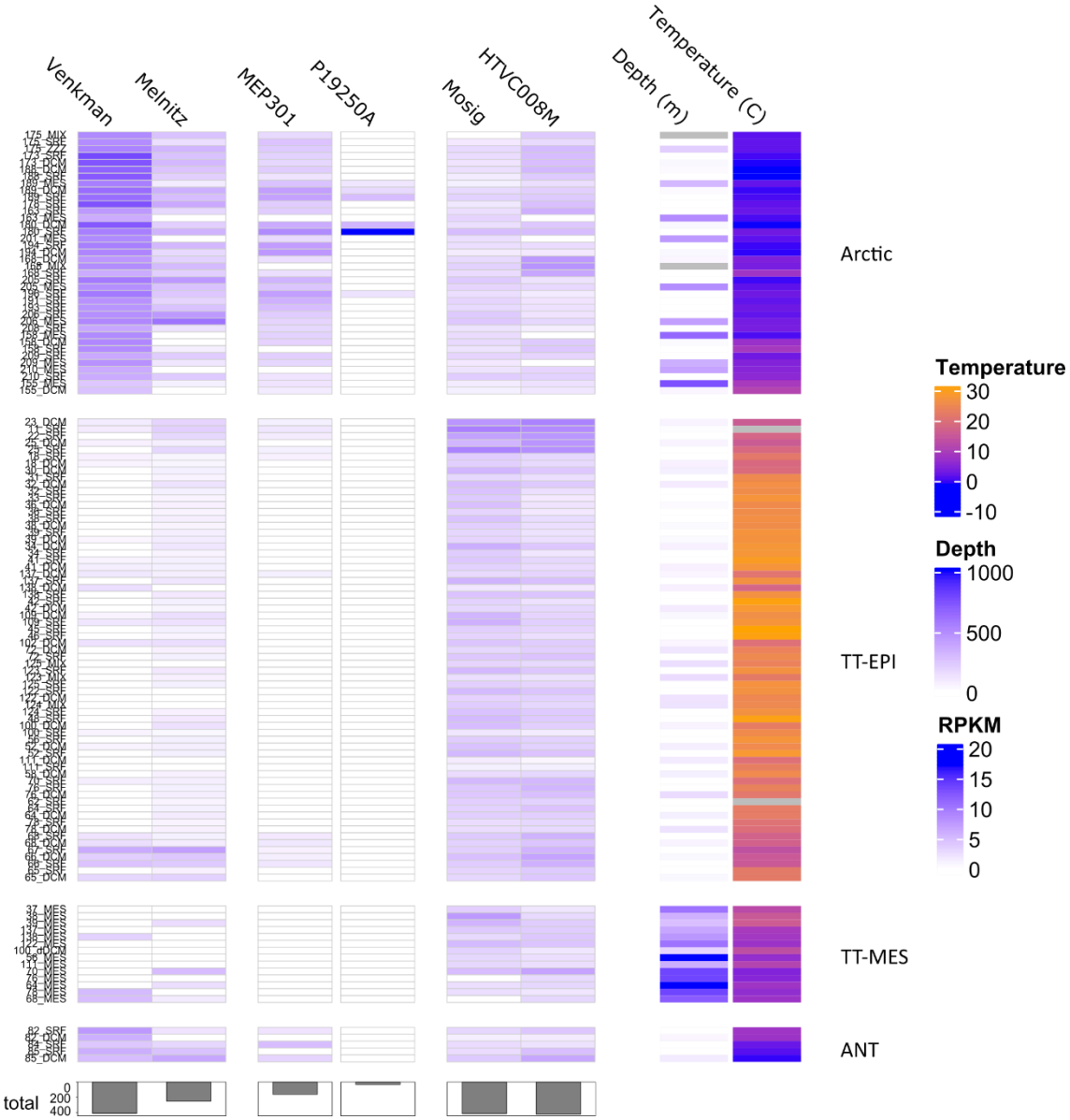
Supplementary Figure 3.2 Neighbour-joining tree (100 bootstraps) of a tail sheath encoding gene found in Melnitz related myophages. Branch support values <1 are indicated by circle colours on the tree. The cyanophage branch was used to root the tree.



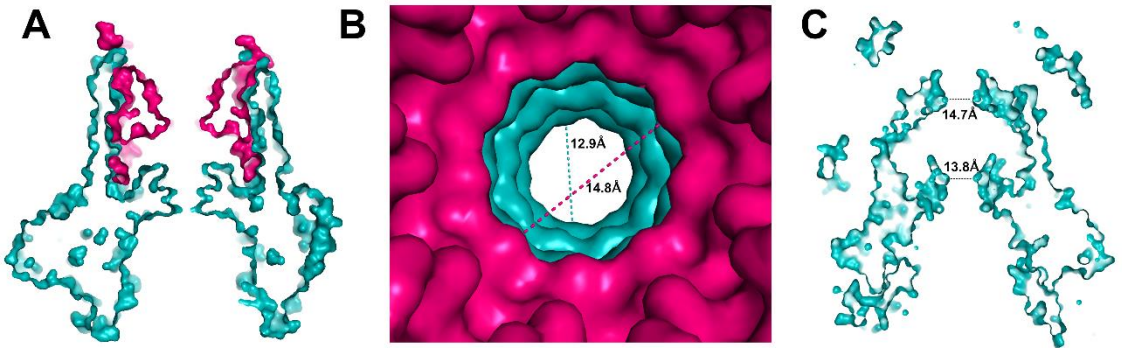
Supplementary Figure 3.3 Neighbour-joining tree (100 bootstraps) of the *TerL* gene found in Melnitz related myophages. Branch support values <1 are indicated by circle colours on the tree. The cyanophage branch was used to root the tree.



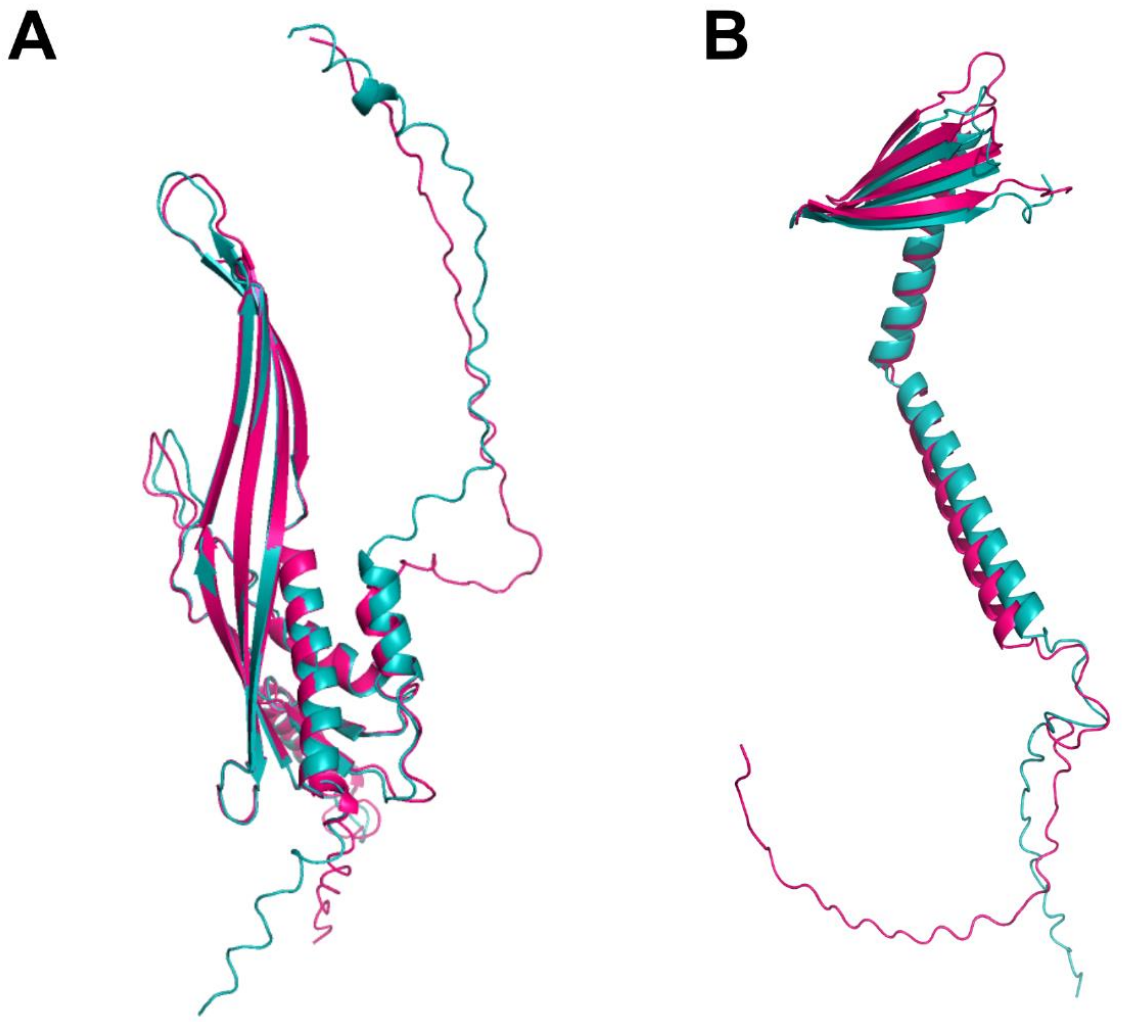
Supplementary Figure 3.4 Neighbour-joining tree (100 bootstraps) of a capsid scaffolding gene found in Melnitz related myophages. Branch support values <1 are indicated by circle colours on the tree. The cyanophage branch was used to root the tree.



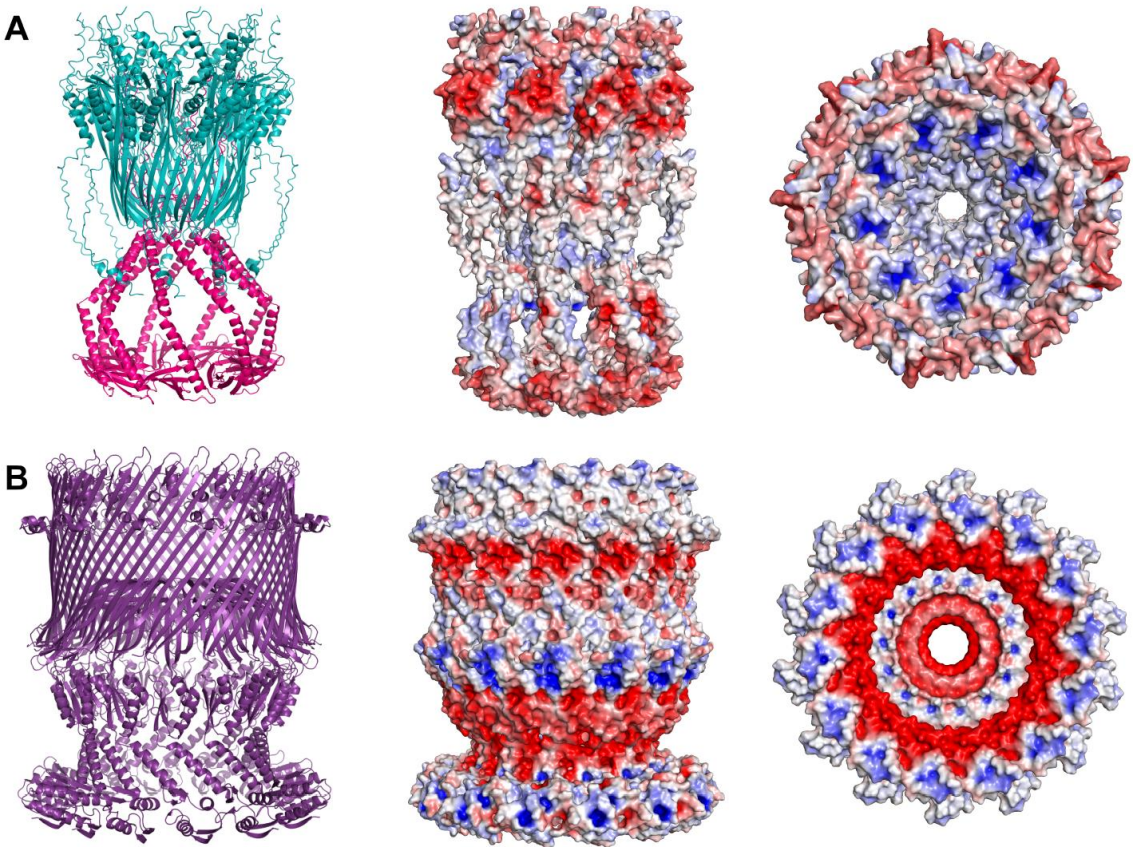
Supplementary Figure 3.5 Global abundance of Melnitz related phages. Reads recruited per kilobase of contigs per million reads (RPKM) of GOV2 viromes against phages infecting OM43 and LD28 as well as known *Pelagibacter* myophages. Samples are organized by ecological zone: Arctic, TT-EPI temperate-tropical epipelagic, TT-MES temperate-tropical mesopelagic, ANT Antarctic.



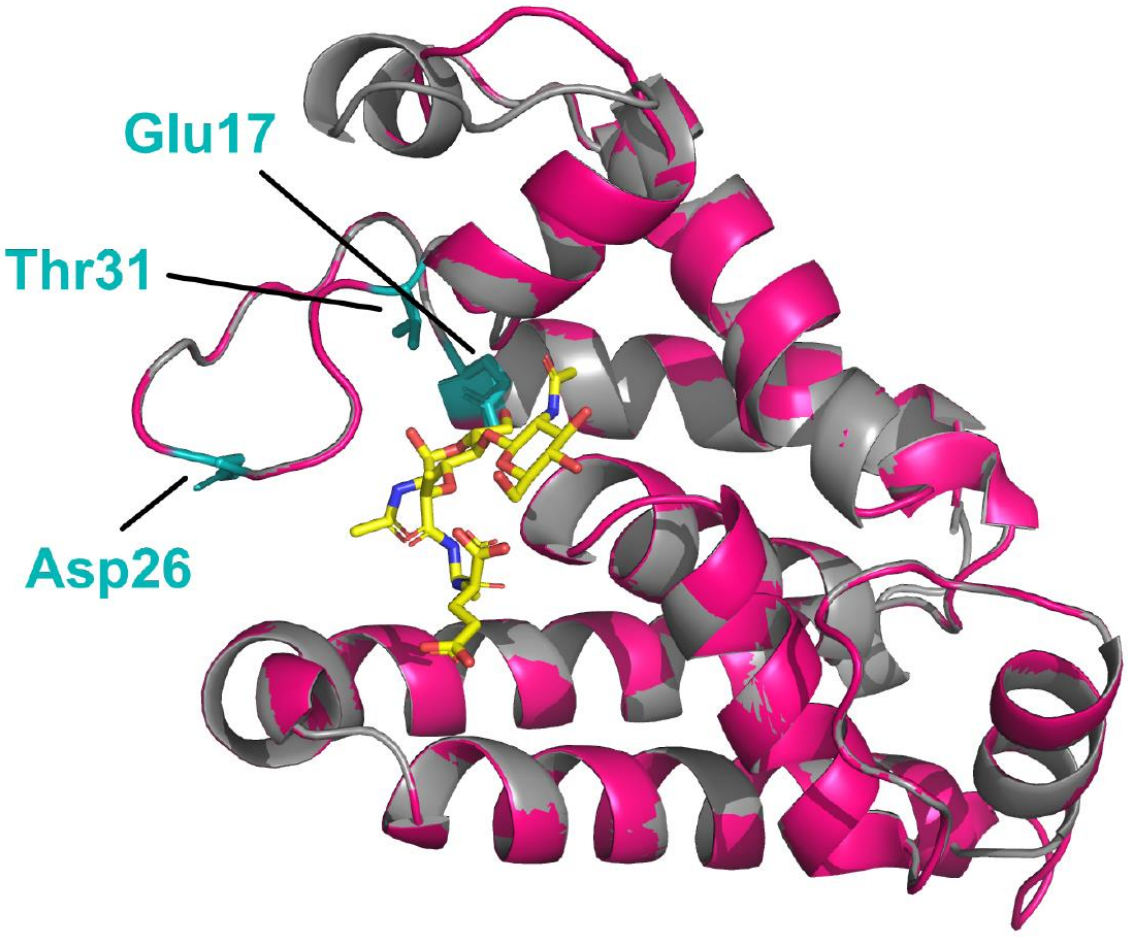
Supplementary Figure 3.6. **Cross sectional structure of CsgGF**. **A** Clipped predicted surface model of internal structure of CsgGF in *E. coli* showing CsgG (teal) and CsgF (pink) showing the internal channel structure. This structure comprises a series of narrowing pores. **B** with CsgF creating a pore 14.8Å in diameter and CsgG creating a smallest pore 12.9Å in diameter. **C** Clipped predicted surface model of internal structure of Melnitz-encoded CsgG comprises two channels of similar size to those of the *E. coli* CsgGF complex.



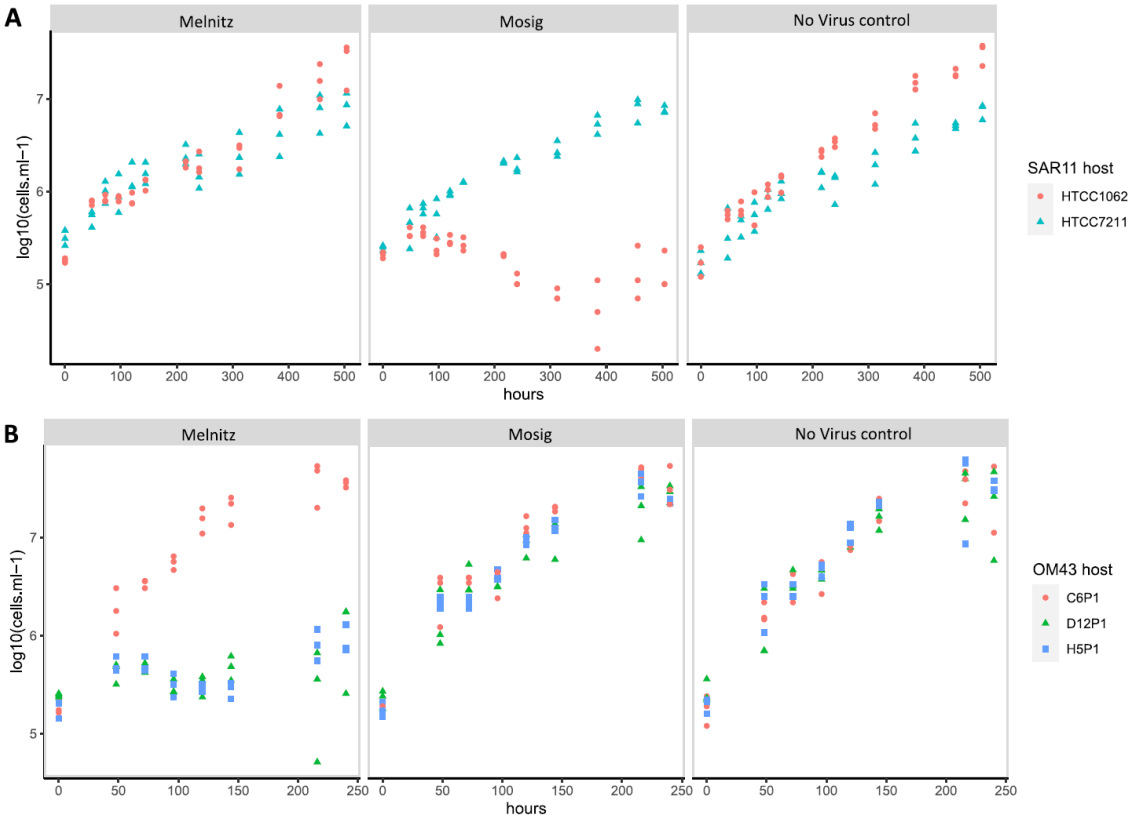
Supplementary Figure 3.7. Structural comparison between **A** CsgG encoded by Melnitz (teal) and HTVC008M (pink); **B** CsgF encoded by Melnitz (teal) and HTVC008M (pink).



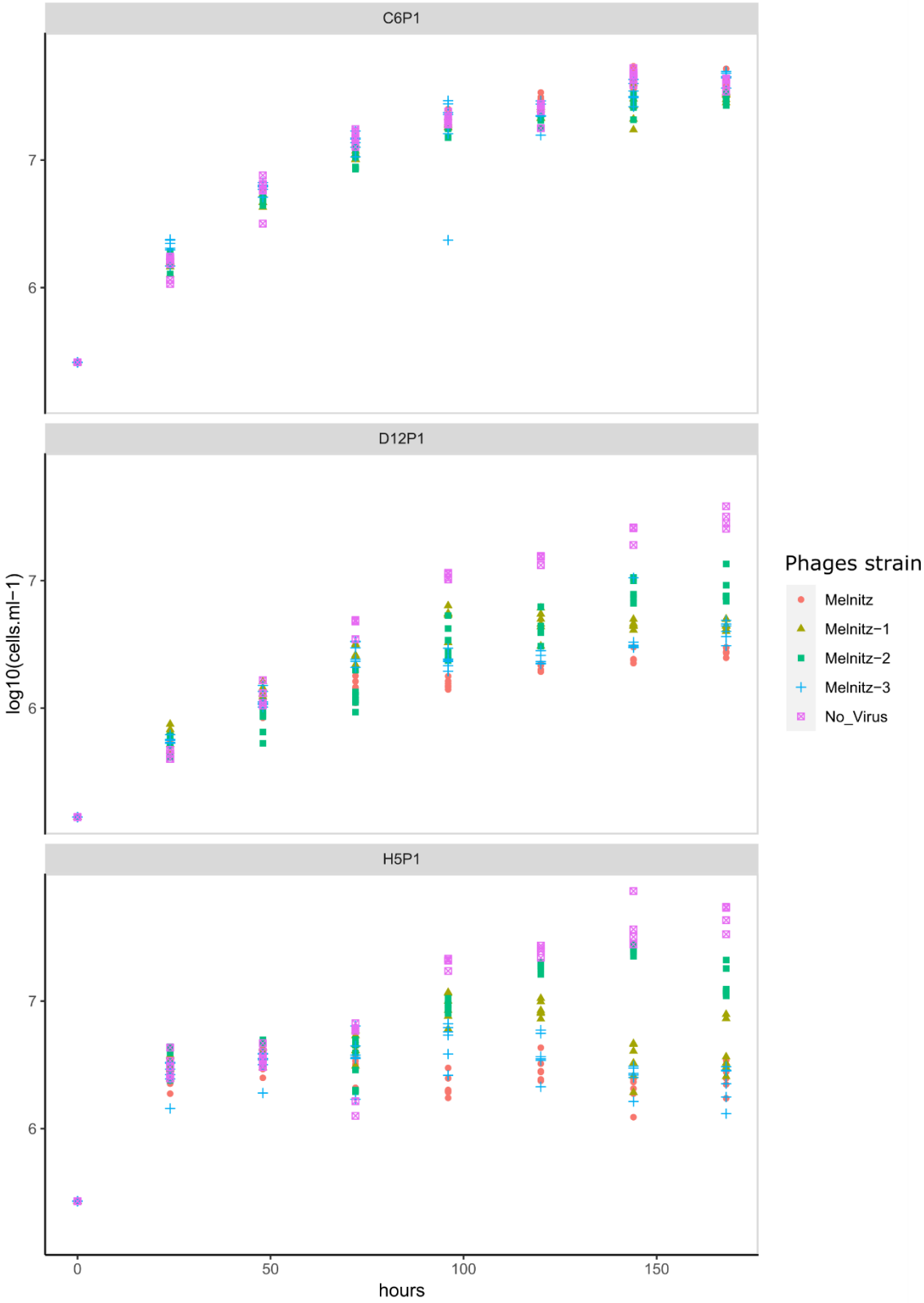
Supplementary Figure 3.8. **A** Predicted structure of Melnitz CsgF in an inverted orientation compared to **B** structure of secretin PilQ (6W6M) from *Vibrio cholerae*. Left to right - cartoon model; electrostatic potential (red = -ve, blue=+ve); top-down view through the barrel.



Supplementary Figure 3.9. **SwissModel alignment of putative endolysin structure** encoded by Melnitz gp67 (hot pink), with a SagA autolysin encoded by Brucella abortus (PDB model 7dnp.1, grey). The conserved peptidoglycan binding site encoded by Glu17, Asp26 and Thr31 is highlighted (teal), with the peptidoglycan substrate (yellow stick representation).



Supplementary Figure 3.10. **Host range assessment of methylophage Melnitz and pelagiphage Mosig**. **A** Growth curve of SAR11 strains HTCC1062 and HTCC7211, treated with OM43 phage Melnitz and SAR11 phage Mosig. **B** Growth curves of OM43 strains C6P1, D12P1 and H5P1 treated with OM43 phage Melnitz and SAR11 phage Mosig



Supplementary Figure 3.11. **Host range assessment of Melnitz strains**. **A** Growth curve of OM43 strains C6P1 and HTCC7211, treated with OM43 phage Melnitz and SAR11 phage Mosig. **B** Growth curves of OM43 strains C6P1, D12P1 and H5P1 treated with OM43 phage Melnitz and SAR11 phage Mosig

Chapter 4: *Pelagibacter* phage Skadi - an abundant polar specialist exemplifies ecotypic niche specificity among the most abundant viruses on Earth.

This chapter introduces the novel and highly abundant *Pelagibacter* phage Skadi, providing an example of ecotypic niches in SAR11 virus-host systems. Based on ICTV guidelines, it shows that Skadi and other HTVC010P-type viruses form a novel viral family, for which we propose the name *Ubiqueviridae*. The chapter is reformatted from a manuscript intended for publication and has been reviewed for submission by my co-authors.

Holger H. Buchholz, Ashley G. Bell, Luis M. Bolaños, Michelle L. Michelsen, Michael J. Allen, Ben Temperton. (2021). *Pelagibacter* phage Skadi - an abundant polar specialist exemplifies niche specificity among the most abundant viruses on Earth. *In prep.*

Author Contributions

- *HHB authored and reviewed the manuscript, conceived and designed experiments, performed experiments, analysed the data, conceived and designed experiments*
- *AGB analysed the data, prepared figures (Figure 4.2)*
- *LMB analysed the data,*
- *MLM performed experiments*
- *MJA provided directions and edits*
- *BT authored and reviewed the manuscript, conceived and designed experiments, analysed data, prepared figures/tables contributed reagents/materials and analysis tools*

4.1 Abstract

Bacteria in the SAR11 clade are the most abundant members of marine bacterioplankton communities and are a critical component of global biogeochemical cycles. Similarly, phages that predate SAR11 (known as pelagiphages) are both ubiquitous and highly abundant in the oceans. These

viruses are predicted to shape microbial community structures and increase carbon turnover throughout the oceans. Yet, the ecological drivers of host and niche specificity of pelagiphage populations are poorly understood. Here we report the isolation and global distribution of a novel pelagiphage named Skadi, named after the Norse goddess of winter. Isolated from the Western English Channel on an ecotypic cold-water SAR11 host, Skadi is closely related to the globally dominant pelagiphage HTVC010P. Along with other HTVC010P-type viruses, Skadi belongs to a distinct viral family within the order *Caudovirales* for which we propose the name *Ubiqueviridae*. Metagenomic recruitment from Global Ocean Viromes (GOV2) identifies Skadi as one of the most abundant pelagiphages on Earth, and reveals that Skadi replaces HTVC010P as the dominant pelagiphage species at high latitudes, marking Skadi as a polar specialist. Correlation analysis and constrained ordinations show that Skadi is most abundant in conditions with low temperature, high nutrient and oxygen concentrations, and high primary productivity. In contrast, the distribution of pelagiphage HTVC010P was correlated to environments with low oxygen and high salinity, but not temperature. The majority of other pelagiphages remain scarce in most marine provinces and 7 out of 28 species that were detectable in viromes belong to discrete ecological niches. Our results suggest that pelagiphage populations occupy the global viral Seed-Bank, with environmental parameters selecting for a few pelagiphage species that dominate global ocean viromes.

4.2 Introduction

Bacteriophages are the most abundant organisms in the oceans and play a critical role in driving microbial global biogeochemical cycles, particularly carbon cycling (Jover *et al.*, 2014; Worden *et al.*, 2015; Puxty *et al.*, 2016, 2018). During an infection, phages hijack their hosts and alter cellular metabolism, changing nutrient acquisition and catabolism rates, and influence growth efficiency and structure of microbial populations (Motegi *et al.*, 2009; Hurwitz *et al.*, 2013; Warwick-Dugdale *et al.*, 2019). Ultimately, phages predate and kill bacterial host cells, releasing the carbon and other cell-bound nutrients back into the environment, thereby recycling organic compounds (Danovaro *et al.*, 2008;

Worden *et al.*, 2015). These compounds are then remineralized by other microbes, forming the '*viral shunt*', which is thought to be an essential part of the marine food web (Wilhelm & Suttle 1999, Worden *et al.*, 2015; Zimmerman *et al.*, 2019). Phages drive microbial evolution and are agents of frequent genetic exchange between hosts as well as other phages, creating highly mosaic genomes (Lima-Mendez *et al.*, 2008; Dion *et al.*, 2020). However, in marine systems, viral tagging has confirmed discernible, discrete viral populations where the majority of sequence deviations occur in the metabolic genes, indicating niche specificity (Deng *et al.*, 2014; Marston and Martiny, 2016). Genomic clusters for T4-like phages infecting Cyanobacteria, were found to exhibit spatial and temporal abundance patterns suggesting that they represent distinct viral ecotypes (Marston and Martiny, 2016). Analysing viral communities at the Bermuda Atlantic Time-Series (BATS) station also revealed recurrent seasonal patterns linking viral communities with stratification and mixing events in the water column (Parsons *et al.*, 2012). Temporal analysis of T4-like phages and their hosts at the San Pedro Ocean Time series station further suggested that the relative abundance of virus populations is tightly linked to those of their associated hosts (Chow *et al.*, 2014). Spatiotemporal distribution and ecotypes of viral populations are therefore likely interlinked and influenced by both physical and biological parameters affecting their host cells (Breitbart *et al.*, 2018).

The ubiquitous alphaproteobacterial order *Pelagibacterales* (clade SAR11) comprise up to one-third of total marine microbial communities, making them likely the most abundant cells on Earth (Morris *et al.*, 2002; Stingl *et al.*, 2007). SAR11 are an ancient and diverse clade, characterised by small, streamlined genomes with low G+C % content, high gene synteny and are phylogenetically organised into specialized ecotypes (Grote *et al.*, 2012; Giovannoni, 2017; Delmont *et al.*, 2019). Phages infecting members of the SAR11 (pelagiphages) are the most abundant viruses in global ocean viromes (Zhao *et al.*, 2013; Zhang *et al.*, 2020). Since their relatively recent discovery in 2013, pelagiphages have been isolated across most major oceans, resulting in 40 different cultivated species, 37 of which were reported as podophage morphotypes (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Buchholz *et al.*, 2021a; Buchholz *et al.*, 2021b; Du *et al.*, 2021), two myophages (Zhao *et al.*, 2013; Buchholz *et al.*, 2021b) and one siphophage (Figure 2.4). Viral single-amplified genomes (vSAGs) and

metagenomically assembled viral genomes (MAVGs) of predicted pelagiphages (without cultured representatives) suggest that current viral isolates only cover a fraction of the available pelagiphage diversity (Mizuno *et al.*, 2013; Martinez-Hernandez *et al.*, 2017; Zaragoza-Solas *et al.*, 2020). Owing to the challenges of culturing SAR11 (Carini *et al.*, 2013, 2014) and isolating phages for them (Buchholz, *et al.*, 2021a), experimental evaluation of predicted pelagiphage ecology is in its infancy. Unlike the clear delineation of their SAR11 hosts along ecological niches, lineages of pelagiphages infecting SAR11 are generally not known to have discernible spatial distribution patterns, despite generally narrow host ranges restricted to specific ecotypes (though some exceptions exist) (Zhao *et al.*, 2018; Zhang *et al.*, 2020). The majority of these confirmed pelagiphage genomes have low abundances, with few notable exceptions e.g. the ubiquitously abundant phage HTVC010P, conforming to the viral Seed-Bank model (Breitbart and Rohwer, 2005), which states that only a fraction of phages are actively replicating, depending on host availability.

Information of host ranges for known pelagiphages are limited by available host cultures. Only two SAR11 host strains have been used by both groups that have reported pelagiphage cultivation to date - the cold-water *Pelagibacter ubique* HTCC1062 and warm-water *Pelagibacter bermudensis* HTCC7211 (Zhao *et al.*, 2018), additionally both have used local a strain from the WEC and the Taiwan Strait respectively. Host range experiments with cyanophages have shown that T7-like phages tend to be specific to a single host, whereas T4-like phages have broader host ranges within a host ecotype, i.e. those isolated on high- or low-light adapted *Prochlorococcus* (Sullivan *et al.*, 2003; Dekel-Bird *et al.*, 2015). Phage specialism or generalism are often considered as a trade-off between broad host range and infection efficiency. The majority of phages are considered to be specialists (Koskella and Meaden, 2013), though the evolutionary drivers behind specialism and generalism are topic of ongoing debate (Ross *et al.*, 2016). However, for pelagiphages, little is known about a possible partition between generalists and specialists.

In this study, we isolated and sequenced eleven strains of a single novel pelagiphage species, the *Pelagibacter* phage Skadi, named after the Norse goddess of winter. We demonstrate that Skadi is relatively closely related to the globally abundant HTVC010P, both in gene sharing networks and single-gene

phylogeny. Together with our previously isolated *Pelagibacter* phages Greip and Lederberg as well as seven other HTVC010P-type pelagiphages, they form a distinct novel viral family, for which we propose the name *Ubiqueviridae*. Recruitment of reads from global marine viromes revealed that Skadi is a polar pelagiphage ecotype that dominates Arctic and Antarctic marine viral communities, and is replaced by HTVC010P at temperate latitudes. At tropical and subtropical latitudes, Skadi is near or below limits of detection, apart from regions with direct influx of polar water, e.g. the Humboldt Current System in the Southern Pacific Ocean. The global distribution pattern in addition to experimentally confirmed host ranges suggest that abundant populations of Skadi replicate in polar waters on cold-water ecotypes of SAR11 and are potentially transported around the globe in currents of sinking polar water which are part of the ocean conveyor belt. Skadi therefore serves as an exemplar of niche specialisation in pelagiphages and a model organism for determining how ocean currents might shape regional viral communities.

4.3 Results & Discussion

4.3.1 *Isolation, morphology and general features of Skadi-like* Pelagibacter *phages*

A total of eleven Skadi-like (ANI > 99 %) *Pelagibacter* phages were isolated on the SAR11 cold-water ecotype *Pelagibacter ubique* HTCC1062 from three coastal surface water samples taken from the Western English Channel (WEC) in September, October and November of 2018 (For environmental parameters see Supplementary Table 2.2). Ten out of eleven genomes were assembled into single circular contigs, with one assembly returning a fragmented, linear genome. Four out of these ten genomes contained a direct terminal repeat of 125bp at the end of the linearized genome (identified with CheckV (Nayfach *et al.*, 2021)) and confirmed through a doubling of read coverage at contig ends (Casjens and Gilcrease, 2009). This is indicative of terminal repeats for genome replication and *pac* headful packaging similar to the T7-like P-SSP7 cyanophage (Sabehi and Lindell, 2012). General features of the genomes are shown in Table 4.1. Genome sizes range from 34,897 base pairs (bp) to 35,392 bp with a G+C content between 31.49 % and 32.19 % and 65 open reading frames (ORFs) for

Table 4.1. Overview of Skadi-like viral isolate genomes.

Phage	Genome Size (bp)	G+C (%)	Terminal Repeat (bp)	ANI (%) to Skadi	ORFs	Collection Date
Skadi	35,392	31.53	125	100	65	Sept 24 2018
Mimir	35,392	31.49	125	99.763	65	Sept 24 2018
Kvasir	35,083	31.50	n.a.	99.506	65	Sept 24 2018
Odr	34,897	31.51	n.a.	99.274	65	Sept 24 2018
Freyja	35,374	31.54	125	99.633	65	Sept 24 2018
Gerda	35,392	31.57	125	99.811	65	Sept 24 2018
Naecken	35,247	31.56	n.a.	99.638	65	Oct 17 2018
Jarnsaxa	35,276	31.53	n.a.	99.554	65	Oct 17 2018
Ulfrun	35,276	31.53	n.a.	99.554	65	Oct 17 2018
Imd	35,267	31.50	n.a.	99.813	65	Oct 17 2018

the full-length genomes (Figure 4.1A). Average nucleotide identity (ANI) between the genomes was > 99.1 % and were therefore considered to be the same viral species (Turner *et al.*, 2021), hereafter referred to as Skadi (named after the Norse goddess of winter), unless a specific strain is specified. Skadi was closely related to Pelagiphage HTVC010P, based on whole-genome phylogenetic analysis for all *Pelagibacter* phages (Figure 4.1C), but had a smaller capsid (39±0.7 nm (Figure 4.1B), compared to 50±3 nm for HTVC010P (Zhao *et al.*, 2013)). The difference in capsid size reported for HTVC010P (Zhao *et al.*, 2013; Zhang *et al.*, 2020) and here for Skadi may be partly due to different operators and protocols for performing transmission electron microscopy (TEM). Like HTVC010P, the short stubby tail of *Skadi* suggested a *Podoviridae* morphology.

4.3.2 *Genome characterisation and structure of Skadi*

Whole-genome alignment showed that the Skadi genome is arranged into early and late genomic regions similar to other HTVC010P-type genomes (Du *et al.*, 2021). A TonB-dependent receptor protein located between the two tail tube proteins A and B and the major capsid protein suggests TonB as a putative receptor for adsorption and initiating infection (Rabsch *et al.*, 2007; Bertozzi Silva *et al.*, 2016). TonB-dependent transporters, likely belonging to SAR11 cells, are one of the dominant proteins found in surface ocean metaproteomes

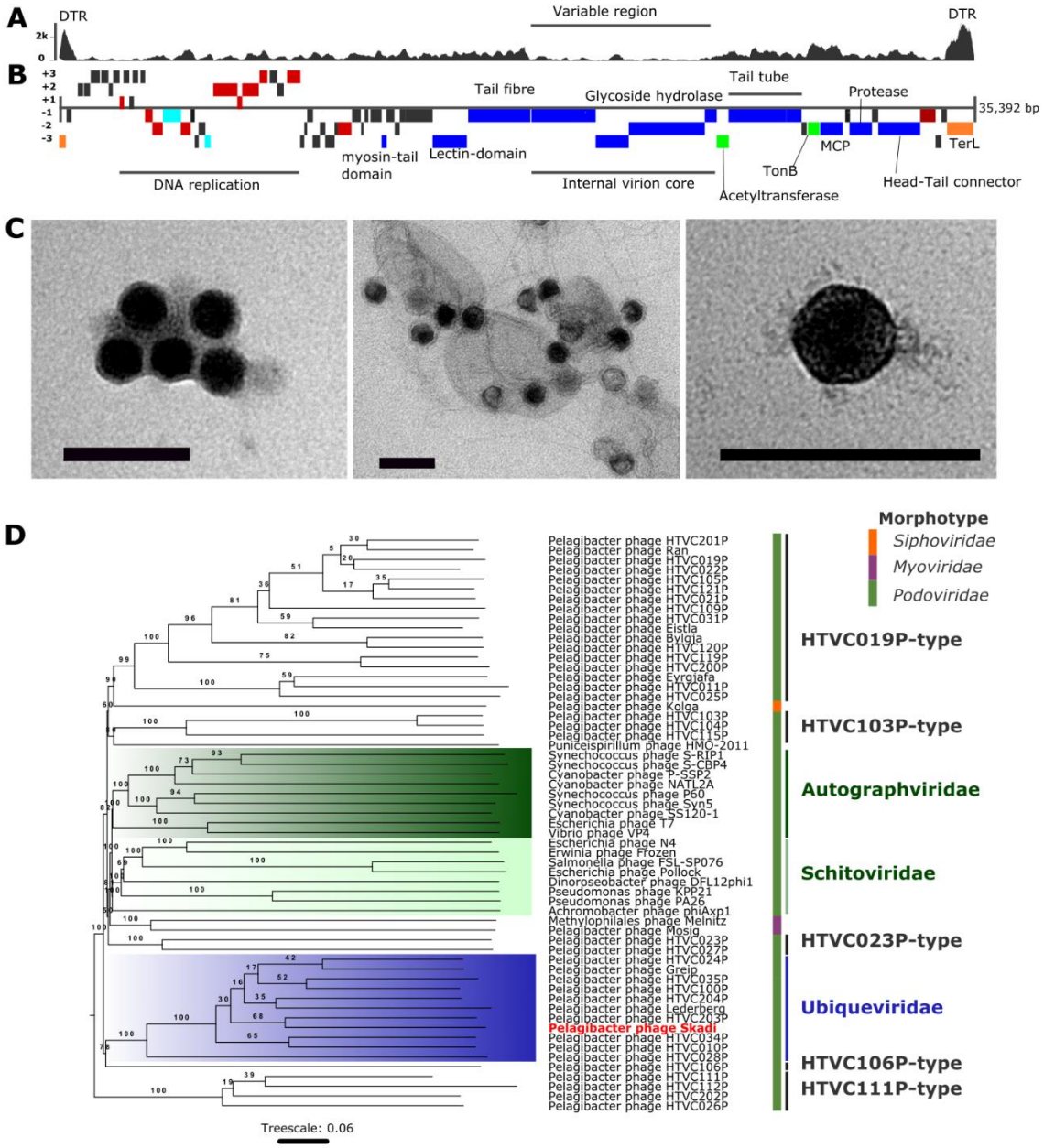


Figure 4.1 General features of *Pelagibacter* phage Skadi. (A) Plot showing per nucleotide coverage of the Skadi genome, high coverage at contig-ends suggest direct terminal repeats; the region encoding for virion proteins has low average coverage, suggestive of a hypervariable region **(B)** Genome map of Skadi, translation frames of Open Reading Frames (ORFs) are indicated in the figure, functional annotations of ORFs are color-coded with red: DNA replication and metabolism; teal: transcription; blue: Virus structural genes; green: virulence related genes; Orange: packaging. **(C)** Transmission Electron Microscopy (TEM) images, black scale bar represent 100 nm, left: icosahedral viral capsids of Skadi with a size of 39± 0.7 nm; middle: Skadi viral particles and cellular debris of lysed HTCC1062 host cells; right: Skadi single viral particle. **(D)** Full-genome phylogenetic tree created in VICTOR (Meier-Kolthoff and Göker, 2017) using known *Pelagibacter* phage genomes, leaf-label for Skadi is highlighted in red. Randomly chosen representative genomes of the *Autographviridae* (light green) and *Schitoviridae* (dark green) families were included and highlighted, the branch with all members of the proposed *Ubiqueviridae* is highlighted in blue. All bootstrap values are shown.

(Morris *et al.*, 2010). Therefore TonB receptors are likely to provide a desirable target for pelagiphages, however, we did not find orthologues of Skadi-like TonB-receptors in other pelagiphages, suggesting that infection strategies vary between species. We also identified an ORF encoding for a lectin-fold containing protein next to a putative tail fibre protein. Lectin-folds in association with phage tail fibre proteins are often involved in host receptor binding and cell attachment as well (Salazar *et al.*, 2019), and might also be used by Skadi to attach to SAR11 cells. After the initial attachment to a susceptible cell, phages need to breach the cell wall and eject viral DNA into the cytoplasm. We identified three internal virion proteins similar to gp14, gp15 and gp16 found in T7. As shown in T7, Skadi likely uses these proteins to assemble into a tubular, spiral ring complex inside the phage capsid during capsid formation. After the virus particle attaches to the cellular outer membrane of a host, the protein complex ejects into the cell envelope, binding and leading the end of the phage DNA strand through the periplasm into the infected cell (Chang *et al.*, 2010; Lupo *et al.*, 2015). The proximate glycoside hydrolase gene found in Skadi is likely a structural component of the internal virion complex, facilitating the injection of phage DNA by breaking glycosidic bonds in the cell wall, as was found to be the role of the glycoside hydrolase (Gp255) in *Bacillus* phage vB_BpuM_BpSp (Yuan and Gao, 2016). Skadi encodes for an acetyltransferase gene downstream of the internal core complex, which in a phiKMV-like virus infecting *Pseudomonas aeruginosa* was responsible for the shutdown of host transcription by cleaving the bacterial RNA polymerase during the early stages of infection (Ceyssens *et al.*, 2020). An additional methylase gene likely provides Skadi with protection from host restriction endonucleases found in *Pelagibacter ubique* HTCC1062, or might help to protect the Skadi virocells against superinfecting DNA from phage competitors as endonucleases are a common feature in pelagiphages (Murphy *et al.*, 2013). Skadi further encodes peptide hydrolase and nuclease genes that are likely involved in breaking down host proteins and genetic recombination during the early infection stage. Skadi does not possess a viral RNA polymerase (RNAP) gene, unlike pelagiphages in the *Autographviridae* family, and has no DNA polymerases. Temporal regulation in the absence of endogenous RNAP is common in T4 and λ-like phages and involves sophisticated mechanisms that rely on early to late promoter sequences for recognition by the host RNAP (Yang

et al., 2014). We identified four σ^{70} promoters and two Pribnow boxes (TATAAT-promoter sequences) in proximity of DNA replication and manipulation genes, but were unable to identify any terminator sequences, tRNA, tmRNA or ribonucleotide switches. Instead, Skadi encodes for two transcriptional regulator proteins, similar to HTVC010P-type phages (Du *et al.*, 2021). Likely these transcriptional regulators are used to hijack the cellular transcriptional machinery, indicating that Skadi, like other HTVC010P-type phages, relies on host RNAP and DNAP proteins for translation and transcription. A *DnaA*-like chromosomal replication initiation gene was identified, which in *Escherichia coli* phage λ is involved in the initiation of DNA replication (Weigel and Seitz, 2006), and most likely fulfils the same function in Skadi. Skadi also encodes a helicase loader protein similar to the λ-like DNA replication machinery, without virally encoded genes for helicases or primases, which in addition to the lack of a terminator sequence, suggests that Skadi utilises a λ-like 'rolling-circle' DNA replication module. The genome region encoding late genes contains virus structure proteins, namely: three internal virion proteins, major capsid protein, tail tube proteins A and B, tail fibre, head-tail connector and prohead-protease. Within the direct terminal repeat region, that was identified by mapping Skadi reads back to the assembled genome, is a conserved TerL gene for packaging. For cell lysis, Skadi encodes a peptidase M15, similar to other HTVC010P-type pelagiphages (Du *et al.*, 2021).

Alignment of HTVC010P-type genomes showed that regions related to DNA replication and metabolism, as well as the terminase and phage head related proteins are conserved, using recommended amino acid identity (AAI) thresholds of >30% (Turner *et al.*, 2021) (Figure 4.2). In contrast, the region of the genome encoding virion core proteins and tail fibre related genes had an AAI identity of <30% between HTVC010P-types and our phages. Previous analyses of long-read sequencing based viromes from the Western English Channel resolved the microdiversity of HTVC010P across niche-defining genomic islands, revealing a hypervariable region of about ~5000 bp in length that contained a ribonuclease and an internal virion protein (Warwick-Dugdale *et al.*, 2019). Such regions of low coverage occur when reads from a diverse population fail to recruit to the genome of a specific strain and are indicative of genetic population variance (Coleman *et al.*, 2006). Mapping short reads from an axenic Skadi culture against the Skadi

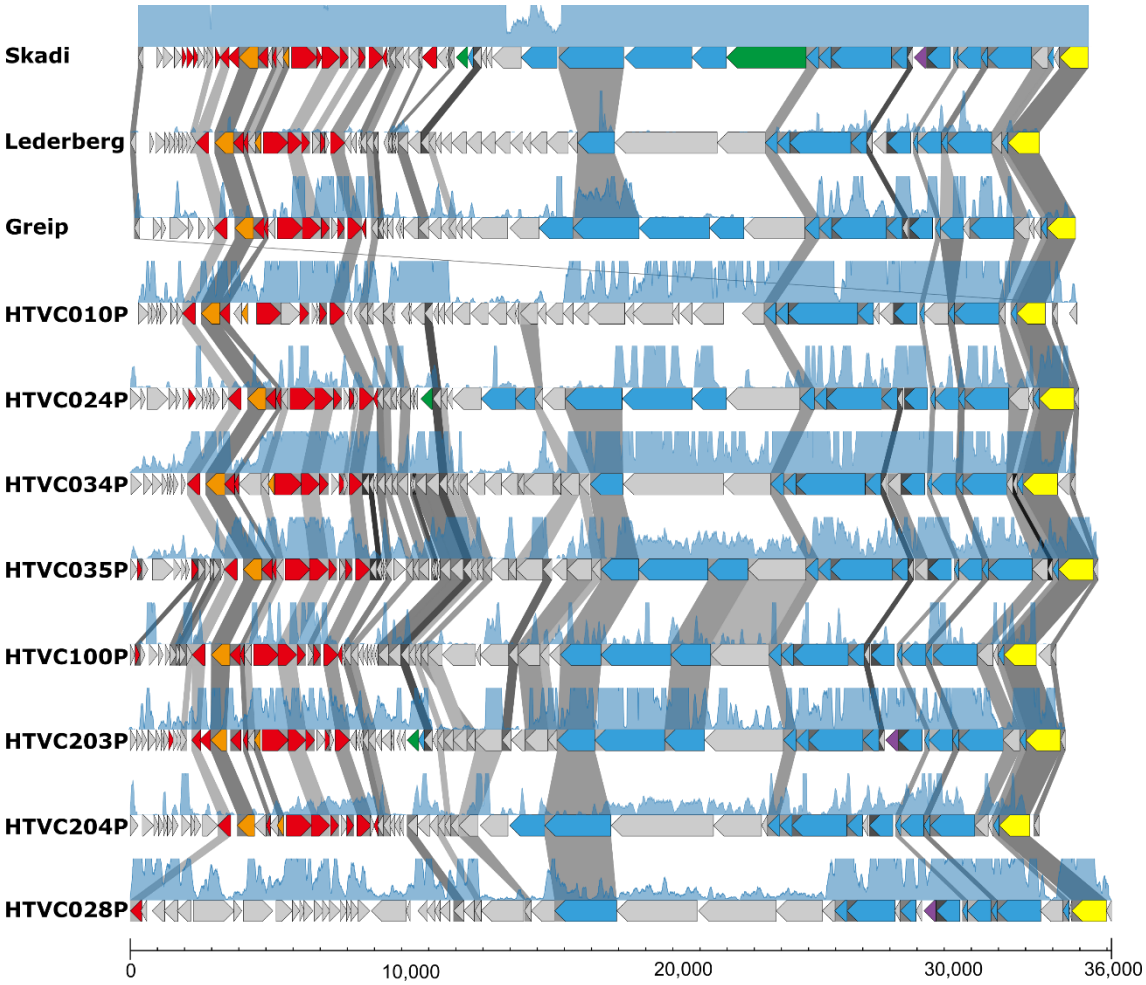


Figure 4.2 Genomic map of all members of the proposed *Ubiqueviridae* (HTVC010P-types and Skadi). Arrow directions represents coding strand (positive/negative), shading connecting the open reading frames (ORFs) indicate identity between shared genes. The filled blue line graph indicates per nucleotide coverage (capped at 100) based on Global Ocean Virome (GOV2) reads mapped against phage genomes.

genome also returned a ~5000 base pairs long region with low coverage that perfectly aligned with three ORFs encoding internal virion core associated proteins, but did not encode a ribonuclease (Figure 4.1A). This suggests that Skadi and other HTVC010P-like phages may generate diversity within their internal virion core at a population level even in seemingly axenic cultures, presumably enabling intra-population mutant strains to either switch to another host or to overcome host resistance through Red Queen dynamics. Mapping reads (randomly subsampled to 5 million reads per virome) from all environmental GOV2 viromes against HTVC010P-type genomes revealed an average per nucleotide coverage of 4,745 for Skadi. The region between 13,646 and 15,713 bp returned per nucleotide coverage of less than 100, marking it as a putative

hypervariable region (HVR) using previously established cut-offs (Mizuno *et al.*, 2014). In contrast to the low coverage determined by mapping culture-derived reads, the internal virion core protein region recruited high coverage of on average 1493 per nucleotide from GOV2 reads, implying that intra-population genomic variety is generated through different mechanisms between cultured and environmental Skadi populations, possibly due to different selection pressures introduced by artificial and natural environments, respectively (Figure 4.2). Coverage of the other HTVC010P-like genomes by GOV2 derived reads was consistently low in the same respective genomic region, but the overall coverage for six out of 11 phages was too low to identify putative HVRs. The putative HVR encoded by Skadi, as determined by GOV read mapping, spanned two ORFs encoding for an uncharacterized phage protein and a structural lectin-domain containing protein. As discussed above the lectin-folds in association with tail fibre proteins have been shown to be involved in host receptor binding and cell attachment (Salazar *et al.*, 2019), which suggests that hypervariable cell recognition and attachment on a species level is a feature of global Skadi populations. Presumably, altered cell binding would allow the phage to maintain an advantage over the defence mechanisms of host populations that can rapidly evolve outer membrane receptors in order to avoid viral predation, or it allows Skadi to infect a broader range of hosts.

4.3.3. *Skadi and other HTVC010P-type pelagiphages constitute a novel viral family within the* Caudovirales order

Analysis of the average nucleotide identity between Skadi and all other known *Pelagibacter* phages showed that Skadi's closest relatives were the phages Greip (ANI 37%) (Table 2.1), Lederberg (ANI 33%) (Appendix 2) and the highly abundant HTVC010P-type pelagiphages (ANI 33%) (Zhao *et al.*, 2013; Du *et al.*, 2021). The average nucleotide identity between all these phages was <60%, which based on established ICTV thresholds suggest that each phage are sole representatives of separate genera (Turner *et al.*, 2021). Other known pelagiphages with podovirus morphology were <5% identical. A phylogenetic tree based on complete genome-based proteomics placed Skadi and HTVC010P-type phages on a monophyletic branch distinct from other phage groups,

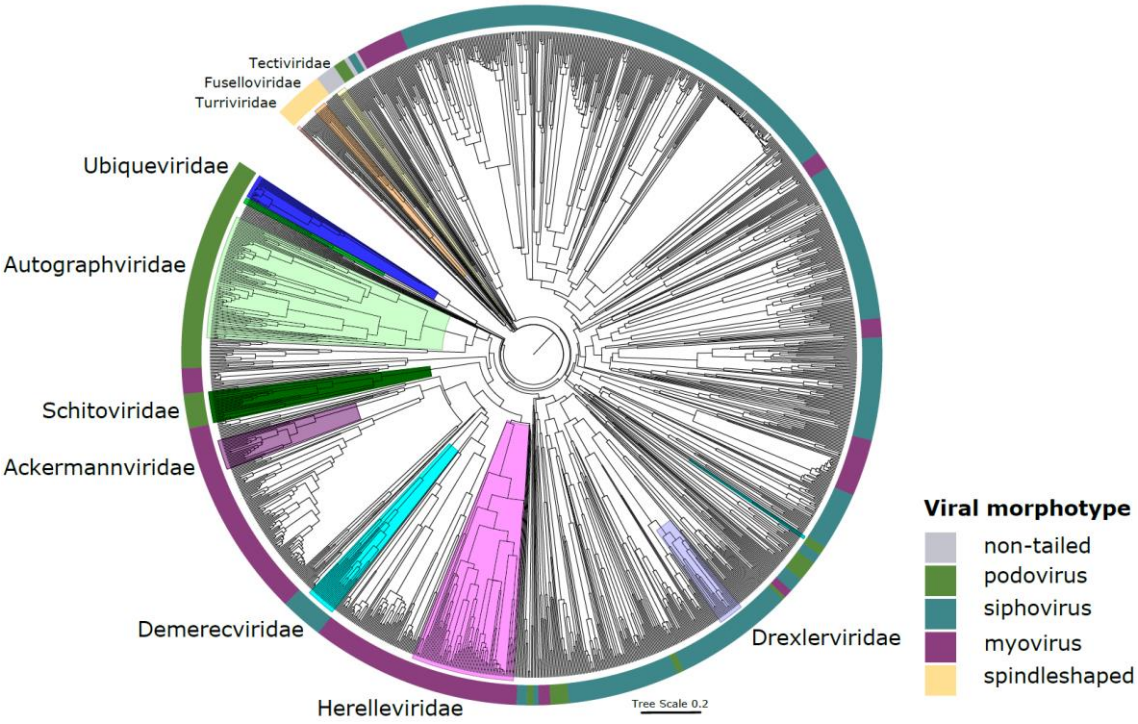


Figure 4.3. **Proteomics tree for *Pelagibacter* phages**. All full length genomes were screened against DB-B:Baltimore Group Ib prokaryotic and archaeal, calculated using GRAViTy 1.1.0 (http://gravity.cvr.gla.ac.uk/, accessed on 2 June 2021). Outer ring represents the main virus morphotypes, branch highlights indicate ICTV recognised viral families. Blue highlighted branch represents the proposed *Ubiqueviridae* family.

providing support for a family-level taxonomic assignment (Figure 4.3). Comparative phylogenomic analysis against all dsDNA prokaryotic and archaeal viruses within the DB-B:Baltimore Group Ib (accessed via http://gravity.cvr.gla.ac.uk/, 2 June 2021) suggested that there are no other cultivated phages in the database which are closely related to Skadi and HTVC010P-type phages (Figure 4.3). Using arbitrarily chosen representatives (17 total) of the *Schitoviridae* and *Autographviridae* families, which are the two main ICTV recognised families with podovirus morphology, together with all 40 *Pelagibacter* phages for full-length genome based phylogenetic analysis also placed Skadi and HTVC010P-type phages on a well-supported branch distinct to all other branches (Figure 4.1D). This suggests that these phages are members of the same family-level taxonomic group, for which we suggest the name *Ubiqueviridae* – derived from the latin word *ubique* (“everywhere”) to reflect both the high abundance and ubiquitous distribution of these phages as well as in recognition of their common host, *Pelagibacter ubique*.

Further evidence that *Ubiqueviridae* represents a distinct taxonomic cluster from other pelagiphages was provided by hypergeometric probability of shared protein clusters (Figure 4.4A), showing high probability of shared genes within groups corresponding to pelagiphage types, with *Ubiqueviridae* forming a single viral cluster. The taxonomic assignment tool vConTACT2 placed each HTVC'-type and the *Ubiqueviridae* into separate viral clusters. Visualising the vConTACT2 shared gene network analysis with all RefSeq phages showed that podophages infecting SAR11, as well as the pelagisiphophage Kolga, share genes between pelagiphage viral clusters at a higher level compared to pelagiphages and viruses of other hosts. PIRATE analysis (Bayliss *et al.*, 2019) identified ten shared core genes within all eleven members of the proposed *Ubiqueviridae* family (~16% of ORFs in Skadi) (Supplementary Table 4.1). *Ubiqueviridae* did not share core genes with podovirus pelagiphages of the HTVC019P-type genomes (part of the *Autographviridae* family) or taxonomically unclassified HTVC111P-types, HTVC103P-types or HTVC023P-types. One shared core gene (*TerL*) was found between *Ubiqueviridae* and HTVC106P representing the singleton HTVC106P-type, however, phylogenetic analysis of *TerL* shows that the *Ubiqueviridae* genes form a monophyletic branch, with HTVC106P as a closely related but separate branch (Supplementary Figure 4.1). Additional phylogenetic analysis on individual structural genes (tail tube and major capsid proteins) that are commonly found in dsDNA podophages also supports the proposed *Ubiqueviridae* as a monophyletic group and places other pelgiphage types onto distinct and well-supported monophyletic branches (Supplementary Figures 4.2 and 4.3). Other pelagiphage types had 15 to 33 shared core genes in their respective groups (Supplementary Table 4.2). No core genes were shared between *Ubiqueviridae* and siphophages (*Pelagibacter* phage Kolga), nor with myophages (*Pelagibacter* phages Melnitz and HTVC008M) that also infect SAR11.

4.3.4. *Global distribution patterns reveal Skadi as a polar pelagiphage ecotype*.

Skadi was observed almost exclusively in high latitude surface waters in both the Arctic and Southern Ocean, despite the fact that both bodies of water are separated by the Atlantic and Pacific Ocean, respectively (Figure Skadi 4.5A). Skadi was near or below limits of detection in viromes from lower latitudes (below

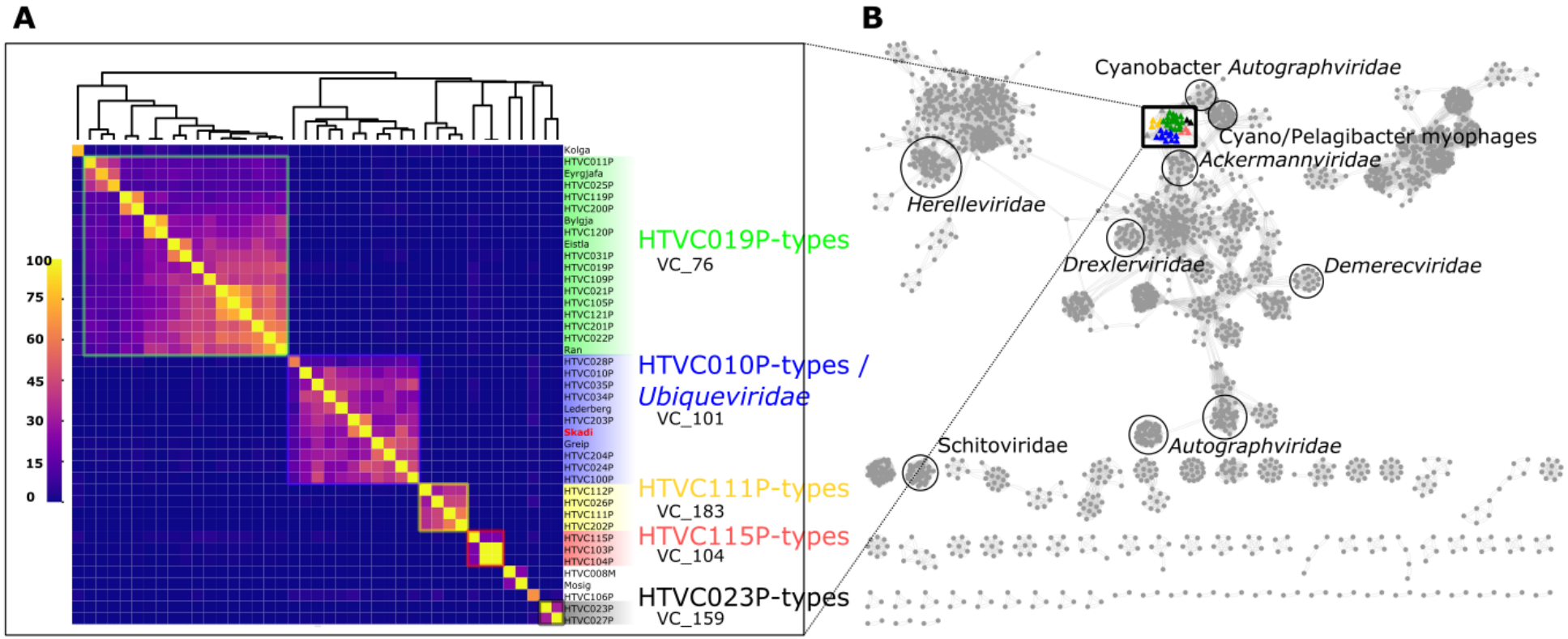


Figure 4.4 *Pelagibacter* phages can be grouped into distinct clusters. **A** Heatmap of hypergeometric probability of shared genes between phages isolated on SAR11 hosts based on protein clusters identified by vConTACT2, highlights indicate corresponding vConTACT2 viral cluster assignments, no highlight indicates singleton viral clusters in vConTACT2 gene sharing analysis. Hypergeometric probability of shared genes between Skadi-like and HTVC010P-like phages suggests both are relatively closely related. **B** Visualised vConTACT2 analysis of shared gene network. The location of genomes of ICTV recognised bacteriophage families and genomes of phages infecting SAR11 with podovirus morphology are indicated, coloured nodes represent viral cluster identified by vConTACT2 and correspond to viral clusters colour coded in (A).

66° absolute latitude). Based on the absence of Skadi populations at lower latitudes, we speculate that it is unable to replicate sufficiently to maintain sizable populations in these waters. Therefore, the Atlantic and Pacific Oceans form an evolutionary barrier separating southern and northern populations. As ^{14}C carbon isotope analysis has shown that a single "parcel" of seawater can take up to 1,000 years to be transported around the world by global ocean circulations (Gebbie and Huybers, 2012), the high abundance of genetically similar Skadi populations in both polar regions (based on read recruitment) suggests that they form conserved populations despite the implied relatively low rates of direct genetic exchange. This supports the viral Seed-Bank model for SAR11 virus-host systems, which suggests that marine viral genotypes are ubiquitously distributed albeit relatively rare, but can become abundant when environmental conditionsenable efficient replication in available suitable hosts (Breitbart and Rohwer, 2005). Based on the study of T4-like marine cyanophages, it has been suggested that the abundance of virus populations reflects that of their hosts (Chow *et al.*, 2014). Using single amino acid variants (SAAVs) Delmont et al. estimated the biogeography and relative abundance of SAR11 populations and identified *Pelagibacter ubique* HTCC1062 as a cold-water ecotype and *Pelagibacter bermudensis* HTCC7211 as a warm-water ecotype (Delmont *et al.*, 2019). Like Skadi and Greip, HTVC010P was isolated on the cold-water *Pelagibacter ubique* HTCC1062, but did not have any discernible spatial distribution patterns and recruited a broadly uniform number of reads across all latitudes (Figure 4.5B). Unfortunately, experimental evidence of the host range for HTVC010P is not available. However, given it was isolated from the subtropical Sargasso Sea, where warm-water SAR11 ecotypes dominate, it is highly likely that HTVC010P infects both warm- and cold-water ecotypes, marking it as a generalist phage which would explain its uniform distribution across the global oceans. In contrast, host range experiments with Skadi and Greip against the warm- and cold-water ecotypes of SAR11 showed that both phages infect the cold-water SAR11 ecotype exclusively (Figure 2.2), supporting the hypothesis that global distributions of Skadi should follow those of its host HTCC1062 (Figure 4.5C), similar to observations made in cyanobacteria virus-host systems (Marston *et al.*, 2013). Low abundance of Skadi at temperate and tropical latitudes can most likely be explained by a lack of suitable hosts for efficient

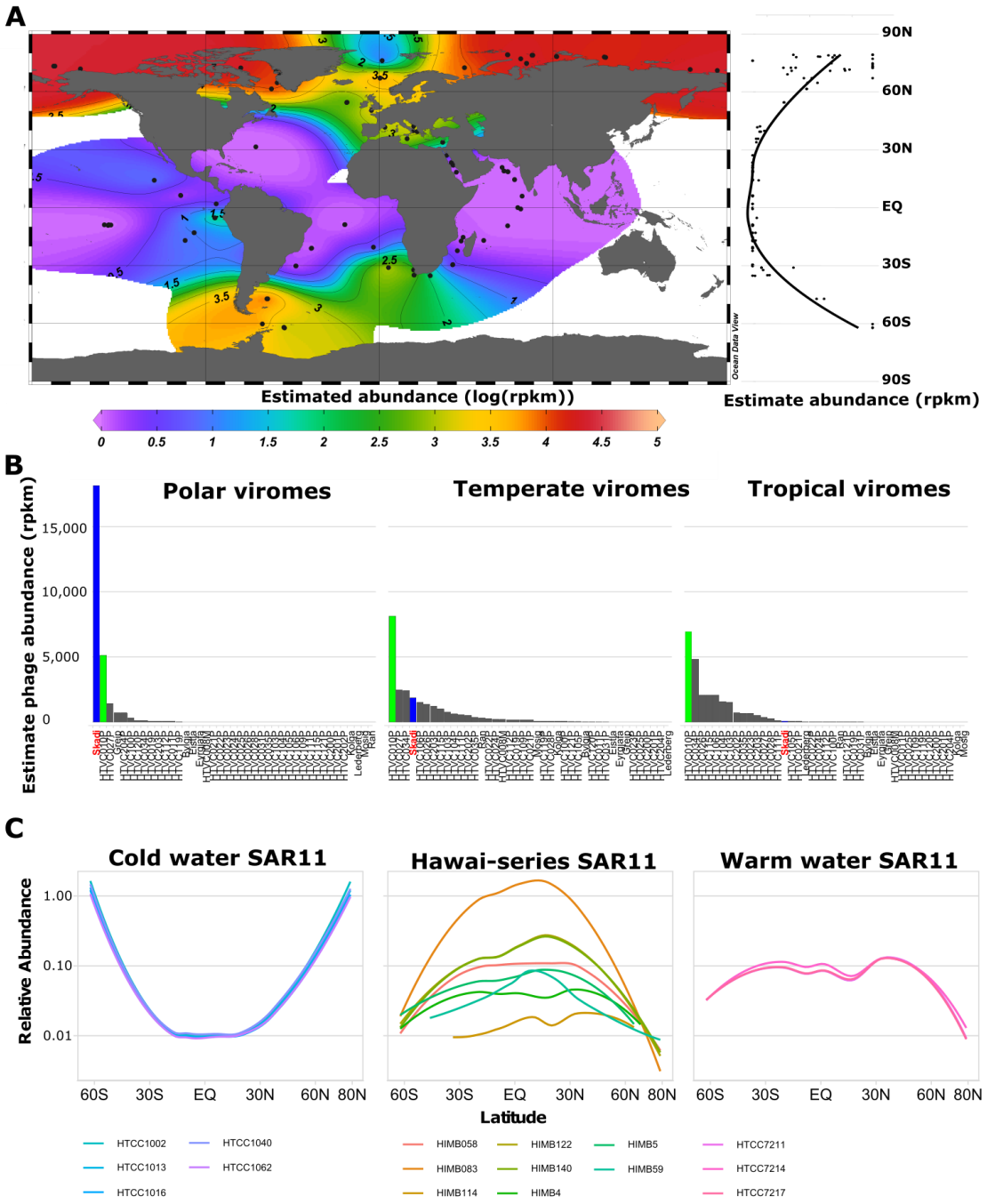


Figure 4.5 Global distribution pattern of *Pelagibacter* phage Skadi A Estimated abundance in reads per genome kilobase mapped per million reads (RPKM) of Skadi's full length genome from the GOV2 dataset; **B** Estimated abundance (RPKM) of isolated phages infecting SAR11 (Skadi is coloured blue, HTVC010P is colour in green) grouped by absolute latitude into "Polar" (90° to 66°), "Temperate" (66° to 33°) and "Tropical" (33° to 0°) regions; **C** Relative abundance of SAR11 bacterial strains based on single amino acid variants from Delmont et al. (Delmont *et al.*, 2019).

replication. Therefore, successful, high-abundance pelagiphages can be categorized as either generalist phages (HTVC010P) or specialist phages such as Skadi.

We used competitive recruitment of short reads from marine viromes (GOV2) to evaluate relative abundances of all 41 isolate genomes of phages infecting SAR11 across global oceans (≥95% nucleotide identity over 90% of read length, with 40% minimum genome coverage threshold to reduce false positives (Roux *et al.*, 2017)). Skadi recruited a mean 6,442 RPKM (4556 – 8322, 95% CI, bootstrapped mean; n=1000) from all GOV2 viromes compared to the mean 6,561 RPKM recruited by HTVC010P (5316 – 7893, 95%, bootstrapped mean; n=1000). A value at least as extreme as the measured difference was observed in ~45% of bootstraps where the null-hypothesis (no difference) was assumed. The bootstrapped 119 RPKM difference in means was not significant (-2289 – 2444, 95% CI, n=1000). The median RPKM of HTVC010P (3100 - 7713, 95% CI, bootstrapped mean; n=1000) was significantly higher compared to the median 271 RPKM in Skadi (0 – 1673, 95% CI, bootstrapped mean; n=1000) by an average median of 5,141 RPKM (2699-7530 95% CI, bootstrapped mean; n=1000). Therefore, Skadi and HTVC010P have similar global relative abundance, but this is driven by the extremely high relative abundance of Skadi at the pols compared to HTVC010P. In contrast, mean read recruitment by the 39 remaining pelagiphage genomes (269 RPKM on average) ranged from 0 to 2,562 RPKM. Mean abundance of Skadi was significantly higher compared to the mean 2,562 RPKM of HTVC034P (1887 – 3323 RPKM, 95% CI; bootstrappted mean; n=1000), as the next highest recruiting pelagiphage. The mean difference between Skadi and HTVC034P was 3,902 RPKM (2088 – 5776 RPKM, 95% CI, bootstrapped mean; n=1000), indicating that Skadi and HTVC010P are the two most abundant pelagiphages in the oceans by about two to up to 20-fold.

When plotting relative pelagiphage species abundance (normalized by cubic root) as ranked abundance curve (Figure 4.6), two points became evident: (1) Polar pelagiphage diversity appears low compared to low latitudes, with only three pelagiphage species above detection limits in the Southern Ocean. In contrast, general polar microbial and viral diversity has been reported as unusually high (de Pascale *et al.*, 2012; Gong *et al.*, 2018; Gregory *et al.*, 2019), what might hint at a large reservoir of undiscovered pelagiphage diversity in polar regions; and (2) the ranked abundance curve is heavily right skewed, suggesting that there are ecological constraints which prevent a large proportion of pelagiphage species to become dominant. Skadi relative abundance was significantly

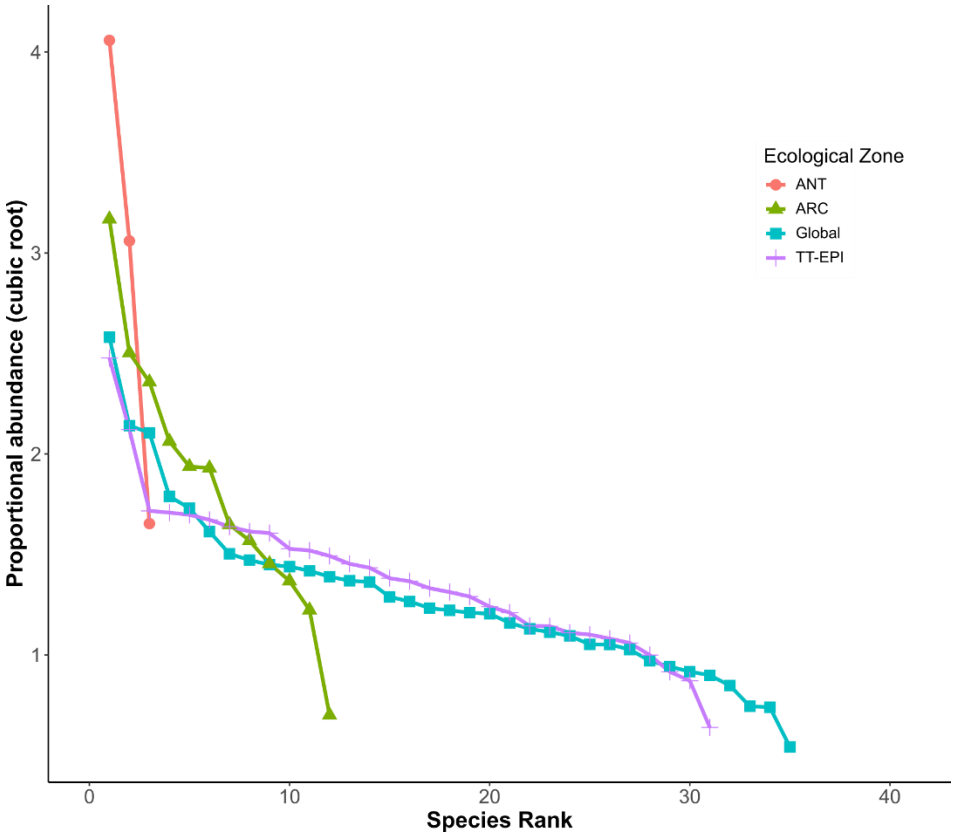


Figure 4.6 Pelagiphage species rank by ecological zone. Ranked Relative abundance curve for all pelagiphage species above detection limit in GOV2 viromes, partitioned into ecological zones as assigned in GOV2 metadata. Relative abundance (normalised by cubic root) in the Arctic (ARC) with 12 species, and Antarctic with three species (ANT), suggestive of low pelagiphage diversity and might indicate hidden polar viral ecotypes exist.

negatively correlated to temperature (RPKM~temperature linear regression, $p < 0.001$, $R^2=0.69$) (Supplementary Figure 4.4). In contrast, relative abundance of HTVC010P and the majority of other pelagiphages (21 out of 28 pelagiphages with non-zero RPKM in >3 viromes) was not significantly correlated to temperature ($p= 0.06$, $R^2=0.08$ for HTVC010P, $R^2 < 0.5$, $p \geq 0.05$ for other pelagiphages). *Pelagibacter* phage HTVC027P was the only phage other than Skadi to have a significant, but weak, negative correlation with temperature ($p < 0.001$, $R^2=0.27$). *Pelagibacter* phage HTVC023P significantly and positively correlated with high temperatures ($p=0.003$, $R^2=0.59$), suggesting an association with warm-water SAR11 ecotypes, though the original host is the cold-water HTCC1062, indicating potential generalism in HTVC023P. Noticeably, the abundance of *Pelagibacter* phage Greip, which was identified as a potential polar ecotypic phage previously (Section 2.2), did not significantly correlate to temperature ($p=0.38$, $R^2=0.08$), yet was found only in samples taken in the Arctic

with temperatures of <5 °C. It is worth highlighting that Greip was isolated on HTCC1062 in the WEC at an ambient surface temperature of 14.8 °C, and was below detection limit in a WEC virome (Warwick-Dugfale *et al.*, 2019), hence likely does not replicate on local SAR11 populations in this environment. Yet, despite the high abundance of Skadi and HTVC010P as competitors, the deployed enrichment method managed to capture this locally rare phage in cultures kept at 15 °C. This also means that temperature in itself does not inhibit viral replication for this particular putative ecotypic specialists. Overall, the correlation analysis suggests that Skadi is a polar pelagiphage ecotype similar to the genomically related *Pelagibacter* phage Greip, whereas the majority of other pelagiphages do not display strong niche-associated geographical distribution.

An intriguing exception to the otherwise polar distribution pattern for Skadi was found in three GOV2 samples (100_DCM, 102_DCM, 102_MES), from which Skadi recruited between 517 and 912 RPKM. These samples were taken in the eastern parts of the Pacific Equatorial Divergence Province (PEOD) and the South Pacific Subtropical Gyre Province (SPSG) in Autumn 2011 (Pesant *et al.*, 2015). An important feature of this region is the Humboldt Current System (HCS), which is a cold-water current flowing from Antarctica in parallel to the South American west coast. This causes upwelling of nutrients along the coastline, creating one of the most productive regions in the ocean, but also feeds cold water from the Southern Ocean into the Pacific Ocean (Montecino and Lange, 2009; Igarza *et al.*, 2019). Speculatively, dominant Skadi populations in the Southern Ocean were transported via the HCS into the Southern Pacific Ocean, as HCS water and microbiomes mix with the surrounding water, Skadi populations are replaced by other phages like HTVC010P as the most dominant pelagiphages (4,829 to 5,191 RPKM).

As Skadi was only the 8th to 14th most abundant pelagiphage in the Southern Pacific, compared to the Southern Ocean where Skadi was the most abundant pelagiphage, it is likely that Skadi populations are either outcompeted by phages that are better adapted to local conditions, or that the range of regional SAR11 population do not provide enough host cells for Skadi to maintain its dominance. In this scenario, if the population dynamics conform to the Royal-Family model (Breitbart *et al.*, 2018), the Southern Pacific Skadi population is likely in a state of decay and being succeeded by the closely related HTVC010P. This is an

intriguing insight into how ocean currents and viral communities might interact, providing evidence for a viral Royal-Family Seed-Bank model controlled by ecological niche specialisation. We suggest that Skadi serves as an exemplar model organism for how ocean currents shape ecotypic viral communities under viral Seed-Bank dynamics.

4.3.5. *Ecotypic pelagiphages may be an important aspect of global viral communities*

Despite using water from a temperate and coastal site, population variants of Skadi were isolated eleven times from three water samples, suggesting even at lower latitudes, it was sufficiently abundant in the viral Seed-Bank to be successfully isolated through enrichment. In contrast, HTVC010P was not isolated, either in this study, or in other efforts to isolate viruses from waters around the world on SAR11 host HTCC1062 (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Buchholz *et al.*, 2021a; Buchholz *et al.*, 2021b; Du *et al.*, 2021), despite being sufficiently abundant enough to be detected in 75% of all marine viromes. Assuming the probability to isolate any phage is proportional to its abundance, and isolation can only be a success or failure, using the global average abundance from GOV2 samples as isolation likelihood and number of isolates on HTCC1062 as trials, a Bernoulli trial test suggests there is only a 3% likelihood that isolations randomly missed HTVC010P. Though this is an oversimplification, it does raise a curious question: Why have we not re-isolated the most abundant pelagiphage on its original host? Since HTCC1062 has been used as bait for over a decade, it is likely that other factors have prevented the re-isolation of HTVC010P. Ordinations, using normalised (cubic root) RPKM values from the GOV2 data for pelagiphages, constrained with environmental parameters showed that HTVC010P distribution was driven by high salinity, low oxygen and low nutrient concentrations (Figure 4.7). These are commonly observed parameters of ultraoligotrophic subtropical gyres, which cover up to 40% of the planet (Signorini and McClain, 2012). This may indicate that HTVC010P is a specialist for pelagic/ultraoligotrophic SAR11 rather than a generalist as the global distribution would indicate. This might help explain the failure to re-isolate

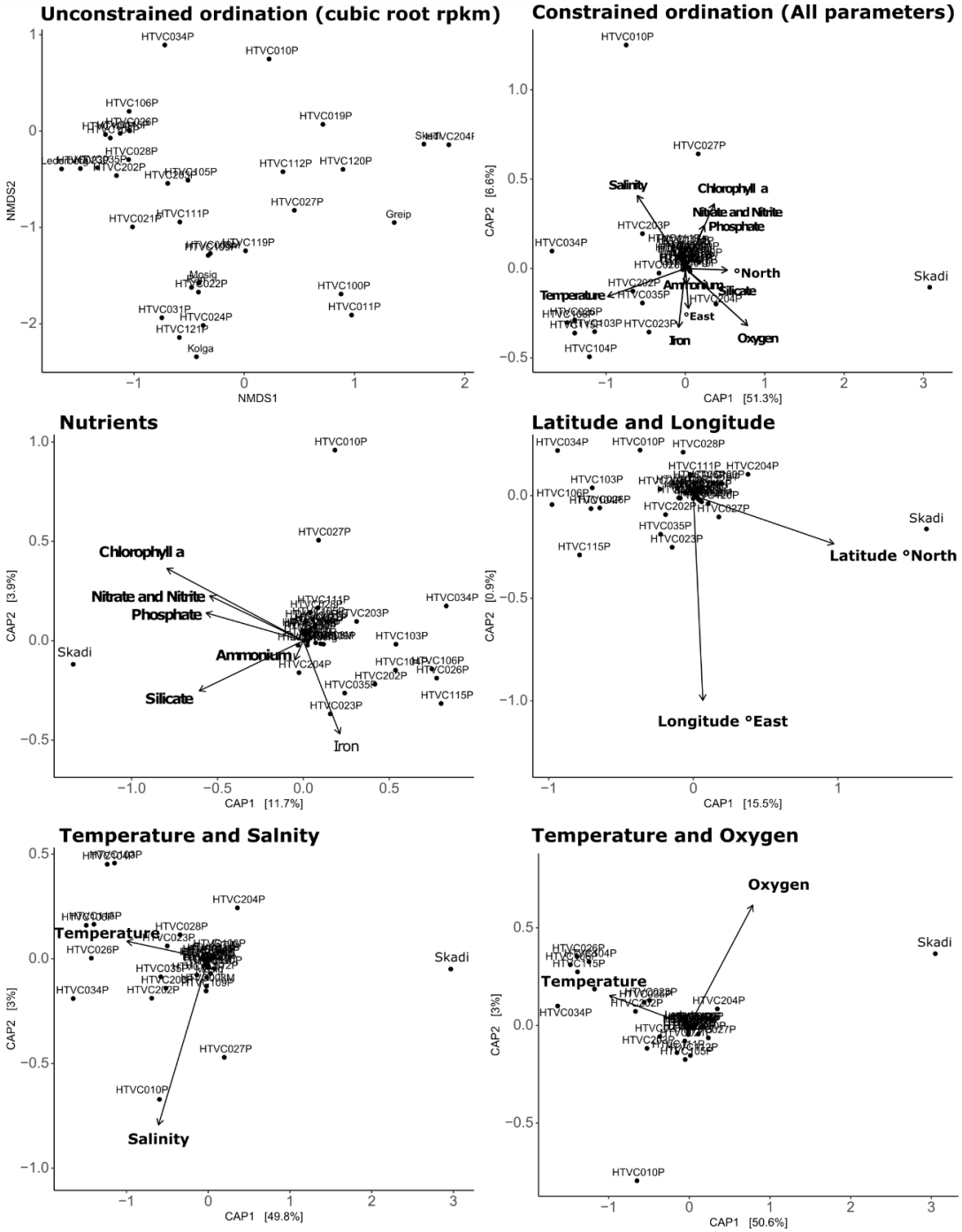


Figure 4.7 Skadi's abundance is driven by typical cold-water parameters. Plot shows ordinations of RPKM values obtained from GOV2 viromes constrained with different environmental parameters provided by the GOV2 metadata. Skadi abundance is driven by low temperature, high latitude and high nutrient concentrations, HTVC010P abundance is driven by low oxygen and high salinity, but not temperature

HTVC010P, as laboratory conditions generally provide an excess of all required nutrients, which could favour other phages over a putative ultraoligotropic HTVC010P ecotype. This might fit the cosmopolitan distribution patterns observed for HTVC010P as well (Figure 4.5B) and would help explain the

dominance of a single species across most oceans - with the exception of polar oceans, which are generally considered high-nutrient regions but also supported a large HTVC010P population. An alternative explanation is that the original host of HTVC010P (HTCC1062) is not reflective of the preferred host in the environment. HTCC1062 has been continuously cultured since its isolation in 2002 (Rappé *et al.*, 2002), and by adapting to laboratory conditions (Pratt and Waddell, 1959), might be have become a sufficiently different genotype compared to the wild-type strain to allow infection by a putative rare mutant of a generalist phage HTVC010P. However, without the original culture of HTVC010P, which has been lost (personal correspondence), the host range will likely remain unknown.

Cold-water environments are often defined as areas with average annual temperatures below 15 °C (Miller and Whyte, 2013), with polar oceans temperatures between +5 °C and -2 °C, which allows for up to double the available oxygen concentrations compared to water at 20 °C (Still *et al.*, 2014). The Arctic Ocean also experiences high riverine influx of terrestrial nutrients (Terhaar *et al.*, 2021), and subsequently high primary production (Arrigo and van Dijken, 2015) (for which chlorophyll *a* serves as a proxy). Constrained ordinations provide further evidence that Skadi abundance is correlated with high latitude and/or associated low temperature (Figure 4.7), but also suggest that high oxygen and increased concentration of nutrients and chlorophyll *a* correlate with increasing Skadi abundance. Absolute longitude did not impact ordinations for any pelagiphage, indicating that distribution patterns are a function of shared environmental conditions associated with geographical provinces, but not the marine geographical provinces themselves. This is in line with our observations of bi-polar Skadi populations that are putatively connected via the global conveyor belt, as ocean currents would continuously mix global viral communities, maintaining the Seed-Bank. These results therefore support Skadi as a polar specialist. High temperatures correlated positively with relative abundance of five other (lower abundance) pelagiphages in addition to HTVC023P (HTVC103P, HTVC104P, HTVC115P, HTVC106P, HTVC026P). However, 18 out of 41 pelagiphage genomes did not recruit reads from enough viral metagenome samples (non- zero RPKM in five or more metagenomes) to establish any robust patterns, indicating that these phages were part of the respective Seed-Bank at

sampling locations. Considering that the Seed-Bank model is addressing the turnover of dominant phages, the fact that across space and time both HTVC010P and Skadi remain dominant, suggesting that the aforementioned ecological constraints maintain the host population and subsequently the dominance of a small group of phages – conforming with the Royal-Family model (Breitbart *et al.*, 2018). Deep-sequencing of viral *phoH* genes at the same site in the Sargasso Sea also suggested that the majority of operational taxonomic units (OTUs) remained rare throughout the sampling period, with a small number of OTUs dominating throughout the seasons, depths and years (Goldsmith *et al.*, 2015). Similarly, in SAR11 virus-host systems our results could mean that abiotic factors shaping SAR11 ecotypic niche partitioning also indirectly maintain the dominance of a small number of phages out of a large diverse pelagiphage community akin to pelagiphage ecotypes.

4.4. Conclusion

In this study we isolated and characterized the highly abundant pelagiphage Skadi, marking it as one of the most abundant phages infecting the ubiquitous SAR11 clade. Phylogenetic analysis and comparison to other pelagiphages shows that Skadi and HTVC010P-type phages are a monophyletic and distinct taxonomic unit for which we propose the creation of a novel viral family named *Ubiquiviridae*. Using Skadi and HTVC010P as examples, we show that both putative generalist and specialist pelagiphages can be successful and dominant. Our metagenomic analysis of the GOV2 viromes shows that pelagiphage distribution is ubiquitous as thought previously, but the majority of pelagiphage species remain in the low abundance fraction, conforming to the Seed-Bank model. Skadi abundance is strongly correlated to polar environmental conditions, marking it as a pelagiphage cold-water specialist. The high abundance of Skadi at both poles further suggests that SAR11 ecotypic niche partitioning indirectly shape the pelagiphage community akin to pelagiphage ecotypes, but also that the dominance of this viral species is being maintained across space and time as suggested by the Royal-Family model. Furthermore, considering the virome sampling locations across global oceans, observations of increased abundance for Skadi populations along the Humboldt Current Systems demonstrates how

ocean currents could potentially maintain the viral Seed-Bank globally. If polar oceans keep increasing in temperature due to escalating climate concerns, a shift in the polar microbial community is likely (Comeau *et al.*, 2011). By uncovering polar ecotypic distribution of an abundant pelagiphage, our results suggest that polar pelagiphage communities are under threat from climate change as well. The majority of reports concerning climate change focus on macroecology, though reports considering viral community relations under changing climate (Danovaro *et al.*, 2011) suggest that virus-host dynamics will shift. This could potentially lead to wide-scale alterations of biogeochemical cycles as well as the loss of viral diversity.

4.5. Materials and Methods

4.5.1. *SAR11 hosts and growth conditions*

The SAR11 strains *Pelagibacter* HTCC1062 and HTCC7211 were kindly provided by the Giovannoni lab (Oregon State University, USA) and grown in the Artificial Seawater Medium (ASM1) (Carini *et al.*, 2013) amended with 1 mM NH_4Cl, 10 µM KH_2PO_4, 1 µM $FeCl_3$, 100 µM pyruvate, 25 µM glycine, 25 µM methionine as well as 1 nM HMP, pantothenate, Biotin, Pyrroloquinoline quinone PQQ and Vitamin B12 each (Carini *et al.*, 2014). Continuous cultures were cultivated in 50 mL acid-washed (10% HCl) polycarbonate flasks at 15 °C without shaking.

4.5.2. *Water sampling and viral isolation*

1 L of seawater was collected in rosette-mounted Niskin bottles at a depth of 5 m from the Western English Channel Observatory coastal station L4 (WCO; http://www.westernchannelobservatory; 50°15.00N; 4°13.00W) on the 2018-09-24. Seawater was transferred immediately to a clean 2 L acid-washed polycarbonate (PC) Nalgene bottles (Thermo Fisher Scientific, Waltham, USA) and placed in a cooler box (Igloo, Katy, USA) at ambient temperature. Upon return to shore, water was transported to the University of Exeter for immediate processing (Two hours maximum duration from collection to processing). The sample was then processed and used for viral isolation on the host strain

HTCC1062 as described previously (see Section 2.5). Briefly, larger plankton was removed from the sample using a series of filter (GF/D 2.7 µm filters, 142mm 0.2 µm PC filter), and concentrated to ~50 mL with a 50R VivaFlow tangential flow filtration unit with a 100 kDa Hydrosart membrane (Sartorius instruments, Goettingen, Germany). The concentrate was filtered through a 0.1 µm pore PC syringe filter to remove remaining small bacteria, and used as inoculum (10% v/v) to infect exponentially growing HTCC1062 cultures in 96-well Teflon plates (Radleys, UK), prepared with ASM1. Growth of all wells was monitored using standard flow cytometry on 20 µL sub-samples diluted in 180 of µL 1x PBS, stained with SYBR Green. No-virus controls were compared to virus-treated samples to spot lysis. Cells in wells where lysis was observed were removed using 0.1 µm pore PC syringe filters, and used to prepare a dilution-to-extinction series in 10-fold dilution steps, that were then used to infect freshly prepared host cultures in 96-well Telfon plates, prepared as described above. This process was repeated three times, before a viral culture was considered ready for downstream applications.

4.5.3 *Transmission Electron Microscopy of viral isolates*

For ultrastructural analysis, virus particles were transferred onto pioloform-coated electron microscopy (EM) copper grids (Agar Scientific, Standsted, UK) by floating the grids on droplets of virus-containing suspension for 3 min. Following a series of four washes on droplets of deionized water, the bound virus particles were contrasted with 2 % (w/v) uranyl acetate in 2 % (w/v) methyl cellulose (mixed 1:9) on ice for 8 min and the grids then air-dried on a wire loop after carefully removing excess staining solution with a filter. Dried grids were inspected with a JEOL JEM 1400 transmission electron microscope operated at 120 kV and images taken with a digital camera (ES 1000W CCD, Gatan, Abingdon, UK).

4.5.4. *DNA isolation and sequencing*

HTCC1062 cultures in 300 mL ASM1 medium (prepared as described above) were grown in 1 L PC flasks at 15 °C without shaking. When cell density reached 1×10^6 cells/mL, 30 mL of viral lysate were added. Infected cultures were

incubated for about one week. The cultures were transferred were transferred to 50 mL falcon tubes and centrifuged for 120 minutes (GSA rotor, Thermo Scientific 75007588) at 8,500 rpm/10,015 × g to remove larger cellular debris. To remove remaining cell fragments, supernatant was filtered through pore-size 0.1 µm PVVDF syringe filters. Viral particles were precipitated using a modified PEG8000/NaCl DNA isolation method (Solonenko, 2016). Briefly, per 50 mL lysate, 5 g PEG8000 and 3.3 g NaCl were dissolved using Falcon tubes and incubated on ice overnight. The precipitate was pelleted by centrifugation at 8,500 rpm/10,015 × g for 90 min at 4 °C. After discarding the supernatant, the pelleted virus particles were re-suspended by rinsing the tubes twice with 1 mL SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-Cl). DNA was cleaned and extracted using the Wizard DNA Clean-Up system (Promega) following manufacturer's instructions, and eluted from the resin-columns using nuclease free water (60 °C). DNA libraries were prepared and sequenced using the services of MicrobesNG (Birmingham, UK), opting for standard Nextera protocols, targeting minimum 30-fold coverage with Illumina paired-end [2 × 250 bp] sequencing on the HiSeq 2500. Raw reads were trimmed, quality controlled and error corrected using bbmap and tadpole (Bushnell *et al.*, 2017). Viral genomes were assembled using SPAdes v3.13 (Bankevich *et al.*, 2012) and evaluated with QUAST (Gurevich *et al.*, 2013). The trimmed reads were mapped back against the contigs for scaffolding using bowtie2 (Langmead and Salzberg, 2012), BamM (alignment 0.9, identity 0.95) (available at https://github.com/minillinim/BamM) and samtools (Li *et al.*, 2009).

4.5.5. *Genome annotation*

Viral contigs were confirmed with VirSorter (categories 1 or 2, >15kbp) and only accepted if the contig showed mean depth coverage an order of magnitude greater than the next highest recruiting contig (to filter out any cellular DNA carryover). Gene calls returned by VirSorter were imported into DNA Master for manual assessment and curation (Salisbury and Tsourkas, 2019). Additional gene calls were made using GenMark (Borodovsky and McIninch, 1993), GenMarkS (Besemer *et al.*, 2001), GenMarkS2 (Lomsadze *et al.*, 2018), GenMark.hmm (Zhu *et al.*, 2010), GenMark.heuristic (Besemer and Borodovsky,

1999), Glimmer v.3.02 (Delcher *et al.*, 2007), and Prodigal v.2.6.3 (Hyatt *et al.*, 2010). All gene calls were tabulated and compared using a scoring system which evaluates gene length, gene overlap and coding potential of ORFs (Salisbury and Tsourkas, 2019). ORFs were then annotated using BLASTp against the NCBI's non-redundant protein sequences (Pruitt *et al.*, 2007), Phmmer (Potter *et al.*, 2018) against UniProts UniProtKB and uniprotrefprot (UniProt Consortium, 2019), Swissprot (Bairoch and Apweiler, 2000) as well as InterProScan (Jones *et al.*, 2014) and Pfam (Finn *et al.*, 2014). Virally-encoded tRNAs were searched for with the web applications of tRNAScan-SE v2.0 (Lowe and Chan, 2016) and ARAGORN (Laslett and Canback, 2004). Genomes were scanned for riboswitches using the web application of RiboswitchScanner (Mukherjee and Sengupta, 2016). FindTerm (energy score < -11) and BPROM (LDF > 2.75) from the Fgenesb_annotator pipeline (Solovyev and Salamov, 2011) were used to predict promoter and terminator sequences, using default parameters. Promoter sequence 5'-TATAAAT-3' (Miller *et al.*, 2003; Geiduschek and Kassavetis, 2010) and MotA box (TGCTTtA) dependent promoters were predicted by aligning these sequences against the whole genome using a BLASTn search. Predicted promoter/terminator sequences that were intergenic or within 10 bp of the start/stop of ORFs were disregarded.

4.5.6. *Phylogenetic analysis*

For the phylogenetic analysis we followed the roadmap for phage taxonomy (Turner *et al.*, 2021), which was based on the most recent ICTV guidelines for viral taxonomy (Adriaenssens *et al.*, 2020). Species level genome comparisons were done using VIRIDIC v1.0 (Moraru *et al.*, 2020) with a species similarity threshold of 95% and genus threshold of 70%. Whole-genome based phylogenetic trees were constructed with VICTOR (formula d_6) through the online application available at https://ggdc.dsmz.de/victor.php (Meier-Kolthoff and Göker, 2017). Additionally, a proteomics tree for Pelagibacter phages against DB-B:Baltimore Group Ib prokaryotic and archaeal was created using GRAViTy 1.1.0 (http://gravity.cvr.gla.ac.uk/). For the shared gene network analysis, genes of all 41 known pelagiphage isolate genomes were called using default prodigal v2.6.3 (Hyatt *et al.*, 2010) and imported into the Discovery Environment 2.0 at

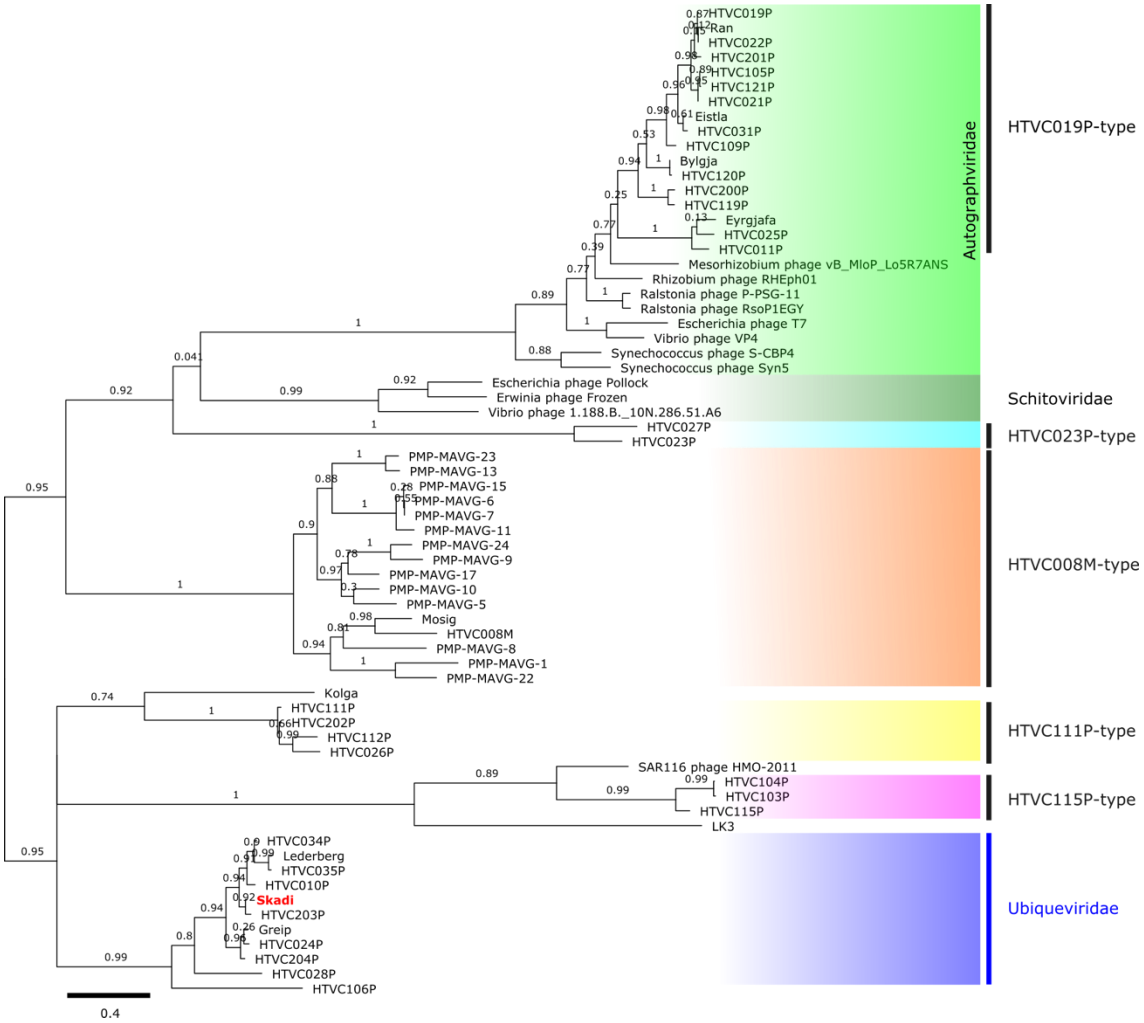
https://de.cyverse.org/, where vContact-Gene2Genome 1.1.0 was used to prepare protein sequences for analysis with vConTACT2 (v.0.9.8.) in default settings (Bolduc *et al.*, 2017). Cytoscape v3.7.2 (Shannon *et al.*, 2003) was used to visualize the results from vConTACT2. Single-gene based phylogeny was done within the phylogeny.fr webserver (Dereeper *et al.*, 2008), opting for default MUSCLE alignment (Edgar, 2004) and the built-in curer removing only positions with gaps to conserve more positions for constructing phylogenetic trees. For the phylogenetic tree building, maximum likelihood trees were constructed with PhyML (100 bootstraps, unless stated otherwise) (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006) and visualized using FigTree (v1.4.4 available at http://tree.bio.ed.ac.uk/software/figtree/). Prokka v1.14.6 (Seemann, 2014) in standard settings was used to create genome feature files (gff) of all Pelagibacter phages with podophage morphology, used as input for PIRATE v1.0.4 (Bayliss *et al.*, 2019) for identifying orthologous core genes shared between and within family level groups using 'PIRATE --steps "30,40,50,60,70,80,90" -k "--e 1E-5 --hsp-length 0.5" -r –a' flags based on ICTV recommended thresholds of >30% identity and >50% coverage for constructing a pangenome (Turner *et al.*, 2021). To identify variable regions in the genome, bowtie-2 (v.2.3.5.1) (Langmead and Salzberg, 2012) and sorted with samtools (v.1.11) (Li *et al.*, 2009) were used to map reads from the GOV2 dataset (Gregory *et al.*, 2019) with flags: 'bowtie2 –-seed 42 –-nondeterministic | samtools view –F 4 –bS' against full length genomes, which were rearranged so that the region encoding TerL was last in each genome respectively. Results were visualized with python 3.0 (www.python.org) and packages imported from the biopython project (Cock *et al.*, 2009). All figures were edited in Inkscape (www.inkscape.org) for aesthetics. Metagenomic reads were subsampled to 5 million reads for each virome using the reformat.sh within the bbmap suite. Bowtie2 v.2.3.5.1 (Langmead and Salzberg, 2012) with the flags 'bowtie2 –-seed 42 –-nondeterministic' was used to created indexes from all known pelagiphage genomes (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Buchholz, M. L. Michelsen, *et al.*, 2021; Buchholz, M. Michelsen, *et al.*, 2021; Du *et al.*, 2021) and sorted with samtools (v.1.11) using 'samtools view –F 4 –bS' flags (Li *et al.*, 2009).

4.5.7. *Metagenomic read mapping and global abundance analyses*

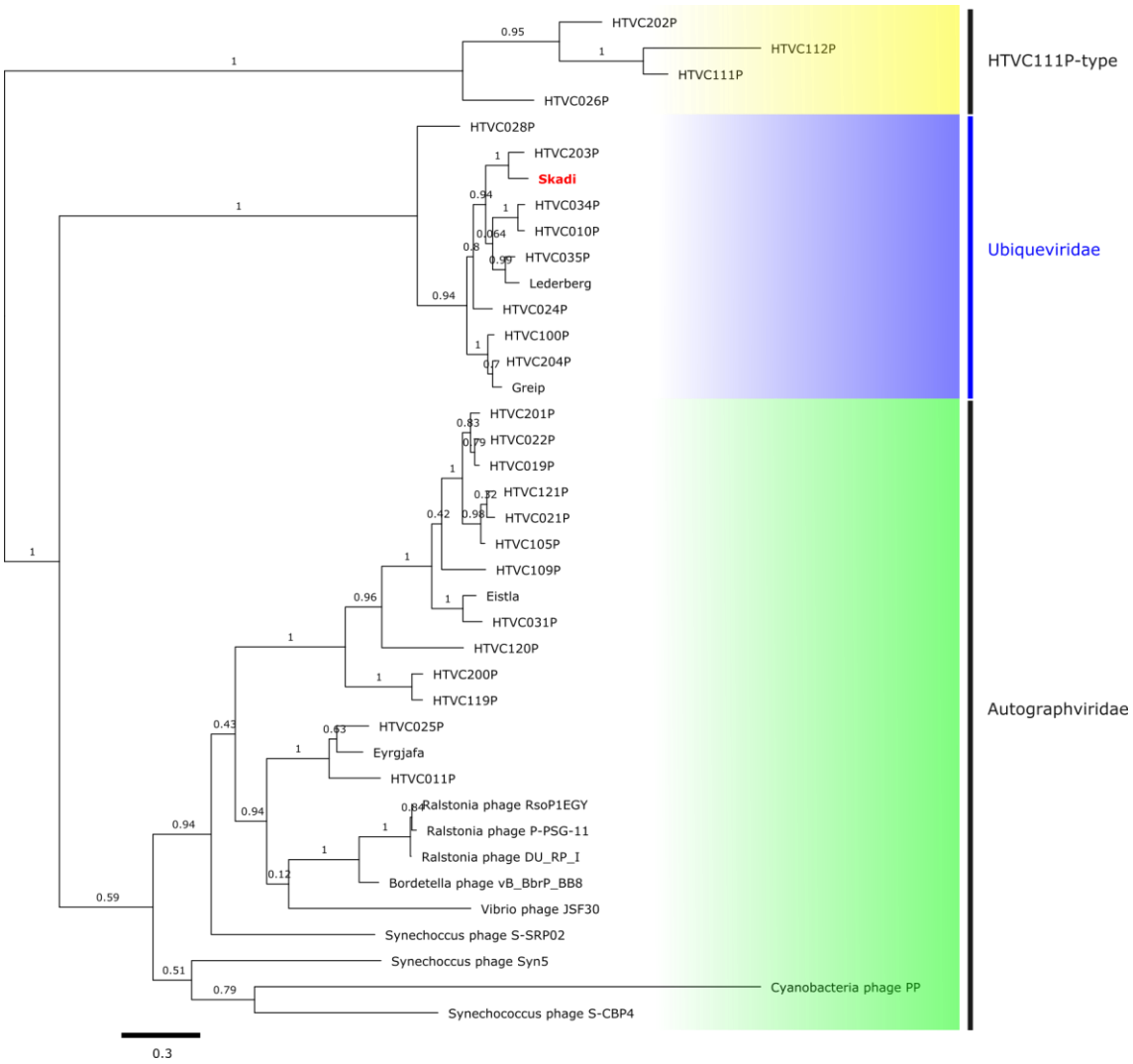
Global Oceans Viromes (GOV2) were used to recruit reads against pelagiphage genomes for assessing global abundances (Gregory *et al.*, 2019). Coverage and Reads Per Kilbobase of each contig per Million Reads (RPKM) were calculated using coverm (available at https://github.com/wwood/CoverM) with thresholds on minimum read percent identity 90% and minimum covered fraction of 40% 'coverm contig –bam-files *.bam –min-read-percent-identity 0.9 –methods RPKM –min-covered-fraction 0.4'. To compare coverage threshold between HTVC010P and Skadi the coverm was used with minimum coverage thresholds set to 0 to 100%. The Ocean Data Viewer (ODV) v.5.2.0 available at (https://odv.awi.de/) was used to visualize read recruitment for Skadi. Ordination analysis was performed using the coverm RPKM values and GOV2 metadata and the phyloseq (McMurdie and Holmes, 2013) package for R v.4.0.3 (https://www.r-project.org/).

4.6. Acknowledgements

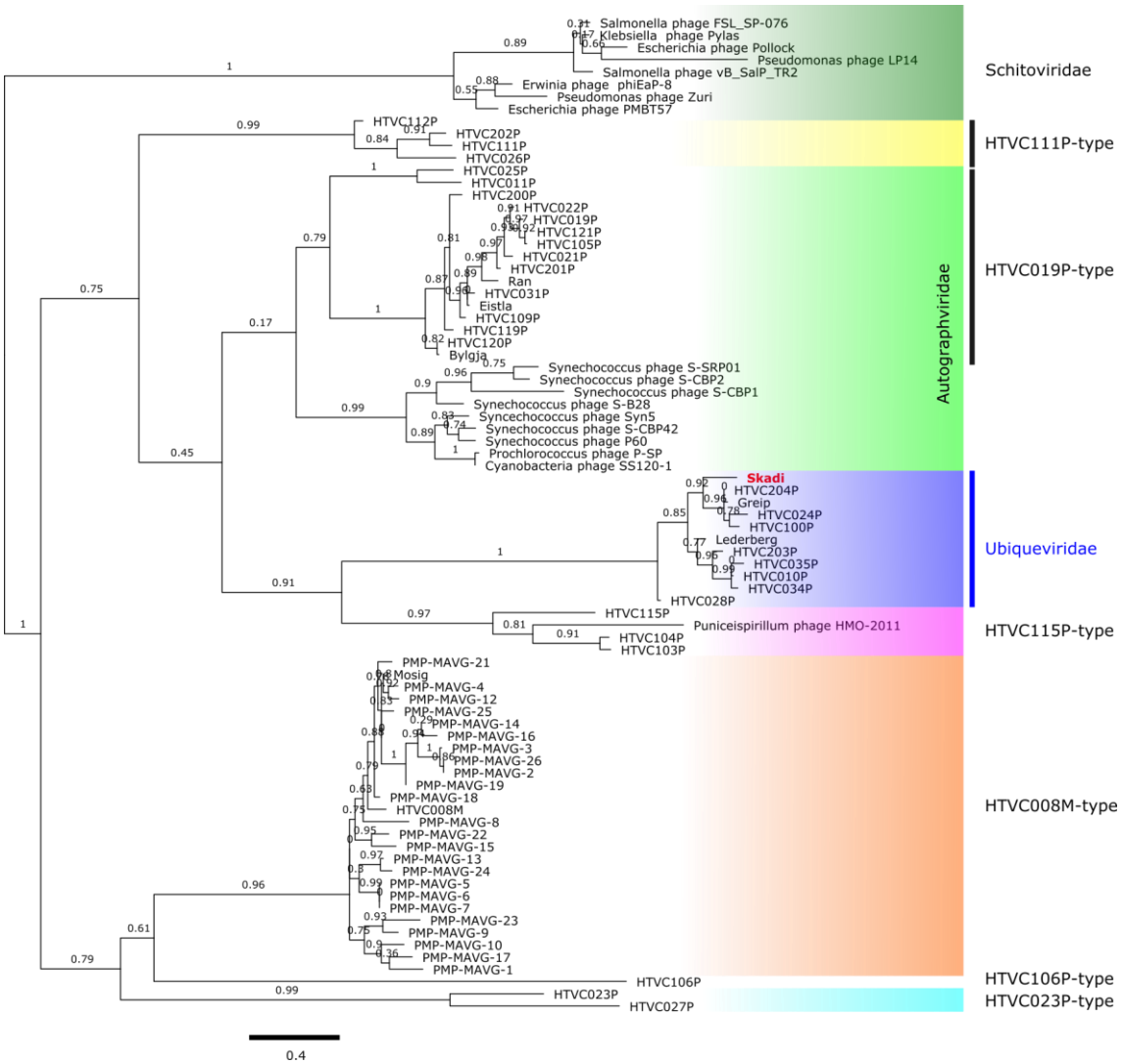
We thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing TE microscopy and imaging. We would also like to thank the crew of the R/V *Plymouth Quest* and our collaborators at the Plymouth Marine Laboratory for providing water samples. We acknowledge the IT team at the University of Exeter and using the High-Performance Computing (HPC) facilities. Phage genome sequencing was provided by the Exeter Sequencing Service. Holger H. Buchholz was funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training program. LMB and BT were funded by NERC (NE/ R010935/1) and the Simons Foundation BIOS-SCOPE program.



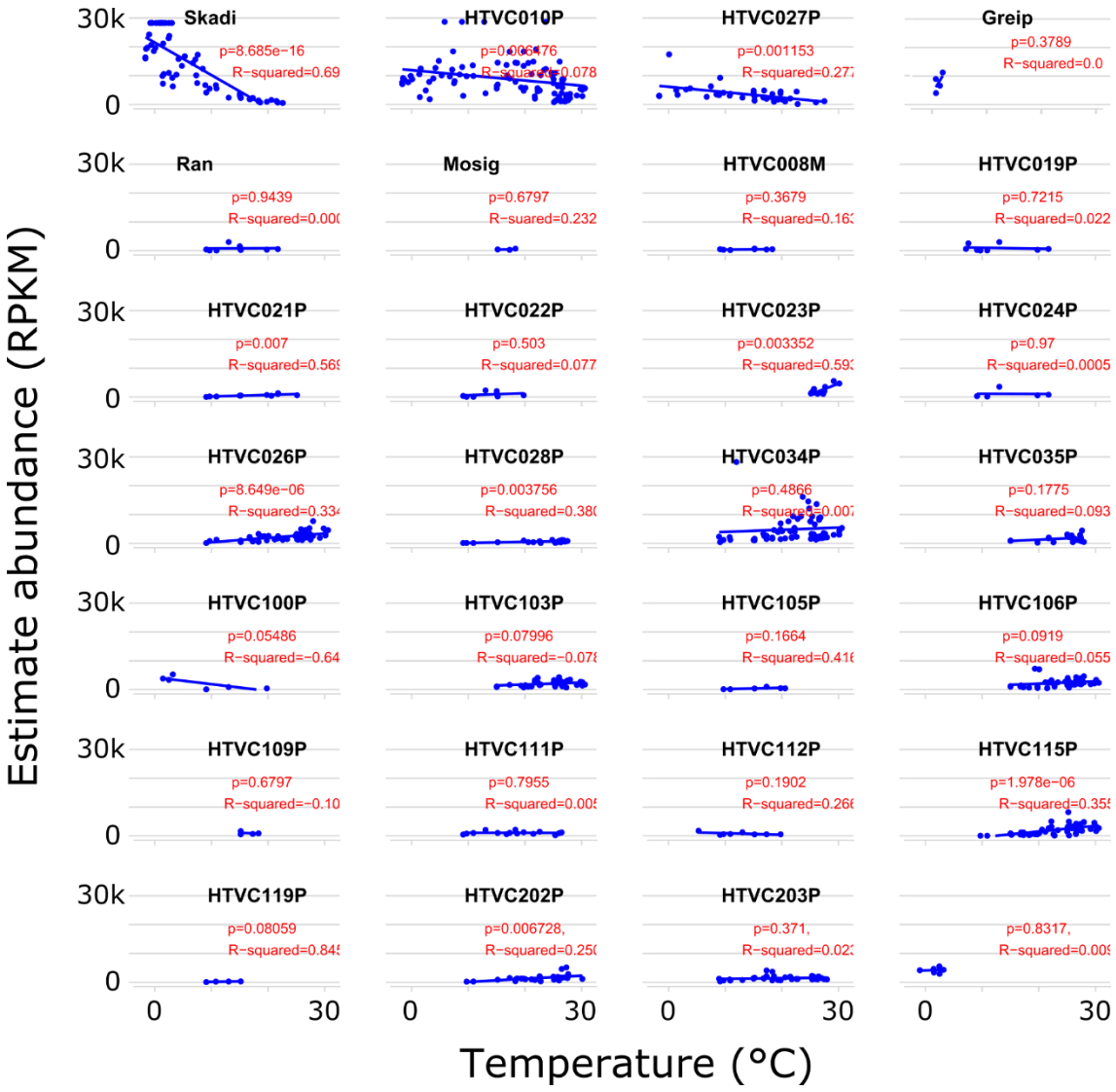
Supplementary Figure 4.1. **Phylogeny of pelagiphage *TerL* genes**. Unrooted neighbour-joining tree (100 bootstraps) of the *TerL* gene found in Pelagibacter phages and representatives of other viral families. Branches are coloured to highlight the different taxonomic groups in relation to the proposed *Ubiqueviridae*. Pelagibacter phage genome types are marked by black bars.



Supplementary Figure 4.2. Phylogeny of pelagiphage tail tube B genes. Unrooted neighbour-joining tree (100 bootstraps) of the tail genes related to Skadis tail tube B protein found in *Pelagibacter* phages and representatives of other viral families. Branches are coloured to highlight the different taxonomic groups in relation to the proposed *Ubiqueviridae*. *Pelagibacter* phage genome types are marked by black bars.



Supplementary Figure 4.3. Phylogeny of pelagiphage major capsid protein genes. Unrooted neighbour-joining tree (100 bootstraps) of the major capsid protein found in Skadi, *Pelagibacter* phages and representatives of other viral families. Branches are coloured to highlight the different taxonomic groups in relation to the proposed *Ubiqueviridae*. *Pelagibacter* phage genome types are marked by black bars.



Supplementary Figure 4.4. **Pelagiphage Skadi correlates with Latitude**. Estimated abundances (RPKM) of isolated *Pelagibacter* phages from the GOV2 dataset with linear regression lines, samples with estimated abundances between zero and three RPKM in all viromes were excluded for clarity.

Supplementary Table 4.1 Shared core genes identified in each genome of the *Ubiqueviridae*.

	Gene ID number of phage strain:										
Core Gene Annotation	Greip	HTVC010P	HTVC024P	HTVC028P	HTVC034P	HTVC035P	HTVC100P	HTVC203P	HTVC204P	Lederberg	Skadi
Tetratricopetide repeat	8	15	48	37	50	55	50	54	45	10	50
DNA binding	9	14	49	38	51	56	51	55	46	11	51
Acyltransferase	10	13	50	40	52	57	52	56	47	12	52
Tail tube B	12	10	52	43	55	60	54	59	49	15	55
Tail tube A	13	9	53	44	56	61	55	60	50	16	56
MCP	14	8	54	45	57	62	56	61	51	17	57
Tetratricopetide repeat	15	7	55	46	58	63	57	62	52	18	58
Portal	16	6	56	47	59	64	58	63	53	19	59
Hypothetical	20	4	59	49	62	67	61	65	56	21	61
TerL	21	3	60	51	63	68	62	66	57	22	63

Supplementary Table 4.2 Overview of shared core genes per pelagiphage type.

Taxon / Group	morphotype	core genes	Species
Ubiqueviridae	podophage	10	11
Autographviridae	podophage	15	16
Kolga-type	siphophage	n.a.	1
Mosig-type	myophage	79	2
HTVC111P-type	podophage	28	4
HTVC103P-type	podophage	31	3
HTVC023P-tpye	podophage	33	2
HTVC106P-type	podophage	n.a.	1

Chapter 5: General Discussion

This chapter concludes my thesis by reflecting on the main findings, highlighting the advantages and limitations of culture-based virology in contrast to culture-independent approaches. I consider how my research has advanced the field of marine viral ecology, and I bring forth the aspects of my science that are the most intriguing and promising areas for future work.

5.1. Background

Microorganisms control oceanic biogeochemical cycles that form a fundamental part of ecosystems on Earth (Falkowski *et al.*, 2008). Viral predation influences microbial communities directly through viral lysis, which kills up to a quarter of the microbial standing stock each day. The study of virus-host model systems also exposed the crucial role of viral cell lysis on biogeochemical cycles (Suttle, 2007), revealing phages as important agents of coevolution (Lindell *et al.*, 2007) and genetic exchange between hosts and other viruses (Pedulla *et al.*, 2003; Belcaid *et al.*, 2010; Duhaime *et al.*, 2011). Through the metabolic study of cultivated phages it has been revealed that during the infection phages 'hijack' cells, reprogramming them either to optimise viral progeny production ('pillage & plunder'), or to maximise host fitness when environmental conditions become unfavourable – they 'batten down the hatches' (reviewed in(Warwick-Dugdale *et al.*, 2019)). The virocell concept has been proposed to emphasize the importance of the intracellular part of the viral life cycle (Forterre, 2011, 2012, 2013). Infected cells might be considered as different biological entities compared to virions and uninfected cells – thus virocells may occupy different ecological niches as they fulfil different metabolic functions (Howard-Varona *et al.*, 2020).

To study the metabolism of virocells, cultured virus-host systems are a prerequisite. Due to the inherent challenges of culturing streamlined marine microbes (Rappé *et al.*, 2002; Steen *et al.*, 2019) and subsequently the difficulties of isolating viruses for these fastidious hosts, culture-independent methods have become the main tool of marine virology. Over the past two decades metagenomics and other culture-independent techniques have evolved rapidly, profoundly expanding our understanding of viral diversity by providing extensive sequencing data and genome data banks from across all oceans (DeLong, 2009;

Hurwitz and Sullivan, 2013; Mizuno *et al.*, 2013). However, through these studies it has also become evident that our holistic understanding of marine viruses is bottlenecked by a lack of suitable representative phages, thus large amounts of data cannot be interpreted fully due to a lack of references. This viral "dark matter" includes 63-93% of surveyed sequence space, drastically limiting the inferential power for developing ecological models from meta'omics data (Brum and Sullivan 2015; Brum *et al.*, 2013; Hurwitz and Sullivan, 2013). Supporting technologies such as metaproteomics (Brum *et al.*, 2016) have helped to confirm expression of some viral proteins, though ultimately even the most advanced prediction tools, e.g. for host prediction and genome annotations, are based on – and limited by – experimentally derived evidence from cultures. With the isolation of the first highly abundant viruses infecting SAR11 in 2013 (Zhao *et al.*, 2013) it became clear that short-read sequence-based metagenomes could miss even the most abundant and important viruses. Hypervariable regions and high microdiversity in these genomes are likely detrimental to short-read genome assemblies (Roux *et al.*, 2017; Warwick-Dugdale *et al.*, 2019). Long-read sequencing has alleviated some of these issues (Mizuno *et al.*, 2013; Martinez-Hernandez *et al.*, 2017), but even with that technology, the number of partial genome fragments far exceeds that of complete, high-quality genomes. Acquiring complete viral genomes from cultures will help resolve biases in metagenome assemblies.

Even though phages for SAR11 have been isolated over the past decade, only two research groups are known to cultivate pelagiphages, and for many more ecologically important bacteria no viral isolates exist at all. With the resurgence of microbial cultivation efforts over the past years, improved artificial seawater-based media for broad-range cultivation of marine bacteria (Henson *et al.*, 2016), as well as targeted media for specific clades (Carini *et al.*, 2013, 2014) have emerged. These efforts have drastically increased the diversity of bacterial species in culture collections (Henson *et al.*, 2020), and brought many ecologically important groups into culture, e.g. the first members of SAR86, LD12 and LD28 (Salcher *et al.*, 2015; Henson *et al.*, 2018), as well as hundreds of novel species for SAR11, OM43, Roseobacteria and Cyanobacteria. Despite the proof of concept by the first isolation of viruses for SAR11 (Zhao *et al.*, 2013), advances in microbial cultivation have not been matched with equally successful viral

isolations, and viruses for many ecologically important clades such as OM43 are uncharted. It is clear that the ecology of marine viruses and bacteria are intertwined, still the vast majority of projects target either the host or the virus, but rarely both. As has been shown for cyanophages, local and regional biogeography is an important consideration for ecology (Marston and Martiny, 2016), and what is more, a large percentage of phages exhibit high host specificity (Koskella and Meaden, 2013). Therefore, single virus-host systems for any given taxon are likely insufficient to represent globally distributed microbial communities and possibly endemic systems of other regions. Developing improved isolation methods will help fill the gaps, as it will improve the accessibility of virus-host systems for the wider research community, and likely allow for broader representation in culture collections. It remains of critical importance to bring representative virus-host model systems into culture to ground-truth culture-independent methods and models and test associated hypotheses.

5.2. Key solutions for high-throughput viral isolation

As the first goal of my project, I set out to use established bacterial DtE methods (Connon and Giovannoni, 2002) to obtain local WEC strains of the OM43 and SAR11 clade, to use as bait as the model organisms in the development of a workflow for isolating virus-host systems. Using continuously cultivated hosts as models for ocean-wide systems could potentially result in biased observations due to bacteria rapidly adapting to the different selection pressures of artificial media (Pratt and Waddell, 1959). Therefore, a steady supply of recently isolated microbes should be used for maximising diversity in cultured phage genomes. The methodology utilised the advancements from increasing bacterial cultivation efforts, due to which a large variety of isolation and cultivation methods exist (Connon and Giovannoni, 2002; Stingl *et al.*, 2007; Carini *et al.*, 2014; Henson *et al.*, 2016, 2020). This resulted in three novel OM43 strains and a novel HTCC7211-like SAR11 strain, expanding our knowledge about WEC microbial communities, and providing model microbes for viral isolation (curiously, but not explored in this thesis, the HTCC7211 variant isolated from the WEC (H2P3α) was nearly identical to the original isolate genome from BATS, with only a few

single nucleotide polymorphisms (SNPs) difference over the full genome, despite nearly a decade between the isolation events). With more efficient viral isolation methods, minimal additional effort and resources are required to isolate viruses alongside bacteria. These efforts will be rewarded with valuable representative genomes that will improve the interpretation of data from culture-independent tools. By targeting both host and viral isolation in my project, I provide a valuable example of efficient virus-host system isolation that can be easily followed and adapted for future projects.

The main struggle to isolate viruses for many aquatic microorganisms is that they do not form colonies on agar plates, and therefore require liquid systems. Plate-independent DtE is an established concept for the isolation of algae and cyanobacterial viruses (Nagasaki and Bratbak, 2010). DtE has also been shown to be a viable option for the isolation of heterotrophic viruses including those of streamlined, fastidious heterotrophs (Zhao *et al.*, 2013; Moon *et al.*, 2017). Steps commonly involved in liquid-culture viral isolation (reviewed for phototrophic viruses by (Nagasaki and Bratbak, 2010) can be summarized as:

(1) Sample preparation, where the bacterial fraction from natural (water) samples are removed, typically using 0.2 µm filters, in order to obtain a cell free natural viral sample.

(2) Inoculum concentration and enrichment, the concept behind this step is to concentrate viral particles with low-density, in order to increase the chance of a successful infection. Many methods have been put forth, including culture based enrichment, ultracentrifugation, ultrafiltration ('dead-end' or continuous) or continuous flow centrifugation.

(3) Screening for viruses, i.e. the means to observe viral infections. In liquid systems, this can be achieved by recording host growth with optical density or flow cytometry; in phototrophic organisms fluorometers can be used. More elaborate methods count viral particles directly with cytometry or microscopy.

(4) Viral isolation/purification for separating single virus strains into axenic cultures, which is a prerequisite for most downstream applications. In liquid culture Dilution-to-Extinction is the commonly used method.

(5) Lysate filtration and storage, where successfully isolated viruses are prepared for downstream applications and long-term storage. Effective storage to keep viral particles infectious can be difficult and evolved (Petsong *et al.*, 2019), and are typically stored without the cellular fraction between 4 °C and -80 °C.

Sample preparation is essential to avoid contamination by microbes from the inoculum. Separating the viral fraction from the bacterial fraction in SAR11 and OM43 required 0.1 µm filters instead of commonly used 0.2 µm pore size filters as their cells are extremely small (~1 µm × 0.2 µm for SAR11 (Zhao *et al.*, 2017) and were not removed from the viral fraction reliably with 0.2 µm filters. However, 0.1 µm filters might also remove large-size viruses, e.g. giant viruses (giruses) (Van Etten *et al.*, 2010), jumbo-phages (Yuan and Gao, 2017) or other larger phages (Sulcius *et al.*, 2011). Therefore, any workflow, including the one presented here (Figure 2.1), needs to be adapted to whatever the targeted microbes and viruses of interest are. The isolation of the *Pelagibacter* myophages Mosig (Appendix 2,(Buchholz *et al.*, 2021)) and *Methylophilales* myophage Melnitz (Chapter 3) and the *Pelagibacter* siphophage Kolga (Figure 2.4), demonstrated that a 0.1 µm filter size did not preclude the longer tailed viruses and I have shown that all three main bacteriophage morphotypes in marine systems (siphophage, myophage, podophage) could be isolated with this approach. Similarly, in step (2), physical stress from the chosen tangential flow filtration (TFF) approach did not appear to inhibit the isolation of phages with non-contractile tails, further supporting the presented methods as suitable for viruses of the main morphotypes. Though unlikely due to the small host cell sizes, this does not exclude the possibility of the existence of other unknown phages for these model hosts, which could have been biased against by the applied mechanical factors.

Many studies using liquid-culture viral isolation include ultracentrifugation, ultrafiltration (‘dead-end’ or continuous) or continuous flow centrifugation for concentrating viruses in the inoculum (Step (2)). Ultracentrifugation and continuous flow centrifugation are limited by their reliance on equipment that is often not available for less well funded and equipped labs (or suitable for

deployment on boats). To truly unlock virus-host isolation for the international research community, methods and equipment need to be obtainable for any research group. Many ultracentrifuge models are also limited to 50 mL falcon tubes (or smaller), this limits the sample volumes that can be processed within a reasonable amount of time. In contrast, TFF cassettes are relatively cheap and reliable and can be used for large volumes. For example, Bettarel et al. used TFF to concentrate 50 l of seawater to isolate viruses for the diatom *Chaeoceros gracilis* (Bettarel *et al.*, 2005). Whilst using TFF (and other methods) for concentrating the inoculum is well established for phototrophs, the few examples of viral isolations on fastidious marine heterotrophs do largely not utilise any inoculum concentration. All previously (and contemporaneously to this project) reported pelagiphages (by the Zhao group) use the viral fraction of natural seawater amended with nutrients for initially capturing phages instead (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Du *et al.*, 2021). Similar methods were used for the isolation of the first (and to date only) phage for SAR116 (Kang *et al.*, 2013), the first and only phage for LD28 (Moon *et al.*, 2017), many Roseobacteria phages (Liang *et al.*, 2019) and the OM43 phage MEP301 (Yang *et al.*, 2021), which is the only other known phage of the OM43 clade, isolated shortly after the first OM43 phages were reported (Figure 2.5,(Buchholz *et al.*, 2021)). Arguably, this indicates that there is no need for inoculum concentration. Though it is evidently possible to isolate phage this way, there are at least two major drawbacks: (1) Logistical problems caused by the reliance on access to relatively large volumes of seawater; and (2) Filter-sterilising larger volumes of sample to be turned into media for the initial incubation, is more susceptible to contamination – smaller volumes are easier to handle. A possible third major drawback is that without concentrating low-abundance viruses using such concentration steps, the probability of isolating only the most abundant phages is higher. For example, in Chapter 3 the isolation of Melnitz at BATS using TFF concentration is described (both Melnitz and its host were below detection limits), suggesting that with the TFF step even the rarest virus-host systems can be isolated. Though keeping to the principle of pre-incubating phage prior to purification reported by the Zhao group, I chose a modification similar to the sequential enrichment methods that resulted in the isolation of the first

crAssphages of the human-gut associated *Bacteroides intestinalis* (Shkoporov *et al.*, 2018), and has been reported for phototrophic phage isolation (Nagasaki and Bratbak, 2010). Combining this enrichment step with TFF inoculum concentration likely helped achieve the isolation success rate of up to 96.6% here (Figure 2.2). It is noteworthy, that for many phages isolated here, up to three rounds of enrichment were required before a viral infection could be spotted (Supplementary Table 2.5), which was not reported for the aforementioned streamlined heterotrophic phages. These steps constitute the basis for my improved high-throughput viral isolation methods for streamlined heterotrophic virus-host systems.

The next challenge to overcome was finding the means to spot viral infections. Most methods use lysis of the host cells or enumeration of viral particles as indicators for a successful infection (Thyrhaug *et al.*, 2003). Methods for direct viral enumeration, that can be used to confirm viral infection include: (1) most-probable-number assays (MPNs); (2) plaque assays; (3) Transmission electron microscopy (TEM); (4) epifluorescence microscopy; (5) qPCR; and (6) flow cytometry (more methods were reviewed here (Ács *et al.*, 2020)). Many of these techniques are either inapplicable to liquid-cultures, expensive and/or labour extensive (Suttle, 1993), and therefore ill-suited for high-throughput viral isolation. Instead, viral infections were screened for via cell lysis. Oligotrophic SAR11 and OM43 neither grow on agar, nor do these microbes reach cell densities that could be recorded using optical density (Connon and Giovannoni, 2002). Using flow cytometry is already required for cellular enumeration of the host and has been shown to work for SAR11 phages (Zhao *et al.*, 2013). Using subtle changes of infected cultures compared to no-virus controls, I show that growth retardation as indication of infection can be summarized as: (1) Increased background noise; (2) Decreased maximum cell density; (3) Increased green fluorescence signals; and/or (4) “Blurring” of cell population signals in the cytograms (increased range of green fluorescence and/or side scatter signals) – see the protocol available at dx.doi.org/10.17504/protocols.io.bb73irqn for examples of cytograms. As analysing a single sample with standard commercially available flow cytometers only takes about one minute, this approach is both reliable for spotting even incomplete viral infections, as well as efficient in time and resource demands.

Importantly, many reported isolations of viruses infecting streamlined heterotrophs, often based on the works of Zhao et al. (2013), use epifluorescence microscopy to confirm viral infection, which is costly and labour intensive. My results show that epifluorescence microscopy or other direct viral enumerations are not necessary at any stage. Though TEM was used in a similar way for sequenced viral isolates here, this step could also be omitted, but is recommended to confirm morphotypes. Using only flow cytometry could significantly improve the throughput of future viral isolations.

Persistent phage infections can be difficult to discover, e.g. chronic infections caused by filamentous phages do not induce lysis as part of their natural infection cycle (Smeal *et al.*, 2017a, 2017b). Relatively few filamentous bacteriophages are known (Ackermann, 2007), and they are often considered fascinating, but extreme strategists, and with limited ecological impact even though chronic infections in the environment might have the highest reproductive success rates (Weitz *et al.*, 2019). Lysogeny in pelagiphages may be common (Zhao *et al.*, 2018), but even within the same types of viruses, strategies can be different, such as the lytic and lysogenic lineages within HTVC010P-type viruses (Du *et al.*, 2021). Utilising subtle changes in host growth rates and dynamics as observed with flow cytometry could be coupled with machine learning to eliminate human bias from interpreting cytometry data and reveal subtle differences between infected cultures and controls, and thereby assist with the capture of viruses that do not cause immediate or only partial lysis. This would enable future projects to capture a broader range of phages, possibly including lysogenic viruses without requiring elaborate (and resource intensive) methods to induce lysis e.g. with UV-treatment or inducing agents such as mitomycin C (Suttle *et al.*, 1991; Jacobsen *et al.*, 1996).

Overall, using TFF viral inoculum concentration in combination with a sequential enrichment step (Figure 2.1) increased the viral isolation success rate to up to 96% depending on the host strain e.g. for the local SAR11 *Candidatus* Pelagibacter bermudensis H2P3α (Figure 2.2). Combining this efficiency with flow cytometry techniques streamlined the process to be suitable for high-throughput isolation. This led to 117 viral cultures from the Western English Channel alone, though resource and time limitations did not allow me to

sequence the genomes of all isolates, this more than tripled the number of pelagiphages isolates (32 reported species) gathered over the past decade (Zhao *et al.*, 2013, 2018; Zhang *et al.*, 2020; Du *et al.*, 2021).

Direct comparison to success rates from other studies is challenging due to the unfortunate tendency of published work to not include negative results; nonetheless there are several noteworthy advantages in my presented workflow:

(1) Logistics, only relatively small volumes of seawater are required (two litres) that can be easily processed on site, transported and/or stored long-term.

(2) High yield per sample, with up to 30 isolates (out of 48 trials) from the March 2019 sample.

(3) High efficiency, with up to 96.6% efficiency on host H2P3α.

(4) Isolation of both extremely rare and abundant phages is possible, both the highly abundant phage Skadi as well as six phage species below detection limits in the WEC were isolated.

(5) Scalable due to low time/resource costs, we estimated £70+ per phage including full genome sequencing, and ~4h handling time per 96-well plate.

Importantly, this demonstrates that my project has achieved a cost-efficient, flexible, and high-throughput workflow that will help unlock virus-host system isolation for the wider research community. Enabling a steady and reliable supply of ecologically relevant model systems is crucial to capture the full range of diversity in microbes and their strategies in the oceans. With every discovered genome from a viral culture, our understanding of MAVGs for which we previously had no host data, will also improve. Subsequently will better host predictions and increase the interpretive power of metagenomics, benefiting the entire field of viral ecology.

5.3 Possible Targets for high-throughput virus-host isolation

Bacterial cultures are a prerequisite for viral isolation, thus could be considered the main restriction for cultivating phages. Yet there are many ecologically important marine heterotrophs that have been in culture for years, yet no or only

single viral isolates are known. Therefore, the lack of viral isolates for many bacteria is likely due to a combination of lack of access to hosts as well as insufficient methods. With improved high-throughput isolation presented here, I hope that renewed vigour for viral isolation lead to the discovery of model system from this brief non-exhaustive list of potential targets for environmentally relevant aquatic heterotrophic virus-host systems isolations:

(1) SAR11, though 38 podophage species have been brought into culture, only two myophages and one siphophage (Kolga, Figure 2.4) have been cultivated.
(2) SAR86 is an abundant and ubiquitous surface ocean Gammaproteobacteria with high levels of genetic diversity (Hoarfrost *et al.*, 2020). Isolation of SAR86 has been announced at the Ocean Sciences Meeting, San Diego 2020 by Rappé *et al*.; no viruses for SAR86 are known.
(3) SAR116 is an Alphaproteobacteria that constitutes up to 10% of surface ocean bacteria and was first isolated in 2010 (Oh *et al.*, 2010). In 2013 the first and only podophage infecting SAR116 was isolated and shown to be closely related to SAR11 phages (Kang *et al.*, 2013).
(4) SAR202 of the phylum *Chloroflexi* remain uncultured, but are hypothesised to specialise in the degradation of recalcitrant carbon compounds, and are thus extremely important in deep sea carbon cycles (Saw *et al.*, 2020).
(5) LD12 are freshwater Alphaproteobacteria within the SAR11 clade, occupying a similar niche as marine SAR11. LD12 has been brought into culture in 2018 (Henson *et al.*, 2018). Prophage genomic sequences have been acquired, yet no phage has been isolated (Chen *et al.*, 2019).
(6) LD28 are methylotrophic Gammaproteobacteria related to the marine OM43 Methylotrophs. A single phage has been isolated in 2017 (Moon *et al.*, 2017).
(7) OM43 are important methylotrophic Gammaproteobacteria that have been in culture since 2002 (Connon and Giovannoni, 2002), all three OM43 phages were only isolated in 2021, the first two of which reported

in chapters 2 and 3 (Buchholz *et al.*, 2021a; Buchholz *et al.*, 2021c; Yang *et al.*, 2021).

Another possible application for improved viral-isolation methods could be biomedical research. Antimicrobial Resistance (AMR) is likely to become one of the most serious health problems of the 21st century as bacterial infections are increasingly resistant against common antibiotics (Prestinaci *et al.*, 2015). Phage therapy is the concept of treating resistant infections with bacteriophages (history of phage therapy was reviewed by (Salmond and Fineran, 2015). Though phage therapy has achieved promising results against AMR (Brives and Pourraz, 2020), resistance against phages can evolve rapidly, causing phage treatments to fail (Oechslin, 2018; Castledine *et al.*, 2021). However, phage suitable for treatment of many important pathogenic taxa, such as *Haemophilus influenzae* and *Mycobacterium abscessus* are poorly represented in culture collections. In part, this is due to challenges associated with cultivating their hosts, such as highly specialised growth conditions and/or slow growth rates. Isolation of clinically relevant phages for these World Health Organisation priority pathogens (Tacconelli *et al.*, 2018) could greatly benefit from the methodology developed here.

The following project idea has now been funded: "*Viral Impacts on volatile organic carbon and energy cycling in the ocean*" by the Life Sciences-Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology (Award ID:879226). This section has been modified from the proposal's case of support text. The original text has been reviewed and approved by Dr. Steven Giovannoni and Dr. Kimberly Halsey and is available in the appendices (Appendix 5):

Investigating impacts of phages on volatile organic compounds (VOC) and energy cycling in the ocean is important future work that is now possible as a direct result of my efforts. The oceans are a major source of climate active VOCs in the atmosphere (Fischer *et al.*, 2012). VOCs are also proposed to support specific synergistic or antagonistic interactions with other microorganisms (Amin *et al.*, 2015; van Tol *et al.*, 2017). The turnover of VOCs is thought to involve phototrophic VOC producers (Sun *et al.*, 2011; Halsey *et al.*, 2012, 2017a; Thrash

et al., 2014; Beale *et al.*, 2015), and methylotrophic consumers, most importantly by the methylotroph OM43, VOC metabolism in global oceans is therefore critical link between primary production and carbon oxidation, but little is known about the influence of viral infection on this process that forms an important component of the "viral shunt". Utilising proton-transfer mass spectrometry (PTR-MS) coupled with environmentally controlled dynamic stripping chambers that force liquid and gas phases into a steady state allows measurements of VOCs production of consumption (Halsey *et al.*, 2017b; Davie-Martin *et al.*, 2020), therefore could be used to investigate VOCs metabolism in OM43 virus-host systems. Combining this knowledge and established cultures of phototrophic systems with my advances in culturing methylotrophic OM43 virus-host systems, facilitates the quantification of VOC cycling by investigating the impact of viral infection on phototrophic VOC production, and methylotrophic VOC metabolism respectively. This has large-scale implications for global marine carbon cycles and will be important for climate research.

5.4 The wheel of culture-dependent and independent methods

Both culture-independent and culture-dependent studies often use each other's limitations as justification for their own significance. Whilst culture-dependent investigation of e.g. the viral impact on metabolic rates (Waldbauer *et al.*, 2019) or transcriptomic responses to abiotic factors (Doron *et al.*, 2016) are difficult to replace by *in silico* estimations with current technology, culture based viral ecology lacks the scalability of culture-independent methods. Both approaches should therefore not be seen as opposites, but used in concert and enhance each other.

Based on the experience from culture-based molecular work, sequencing on a community level such as bacterial community 16S rRNA and viral and bacterial meta'omics have been made possible (Schmidt *et al.*, 1991; Stein *et al.*, 1996; Breitbart *et al.*, 2002). Metagenomics has unveiled a huge amount of genomic diversity in viral communities that could not be annotated, coining the term viral "dark matter" (Reyes *et al.*, 2012). Even though a large proportion of sequence data remains unassigned, the culturing of for example HTVC010P showed that within the "dark matter" extraordinarily abundant and ecologically important

viruses were hiding in plain sight (Zhao *et al.*, 2013). Yet, relying solely on evidence from cultures for annotating meta'omics data would require practically insurmountable cultivation efforts. This highlighted the need to devise culture-independent methodologies to annotate existing databases. The difficulties associated with culturing (discussed in Chapter 2 and 5.2) in combination with increasingly powerful computational capacities, have led to a one-sided explosion of novel culture-independent methods, which are attempting to link viral sequences to host organisms. However, these tools still require information from phage cultures for controlling error rates, estimating cut-offs and re-interpretation of historic data. Some host-prediction tools that would benefit from or rely on culture collections are:

(1) Viral tagged metagenomics, where fluorescently labelled wild viruses are used on 'bait' hosts enabled flow sorting to separate adsorbed viruses for DNA sequencing (Deng *et al.*, 2012, 2014). The development relied on model *Pseudoalteromonas* and *Synechococcus* phages, which are not representative of most fastidious marine heterotrophs, and requires traditional growth assays to confirm if adsorption equals infection in the oceans.

(2) Single-virus tagging, where fluorescently labelled wild viruses are sorted into individual virus-host pairs as they adsorb to unknown host cells (Džunková *et al.*, 2019). Similar to (1) this relies on culture-based confirmation of adsorption vs. infection, and also needs confirmed sequences from cultures to distinguish between adsorbed viruses and pre-existing prophages and cryptic infections.

(3) RaFAH (Random Forest Assignment of Hosts), is based on machine learning and uses available virus-host pairs in databases (Coutinho *et al.*, 2021), which are hugely biased towards host-virus systems from clinical pathogens. Any approach that uses confirmed experimental data to train classifiers will benefit from a large increase in culturally confirmed virus-host systems.

(4) EpicPCR (emulsion paired isolation-concatenation PCR (Spencer *et al.*, 2016), has been used to fuse viral marker genes (ribonucleotide

reductase) with host 16S RNA within emulsion droplets (Sakowski *et al.*, 2021). The authors highlighted that primer biases remain an issue, and that primers designed for different groups could be necessary. Thus epicPCR for diverse viral groups without known representatives is likely biased.

These examples demonstrate that every single cultured virus-host system has the potential to improve existing culture-independent methods. With the isolation of *Pelagibacter* phage Kolga, the first siphovirus for SAR11, and *Methylophilales* phages Melnitz and Venkman, the first myovirus and first siphovirus for OM43, I provide at least three singleton representatives for novel family-level viral groups. Identifying novel viral groups will enable re-interpretation of historic datasets and identification of MAVGs, and help illuminate the viral "dark matter". Future work will undoubtedly be able to raise new hypotheses, which can be tested experimentally now.

5.5. Unusual virus-host dynamics are prevalent in heterotrophic phages.

As demonstrated here, flow cytometry allowed for the observation of viral infection without directly observing viral particles. Out of 117 viral isolates only 16 viral infections reduced host abundance below the detection limit (Supplementary Figure 2.13). Recovery after virus-induced lysis is a common occurrence in algae, and is possibly enabled by phenotypic plasticity of algal susceptibility, where inhibitory compounds released by cell lysis allow a subpopulation to resist predation and repopulate (Thyrhaug *et al.*, 2003). Studies using experimentally evolved cyanobacteria also show that emerging resistances against viruses are often caused by passive mutations in cell-surface proteins that serve as receptors and attachment sites for the virus (Stoddard *et al.*, 2007; Avrani *et al.*, 2011). Thereby a subpopulation might be inherently resistant to any single virus strain. Other unknown extrinsic and intrinsic factors could further influence the virulence of viruses and render enough infections unsuccessful to allow a microbial population to survive in coexistence with viruses (Horas *et al.*, 2018). Whatever the survival mechanism, in SAR11 the addition of viruses frequently did not result in the expected initial decrease of the population through cell lysis (cell lysis

followed by population death or recovery), but rather a steady-state of the host culture. It is possible that most pelagiphages are capable of temperate infections using unknown genes. However, the pelagiphages Greip is related to the HTVC010P-type lineage that was shown to be purely lytic (Du *et al.*, 2021), and still caused a steady-state (Figure 2.1Bii). Yet, the function of most phage genes remains hypothetical. In Melnitz and HTVC008M, curli genes were found, that speculatively might be involved in facilitating chronic infections (See Section 3.3.5), but no such genes were found in the podophages with steady-state growth curves. We speculated that either phenotypic bistability or plasticity similar to algae-virus systems allows a subpopulation to survive viral predation pressures. The survivors might be able to subsist on intra-population nutrient recycling, i.e. taking up nutrients released from the cells of the lysed fraction. A combination of both mechanisms might explain the observed steady-state growth of infected SAR11, and led to our proposed moniker 'Soylent Green Hypothesis' (See Section 2.4.9), where a resistant subpopulation able to maintain a steady-state by utilising lysed intra-population cells for nutrients. Intra-population nutrient recycling would help explain the predicted limited influence of SAR11 virus-host systems on carbon export (Guidi *et al.*, 2016), as a certain proportion of carbon released by lysis would be recycled within the SAR11 population itself. My work therefore unveiled a possible mechanism that would influence global carbon cycling and with further evidence should be considered for ecological models. To provide quantitative evidence, future work needs to include free viral particle counts in combination with re-growth/re-infection growth experiments, which will help to determine growth dynamics between virus and host and improve ecological models. Using my model systems for stable isotope labelled carbon incorporation experiments with additional metabolomics/transcriptomics measurements would allow future projects to quantify intra-population carbon recycling under Soylent Green dynamics as well as identify the cellular mechanisms involved. This data will be critical for modelling global carbon cycling, and subsequently climate models.

5.6 Curli genes found in the *Methylophilales* phage Melnitz hint at unexpected replication strategies

The genomic annotation and analysis of the novel phage Melnitz infecting members of the OM43 clade, found two predicted proteins for a Curli CsgGF complex (Figure 3.3). In the bacteria *E. coli* curli genes are part of the secretion VIII system, which is responsible for the extracellular formation of amyloid fibres (Yan *et al.*, 2020). Amyloid fibres have been shown to promote the formation of biofilms in their natural environment (Barnhart and Chapman, 2006). Typically, the structure of the pore is determined by barrel protein CsgG, with CsgF guiding the excretion of CsgA proteins that form into a fibre structure outside the outer membrane (Figure 3.4). Other CsgGF associated genes were also identified in the SAR11 myophage HTVC008M and MAVGs phylogenetically related to it (Zaragoza-Solas *et al.*, 2020). However, no curli associated genes were found within the OM43 or SAR11 clades, nor has biofilm formation been observed in their members. No other phage genomes encoding for curli are known, unveiling it as an important (as several non-essential or residual genes spanning a family level viral group would be unlikely) and a unique feature in myophages infecting streamlined heterotrophic marine microbes.

A list of possible explanations for curli genes in marine phages that have been suggested are:

(1) Amyloid fibre production, where the phage provides the host with the ability to produce amyloid fibre.

(2) Sibling capture, where phage amyloid fibres are used to capture similar cells, forming cell conglomerates that allow phage progeny to find new hosts more easily (Zaragoza-Solas *et al.*, 2020).

(3) Macronutrient uptake, where curli pores are used to allow the infected cell to take up macronutrients for powering progeny production (Zaragoza-Solas *et al.*, 2020).

(4) Lysis control as a pinholin, where curli performs a similar function as pinholins encoded by *E. coli* phage λ (Section 3.3.5, (Buchholz *et al.*, 2021c))

(5) Superinfection protection, where amyloid fibres are used to create a defensive structure that physically wards off competing phages (Fabian Wittmers, personal correspondence and BIOS-Scope meeting 2021)

(6) Lysis control as a secretin, where curli genes form a secretin channel homologous to the pIV mechanism in f1 filamentous phage (Section 3.3.5, (Buchholz *et al.*, 2021c))

In chapter (3.3.5) I speculated that curli pores might be used for lysis control, however, protein structure models (Supplementary Figure 3.8) also resembled the outer membrane secretin channel pIV of the filamentous bacteriophage f1, which is involved in a mechanisms that allowed chronic infection (Conners *et al.*, 2021). It is not impossible that convergent evolution between CsgGF and pIV occurred, but challenging this hypothesis is that chronic infection in T4-like phages such as Melnitz are unprecedented, and it is unclear how such a mechanism would work for large myophage virions. Utilising my novel model virus-host system, it would be possible to resolve this intriguing possibility. Assuming phage-encoded CsgGF is functionally analogous to pIV, visible curli proteins in the outer membrane of infected cells but not in uninfected controls would be expected. As a first step, negative stain EM to visualise the cell surface of infected and uninfected OM43 could allow for the observation of phage-associated pore formation. Alternatively, the same could be achieved through cryo-electron tomography, which is more difficult to perform but would resolve changes in the cell surface at a higher resolution and would provide evidence for an extracellular function of CsgGF (Van Gerven *et al.*, 2015). If no cell surface change can be detected in virally treated cultures, the hypothesis of pIV-like chronic infection, and alternative hypotheses involving amyloid fibre formation, would have to be rejected. If pore formation can be observed, this would justify investing into Cryo-EM to resolve the protein complex structure, though technically challenging as for the low cell yield of OM43 cultures (max. ~10^8 cells/mL) and no previous information for EM work on OM43 cells, similar to what has been done for SAR11 cells (Zhao *et al.*, 2017). Should the pIV hypothesis prove correct, this could also help explain the low lytic activity of pelagiphages in coastal marine waters (Alonso-Sáez *et al.*, 2018), as well as provide the first example of chronic infection in T4-like phage systems, which has large-scale implications for interpreting the role of these viruses in microbial communities.

5.7 Phage isolation creates possible targets for future studies elucidating the function of phage genes and their role in marine microbial communities

In all three results chapters, speculations about possible gene functions were made, and many more ORFs could not be annotated at all. Elucidating the function of phage genes can be a long and difficult process, but it is necessary in order to test hypotheses of fundamental phage biology and the phage's ecological role in marine microbial communities. Though inherently speculative to a certain degree, annotation of ORFs allows for more targeted approaches for the confirmation of putative genes. For example, the expression of specific genes can be tracked during infection using reverse transcriptase PCR and Western blot detection of the predicted product (Ou *et al.*, 2015). For a small number of genes of interest, it is also possible to synthesize and amplify them, as has been done for a heat-shock protein encoded by cyanophage S-ShM2 (Bourrelle-Langlois *et al.*, 2016). This heatshock protein was then cloned by insertion through a vector into competent *E. coli* cells, which can then either be used to test the effects of the protein in *E. coli*, or to produce larger quantities of the protein for downstream analysis. A worthy target found in this thesis could be the glutamine riboswitch identified on the *ssrA* gene encoded by Melnitz (compare chapter 3). By investigating conformational changes in response to different glutamine concentrations in the growth media, it might be possible to establish the role of viral riboswitches in the replication of the types of oligotrophic phages. Whilst this approach could potentially be useful to confirm selected putatively annotated genes, it is not well suited to identify the function of completely unknown genes, as the required amount of tests for untargeted investigations would be unfeasible. Furthermore, tools used for genome editing of phages in copiotrophic organisms, such as Bacteriophage Recombineering of Electroporated DNA, BRED, (Marinelli *et al*., 2008), or rebooting of phage genomes that were synthesized *de novo* (Ando *et al.*, 2015), require host cells that are amendable to plasmid uptake or electroporation of synthesised DNA, respectively. Neither of these methods are currently available for fastidious oligotrophs such as SAR11 and OM43.

To investigate genes with putative or entirely unknown function, a first step can be to analyse the transcriptome of infected compared to uninfected cultures

throughout the early, middle and late infection stages by clustering phage genes into expression profiles at multiple time points (Howard-Varona *et al.,* 2017). This could allow for an estimation of what genes might be used for, based on the infection stage and what other known genes are expressed at the same time. Targeted manipulation of specific environmental conditions could also be used to test under what conditions certain genes are expressed, and thereby elucidating a potential function of the gene, and by extension could provide information on the potential ecological impact of the phage.

Metabolomics in combination with transcriptomics and/or proteomics could provide further evidence of genes with metabolism related functions (Zhao *et al.*, 2016), which might also allow speculations about the potential ecological impact of the phage on host populations in nature. However, without further tests, these experiment would be unlikely to confirm specific functions of single genes as typically a multitude of genes being expressed at any given time point has to be expected.

To confirm structural virion proteins predicted by genome analysis or identify ORFs encoding structural genes without homologues in databases, MS-based proteomics has been shown to be effective in identifying structural proteins in the cyanophage S-PM2 (Clokie *et al.*, 2008). This approach could for example be used to identify structural proteins in the SAR11 siphophage *Kolga* (Figure 2.4), as it lacked homologues of e.g. tail tip proteins K, L and M or any tail tube proteins present in many other siphophages (Sullivan *et al.,* 2009).

When a specific gene (or small number of genes forming a protein complex) of interest has been identified, for example with transcriptomics, often the next step is to use x-ray crystallography, cryo-electron microscopy or nuclear magnetic resonance to determine the biomolecular structures of the protein(s) produces by that gene. X-ray crystallography has for example been used to confirm the baseplate protein of *E. coli* phage Mu (Kondou *et al*., 2005), to identify the T4 lysozyme complex binding sites to the periplasmic Spackle protein (Shi *et al.,* 2020) and to identify receptor binding domains in tail fiber proteins (Salazar *et al.*, 2019). Whilst very useful in determining how specific genes and proteins work, these methods require large amounts of time and resources, therefore are not well suited to screen the function of many different unknown genes or proteins.

As discussed in section 5.6, the Curli genes found in Melnitz would provide a valid target to confirm the putative annotation with X-ray crystallography or Cryo-EM similar to what has been done for a filamentous phage secretin protein (Conners *et al.*, 2021). As deep-learning approaches for the prediction of protein structures improve (e.g. AlphaFold2 used in Chapter 3), it may become possible to glean information about phage protein functions from structures alone, e.g. by computational analysis of binding sites. Nonetheless, experimental confirmation e.g. via aforementioned cloning approaches would still be necessary. All these approaches for the confirmation of phage gene function have in common that cultured strains of both the host and virus are a pre-requisite, thus the methods for bringing difficult to grow virus-host systems into culture introduced here could be considered the first step for determining unknown phage genes as well.

Investigating single genes of interest in cultured virus-host systems can be powerful in determining what unknown genes are and how they work. Combining culture experiments and nutrient measurements with transcriptomics can be especially useful to determine how phage AMGs influence the rate and utilisation of nutrients such as nitrogen (Waldbauer *et al.*, 2019), carbon (Thompson *et al.,* 2011) or other nutrients (Nilsson *et al.,* 2022). This opens up discussions about the role of phages in global nutrient cycles (Zimmerman *et al.*, 2019). Without cultured virus-host systems experiments like these would not be possible, and the methods presented here will enable future studies to obtain required strains. However, axenic cultures are likely insufficient to accurately capture the complexity of microbial communities and the impacts of changing conditions thereupon. To directly test the influence of certain pressures such as phage predation or environmental conditions on microbial community dynamics and ecosystem services, mesocosm and other community culture experiments are an effective approach (Wilson *et al.,* 1998, Martínez Martínez *et al.*, 2007).

The strength of mesocosm experiments is that they allow for observation of (induced) changes to the microbial and viral community directly, and are often used to observe e.g. community responses and dynamics as well as nutrient flow after inducing phytoplankton blooms (Sinha *et al.*, 2007, Joassin *et al.,* 2011) or changes in environmental conditions (Highfield *et al.*, 2017) . However, manipulating members of the community individually is often not possible, as

most abiotic factors will impact the community in its entirety. Combining high-throughput isolations with mesocosms might help bridge the gap between axenic culture experiments and mesocosms, for example Hoetzinger *et al.* (2021) isolated 121 phage strains for three microbial community members of interest from mesocosms, allowing for more in depth investigations of temporal changes in diversity and abundance within the mesocosms. The high-throughput methods described in my thesis could allow for similar investigations, and other possible application of high-throughput viral isolation in combination with mesocosms and community cultures, for example (1) Isolating and thereby identifying and confirming a large number of associated phages specific to certain SAR11 strains would enable improved observations of the viral community dynamics by counting these specific phages over time within their environment (Baran *et al.*, 2018). Using this approach in combination with mesocosm set-ups would allow a high-resolution investigation of SAR11 virus-host community dynamics in response to e.g. nutrient availability, changing temperatures (which is important in light of climate change), ocean acidification or anoxic conditions. (2) Isolate cultures of a range of phages for environmentally important microbes could be used to produce a concentrated 'cocktail' of phages that has been confirmed to infect the targeted microbes. For example, adding a cocktail of SAR11 specific phages to a natural community could simulate a "viral outbreak" for SAR11.This would enable direct observations of how phage predation on SAR11 cascades through the community using 16S rRNA analysis over time. Additional VOC and/or DOC measurements would provide the opportunity to directly measure the impact of SAR11 phage predation on marine microbial community structure and nutrient cycles. (3) Cultured viral isolates obtained with my high-throughput isolation method could be used to create ^{13}C isotopically labelled pelagiphages. When added to a marine microbial community, tracing the labelled carbon might allow us to determine the fate of pelagiphage bound carbon, potentially identifying what marine microbes consume pelagiphage particles. A similar approach has identified bacterial consumption of T4 phages in wastewater, which might be useful for removing harmful viruses in water treatment or patients (Godon *et al.*, 2021).

Investigating microbial communities and nutrient cycles can be challenging, but reliable and high-throughput isolation methods such as the ones presented in this work will be able provide a large and diverse collection of cultures for downstream experimentation. My method therefore is an important first step for many possible future studies of microbial community dynamics.

5.8 Genomic evidence for inter-class host transition between myophages infecting SAR11 and OM43

Another intriguing feature of the OM43 phage Melnitz was found in the early evidence for a possible host-transition event between SAR11 and OM43. We argued that this was supported by the phylogeny of the shared *SsrA* gene (Figure 3.5), which placed Melnitz in the alphaproteobacterial lineage rather than the Gammaproteobacterial lineage of OM43, as well as mismatches in gene alignments of tRNAs found in SAR11, OM43 and their respectively associated myophages (Section 3.3.6). We argued that even extremely rare events in natural communities are possible due to large population sizes of marine microbes and viruses, which would support a putative mutant myophage ancestor transitioning from SAR11 to OM43. If that were the case, evolution in streamlined heterotrophic virus-host systems is likely driven by host transition and expansion across broad taxonomic groups, increasing mosaicism and genetic exchange. However, finding hard evidence might be challenging. The 'Appelmans protocol' was created to generate therapeutic phages with altered host ranges, but could also be used to recreate host transitioning in OM43 and SAR11 systems (Burrowes *et al.*, 2019), providing experimental evidence as well as mutant strains to investigate the mechanisms behind host transitioning in streamlined systems. Alternatively, large scale host range experiments would help characterise virus-host dynamics for this group and potentially capture generalists able to cross class boundaries, however, this is hampered by the small number of available model systems (only six OM43 species have been reported in culture, including the three strains presented here). Alternatively, modified single-virus tagging, as discussed in section 5.3 (Džunková *et al.*, 2019)), where the fluorescently tagged myophages Melnitz and Mosig could be

used to challenge natural bacterial communities, allowing for the identification of class-level host ranges. A different approach could utilise single-cell genomics (Martinez-Hernandez *et al.*, 2017), as the isolation of 100% ANI Melnitz strains across two different marine provinces (Table 3.1) suggested that Melnitz genome variation on a global scale is minimal. If Melnitz-like genomes can be found within SAR11 cells, this would provide strong evidence for the hypothesised host range expansion. A culture-independent approach could use recent advances in host prediction based on receptor-binding proteins (Boeckaerts *et al.*, 2021), as host ranges are affected by compatible cell surface receptors. If receptor-binding proteins in myophages of OM43 and SAR11 are mismatched similar to tRNAs (Figure 3.6), this would provide additional evidence of host range expansion events in this group. Overall, there are several possible experimental approaches that could be used to acquire further supporting evidence of the host-range expansion hypothesis, which if accurate would have considerable implications for marine viral ecology.

Bacteriophages are typically considered to be relatively specific, but evidence for broad host range selection exists (Ross *et al.*, 2016), seemingly favouring our host-range expansion hypothesis. However, the host range experiments with SAR11/OM43 myophages indicated no simultaneous infection between classes was possible (Supplementary Figure 3.10), therefore the ability to infect across classes was likely lost in their evolutionary history, which suggests selection pressure towards specialism. Such randomly mutated strains are most likely extremely rare, and considering the differences in SAR11/OM43 population sizes, the probability of emerging mutations in abundant SAR11 lineages would be higher. This host-range expansion hypothesis hence aligns with suggested fluctuation selection dynamics over arms-race and kill-the-winner dynamics, as the putative ancestral strain would have transitioned from high abundance SAR11 hosts to low abundance SAR11 host (Avrani *et al.*, 2012). The question is, if this SAR11 to OM43 host transition occurred, does this happen between SAR11 and other marine Proteobacteria as well? In this scenario, SAR11 would likely be the source of the transitioning phage mutant again, because SAR11 is the by far most abundant group within the marine Proteobacteria (Morris *et al.*, 2002). Considering that viruses are considered agents of genetic exchange,

possibly across superkingdoms of Life (Malik *et al.*, 2017), SAR11 virus-host systems might act as a "central hub" of genetic exchange for marine Proteobacteria, or possibly beyond phylum boundaries, driven by host-range expansion. Though this might seem unlikely, vaguely similar events are speculated to have produced a symbiosis involving a SAR11 ancestor which formed the first eukaryotes (Thrash *et al.*, 2011). Building onto the "genetic exchange hub" hypothesis, the omnipresence of SAR11 in epipelagic environments would in turn imply that random mutants of viruses from other microbial groups could exist and transition into SAR11 lineages; even if individual groups might be less abundant than SAR11. In summation their contribution to the mutant "pool" could still be significant. Investigations of high SAR11 genomic diversity suggests that they have unusually high rates of mutation accumulation and recombination compared to other groups (López-Pérez *et al.*, 2020), but would align with the "genetic exchange hub". As an example, functional curli systems are widespread and diverse (Dueholm *et al.*, 2012), and "genetic exchange hub" dynamics would offer an explanation how this operon might have been picked up by related phage infecting OM43 and SAR11 respectively, despite the lack of curli genes in both host groups. Speculatively, an ancestral strain acquired curli genes from an unknown third host group, a random mutant of which transitioned into SAR11, diversified, and subsequently transition to OM43, which could have been aided by rapid adaptation to a suboptimal host, as observed for myophages of Cyanobacteria (Enav *et al.*, 2018). The idea of SAR11 systems as a virally mediated "genetic exchange hubs" would elevate the importance of this clade beyond its extraordinary abundance. An intriguing follow-up question is: Would genome streamlining of SAR11 apply to genes exchanged via the hypothesised "genetic exchange hub" and thereby increase streamlining on a microbial community level? Evidence could be provided by searching for SAR11-like genes in genomes of other marine bacteria.

Overall, the genetic evidence found in Melnitz is intriguing, and might support a SAR11 "genetic exchange hub" model, which implies that phylogenetic boundaries might be much less restrictive for genetic exchange as previously thought.

5.9. Polar ecotypic niche variation for abundant marine phages in future polar oceans

Many cyanophage taxa exhibit spatiotemporal distribution patterns, following the niche defining biogeography of their hosts (Chow and Fuhrman, 2012; Marston *et al.*, 2013; Marston and Martiny, 2016). Interaction with either generalist or specialist phages might change the range and rates of horizontal gene transfer (Zborowsky and Lindell, 2019). Such viral niche variation is therefore critical for structuring genomic phage diversity and influences bacterial community structures, which could alter the impact of bacteria within the microbial community. Despite reports of ubiquitous and abundant pelagiphages, infecting highly dynamic and ecotypically distributed SAR11 hosts (Vergin *et al.*, 2013; Delmont *et al.*, 2019), pelagiphages are frequently reported as widely distributed across all major oceans (Zhang *et al.*, 2020; Du *et al.*, 2021). A notable exception are the HTVC023P-type phages that were significantly more abundant below the photic zone (Zhang *et al.*, 2020), whereas all other known pelagiphages are predominantly found in the epipelagic zone, hinting at vertically distributed pelagiphages in the water column. This might be related to reports of significant vertical fluxes of photic zone viruses to the deep sea (Hurwitz *et al.*, 2015). Pelagiphages presented in this thesis and previously reported, tend to infect only warm-water or cold-water SAR11 1a ecotypes. This contradiction of apparent specialist pelagiphages with global and ubiquitous distribution has to be addressed in order to understand SAR11 virus-host dynamics.

Pelagibacter phage Skadi is the most abundant pelagiphage in the polar oceans and likely one of the most abundant viruses on the planet (Chapter 2). Skadi as the example of a successful polar pelagiphage revealed that pelagiphages can exhibit strong ecotypic distribution patterns (Figure 4.5), similar to local and regional biogeography observed in marine cyanophages (Marston *et al.*, 2013). Contrasting distribution patterns of Skadi with the equally abundant HTVC010P suggested that Skadi represents a successful specialist phage, whereas HTVC010P is likely a successful generalist (Section 4.). For *Synechococcus* phage S-B43, AMGs within its genome were reflective of the environmental adaptation of the host (Wang *et al.*, 2020), but we were not able to identify genomic adaptations that could be used to address the success of Skadi in cold-

water environments. Therefore, the polar distribution might be solely attributed to specialisation on SAR11 polar ecotypic hosts (Kraemer *et al.*, 2020) and follow host distribution, not necessarily virally-encoded ecotypic adaptations (Marston *et al.*, 2013). Environmentally-controlled one-step growth experiments with additional transcriptomics measurements might be necessary to resolve, for example, if polar conditions impact replication efficiency and/or if any unidentified AMGs are up/downregulated in the transcriptome. Furthermore, the results presented here indicate that abundance of seven out of 28 pelagiphages was correlated with temperature (Supplementary Figure 4.4), providing additional evidence that niche variation in SAR11 virus-host systems is ecologically significant and not an exception. However, this further highlights the remaining 21 species whose distribution drivers remain unclear. Niche partitioning in viruses of SAR11 as the most abundant marine heterotroph could have implication for global carbon remineralization rates, similar to how phototrophic niche partitioning has been linked to CO_2-fixation rates across the oligotrophic South Pacific Subtropical Ocean (Duerschlag *et al.*, 2021). Improved host-range prediction, both culture- dependent and independent, would likely help illuminate how generalist and specialist pelagiphage species might contribute to niche variation and SAR11 virus-host population structures.

Pelagibacter phage Skadi as a highly abundant example for polar viral niche variation could allude to potentially threatened viral communities in the Arctic Ocean. In the present day, the Arctic Ocean is characterized by high stratification, low salinity and temperatures, and high riverine nutrient influx. Increasing climate concerns are causing a significant increase in Arctic sea surface temperatures, which are already 1-3 °C warmer than the mean temperature a decade ago (NOAA Arctic Report Card 2020, Timmermans & Labe). Heating polar oceans are likely to cause a shift in the microbial communities (Comeau *et al.*, 2011), and changing phytoplankton diversity with increasingly unstable population structures (Henson *et al.*, 2021). Arctic SAR11 communities are dominated by both bi-polar and Arctic endemic strains, which are likely targets for Skadi infection (Kraemer *et al.*, 2020). Though the lack of seasonal information makes specific predictions about SAR11 successions with changing temperatures impossible, continued changes in the Arctic environment will likely have drastic impacts, and might lead

to the loss of polar microbial diversity. Rank abundance curves (Figure 4.6) suggest that either pelagiphage diversity is drastically underrepresented in culture collections, or that that diversity is low, which is unlikely as the Arctic have been reported as a hotspot of viral diversity (de Cárcer *et al.*, 2015; Kim *et al.*, 2016; Gregory *et al.*, 2019). Danovaro et al. reviewed eight case studies of how climate change could influence virus-host systems, which are too extensive to repeat here (Danovaro *et al.*, 2011). However, the authors came to the conclusion that we cannot yet predict how viruses will influence the ocean's feedback on climate change, but highlighted that testing the effects of environmental conditions on Arctic virus-host systems will be the first step to resolve the multifaceted role of virus-host systems under future Arctic conditions. As Skadi is extremely abundant in both polar oceans, it is the perfect model system to bring pelagiphages into climate change research.

5.10 Summary of suggested future work

The improved high-throughput viral system isolation methods described here have resulted in the isolation of several important virus-host systems, including the first viruses for the OM43 clade. The genomic and ecological evaluation of these virus-host systems raised several questions, some of the most promising areas for future projects, as discussed above, should include:

(1) *Develop machine learning for the interpretation of flow cytometry data*. Improved screening for viral infections will help to capture phages that cause no or only partial lysis, which may include chronic and lysogenic infections, and lytic infections on suboptimal hosts.

(2) *Apply viral particle enumeration to 'steady-state' one-step experiments*. Such a study will provide evidence for 'Soylent Green' dynamics in SAR11 virus-host systems, and help determine ecological dynamics in oligotrophic microbial communities.

(3) *Conduct carbon isotope labels with additional metabolomes to quantify intra-population carbon cycling*. This project would allow estimations of carbon fluxes within SAR11 systems, and would help identify viral infection strategies, as well as improve modelling the SAR11 component of global carbon budgets.

(4) *Resolve cell-surface and curli protein structure for OM43 infected with Melnitz*. Using cryo-electron tomography should be able to determine if the viral curli genes create pore-like or amyloid-fibre-like extracellular structures, thereby provide evidence for the function of the highly unusual curli operon observed in myophages of SAR11 and OM43.

(5) *Re-create viral inter-class host transition between streamlined heterotrophs*. This could be achieved by adapting the Appelmans protocol with the isolated OM43/SAR11 virus-host systems, expand isolation campaigns or utilise single-cell genomics. This should provide evidence for the evolutionary and ecological dynamics of the viruses in streamlined marine heterotrophs, and might provide evidence for the proposed "Genetic Exchange Hub" hypothesis.

(6) *Assess impacts of viral infection on volatile organic carbon and energy cycling in marine heterotrophs*. Utilising the isolated OM43 virus-host systems together with recent advances in PTR-MS coupled with dynamic stripping chambers should facilitate the first quantitative measurements of virus-host system VOC cycles. This project has been funded as a Postdoctoral Simons Foundation Fellowship for 2022 (Award ID: 879226).

(7) *Measure the impact of environmental conditions on polar pelagiphages*. Because polar oceans are under serious threat from climate change, measuring how the SAR11 virus-host systems respond to possible changes is critical for predicting future microbial communities. This can be achieved using *Pelagibacter* phage Skadi, and simple one-step growth experiments under controlled conditions, such as temperature, O_2 concentrations and salinity.

(8) *Apply High-throughput isolation to underrepresented microbial groups*. The workflow provided here should be adapted on SAR11 (*Siphoviridae* and *Myoviridae*), SAR86, SAR116, SAR202, LD12, LD28 and OM43, which will provide much needed reference strains for these ecologically significant virus-host systems, improving culture-independent tool development and improve interpretations of historic meta'omics datasets.

(9) *Sequence the remaining 86 viral isolates for SAR11 and OM43*. This could be done with no additional experimental work, but will require between £6020-£8400 and a considerable amount of time. Any amount of additional reference genomes

would improve culture-independent tools, and any single isolates has the potential of being a novelty that could revolutionise the field.

(10) *Utilise phage cultures to elucidate the function of unknown phage genes.* The viral isolates from this study can be used for transcriptomics and MS-based proteomics to elucidate possible functions of un-annotated ORFs, for example *Kolga* as the first and only represetative of SAR11 associated siphophages has many unidentified genes. Improving public databases would greatly enhance our capabilities to annotate virome contigs and other isolate genomes.

(11) *High-throughput isolation of phage 'cocktails' for SAR11 for mesocosm experiments*. The model SAR11 strain HTCC7211 could be used to isolate more specific viruses for it, which are then added to natural communities from the Sargasso Sea. With 16S rRNA analysis, cascading lysis through the community could be monitored over time. In combination with DOC measurements this would allow for investigations of the pelagiphage predation impact on marine microbial community structure and carbon cycles.

Recent improvements in bacterial cultivation techniques (Henson *et al.*, 2020) have unlocked a large range of diverse bacterioplankton for cultivation. Together with the advancements in viral isolation for fastidious marine heterotrophs presented here, my research will support all efforts in isolation and cultivation of difficult-to-work with virus-host system. The work here provides examples of at least three important novel virus host system: the first siphovirus for SAR11, the first viruses for OM43 and an extremely abundant pelagiphages the ecogenomic analysis of which exemplified ecotypic distribution pattern for the most abundant virus-host system on Earth. Each one of these systems has the potential to play its part in improving global carbon models and in the light of climate change, help save the environment. I hope that these examples will encourage much needed viral cultivation campaigns that will better our understanding of the field of marine viral ecology.

References

Acheson, N.E. & Tamm, I. (1970). Ribonuclease sensitivity of Semliki Forest virus nucleocapsids. Journal of Virology, 5(6), 714-717, http://dx.doi.org/10.1128/JVI.5.6.714-717.1970

Ackermann, H.-W. (2007). 5500 Phages examined in the electron microscope. *Archives of Virology*, *152*(2), 227–243. http://dx.doi.org/10.1007/s00705-006-0849-1

Ács, N., Gambino, M., & Brøndsted, L. (2020). Bacteriophage Enumeration and Detection Methods. *Frontiers in Microbiology*, *11*, 594868. http://dx.doi.org/10.3389/fmicb.2020.594868

Adriaenssens, E. M., Sullivan, M. B., Knezevic, P., van Zyl, L. J., Sarkar, B. L., Dutilh, B. E., Alfenas-Zerbini, P., Łobocka, M., Tong, Y., Brister, J. R., Moreno Switt, A. I., Klumpp, J., Aziz, R. K., Barylski, J., Uchiyama, J., Edwards, R. A., Kropinski, A. M., Petty, N. K., Clokie, M. R. J., & Krupovic, M. (2020). Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Archives of Virology*, *165*(5), 1253–1260. http://dx.doi.org/10.1007/s00705-020-04577-8

Aguirre de Cárcer, D., López-Bueno, A., Pearce, D. A., & Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Science Advances*, *1*(5), e1400127. http://dx.doi.org/10.1126/sciadv.1400127

Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Research*, *45*(1), 39–53. http://dx.doi.org/10.1093/nar/gkw1002

Alonso-Sáez, L., Morán, X. A. G., & Clokie, M. R. (2018). Low activity of lytic pelagiphages in coastal marine waters. *The ISME Journal*, *12*(8), 2100–2102. http://dx.doi.org/10.1038/s41396-018-0185-y

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2

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. http://dx.doi.org/10.3389/fgene.2018.00304

Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K.

R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A., & Armbrust, E. V. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, *522*(7554), 98–101. http://dx.doi.org/10.1038/nature14488

Ando, H., Lemire, S., Pires, D. P., & Lu, T. K. (2015). Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Systems*, *1*(3), 187-196, http://dx.doi.org/10.1016/j.cels.2015.08.013

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, *55*(4), 539–552. http://dx.doi.org/10.1080/10635150600755453

Arrigo, K. R., & van Dijken, G. L. (2015). Continued increases in Arctic Ocean primary production. *Progress in Oceanography*, *136*, 60–70. http://dx.doi.org/10.1016/j.pocean.2015.05.002

Avrani, S., Schwartz, D. A., & Lindell, D. (2012). Virus-host swinging party in the oceans: Incorporating biological complexity into paradigms of antagonistic coexistence. *Mobile Genetic Elements*, *2*(2), 88–95. http://dx.doi.org/10.4161/mge.20031

Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., & Lindell, D. (2011). Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature*, *474*(7353), 604–608. http://dx.doi.org/10.1038/nature10172

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., & Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series*, *10*, 257–263. DOI:10.3354/meps010257

Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews. Microbiology*, *5*(10), 782–791. http://dx.doi.org/10.1038/nrmicro1747

Bailly-Bechet, M., Vergassola, M., & Rocha, E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Research*, *17*(10), 1486–1495. http://dx.doi.org/10.1101/gr.6649807

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, *28*(1), 45–48. http://dx.doi.org/10.1093/nar/28.1.45

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: A Journal of Computational Molecular Cell Biology*, *19*(5), 455–477. http://dx.doi.org/10.1089/cmb.2012.0021

Baran, N., Goldin, S., Maidanik, I., & Lindell, D. (2018). Quantification of diverse virus populations in the environment using the polony method. *Nature Microbiology, 3*(1), 62-72, http://dx.doi.org/10.1038/s41564-017-0045-y

Barnhart, M. M., & Chapman, M. R. (2006). Curli biogenesis and function. *Annual Review of Microbiology*, *60*, 131–147. http://dx.doi.org/10.1146/annurev.micro.60.080805.142106

Bartelme, R. P., Custer, J. M., Dupont, C. L., Espinoza, J. L., Torralba, M., Khalili, B., & Carini, P. (2020). Influence of Substrate Concentration on the Culturability of Heterotrophic Soil Microbes Isolated by High-Throughput Dilution-to-Extinction Cultivation. *mSphere*, *5*(1). https://doi.org/10.1128/mSphere.00024-20

Barylski, J., Enault, F., Dutilh, B. E., Schuller, M. B. P., Edwards, R. A., Gillis, A., Klumpp, J., Knezevic, P., Krupovic, M.; Kuhn, J. H., Lavigne, R., Oksanen, H. M., Sullivan, M. B., Jang, H. B., Simmonds, P., Aiewsakun, P., Wittmann, J., Tolstoy, I., Brister, J. R.; Kropinski, A. M. & Adriaenssens, E. M. (2020) Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages. *Systematic Biology*, *69*(1), 110-123, http://dx.doi.org/10.1093/sysbio/syz036

Bayliss, S. C., Thorpe, H. A., Coyle, N. M., Sheppard, S. K., & Feil, E. J. (2019). PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience*, *8*(10). https://doi.org/10.1093/gigascience/giz119

Beale, R., Dixon, J. L., Smyth, T. J., & Nightingale, P. D. (2015). Annual study of oxygenated volatile organic compounds in UK shelf waters. *Marine Chemistry*, *171*, 96–106. http://dx.doi.org/10.1016/j.marchem.2015.02.013

Becker, J. W., Hogle, S. L., Rosendo, K., & Chisholm, S. W. (2019). Co-culture and biogeography of Prochlorococcus and SAR11. *The ISME Journal*. https://doi.org/10.1038/s41396-019-0365-4

Beier, S., Gálvez, M. J., Molina, V., Sarthou, G., Quéroué, F., Blain, S., & Obernosterer, I. (2015). The transcriptional regulation of the glyoxylate cycle in SAR11 in response to iron fertilization in the Southern Ocean. *Environmental Microbiology Reports*, *7*(3), 427–434. http://dx.doi.org/10.1111/1758-2229.12267

Belcaid, M., Bergeron, A., & Poisson, G. (2010). Mosaic graphs and

comparative genomics in phage communities. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *17*(9), 1315–1326. http://dx.doi.org/10.1089/cmb.2010.0108

Bergh, O., Borsheim, K., Bratbakt, G. & Heldal, M.. (1989). High abundance of viruses found in aquatic environments. *Nature*, *340*(10), 467–468.

Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters*, *363*(4). https://doi.org/10.1093/femsle/fnw002

Berube, P. M., Biller, S. J., Hackl, T., Hogle, S. L., Satinsky, B. M., Becker, J. W., Braakman, R., Collins, S. B., Kelly, L., Berta-Thompson, J., Coe, A., Bergauer, K., Bouman, H. A., Browning, T. J., De Corte, D., Hassler, C., Hulata, Y., Jacquot, J. E., Maas, E. W., … Chisholm, S. W. (2018). Single cell genomes of Prochlorococcus, Synechococcus, and sympatric microbes from diverse marine environments. *Scientific Data*, *5*, 180154. http://dx.doi.org/10.1038/sdata.2018.154

Besemer, J., & Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Research*, *27*(19), 3911–3920. http://dx.doi.org/10.1093/nar/27.19.3911

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, *29*(12), 2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607

Bettarel, Y., Kan, J., Wang, K., Williamson, K. E., & Coats, D. W. (2005). Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom Chaetoceros cf. gracilis. *Aquatic Microbial Ecology: International Journal*, *40*(2). https://doi.org/10.3354/ame040103

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. http://dx.doi.org/10.1038/s41587-019-0100-8

Boeckaerts, D., Stock, M., Criel, B., Gerstmans, H., De Baets, B., & Briers, Y. (2021). Predicting bacteriophage hosts based on sequences of annotated receptor-binding proteins. *Scientific Reports*, *11*(1), 1467. http://dx.doi.org/10.1038/s41598-021-81063-4

Bolduc, B., Jang, H. B., Doulcier, G., You, Z.-Q., Roux, S., & Sullivan, M. B.

(2017). vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ*, *5*, e3243. http://dx.doi.org/10.7717/peerj.3243

Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., & Maxwell, K. L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal*, *10*(12), 2854–2866. http://dx.doi.org/10.1038/ismej.2016.79

Borodovsky, M., & McIninch, J. (1993). GenMark: Parallel gene recognition for both DNA strands. *Computers Chemistry*, *17*(2), 123–133.

Bourrelle-Langlois, M., Morrow, G., Finet, S.,& Tanguay, R. M. (2016). In Vitro Structural and Functional Characterization of the Small Heat Shock Proteins (sHSP) of the Cyanophage S-ShM2 and Its Host, Synechococcus sp. WH7803. *PLoS One, 11*(9), e0162233, http://dx.doi.org/10.1371/journal.pone.0162233

Breitbart, M. (2012). Marine viruses: truth or dare. *Annual Review of Marine Science*, *4*, 425–448. 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. http://dx.doi.org/10.1038/s41564-018-0166-y

Breitbart, M., Miyake, J. H., & Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiology Letters*, *236*(2), 249–256. http://dx.doi.org/10.1016/j.femsle.2004.05.042

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. http://dx.doi.org/10.1016/j.tim.2005.04.003

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(22), 14250–14255. http://dx.doi.org/10.1073/pnas.202488399

Brinkman, N.E., Villegas, E. N., Garland, J.L., & Keely, S.P. (2018). Reducing inherent biases introduced during DNA viral metagenome analyses of municipal wastewater. *PLoS One*, *13*(4), e0195350, http://dx.doi.org/10.1371/journal.pone.0195350

Brister, J. R., Ako-Adjei, D., Bao, Y., & Blinkova, O. (2015). NCBI viral genomes resource. *Nucleic Acids Research*, *43*(Database issue), D571–D577.

http://dx.doi.org/10.1093/nar/gku1207

Brives, C., & Pourraz, J. (2020). Phage therapy as a potential solution in the fight against AMR: obstacles and possible futures. *Palgrave Communications*, *6*(1), 1–11. http://dx.doi.org/10.1057/s41599-020-0478-4

Brown, M. V., Lauro, F. M., DeMaere, M. Z., Muir, L., Wilkins, D., Thomas, T., Riddle, M. J., Fuhrman, J. A., Andrews-Pfannkoch, C., Hoffman, J. M., McQuaid, J. B., Allen, A., Rintoul, S. R., & Cavicchioli, R. (2012). Global biogeography of SAR11 marine bacteria. *Molecular Systems Biology*, *8*, 595. http://dx.doi.org/10.1038/msb.2012.28

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. http://dx.doi.org/10.1038/ismej.2015.125

Brum, J. R., Ignacio-Espinoza, J. C., Kim, E.-H., Trubl, G., Jones, R. M., Roux, S., VerBerkmoes, N. C., Rich, V. I., & Sullivan, M. B. (2016). Illuminating structural proteins in viral “dark matter” with metaproteomics. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(9), 2436–2441. http://dx.doi.org/10.1073/pnas.1525139113

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J. M., Gorsky, G., Gregory, A. C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B. T., … Sullivan, M. B. (2015). Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science*, *348*(6237), 1261498. http://dx.doi.org/10.1126/science.1261498

Brum, J. R., Schenck, R. O., & Sullivan, M. B. (2013). Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *The ISME Journal*, *7*(9), 1738–1751. http://dx.doi.org/10.1038/ismej.2013.67

Brum, J. R., & Sullivan, M. B. (2015). Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nature Reviews. Microbiology*, *13*(3), 147–159. https://doi.org/10.1038/nrmicro3404

Brussaard, C. P. D., Wilhelm, S. W., Thingstad, F., Weinbauer, M. G., Bratbak, G., Heldal, M., Kimmance, S. A., Middelboe, M., Nagasaki, K., Paul, J. H., Schroeder, D. C., Suttle, C. A., Vaqué, D., & Wommack, K. E. (2008). Global-scale processes with a nanoscale drive: the role of marine viruses. *The ISME Journal*, *2*(6), 575–578. http://dx.doi.org/10.1038/ismej.2008.31

Bryan, M. J., Burroughs, N. J., Spence, E. M., Clokie, M. R. J., Mann, N. H., &

Bryan, S. J. (2008). Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. *PloS One*, *3*(4), e2048. http://dx.doi.org/10.1371/journal.pone.0002048

Buchholz, H. H., Michelsen, M. L., Bolaños, L. M., Browne, E., Allen, M. J., & Temperton, B. (2021a). Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria. *The ISME Journal*, *15*, 1585–1598. http://dx.doi.org/10.1038/s41396-020-00872-z

Buchholz, H. H., Michelsen, M., Parsons, R. J., Bates, N. R., & Temperton, B. (2021b). Draft Genome Sequences of Pelagimyophage Mosig EXVC030M and Pelagipodophage Lederberg EXVC029P, Isolated from Devil's Hole, Bermuda. *Microbiology Resource Announcements*, *10*(7). https://doi.org/10.1128/MRA.01325-20

Buchholz, H. H., Bolaños, L. M., Bell, A. G., Michelsen, M. L., Allen, M. J., & Temperton, B. (2021c). Genomic evidence for inter-class host transition between abundant streamlined heterotrophs by a novel and ubiquitous marine Methylophage. *bioRxiv*. https://doi.org/10.1101/2021.08.24.457595

Burrowes, B. H., Molineux, I. J., & Fralick, J. A. (2019). Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans Protocol. *Viruses*, *11*(3). https://doi.org/10.3390/v11030241

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PloS One*, *12*(10), e0185056. http://dx.doi.org/10.1371/journal.pone.0185056

Button, D. K. (1993). Nutrient-limited microbial growth kinetics: overview and recent advances. *Antonie van Leeuwenhoek*, *63*(3-4), 225–235. http://dx.doi.org/10.1007/BF00871220

Cai, X., Tian, F., Teng, L., Liu, H., Tong, Y., Le, S. & Zhang, T. (2021) Cultivation of a Lytic Double-Stranded RNA Bacteriophage Infecting Microvirgula aerodenitrificans Reveals a Mutualistic Parasitic Lifestyle. *Journal of Virology. 95*(17), e0039921, http://dx.doi.org/10.1128/JVI.00399-21

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. http://dx.doi.org/10.1038/nmeth.3869

Callanan, J., Stockdale, S. R., Shkoporov, A., Draper, L. A., Ross, R. P., & Hill, C. (2018) RNA Phage Biology in a Metagenomic Era. *Viruses*, *10*(7), 386, http://dx.doi.org/10.3390/v10070386

Callanan, J., Stockdale, S. R., Shkoporov, A., Draper, L. A., Ross, R. P., & Hill, C. (2020) Expansion of known ssRNA phage genomes: From tens to over a thousand. *Science Advances*, *6*(6), eaay5981, http://dx.doi.org/10.1126/sciadv.aay5981

Carini, P., Campbell, E. O., Morré, J., Sañudo-Wilhelmy, S. A., Thrash, J. C., Bennett, S. E., Temperton, B., Begley, T., & Giovannoni, S. J. (2014). Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. *The ISME Journal*, *8*(8), 1727–1738. http://dx.doi.org/10.1038/ismej.2014.61

Carini, P., Steindler, L., Beszteri, S., & Giovannoni, S. J. (2013). Nutrient requirements for growth of the extreme oligotroph "Candidatus Pelagibacter ubique" HTCC1062 on a defined medium. *The ISME Journal*, *7*(3), 592–602. http://dx.doi.org/10.1038/ismej.2012.122

Carini, P., Van Mooy, B. A. S., Thrash, J. C., White, A., Zhao, Y., Campbell, E. O., Fredricks, H. F., & Giovannoni, S. J. (2015). SAR11 lipid renovation in response to phosphate starvation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(25), 7767–7772. http://dx.doi.org/10.1073/pnas.1505034112

Carini, P., White, A. E., Campbell, E. O., & Giovannoni, S. J. (2014). Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nature Communications*, *5*, 4346. http://dx.doi.org/10.1038/ncomms5346

Carlson, C. A., Morris, R., Parsons, R., Treusch, A. H., Giovannoni, S. J., & Vergin, K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *The ISME Journal*, *3*(3), 283–295. http://dx.doi.org/10.1038/ismej.2008.117

Casey, A., Jordan, K., Neve, H., Coffey, A., & McAuliffe, O. (2015). A tail of two phages: genomic and functional analysis of Listeria monocytogenes phages vB_LmoS_188 and vB_LmoS_293 reveal the receptor-binding proteins involved in host specificity. *Frontiers in Microbiology*, *6*, 1107. http://dx.doi.org/10.3389/fmicb.2015.01107

Casjens, S. R., & Gilcrease, E. B. (2009). Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions. In K. A. M. Clokie M. R. (Ed.), *Bacteriophages: Methods and Protocols* (Vol. 502, pp. 91–111). Humana Press.

Castledine, M., Padfield, D., Sierocinski, P., Pascual, J. S., Hughes, A., Mäkinen, L., Friman, V.-P., Pirnay, J.-P., Merabishvili, M., De Vos, D., &

Buckling, A. (2021). Parallel evolution of phage resistance - virulence trade - offs during in vitro and nasal Pseudomonas aeruginosa phage treatment. In *bioRxiv* (p. 2021.09.06.459069). https://doi.org/10.1101/2021.09.06.459069

Catalão, M. J., Gil, F., Moniz-Pereira, J., São-José, C., & Pimentel, M. (2013). Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiology Reviews*, *37*(4), 554–571. http://dx.doi.org/10.1111/1574-6976.12006

Ceyssens, P.-J., De Smet, J., Wagemans, J., Akulenko, N., Klimuk, E., Hedge, S., Voet, M., Hendrix, H., Paeshuyse, J., Landuyt, B., Xu, H., Blanchard, J., Severinov, K., & Lavigne, R. (2020). The Phage-Encoded N-Acetyltransferase Rac Mediates Inactivation of Pseudomonas aeruginosa Transcription by Cleavage of the RNA Polymerase Alpha Subunit. *Viruses*, *12*(9). https://doi.org/10.3390/v12090976

Chang, C.-Y., Kemp, P., & Molineux, I. J. (2010). Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. *Virology*, *398*(2), 176–186. http://dx.doi.org/10.1016/j.virol.2009.12.002

Chen, L.-X., Zhao, Y., McMahon, K. D., Mori, J. F., Jessen, G. L., Nelson, T. C., Warren, L. A., & Banfield, J. F. (2019). Wide Distribution of Phage That Infect Freshwater SAR11 Bacteria. *mSystems*, *4*(5). https://doi.org/10.1128/mSystems.00410-19

Chiang, S. M., & Schellhorn, H. E. (2010). Evolution of the RpoS regulon: origin of RpoS and the conservation of RpoS-dependent regulation in bacteria. *Journal of Molecular Evolution*, *70*(6), 557–571. http://dx.doi.org/10.1007/s00239-010-9352-0

Chow, C.-E. T., & Fuhrman, J. A. (2012). Seasonality and monthly dynamics of marine myovirus communities. *Environmental Microbiology*, *14*(8), 2171–2183. http://dx.doi.org/10.1111/j.1462-2920.2012.02744.x

Chow, C.-E. T., Kim, D. Y., Sachdeva, R., Caron, D. A., & Fuhrman, J. A. (2014). Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. *The ISME Journal*, *8*(4), 816–829. http://dx.doi.org/10.1038/ismej.2013.199

Clokie, M. R. J., Thalassinos, K., Boulanger, P., Slade, S. E., Stoilova-McPhie, S., Cane, M., Scrivens, J. H., & Mann, N. H. (2008). A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. *Microbiology, 154*(6), 1775-1782, http://dx.doi.org/10.1099/mic.0.2007/016261-0

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423. http://dx.doi.org/10.1093/bioinformatics/btp163

Cohan, F. M. (2006). Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1475), 1985–1996. http://dx.doi.org/10.1098/rstb.2006.1918

Coleman, M.L., Sullivan, M.B, Martiny, A., Steglich, C., Barry, K., DeLong, E.F., Chisholm, S.W. (2006). Genomic islands and the ecology and evolution of Prochlorococcus. *Science*, *311*(5768), 1768-1770. http://dx.doi.org/10.1126/science.1122050

Comeau, A. M., Li, W. K. W., Tremblay, J.-É., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PloS One*, *6*(11), e27492. http://dx.doi.org/10.1371/journal.pone.0027492

Conners, R., McLaren, M., Łapińska, U., Sanders, K., Stone, M. R. L., Blaskovich, M. A. T., Pagliara, S., Daum, B., Rakonjac, J., & Gold, V. A. M. (2021). CryoEM structure of the outer membrane secretin channel pIV from the f1 filamentous bacteriophage. *Nature Communications, 12*(1), 6316, http://dx.doi.org/10.1038/s41467-021-26610-3 Connon, S. A., & Giovannoni, S. J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied and Environmental Microbiology*, *68*(8), 3878–3885. http://dx.doi.org/10.1128/AEM.68.8.3878-3885.2002

Cook, R., Brown, N., Redgwell, T., Rihtman, B., Barnes, M., Clokie, M., Stekel, D. J., Hobman, J., Jones, M. A., & Millard, A. (2021). INfrastructure for a PHAge REference Database: Identification of large-scale biases in the current collection of phage genomes. In *bioRxiv* (p. 2021.05.01.442102). https://doi.org/10.1101/2021.05.01.442102

Coutinho, F. H., Silveira, C. B., Gregoracci, G. B., Thompson, C. C., Edwards, R. A., Brussaard, C. P. D., Dutilh, B. E., & Thompson, F. L. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications*, *8*(15955), 1–12.

Coutinho, F. H., Zaragoza-Solas, A., López-Pérez, M., Barylski, J., Zielezinski, A., Dutilh, B. E., Edwards, R., & Rodriguez-Valera, F. (2021). RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content.

Patterns of Prejudice, 100274. http://dx.doi.org/10.1038/ncomms15955

Cremonese, M., Giampaoli, C., Matzeu, M., & Onori, G. (1969) Cesium chloride density gradient study of some modifications induced by ultraviolet radiation on the denatured DNA molecule of phage T2. *Biophysical Journal, 9*(12), 1451-1463, http://dx.doi.org/10.1016/S0006-3495(69)86465-9

Cross, K. L., Campbell, J. H., Balachandran, M., Campbell, A. G., Cooper, S. J., Griffen, A., Heaton, M., Joshi, S., Klingeman, D., Leys, E., Yang, Z., Parks, J. M., & Podar, M. (2019). Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nature Biotechnology*, *37*(11), 1314–1321. http://dx.doi.org/10.1038/s41587-019-0260-6

Culley, A.I., Lang, A.S., & Suttle, C.A. (2006) Metagenomic Analysis of Coastal RNA Virus Communities. *Science*, *312*(5781), 1795-1798, http://dx.doi.org/10.1126/science.1127404

Danovaro, R., Corinaldesi, C., Dell'anno, A., Fuhrman, J. A., Middelburg, J. J., Noble, R. T., & Suttle, C. A. (2011). Marine viruses and global climate change. *FEMS Microbiology Reviews*, *35*(6), 993–1034. http://dx.doi.org/10.1111/j.1574-6976.2010.00258.x

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., & Weinbauer, M. (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*, *454*(7208), 1084–1087. http://dx.doi.org/10.1038/nature07268

Davern, C.I. (1963) The isolation and characterization of an RNA bacteriophage. *Australian Journal of Biological Sciences, 17*(3), 719-725, http://dx.doi.org/10.1071/bi9640719

Davie-Martin, C. L., Giovannoni, S. J., Behrenfeld, M. J., Penta, W. B., & Halsey, K. H. (2020). Seasonal and Spatial Variability in the Biogenic Production and Consumption of Volatile Organic Compounds (VOCs) by Marine Plankton in the North Atlantic Ocean. *Frontiers in Marine Science*, *7*, 1144. http://dx.doi.org/10.3389/fmars.2020.611870

de Jonge, P. A., Nobrega, F. L., Brouns, S. J. J., & Dutilh, B. E. (2019). Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends in Microbiology*, *27*(1), 51–63. http://dx.doi.org/10.1016/j.tim.2018.08.006

Dekel-Bird, N. P., Sabehi, G., Mosevitzky, B., & Lindell, D. (2015). Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. *Environmental Microbiology*, *17*(4), 1286–1299. http://dx.doi.org/10.1111/1462-2920.12569

Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, *23*(6), 673–679. http://dx.doi.org/10.1093/bioinformatics/btm009

Del Giudice, M. G., Ugalde, J. E., & Czibener, C. (2013). A lysozyme-like protein in Brucella abortus is involved in the early stages of intracellular replication. *Infection and Immunity*, *81*(3), 956–964. http://dx.doi.org/10.1128/IAI.01158-12

Delmont, T. O., Kiefl, E., Kilinc, O., Esen, O. C., Uysal, I., Rappé, M. S., Giovannoni, S., & Eren, A. M. (2019). Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *eLife*, *8*. https://doi.org/10.7554/eLife.46497

DeLong, E. F. (2009). The microbial ocean from genomes to biomes. *Nature*, *459*(7244), 200–206. http://dx.doi.org/10.1038/nature08059

Deng, L., Gregory, A., Yilmaz, S., Poulos, B. T., Hugenholtz, P., & Sullivan, M. B. (2012). Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *mBio*, *3*(6). https://doi.org/10.1128/mBio.00373-12

Deng, L., Ignacio-Espinoza, J. C., Gregory, A. C., Poulos, B. T., Weitz, J. S., Hugenholtz, P., & Sullivan, M. B. (2014). Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. *Nature*, *513*(7517), 242–245. http://dx.doi.org/10.1038/nature13459

de Pascale, D., De Santi, C., Fu, J., & Landfald, B. (2012). The microbial diversity of Polar environments is a fertile ground for bioprospecting. *Marine Genomics*, *8*, 15–22. http://dx.doi.org/10.1016/j.margen.2012.04.004

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., & Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(Web Server issue), W465–W469. http://dx.doi.org/10.1093/nar/gkn180

Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews. Microbiology*, *18*(3), 125–138. http://dx.doi.org/10.1038/s41579-019-0311-5

Doron, S., Fedida, A., Hernández-Prieto, M. A., Sabehi, G., Karunker, I., Stazic, D., Feingersch, R., Steglich, C., Futschik, M., Lindell, D., & Sorek, R. (2016). Transcriptome dynamics of a broad host-range cyanophage and its hosts. *The ISME Journal*, *10*(6), 1437–1455.

http://dx.doi.org/10.1038/ismej.2015.210

Dueholm, M. S., Albertsen, M., Otzen, D., & Nielsen, P. H. (2012). Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. *PloS One*, *7*(12), e51274. http://dx.doi.org/10.1371/journal.pone.0051274

Duerschlag, J., Mohr, W., Ferdelman, T. G., LaRoche, J., Desai, D., Croot, P. L., Voß, D., Zielinski, O., Lavik, G., Littmann, S., Martínez-Pérez, C., Tschitschko, B., Bartlau, N., Osterholz, H., Dittmar, T., & Kuypers, M. M. M. (2021). Niche partitioning by photosynthetic plankton as a driver of CO2-fixation across the oligotrophic South Pacific Subtropical Ocean. *The ISME Journal*, 1–12. http://dx.doi.org/10.1038/s41396-021-01072-z

Duhaime, M. B., Wichels, A., Waldmann, J., Teeling, H., & Glöckner, F. O. (2011). Ecogenomics and genome landscapes of marine Pseudoalteromonas phage H105/1. *The ISME Journal*, *5*(1), 107–121. http://dx.doi.org/10.1038/ismej.2010.94

Duhaime, M.B., Deng L., Poulos, B., & Sullivan, M.B. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method, *Environmental Microbiology*, *14*(9), 2526-2537, http://dx.doi.org/10.1111/j.1462-2920.2012.02791.x

Du, S., Qin, F., Zhang, Z., Tian, Z., Yang, M., Liu, X., Zhao, G., Xia, Q., & Zhao, Y. (2021). Genomic diversity, life strategies and ecology of marine HTVC010P-type pelagiphages. *Microbial Genomics*, *7*(7). https://doi.org/10.1099/mgen.0.000596

Džunková, M., Low, S. J., Daly, J. N., Deng, L., Rinke, C., & Hugenholtz, P. (2019). Defining the human gut host-phage network through single-cell viral tagging. *Nature Microbiology*, *4*(12), 2192–2203. http://dx.doi.org/10.1038/s41564-019-0526-2

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. http://dx.doi.org/10.1093/nar/gkh340

Edwards, R. A., McNair, K., Faust, K., Raes, J., & Dutilh, B. E. (2016). Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews*, *40*(2), 258–272. http://dx.doi.org/10.1093/femsre/fuv048

Enav, H., Kirzner, S., Lindell, D., Mandel-Gutfreund, Y., & Béjà, O. (2018). Adaptation to sub-optimal hosts is a driver of viral diversification in the

ocean. *Nature Communications*, *9*(1), 4698. http://dx.doi.org/10.1038/s41467-018-07164-3

Engelberg-Kulka, H., Hazan, R., & Amitai, S. (2005). mazEF: a chromosomal toxin-antitoxin module that triggers programmed cell death in bacteria. *Journal of Cell Science, 118*(19), 4327-4332, http://dx.doi.org/10.1242/jcs.02619

Eren, A. M., Borisy, G. G., Huse, S. M., & Mark Welch, J. L. (2014). Oligotyping analysis of the human oral microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), E2875–E2884. http://dx.doi.org/10.1073/pnas.1409644111

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. http://dx.doi.org/10.7717/peerj.1319

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, *320*(5879), 1034–1039. http://dx.doi.org/10.1126/science.1153213

Feng, X., Chu, X., Qian, Y., Henson, M. W., Celeste Lanclos, V., Qin, F., Zhao, Y., Cameron Thrash, J., & Luo, H. (2021). Mechanisms Driving Genome Reduction of a Novel Roseobacter Lineage Showing Vitamin B12 Auxotrophy. In *Cold Spring Harbor Laboratory* (p. 2021.01.15.426902). https://doi.org/10.1101/2021.01.15.426902

Fiebig, A., Herrou, J., Willett, J., & Crosson, S. (2015). General Stress Signaling in the Alphaproteobacteria. *Annual Review of Genetics*, *49*, 603–625. http://dx.doi.org/10.1146/annurev-genet-112414-054813

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*, *281*(5374), 237–240. http://dx.doi.org/10.1126/science.281.5374.237

Field, K. G., Gordon, D., Wright, T., Rappé, M., Urback, E., Vergin, K., & Giovannoni, S. J. (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Applied and Environmental Microbiology*, *63*(1), 63–70. http://dx.doi.org/10.1128/aem.63.1.63-70.1997

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., & Ysebaert, M. (1976). Complete nucleotide

sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, *260*(5551), 500-507, http://dx.doi.org/10.1038/260500a0

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., & Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, *42*(Database issue), D222–D230. http://dx.doi.org/10.1093/nar/gkt1223

Fischer, E. V., Jacob, D. J., Millet, D. B., Yantosca, R. M., & Mao, J. (2012). The role of the ocean in the global atmospheric budget of acetone. *Geophysical Research Letters*, *39*(1). https://doi.org/10.1029/2011gl050086

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., & Martiny, A. C. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(24), 9824–9829. http://dx.doi.org/10.1073/pnas.1307701110

Forterre, P. (2011). Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Comptes Rendus Chimie*, *14*(4), 392–399. http://dx.doi.org/10.1016/j.crci.2010.06.007

Forterre, P. (2012). The Virocell Concept. In *The Encyclopedia of Life Science*. John Wiley & Sons, Ltd. 10.1002/9780470015902.a0000462.pub2

Forterre, P. (2013). The virocell concept and environmental microbiology. *The ISME Journal*, *7*(2), 233–236. http://dx.doi.org/10.1038/ismej.2012.110

Frias, M. J., Melo-Cristino, J., & Ramirez, M. (2009). The autolysin LytA contributes to efficient bacteriophage progeny release in Streptococcus pneumoniae. *Journal of Bacteriology*, *191*(17), 5428–5440. doi/10.1128/JB.00477-09

Fridman, S., Flores-Uribe, J., Larom, S., Alalouf, O., Liran, O., Yacoby, I., Salama, F., Bailleul, B., Rappaport, F., Ziv, T., Sharon, I., Cornejo-Castillo, F. M., Philosof, A., Dupont, C. L., Sánchez, P., Acinas, S. G., Rohwer, F. L., Lindell, D., & Béjà, O. (2017). A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected Prochlorococcus cells. *Nature Microbiology*, *2*(10), 1350–1357. http://dx.doi.org/10.1038/s41564-017-0002-9

Fuhrman, J. A., Schwalbach, M. S., & Stingl, U. (2008). Proteorhodopsins: an array of physiological roles? *Nature Reviews. Microbiology*, *6*(6), 488–494.

http://dx.doi.org/10.1038/nrmicro1893

Galiez, C., Siebert, M., Enault, F., Vincent, J., & Söding, J. (2017). WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics*, *33*(19), 3113–3114. http://dx.doi.org/10.1093/bioinformatics/btx383

Gebbie, G., & Huybers, P. (2012). The Mean Age of Ocean Waters Inferred from Radiocarbon Observations: Sensitivity to Surface Sources and Accounting for Mixing Histories. *Journal of Physical Oceanography*, *42*(2), 291–305. http://dx.doi.org/10.1175/JPO-D-11-043.1

Geiduschek, E. P., & Kassavetis, G. A. (2010). Transcription of the T4 late genes. *Virology Journal*, *7*, 288. https://virologyj.biomedcentral.com/articles/10.1186/1743-422X-7-288

Ghiglione, J.-F., Galand, P. E., Pommier, T., Pedrós-Alió, C., Maas, E. W., Bakker, K., Bertilson, S., Kirchmanj, D. L., Lovejoy, C., Yager, P. L., & Murray, A. E. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17633–17638. http://dx.doi.org/10.1073/pnas.1208160109

Ghosh, S., Shaw, R., Sarkar, A., & Gupta, S. K. D. (2020). Evidence of positive regulation of mycobacteriophage D29 early gene expression obtained from an investigation using a temperature-sensitive mutant of the phage. *FEMS Microbiology Letters*, *367*(21). https://doi.org/10.1093/femsle/fnaa176

Gibbons, S. M., Caporaso, J. G., Pirrung, M., Field, D., Knight, R., & Gilbert, J. A. (2013). Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(12), 4651–4655. http://dx.doi.org/10.1073/pnas.1217767110

Gifford, S. M., Becker, J. W., Sosa, O. A., Repeta, D. J., & DeLong, E. F. (2016). Quantitative Transcriptomics Reveals the Growth- and Nutrient-Dependent Response of a Streamlined Marine Methylotroph to Methanol and Naturally Occurring Dissolved Organic Matter. *mBio*, *7*(6). https://doi.org/10.1128/mBio.01279-16

Giovannoni, S. J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science*, *9*, 231–255. http://dx.doi.org/10.1146/annurev-marine-010814-015934

Giovannoni, S. J., Bibbs, L., Cho, J.-C., Stapels, M. D., Desiderio, R., Vergin, K. L., Rappé, M. S., Laney, S., Wilhelm, L. J., Tripp, H. J., Mathur, E. J., &

Barofsky, D. F. (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature*, *438*(7064), 82–85. http://dx.doi.org/10.1038/nature04032

Giovannoni, S. J., Cameron Thrash, J., & Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *The ISME Journal*, *8*(8), 1553–1565. http://dx.doi.org/10.1038/ismej.2014.60

Giovannoni, S. J., Delong, t. E. F., Schmidt, t. T. M., & Pace, N. R. (1990). Tangential Flow Filtration and Preliminary Phylogenetic Analysis of Marine Picoplankton. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *56*(8), 2572–2575. DOI:10.1128/aem.56.8.2572-2575.1990

Giovannoni, S. J., Hayakawa, D. H., Tripp, H. J., Stingl, U., Givan, S. A., Cho, J.-C., Oh, H.-M., Kitner, J. B., Vergin, K. L., & Rappé, M. S. (2008). The small genome of an abundant coastal ocean methylotroph. *Environmental Microbiology*, *10*(7), 1771–1782. http://dx.doi.org/10.1111/j.1462-2920.2008.01598.x

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappé, M. S., Short, J. M., Carrington, J. C., & Mathur, E. J. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, *309*(5738), 1242–1245. http://dx.doi.org/10.1126/science.1114057

Giovannoni, S. J., & Vergin, K. L. (2012). Seasonality in ocean microbial communities. *Science*, *335*(6069), 671–676. http://dx.doi.org/10.1126/science.1198078

Giovannoni, S., Temperton, B., & Zhao, Y. (2013). Giovannoni et al. reply [Review of *Giovannoni et al. reply*]. *Nature*, *499*(7459), E4–E5. http://dx.doi.org/10.1038/nature12388

Godon, J.-J., Bize, A., Ngo, H., Cauquil, L., Almeida, M., Petit, M.-A., & Zemb, O. (2021). Bacterial Consumption of T4 Phages. *Microorganisms*, *9*(9), 1852, http://dx.doi.org/10.3390/microorganisms9091852

Goldsmith, D. B., Parsons, R. J., Beyene, D., Salamon, P., & Breitbart, M. (2015). Deep sequencing of the viral phoH gene reveals temporal variation, depth-specific composition, and persistent dominance of the same viral phoH genes in the Sargasso Sea. *PeerJ*, *3*, e997. http://dx.doi.org/10.7717/peerj.997

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.

http://dx.doi.org/10.3389/fmicb.2018.02981

Good, N. M., Moore, R. S., Suriano, C. J., & Martinez-Gomez, N. C. (2019). Contrasting in vitro and in vivo methanol oxidation activities of lanthanide-dependent alcohol dehydrogenases XoxF1 and ExaF from Methylobacterium extorquens AM1. *Scientific Reports*, *9*(1), 4248. http://dx.doi.org/10.1038/s41598-019-41043-1

Grasis, J.A. (2018). Host-Associated Bacteriophage Isolation and Preparation for Viral Metagenomics. In Pantaleo, V., Chiumenti, M. (Ed.), Molecular Biology (Vol. 1746, pp. 1-25). Springer-Verlag, Berlin

Gregory, A. C., Solonenko, S. A., Ignacio-Espinoza, J. C., LaButti, K., Copeland, A., Sudek, S., Maitland, A., Chittick, L., Dos Santos, F., Weitz, J. S., Worden, A. Z., Woyke, T., & Sullivan, M. B. (2016). Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics*, *17*(1), 930. http://dx.doi.org/10.1186/s12864-016-3286-x

Gregory, A. C., Zayed, A. A., Sunagawa, S., Wincker, P., Sullivan, M. B., Temperton, B., Bolduc, B., Alberti, A., Ardyna, M., Arkhipova, K., Carmichael, M., Cruaud, C., Ferland, J., Kandels, S., Liu, Y., Marec, C., Picheral, M., Pisarev, S., Poulain, J., … Roux, S. (2019). Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell*, *177*, 1109–1123. http://dx.doi.org/10.1016/j.cell.2019.03.040

Grote, J., Thrash, J. C., Huggett, M. J., Landry, Z. C., Carini, P., Giovannoni, S. J., & Rappé, M. S. (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio*, *3*(5). https://doi.org/10.1128/mBio.00252-12

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J., Coelho, L. P., Espinoza, J. C. I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., … Gorsky. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465–470. http://dx.doi.org/10.1038/nature16942

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704. http://dx.doi.org/10.1080/10635150390235520

Guindon, S., Lethiec, F., Duroux, P., & Gascuel, O. (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*, *33*(Web Server issue), W557–W559.

http://dx.doi.org/10.1093/nar/gki352

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. http://dx.doi.org/10.1093/bioinformatics/btt086

Halsey, K. H., Carter, A. E., & Giovannoni, S. J. (2012). Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environmental Microbiology*, *14*(3), 630–640. http://dx.doi.org/10.1111/j.1462-2920.2011.02605.x

Halsey, K. H., Giovannoni, S. J., Graus, M., Zhao, Y., Landry, Z., Thrash, J. C., Vergin, K. L., & de Gouw, J. (2017a). Biological cycling of volatile organic carbon by phytoplankton and bacterioplankton. *Limnology and Oceanography*, *62*(6), 2650–2661. http://dx.doi.org/10.1002/lno.10596

Hanauer, D. I., Graham, M. J., SEA-PHAGES, Betancur, L., Bobrownicki, A., Cresawn, S. G., Garlena, R. A., Jacobs-Sera, D., Kaufmann, N., Pope, W. H., Russell, D. A., Jacobs, W. R., Jr, Sivanathan, V., Asai, D. J., & Hatfull, G. F. (2017). An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(51), 13531–13536. http://dx.doi.org/10.1073/pnas.1718188115

Hanson, B. T., Hewson, I., & Madsen, E. L. (2014). Metaproteomic survey of six aquatic habitats: discovering the identities of microbial populations active in biogeochemical cycling. *Microbial Ecology*, *67*(3), 520–539. http://dx.doi.org/10.1007/s00248-013-0346-5

Henson, M. W., Lanclos, V. C., Faircloth, B. C., & Thrash, J. C. (2018). Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *The ISME Journal*, *12*(7), 1846–1860. http://dx.doi.org/10.1038/s41396-018-0092-2

Henson, M. W., Lanclos, V. C., Pitre, D. M., Weckhorst, J. L., Lucchesi, A. M., Cheng, C., Temperton, B., & Thrash, J. C. (2020). Expanding the Diversity of Bacterioplankton Isolates and Modeling Isolation Efficacy with Large-Scale Dilution-to-Extinction Cultivation. *Applied and Environmental Microbiology*, *86*(17). https://doi.org/10.1128/AEM.00943-20

Henson, M. W., Pitre, D. M., Weckhorst, J. L., Celeste Lanclos, V., Webber, A. T., & Cameron Thrash, J. (2016). Artificial Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico. *American Society for Microbiology*, *1*(2), 1–9.

http://dx.doi.org/10.1128/mSphere.00028-16

Henson, S. A., Cael, B. B., Allen, S. R., & Dutkiewicz, S. (2021). Future phytoplankton diversity in a changing climate. *Nature Communications*, *12*(1), 5372. http://dx.doi.org/10.1038/s41467-021-25699-w

Hidaka, T. (1971). Isolation of Marine Bacteriophages from Seawater. *Bulletin Of The Japanese Society Of Scientific Fisheries, 37*(12), 1199-1206, ISSN 0021-5392

Highfield, A., Joint, I., Gilbert, J. A., Crawfurd, K. J., & Schroeder, D. C. (2017). Change in Emiliania huxleyi Virus Assemblage Diversity but Not in Host Genetic Composition during an Ocean Acidification Mesocosm Experiment. *Viruses, 9*(3), 41, http://dx.doi.org/10.3390/v9030041

Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., Ferrera, I., Sarmento, H., Villar, E., Lima-Mendez, G., Faust, K., Sunagawa, S., Claverie, J.-M., Moreau, H., Desdevises, Y., Bork, P., Raes, J., de Vargas, C., Karsenti, E., … Ogata, H. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal*, *7*(9), 1678–1695. http://dx.doi.org/10.1038/ismej.2013.59

Hinton, D. M. (2010). Transcriptional control in the prereplicative phase of T4 development. *Virology Journal*, *7*, 289. http://dx.doi.org/10.1186/1743-422X-7-289

Hirsch, P., Bernhard, M., Cohen, S. S., Ensign, J. C., Jannasch, H. W., & Koch, A. L. (1979). Life under conditions of low nutrient concentrations. In M. Shilo (Ed.), *Strategies of Microbial Life in Extreme Environments: reports of the Dahlem Workshop on Strategy of Life in Extreme Environments, Berlin 1978* (p. 513). Weinheim Verlag Chemie.

Hoarfrost, A., Nayfach, S., Ladau, J., Yooseph, S., Arnosti, C., Dupont, C. L., & Pollard, K. S. (2020). Global ecotypes in the ubiquitous marine clade SAR86. *The ISME Journal*, *14*(1), 178–188. http://dx.doi.org/10.1038/s41396-019-0516-7

Hoetzinger, M., Nilsson, E., Arabi, R., Osbeck, C. M. G., Pontiller, B., Hutinet, G., Bayfield, O. W., Traving, S., Kisand, V., Lundin, D., Pinhassi, J., Middelboe, M., & Holmfeldt, K. (2021). Dynamics of Baltic Sea phages driven by environmental changes. *Environmental Microbiology, 23*(8), 4576-4594, http://dx.doi.org/10.1111/1462-2920.15651

Horas, E. L., Theodosiou, L., & Becks, L. (2018). Why Are Algal Viruses Not Always Successful? *Viruses*, *10*(9). https://doi.org/10.3390/v10090474

Howard-Varona, C., Roux, S., Dore, H., Solonenko, N. E., Holmfeldt, K., Markillie, L. M., Orr, G., & Sullivan, M. B. (2017). Regulation of infection efficiency in a globally abundant marine Bacteriodetes virus. *The ISME Journal, 11*(1), 284-295, http://dx.doi.org/10.1038/ismej.2016.81

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*. https://doi.org/10.1038/s41396-019-0580-z

Huggett, M. J., Hayakawa, D. H., & Rappé, M. S. (2012). Genome sequence of strain HIMB624, a cultured representative from the OM43 clade of marine Betaproteobacteria. *Standards in Genomic Sciences*, *6*(1), 11–20. http://dx.doi.org/10.4056/sigs.2305090

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., & Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, *1*, 16048. http://dx.doi.org/10.1038/nmicrobiol.2016.48

Hurwitz, B. L., Brum, J. R., & Sullivan, M. B. (2015). Depth-stratified functional and taxonomic niche specialization in the “core” and “flexible” Pacific Ocean Virome. *The ISME Journal*, *9*(2), 472–484. http://dx.doi.org/10.1038/ismej.2014.143

Hurwitz, B. L., Hallam, S. J., & Sullivan, M. B. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology*, *14*(11), R123. http://dx.doi.org/10.1186/gb-2013-14-11-r123

Hurwitz, B. & U’Ren, J.M.. (2016). Viral metabolic reprogramming in marine ecosystems. *Current Opinion in Microbiology*, *31*, 161–168. http://dx.doi.org/10.1016/j.mib.2016.04.002

Hurwitz, B. L., Ponsero, A., Thornton, J., Jr, & U’Ren, J. M. (2018). Phage hunters: Computational strategies for finding phages in large-scale ’omics datasets. *Virus Research*, *244*, 110–115. http://dx.doi.org/10.1016/j.virusres.2017.10.019

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. http://dx.doi.org/10.1371/journal.pone.0057355

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*, 119. http://dx.doi.org/10.1186/1471-2105-11-119

Hyun, Y., Baek, Y., Lee, C., Ki, N., Ahn, J., Ryu, S., & Ha, N.-C. (2021). Structure and Function of the Autolysin SagA in the Type IV Secretion System of Brucella abortus. *Molecules and Cells*, *44*(7), 517–528. http://dx.doi.org/10.14348/molcells.2021.0011

Igarza, M., Dittmar, T., Graco, M., & Niggemann, J. (2019). Dissolved Organic Matter Cycling in the Coastal Upwelling System Off Central Peru During an "El Niño" Year. *Frontiers in Marine Science*, *6*, 198. http://dx.doi.org/10.3389/fmars.2019.00198

Ignacio-Espinoza, J. C., Ahlgren, N. A., & Fuhrman, J. A. (2020). Long-term stability and Red Queen-like strain dynamics in marine viruses. *Nature Microbiology*, *5*, 265–271. http://dx.doi.org/10.1038/s41564-019-0628-x

Jacobsen, A., Bratbak, G., & Heldal, M. (1996). Isolation and characterization of a virus infecting Phaeocystis pouchetii (Prymnesiophyceae). *Journal of Phycology*, *32*(6), 923–927. http://dx.doi.org/10.1111/j.0022-3646.1996.00923.x

Jiao, N., Herndl, G.J., Hansell, D.A., Benner, R., Kattner, G., Wilhelm, S.W., Kirchman, D.L., Weinbauer, M.G., Luo, T., Chen, F., Azam, F. (2010) Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Reviews Microbiology*, 8(8), 593-599. http://dx.doi.org/10.1038/nrmicro2386

Jimenez-Infante, F., Ngugi, D. K., Vinu, M., Alam, I., Kamau, A. A., Blom, J., Bajic, V. B., & Stingl, U. (2016). Comprehensive Genomic Analyses of the OM43 Clade, Including a Novel Species from the Red Sea, Indicate Ecotype Differentiation among Marine Methylotrophs. *Applied and Environmental Microbiology*, *82*(4), 1215–1226. http://dx.doi.org/10.1128/AEM.02852-15

Joassin, P., Delille, B., Soetaert, K., Harlay, J., Borges, A. V., Chou, L., Riebesell, U., Suykens, K., & Grégoire, M. (2011). Carbon and nitrogen flows during a bloom of the coccolithophore Emiliania huxleyi: Modelling a mesocosm experiment. *Journal of Marine Systems, 85*(3), 71-85, http://dx.doi.org/10.1016/j.jmarsys.2010.11.007

Joint, I., Mühling, M. & Querellou, J. (2010). Culturing marine bacteria – an essential prerequisite for biodiversity. *Microbial Biotechnology, 3*(5), 564-575. http://dx.doi.org/10.1111/j.1751-7915.2010.00188.x

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031

Jordan, A., Pontis, E., ATTAt, M., Krook, M., GIBERTt, I., BARBtt, J., & Reichard, P. (1994). A second class I ribonucleotide reductase in Enterobacteriaceae: Characterization of the Salmonella typhimurium enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 12892–12896. https://doi.org/10.1073/pnas.91.26.12892

Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W., & Weitz, J. S. (2014). The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nature Reviews. Microbiology*, *12*(7), 519–528. http://dx.doi.org/10.1038/nrmicro3289

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, https://doi.org/10.1038/s41586-021-03819-2

Kalyuzhnaya, M. G., Bowerman, S., Lara, J. C., Lidstrom, M. E., & Chistoserdova, L. (2006). Methylotenera mobilis gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family Methylophilaceae. *International Journal of Systematic and Evolutionary Microbiology*, *56*(Pt 12), 2819–2823. http://dx.doi.org/10.1099/ijs.0.64191-0

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. http://dx.doi.org/10.1093/nar/gkw1092

Kang, I., Oh, H.-M., Kang, D., & Cho, J.-C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(30), 12343–12348. http://dx.doi.org/10.1073/pnas.1219930110

Kauffman, K. M., Hussain, F. A, Yang, J., Arevalo, P., Brown, J. M., Chang, W. K., Van Insberghe, D., Elsherbini, J., Sharma, R. S., Cutler, M. B., Kelly, L.

& Polz, M. F. (2018) A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature, 554*(7690), 118-122, http://dx.doi.org/10.1038/nature25474

Kazanov, M. D., Vitreschak, A. G., & Gelfand, M. S. (2007). Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics*, *8*, 347. http://dx.doi.org/10.1186/1471-2164-8-347

Keiler, K. C., Shapiro, L., & Williams, K. P. (2000). tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in Caulobacter. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(14), 7778–7783. http://dx.doi.org/10.1073/pnas.97.14.7778

Kim, K.-H. & Bae, J.-W. (2011). Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Applied Environmental Microbiology*, 77(21), 7663-7668, http://dx.doi.org/10.1128/AEM.00289-11

Kim, Y., Aw, T. G., & Rose, J. B. (2016). Transporting Ocean Viromes: Invasion of the Aquatic Biosphere. *PloS One*, *11*(4), e0152671. http://dx.doi.org/10.1371/journal.pone.0152671

Kirchman, D. L. (2016). Growth Rates of Microbes in the Oceans. *Annual Review of Marine Science*, *8*, 285–309. http://dx.doi.org/10.1146/annurev-marine-122414-033938

Klähn, S., Bolay, P., Wright, P. R., Atilho, R. M., Brewer, K. I., Hagemann, M., Breaker, R. R., & Hess, W. R. (2018). A glutamine riboswitch is a key element for the regulation of glutamine synthetase in cyanobacteria. *Nucleic Acids Research*, *46*(19), 10082–10094. http://dx.doi.org/10.1093/nar/gky709

Klappenbach, J. A., Dunbar, J. M., & Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, *66*(4), 1328–1333. http://dx.doi.org/10.1128/AEM.66.4.1328-1333.2000

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

Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobian-Guemes, A. G., Coutinho, F. H., Dinsdale, E. A., Felts, B., Furby, K. A., George, E. E., Green, K. T., Gregoracci, G. B., Haas, A. F., Haggerty, J. M., Hester, E. R., Hisakawa, N., Kelly, L. W., Lim, Y. W., … Rohwer, F.

(2016). Lytic to temperate switching of viral communities. *Nature*, *531*(7595), 466–470. http://dx.doi.org/10.1038/nature17193

Kogure, K., Simidu, U., & Taga, N. (1979). A tentative direct microscopic method for counting living marine bacteria. *Canadian Journal of Microbiology*, *25*(3), 415–420. http://dx.doi.org/10.1139/m79-063

Kondou, Y., Kitazawa, D., Takeda, S., Tsuchiya, Y., Yamashita, E., Mizuguchi, M., Kawano, K., & Tsukihara, T. (2005). Structure of the central hub of bacteriophage Mu baseplate determined by X-ray crystallography of gp44. *Journal of Molecular Biology, 352*(4), 976-985, http://dx.doi.org/10.1016/j.jmb.2005.07.044

Koonin, E. V., Makarova, K. S. & Wolf, Y. I. (2017) Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annual Review Microbiology*, *71*, 233-261. http://dx.doi.org/10.1146/annurev-micro-090816-093830

Koonin, E. V., & Makarova, K. S. (2013). CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biology, 10*(5), 679-686, http://dx.doi.org/10.4161/rna.24022

Korlević, M., Pop Ristova, P., Garić, R., Amann, R., & Orlić, S. (2015). Bacterial diversity in the South Adriatic Sea during a strong, deep winter convection year. *Applied and Environmental Microbiology*, *81*(5), 1715–1726. http://dx.doi.org/10.1128/AEM.03410-14

Koskella, B., & Meaden, S. (2013). Understanding bacteriophage specificity in natural microbial communities. *Viruses*, *5*(3), 806–823. http://dx.doi.org/10.3390/v5030806

Kossykh, V. G., Schlagman, S. L., & Hattman, S. (1995). Phage T4 DNA [N]-adenine6Methyltransferase. OVEREXPRESSION, PURIFICATION, AND CHARACTERIZATION. *The Journal of Biological Chemistry*, *270*(24), 14389–14393. http://dx.doi.org/10.1074/jbc.270.24.14389

Kraemer, S., Ramachandran, A., Colatriano, D., Lovejoy, C., & Walsh, D. A. (2020). Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *The ISME Journal*, *14*(1), 79–90. http://dx.doi.org/10.1038/s41396-019-0499-4

Krishnamurthy, S. R., Janowski, A. B., Zhao, G., Barouch, D., & Wang, D. (2016). Hyperexpansion of RNA Bacteriophage Diversity. *PLoS Biology*, *14*(3), e1002409, http://dx.doi.org/10.1371/journal.pbio.1002409

Krupovic, M. & Forterre, P. (2015). Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Annals of the*

New York Academy of Sciences, 1341, 41-53, http://dx.doi.org/10.1111/nyas.12675

Krupovic, M. & Koonin, E. V. (2017) Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences of the United States of America, 114*(12), E2401-E2410, http://dx.doi.org/10.1073/pnas.1621061114

Labonte, J.M. & Suttle, C.A. (2013). Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. *Frontiers in Microbiology,* 4, 404, http://dx.doi.org/10.3389/fmicb.2013.00404

Labonte, J.M., Hallam, S.J. & Suttle, C.A. (2015). Previously unknown evolutionary groups dominate the ssDNA gokushoviruses in oxic and anoxic waters of a coastal marine environment. *Frontiers in Microbiology,* 6, 315, http://dx.doi.org/10.3389/fmicb.2015.00315

Lang, A. S., Rise, M. L., Culley, A. I., & Steward, G. F. (2009). RNA viruses in the sea. *FEMS Microbiology Reviews, 33*(2), 295-323, http://dx.doi.org/10.1111/j.1574-6976.2008.00132.x

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. http://dx.doi.org/10.1038/nmeth.1923

Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*(1), 11–16. http://dx.doi.org/10.1093/nar/gkh152

Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., DeMaere, M. Z., Ting, L., Ertan, H., Johnson, J., Ferriera, S., Lapidus, A., Anderson, I., Kyrpides, N., Munk, A. C., Detter, C., Han, C. S., Brown, M. V., Robb, F. T., Cavicchioli, R. (2009). The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15527–15533. http://dx.doi.org/10.1073/pnas.0903507106

Lewis, L. O., & Yousten, A. A. (1988). Bacteriophage attachment to the S-layer proteins of the mosquito-pathogenic strains of Bacillus sphaericus. *Current Microbiology*, *17*(1), 55–60. http://dx.doi.org/10.1007/bf01568820

Liang, Y., Wang, L., Wang, Z., Zhao, J., Yang, Q., Wang, M., Yang, K., Zhang, L., Jiao, N., & Zhang, Y. (2019). Metagenomic Analysis of the Diversity of DNA Viruses in the Surface and Deep Sea of the South China Sea. *Frontiers in Microbiology*, *10*, 1951.

http://dx.doi.org/10.3389/fmicb.2019.01951

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352

Lima-Mendez, G., Van Helden, J., Toussaint, A., & Leplae, R. (2008). Reticulate representation of evolutionary and functional relationships between phage genomes. *Molecular Biology and Evolution*, *25*(4), 762–777. http://dx.doi.org/10.1093/molbev/msn023

Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., Kettler, G., Sullivan, M. B., Steen, R., Hess, W. R., Church, G. M., & Chisholm, S. W. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, *449*(7158), 83–86. http://dx.doi.org/10.1038/nature06130

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M., & Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*, *438*, 86–89. http://dx.doi.org/10.1038/nature04111

Loeb, T. & Zinder, N.D. (1961). A Bacteriophage Containing RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 47, 282–289. https://doi.org/10.1073/pnas.47.3.282

Lomsadze, A., Gemayel, K., Tang, S., & Borodovsky, M. (2018). Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Research*, *28*(7), 1079–1089. http://dx.doi.org/10.1101/gr.230615.117

López-Pérez, M., Haro-Moreno, J. M., Coutinho, F. H., Martinez-Garcia, M., & Rodriguez-Valera, F. (2020). The Evolutionary Success of the Marine Bacterium SAR11 Analyzed through a Metagenomic Perspective. *mSystems*, *5*(5). https://doi.org/10.1128/mSystems.00605-20

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, *44*(W1), W54–W57. http://dx.doi.org/10.1093/nar/gkw413

Lupo, D., Leptihn, S., Nagler, G., Haase, M., J Molineux, I., & Kuhn, A. (2015). The T7 ejection nanomachine components gp15-gp16 form a spiral ring complex that binds DNA and a lipid membrane. *Virology*, *486*, 263–271. http://dx.doi.org/10.1016/j.virol.2015.09.022

Mahichi, F., Synnott, A. J., Yamamichi, K., Osada, T., & Tanji, Y. (2009). Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity. *FEMS Microbiology Letters*, *295*(2), 211–217. http://dx.doi.org/10.1111/j.1574-6968.2009.01588.x

Malik, S. S., Azem-E-Zahra, S., Kim, K. M., Caetano-Anollés, G., & Nasir, A. (2017). Do Viruses Exchange Genes across Superkingdoms of Life? *Frontiers in Microbiology*, *8*, 2110. http://dx.doi.org/10.3389/fmicb.2017.02110

Malmstrom, R. R., Kiene, R. P., Cottrell, M. T., & Kirchman, D. L. (2004). Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic ocean. *Applied and Environmental Microbiology*, *70*(7), 4129–4135. http://dx.doi.org/10.1128/AEM.70.7.4129-4135.2004

Mann, N. H. (2003). Phages of the marine cyanobacterial picophytoplankton. *FEMS Microbiology Reviews*, *27*(1), 17–34. http://dx.doi.org/10.1016/S0168-6445(03)00016-0

Mann, N. H., Cook, A., Millard, A., Bailey, S., & Clokie, M. (2003). Bacterial photosynthesis genes in a virus. *Nature*, *424*, 741. https://doi.org/10.1038/424741a

Marinelli, L. J., Piuri, M., Swigonová, Z., Balachandran, A., Oldfield, L. M., van Kessel, J. C., & Hatfull, G. F. (2008). BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS One, 3*(12), e3957, http://dx.doi.org/10.1371/journal.pone.0003957

Marston, M. F., & Martiny, J. B. H. (2016). Genomic diversification of marine cyanophages into stable ecotypes. *Environmental Microbiology*, *18*(11), 4240–4253. http://dx.doi.org/10.1111/1462-2920.13556

Marston, M. F., Taylor, S., Sme, N., Parsons, R. J., Noyes, T. J. E., & Martiny, J. B. H. (2013). Marine cyanophages exhibit local and regional biogeography. *Environmental Microbiology*, *15*(5), 1452–1463. http://dx.doi.org/10.1111/1462-2920.12062

Martinez-Hernandez, F., Fornas, O., Lluesma Gomez, M., Bolduc, B., de la Cruz Peña, M. J., Martínez, J. M., Anton, J., Gasol, J. M., Rosselli, R., Rodriguez-Valera, F., Sullivan, M. B., Acinas, S. G., & Martinez-Garcia, M. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nature Communications*, *8*, 15892. http://dx.doi.org/10.1038/ncomms15892

Martinez-Hernandez, F., Fornas, Ò., Lluesma Gomez, M., Garcia-Heredia, I., Maestre-Carballa, L., López-Pérez, M., Haro-Moreno, J. M., Rodriguez-Valera, F., & Martinez-Garcia, M. (2019). Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *The ISME Journal*, *13*(1), 232–236. http://dx.doi.org/10.1038/s41396-018-0278-7

Martínez Martínez J., Schroeder, D. C., Larsen, A., Bratbak, G., Wilson, W. H. (2007). Molecular Dynamics of Emiliania huxleyi and Cooccurring Viruses during Two Separate Mesocosm Studies. *Applied Environmental Microbiology*, *73*(2), 554-562, http://dx.doi.org/10.1128/aem.00864-06

Martínez Martínez, J., Martinez-Hernandez, F., & Martinez-Garcia, M. (2020). Single-virus genomics and beyond. *Nature Reviews. Microbiology*. https://doi.org/10.1038/s41579-020-00444-0

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

Martiny, J. B. H., Riemann, L., Marston, M. F., & Middelboe, M. (2014). Antagonistic Coevolution of Marine Planktonic Viruses and Their Hosts. *Annual Review of Marine Science*, *6*(1), 393–414. https://doi.org/10.1146/annurev-marine-010213-135108

Mavrich, T. N., & Hatfull, G. F. (2017). Bacteriophage evolution differs by host, lifestyle and genome. *Nature Microbiology*, *2*, 17112. http://dx.doi.org/10.1038/nmicrobiol.2017.112

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One*, *8*(4), e61217. http://dx.doi.org/10.1371/journal.pone.0061217

McNair, K., Zhou, C., Dinsdale, E. A., Souza, B., & Edwards, R. A. (2019). PHANOTATE: a novel approach to gene identification in phage genomes. *Bioinformatics*, *35*(22), 4537–4542. http://dx.doi.org/10.1093/bioinformatics/btz265

Meier-Kolthoff, J. P., & Göker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*, *33*(21), 3396–3404. http://dx.doi.org/10.1093/bioinformatics/btx440

Meyer, M. M., Ames, T. D., Smith, D. P., Weinberg, Z., Schwalbach, M. S., Giovannoni, S. J., & Breaker, R. R. (2009). Identification of candidate structured RNAs in the marine organism “Candidatus Pelagibacter ubique.” *BMC Genomics*, *10*, 268. http://dx.doi.org/10.1186/1471-2164-10-268

Middelboe, M., & Jorgensen, N. O. G. (2006). Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. *Journal of the Marine Biological Association of the United Kingdom. Marine Biological Association of the United Kingdom*, *86*(3), 605–612. Doi:10.1017/s0025315406013518

Middelboe, M., & Lyck, P. G. (2002). Regeneration of dissolved organic matter by viral lysis in marine microbial communities. *Aquatic Microbial Ecology: International Journal*, *27*, 187–194. http://dx.doi.org/10.3354/ame027187

Miller, E. S., Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Durkin, A. S., Ciecko, A., Feldblyum, T. V., White, O., Paulsen, I. T., Nierman, W. C., Lee, J., Szczypinski, B., & Fraser, C. M. (2003). Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *Journal of Bacteriology*, *185*(17), 5220–5233. http://dx.doi.org/10.1128/jb.185.17.5220-5233.2003

Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., & Rüger, W. (2003). Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews: MMBR*, *67*(1), 86–156, table of contents. http://dx.doi.org/10.1128/mmbr.67.1.86-156.2003

Miller, R. V., & Whyte, L. G. (Eds.). (2013). *Polar Microbiology - Life in a Deep Freeze*. ASM Press. https://doi.org/10.1128/jmbe.v14i2.666

Mizuno, C. M., Ghai, R., & Rodriguez-Valera, F. (2014). Evidence for metaviromic islands in marine phages. *Frontiers in Microbiology*, *5*, 27. http://dx.doi.org/10.3389/fmicb.2014.00027

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. http://dx.doi.org/10.1038/s41467-019-08672-6

Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., & Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genetics*, *9*(12), e1003987. http://dx.doi.org/10.1371/journal.pgen.1003987

Montecino, V., & Lange, C. B. (2009). The Humboldt Current System: Ecosystem components and processes, fisheries, and sediment studies. *Progress in Oceanography*, *83*(1), 65–79. http://dx.doi.org/10.1016/j.pocean.2009.07.041

Moon, K., Kang, I., Kim, S., Kim, S.-J., & Cho, J.-C. (2017). Genome characteristics and environmental distribution of the first phage that infects

the LD28 clade, a freshwater methylotrophic bacterial group. *Environmental Microbiology*, *19*(11), 4714–4727. http://dx.doi.org/10.1111/1462-2920.13936

Moraru, C., Varsani, A., & Kropinski, A. M. (2020). VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses*, *12*(11). https://doi.org/10.3390/v12111268

Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M., & Zinser, E. R. (2011). Dependence of the cyanobacterium Prochlorococcus on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PloS One*, *6*(2), e16805. http://dx.doi.org/10.1371/journal.pone.0016805

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio*, *3*(2), e00036–12. http://dx.doi.org/10.5061/dryad.7j8c5s5j

Morris, R. M., Longnecker, K., & Giovannoni, S. J. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology*, *8*(8), 1361–1370. http://dx.doi.org/10.1111/j.1462-2920.2006.01029.x

Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., & Rocap, G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *The ISME Journal*, *4*(5), 673–685. http://dx.doi.org/10.1038/ismej.2010.4

Morris, R. M., Rappe, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., & Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, *420*(19/26). https://doi.org/10.1038/nature01281

Motegi, C., Nagata, T., Miki, T., Weinbauer, M. G., Legendre, L., & Rassoulzadegand, F. (2009). Viral control of bacterial growth efficiency in marine pelagic environments. *Limnology and Oceanography*, *54*(6), 1901–1910. http://dx.doi.org/10.4319/lo.2009.54.6.1901

Mukherjee, S., & Sengupta, S. (2016). Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in genomic sequences. *Bioinformatics*, *32*(5), 776–778. http://dx.doi.org/10.1093/bioinformatics/btv640

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., & van Sinderen, D. (2013). Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Applied and Environmental Microbiology*, *79*(24), 7547–7555. http://dx.doi.org/10.1128/AEM.02229-13

Nagasaki, K., & Bratbak, G. (2010). Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. *Manual of Aquatic Viral Ecology. ASLO*, 92–101. http://dx.doi.org/10.4319/mave.2010.978-0-9845591-0-7.92

Nayfach, S., Camargo, A. P., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. (2020). CheckV: assessing the quality of metagenome-assembled viral genomes. In *bioRxiv* (p. 2020.05.06.081778). https://doi.org/10.1101/2020.05.06.081778

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, *39*, 578–585. http://dx.doi.org/10.1038/s41587-020-00774-7

Nechaev, S., Kamali-Moghaddam, M., André, E., Léonetti, J.-P., & Geiduschek, E. P. (2004). The bacteriophage T4 late-transcription coactivator gp33 binds the flap domain of Escherichia coli RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(50), 17365–17370. http://dx.doi.org/10.1073/pnas.0408028101

Nelson, C. E., Carlson, C. A., Ewart, C. S., & Halewood, E. R. (2014). Community differentiation and population enrichment of Sargasso Sea bacterioplankton in the euphotic zone of a mesoscale mode-water eddy. *Environmental Microbiology*, *16*(3), 871–887. http://dx.doi.org/10.1111/1462-2920.12241

Neufeld, J. D., Boden, R., Moussard, H., Schäfer, H., & Murrell, J. C. (2008). Substrate-specific clades of active marine methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Applied and Environmental Microbiology*, *74*(23), 7321–7328. http://dx.doi.org/10.1128/AEM.01266-08

Ngugi, D. K., & Stingl, U. (2012). Combined analyses of the ITS loci and the corresponding 16S rRNA genes reveal high micro- and macrodiversity of SAR11 populations in the Red Sea. *PloS One*, *7*(11), e50274. http://dx.doi.org/10.1371/journal.pone.0050274

Nilsson, E., Li, K., Hoetzinger, M., & Holmfeldt, K. (2022). Nutrient driven transcriptional changes during phage infection in an aquatic Gammaproteobacterium. *Environmental Microbiology*, http://dx.doi.org/10.1111/1462-2920.15904

Noble, R. T., & Fuhrman, J. A. (2000). Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. *Applied and Environmental Microbiology*, *66*(9), 3790–3797.

http://dx.doi.org/10.1128/AEM.66.9.3790-3797.2000

Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., Kambal, A., Monaco, C. L., Zhao, G., Fleshner, P., Stappenbeck, T. S., McGovern, D. P. B., Keshavarzian, A., Mutlu, E. A., Sauk, J., Gevers, D., Xavier, R. J., Wang, D., Parkes, M., & Virgin, H. W. (2015). Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell*, *160*(3), 447–460. http://dx.doi.org/10.1016/j.cell.2015.01.002

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–217. http://dx.doi.org/10.1006/jmbi.2000.4042

Oechslin, F. (2018). Resistance Development to Bacteriophages Occurring during Bacteriophage Therapy. *Viruses*, *10*(7). https://doi.org/10.3390/v10070351

Oh, H.-M., Kwon, K. K., Kang, I., Kang, S. G., Lee, J.-H., Kim, S.-J., & Cho, J.-C. (2010). Complete genome sequence of “Candidatus Puniceispirillum marinum” IMCC1322, a representative of the SAR116 clade in the Alphaproteobacteria. *Journal of Bacteriology*, *192*(12), 3240–3241. http://dx.doi.org/10.1128/JB.00347-10

Olsen, N. S., Hendriksen, N. B., Hansen, L. H., & Kot, W. (2020). A New High-Throughput Screening Method for Phages: Enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential. *PHAGE*, *1*(3), 137–148. https://doi.org/10.1089/phage.2020.0016

Ou, T., Gao, X.-C., Li, S.-H., & Zhang, Q.-Y. (2015). Genome analysis and gene nblA identification of Microcystis aeruginosa myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *Journal of General Virology*, *96*(12), 3681-3697, http://dx.doi.org/10.1099/jgv.0.000290

Pang, T., Savva, C. G., Fleming, K. G., Struck, D. K., & Young, R. (2009). Structure of the lethal phage pinhole. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 18966–18971. http://dx.doi.org/10.1073/pnas.0907941106

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996–1004. http://dx.doi.org/10.1038/nbt.4229

Parsons, R. J., Breitbart, M., Lomas, M. W., & Carlson, C. A. (2012). Ocean

time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *The ISME Journal*, *6*(2), 273–284. http://dx.doi.org/10.1038/ismej.2011.101

Payne, L. J., Todeschini, T. C., Wu, Y., Perry, B. J., Ronson, C. W., Fineran, P. C., Nobrega, F. L., Jackson, S. A. (2021) Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Research*, *49*(19), 10868-10878. https://doi.org/10.1093/nar/gkab883

Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J. G., Jacobs, W. R., Jr, Hendrix, R. W., & Hatfull, G. F. (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell*, *113*(2), 171–182. http://dx.doi.org/10.1016/s0092-8674(03)00233-2

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., Searson, S., Acinas, S. G., Bork, P., Boss, E., Bowler, C., De Vargas, C., Follows, M., Gorsky, G., … Tara Oceans Consortium Coordinators. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, *2*(1), 150023. https://doi.org/10.1038/sdata.2015.23

Petsong, K., Benjakul, S., & Vongkamjan, K. (2019). Evaluation of storage conditions and efficiency of a novel microencapsulated Salmonella phage cocktail for controlling S. enteritidis and S. typhimurium in-vitro and in fresh foods. *Food Microbiology*, *83*, 167–174. http://dx.doi.org/10.1016/j.fm.2019.05.008

Pingoud, A., Fuxreiter, M., Pingoud, V., Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cellular and Molecular Life Sciences*, *62*(6), 685-707. http://dx.doi.org/10.1007/s00018-004-4513-1

Plaut, R. D., Beaber, J. W., Zemansky, J., Kaur, A. P., George, M., Biswas, B., Henry, M., Bishop-Lilly, K. A., Mokashi, V., Hannah, R. M., Pope, R. K., Read, T. D., Stibitz, S., Calendar, R., & Sozhamannan, S. (2014). Genetic evidence for the involvement of the S-layer protein gene sap and the sporulation genes spo0A, spo0B, and spo0F in Phage AP50c infection of Bacillus anthracis. *Journal of Bacteriology*, *196*(6), 1143–1154. http://dx.doi.org/10.1128/JB.00739-13

Polyphasic taxonomy of Marine bacteria. (n.d.). Studylib.net. Retrieved May 11, 2020, from https://studylib.net/doc/5577926/polyphasic-taxonomy-of-marine-bacteria

Pomeroy, L. R. (1974). The Ocean's Food Web, A Changing Paradigm. *Bioscience*, *24*(9), 499–504. https://doi.org/10.2307/1296885

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Research*, *46*(W1), W200–W204. http://dx.doi.org/10.1093/nar/gky448

Pratt, D., & Waddell, G. (1959). Adaptation of Marine Bacteria to Growth in Media lacking Sodium Chloride. *Nature*, *183*, 1208–1209. DOI: 10.1038/1831208a0

Prestinaci, F., Pezzotti, P., & Pantosti, A. (2015). Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and Global Health*, *109*(7), 309–318. http://dx.doi.org/10.1179/2047773215Y.0000000030

Pride, D.T., Salzman, J., Haynes, M., Roher, F., Davis-Long, C., White, R.A., Loomer, P., Armitage, G.C., & Relman, D.A.(2012) Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *The ISME Journal*, *6*(5), 915-926, http://dx.doi.org/10.1038/ismej.2011.169

Pruesse, E., Peplies, J., & Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*(14), 1823–1829. http://dx.doi.org/10.1093/bioinformatics/bts252

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(Database issue), D61–D65. http://dx.doi.org/10.1093/nar/gkl842

Puxty, R. J., Evans, D. J., Millard, A. D., & Scanlan, D. J. (2018). Energy limitation of cyanophage development: implications for marine carbon cycling. *The ISME Journal*, *12*(5), 1273–1286. http://dx.doi.org/10.1038/s41396-017-0043-3

Puxty, R. J., Millard, A. D., Evans, D. J., & Scanlan, D. J. (2016). Viruses Inhibit CO2 Fixation in the Most Abundant Phototrophs on Earth. *Current Biology: CB*, *26*(12), 1585–1589. http://dx.doi.org/10.1016/j.cub.2016.04.036

Quan, T. M., & Repeta, D. J. (2007). Periodate oxidation of marine high molecular weight dissolved organic matter: Evidence for a major contribution from 6-deoxy- and methyl sugars. *Marine Chemistry*, *105*(3), 183–193. http://dx.doi.org/10.1016/j.marchem.2007.01.012

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies,

J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596. http://dx.doi.org/10.1093/nar/gks1219

Quistad, S.D, Lim, Y.W., Silva, G. G. Z., Nelson, C.E., Haas, A.F., Kelly, L.W., Edwards, R.A., & Rohwer, F.L. (2016). Using viromes to predict novel immune proteins in non-model organisms. *Proceedings of the Royal Society Biological Sciences*, *283*(1837), http://dx.doi.org/10.1098/rspb.2016.1200

Rabsch, W., Ma, L., Wiley, G., Najar, F. Z., Kaserer, W., Schuerch, D. W., Klebba, J. E., Roe, B. A., Laverde Gomez, J. A., Schallmey, M., Newton, S. M. C., & Klebba, P. E. (2007). FepA- and TonB-dependent bacteriophage H8: receptor binding and genomic sequence. *Journal of Bacteriology*, *189*(15), 5658–5674. http://dx.doi.org/10.1128/JB.00437-07

Rajaure, M., Berry, J., Kongari, R., Cahill, J., & Young, R. (2015). Membrane fusion during phage lysis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5497–5502. http://dx.doi.org/10.1073/pnas.1420588112

Ramachandran, A., & Walsh, D. A. (2015). Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems. *FEMS Microbiology Ecology*, *91*(10). https://doi.org/10.1093/femsec/fiv105

Rappé, M. S., Connon, S. A., Vergin, K. L., & Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature*, *418*(6898), 630–633. http://dx.doi.org/10.1038/nature00917

Rappé, M. S., Kemp, P. F., & Giovannoni, S. J. (1997). Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnology and Oceanography*, *42*(5), 811–826. http://dx.doi.org/10.4319/lo.1997.42.5.0811

Reintjes, G., Arnosti, C., Fuchs, B., & Amann, R. (2019). Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. *The ISME Journal*, *13*(5), 1119–1132. http://dx.doi.org/10.1038/s41396-018-0326-3

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. http://dx.doi.org/10.1186/s40168-017-0283-5

Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F., & Gordon, J. I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature Reviews. Microbiology*, *10*(9), 607–617. http://dx.doi.org/10.1038/nrmicro2853

Riede, I. (1986). T-even type phages can change their host range by recombination with gene (tail fibre) or gene (head). *Molecular & General Genetics: MGG*, *205*, 160–163. doi: 10.1007/BF02428046.

Rihtman, B., Bowman-Grahl, S., Millard, A., Corrigan, R. M., Clokie, M. R. J., & Scanlan, D. J. (2019). Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environmental Microbiology Reports*, *11*(3), 448–455. http://dx.doi.org/10.1111/1758-2229.12741

Rode, C. K., Melkerson-Watson, L. J., Johnson, A. T., & Bloch, C. A. (1999). Type-specific contributions to chromosome size differences in Escherichia coli. *Infection and Immunity*, *67*(1), 230–236. http://dx.doi.org/10.1128/IAI.67.1.230-236.1999

Ross, A., Ward, S., & Hyman, P. (2016). More Is Better: Selecting for Broad Host Range Bacteriophages. *Frontiers in Microbiology*, *7*, 1352. http://dx.doi.org/10.3389/fmicb.2016.01352

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., … Sullivan, M. B. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, *537*(7622), 689–693. http://dx.doi.org/10.1038/nature19366

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., & Sullivan, M. B. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*, *5*, e3817.

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. http://dx.doi.org/10.7717/peerj.3817

Roux, S., Hallam, S. J., Woyke, T., & Sullivan, M. B. (2015). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife*, *4*, e08490. https://doi.org/10.7554/eLife.08490

Roux, S., Hawley, A. K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., Woyke, T., Hallam, S. J., & Sullivan, M. B. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *eLife*, *3*, e03125.

http://dx.doi.org/10.7554/eLife.03125

Roux, S., Solonenko, N. E., Dang, V. T., Poulos, B. T., Schwenck, S. M., Goldsmith, D. B., Coleman, M. L., Breitbart, M., & Sullivan, M. B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*, *4*, e2777. http://dx.doi.org/10.7717/peerj.2777

Roux, S., Krupovic, M., Daly, R. A., Borges, A. L, Nayfach, S.; Schulz, F., Sharrar, A., Matheus Carnevali, P. B., Cheng, J.-F., Ivanova, N. N., Bondy-Denomy, J., Wrighton, K. C., Woyke, T.; Visel, A., Kyrpides, N. C., & Eloe-Fadrosh, E. A. (2019) Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nature Microbiology, 4*(11), 1895-1906, http://dx.doi.org/10.1038/s41564-019-0510-x

Russo, G., Landi, R., Pezone, A., Morano, A., Zuchegna, C., Romano, A., Muller, M. T., Gottesman, M. E.; Porcellini, A. & Avvedimento, E. V. (2016). DNA damage and Repair Modify DNA methylation and Chromatin Domain of the Targeted Locus: Mechanism of allele methylation polymorphism. *Scientific Reports*, *6*, 33222, http://dx.doi.org/10.1038/srep33222

Sabehi, G., & Lindell, D. (2012). The P-SSP7 cyanophage has a linear genome with direct terminal repeats. *PloS One*, *7*(5), e36710. http://dx.doi.org/10.1371/journal.pone.0036710

Sakowski, E. G., Arora-Williams, K., Tian, F., Zayed, A. A., Zablocki, O., Sullivan, M. B., & Preheim, S. P. (2021). Interaction dynamics and virus-host range for estuarine actinophages captured by epicPCR. *Nature Microbiology*, *6*(5), 630–642. http://dx.doi.org/10.1038/s41564-021-00873-4

Salazar, A. J., Sherekar, M., Tsai, J., & Sacchettini, J. C. (2019). R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold. *PloS One*, *14*(2), e0211432. http://dx.doi.org/10.1371/journal.pone.0211432

Salcher, M. M., Neuenschwander, S. M., Posch, T., & Pernthaler, J. (2015). The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *The ISME Journal*, *9*(11), 2442–2453. http://dx.doi.org/10.1038/ismej.2015.55

Salisbury, A., & Tsourkas, P. K. (2019). A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation. *International Journal of Molecular Sciences*, *20*(14), 3391. http://dx.doi.org/10.3390/ijms20143391

Salmond, G. P. C., & Fineran, P. C. (2015). A century of the phage: past, present and future. *Nature Reviews. Microbiology*, *13*(12), 777–786. http://dx.doi.org/10.1038/nrmicro3564

Salter, I., Galand, P. E., Fagervold, S. K., Lebaron, P., Obernosterer, I., Oliver, M. J., Suzuki, M. T., & Tricoire, C. (2015). Seasonal dynamics of active SAR11 ecotypes in the oligotrophic Northwest Mediterranean Sea. *The ISME Journal*, *9*(2), 347–360. http://dx.doi.org/10.1038/ismej.2014.129

Santos, S. B., Kropinski, A. M., Ceyssens, P.-J., Ackermann, H.-W., Villegas, A., Lavigne, R., Krylov, V. N., Carvalho, C. M., Ferreira, E. C., & Azeredo, J. (2011). Genomic and Proteomic Characterization of the Broad-Host-Range Salmonella Phage PVP-SE1: Creation of a New Phage Genus. *Journal of Virology*, *85*(21), 11265. http://dx.doi.org/10.1128/JVI.01769-10

Sargeant, S. L., Murrell, J. C., Nightingale, P. D., & Dixon, J. L. (2016). Seasonal variability in microbial methanol utilisation in coastal waters of the western English Channel. *Marine Ecology Progress Series*, *550*, 53–64. http://dx.doi.org/10.3354/meps11705

Saw, J. H. W., Nunoura, T., Hirai, M., Takaki, Y., Parsons, R., Michelsen, M., Longnecker, K., Kujawinski, E. B., Stepanauskas, R., Landry, Z., Carlson, C. A., & Giovannoni, S. J. (2020). Pangenomics Analysis Reveals Diversification of Enzyme Families and Niche Specialization in Globally Abundant SAR202 Bacteria. *mBio*, *11*(1). https://doi.org/10.1128/mBio.02975-19

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology*, *11*(8), 2078–2093. http://dx.doi.org/10.1111/j.1462-2920.2009.01929.x

Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology*, *173*(14), 4371–4378. http://dx.doi.org/10.1128/jb.173.14.4371-4378.1991

Schwalbach, M. S., Tripp, H. J., Steindler, L., Smith, D. P., & Giovannoni, S. J. (2010). The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. *Environmental Microbiology*, *12*(2), 490–500. http://dx.doi.org/10.1111/j.1462-2920.2009.02092.x

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* , *30*(14), 2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153

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. http://dx.doi.org/10.1101/gr.1239303

Sharkady, S. M., & Williams, K. P. (2004). A third lineage with two-piece tmRNA. *Nucleic Acids Research*, *32*(15), 4531–4538. http://dx.doi.org/10.1093/nar/gkh795

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PloS One*, *11*(10), e0163962. http://dx.doi.org/10.1371/journal.pone.0163962

Shi, K., Oakland, J. T., Kurniawan, F., Moeller, N. H., Banerjee, S., & Aihara, H. (2020). Structural basis of superinfection exclusion by bacteriophage T4 Spackle. *Communications Biology*, *3*(1), 691, http://dx.doi.org/10.1038/s42003-020-01412-3

Shkoporov, A. N., Khokhlova, E. V., Fitzgerald, C. B., Stockdale, S. R., Draper, L. A., Ross, R. P., & Hill, C. (2018). ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects Bacteroides intestinalis. *Nature Communications*, *9*(1), 4781. http://dx.doi.org/10.1038/s41467-018-07225-7

Signorini, S. R., & McClain, C. R. (2012). Subtropical gyre variability as seen from satellites. *Remote Sensing Letters* , *3*(6), 471–479. http://dx.doi.org/10.1080/01431161.2011.625053

Silveira, C. B., & Rohwer, F. L. (2016). Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms and Microbiomes*, *2*, 16010. doi:10.1038/npjbiofilms.2016

Simonsen, A. K. (2021). Environmental stress leads to genome streamlining in a widely distributed species of soil bacteria. *The ISME Journal*. https://doi.org/10.1038/s41396-021-01082-x

Simu, K., Holmfeldt, K., Zweifel, U. L., & Hagström, A. (2005). Culturability and coexistence of colony-forming and single-cell marine bacterioplankton. *Applied and Environmental Microbiology*, *71*(8), 4793–4800. http://dx.doi.org/10.1128/AEM.71.8.4793-4800.2005

Singh, P., Bandyopadhyay, P., Bhattacharya, S., Krishnamachari, A., & Sengupta, S. (2009). Riboswitch detection using profile hidden Markov models. *BMC Bioinformatics*, *10*, 325. http://dx.doi.org/10.1186/1471-2105-10-325

Sinha, V., Williams, J., Meyerhöfer, M., Riebesell, U., Paulino, A. I., & Larsen, A. (2007). Air-sea fluxes of methanol, acetone, acetaldehyde, isoprene and

DMS from a Norwegian fjord following a phytoplankton bloom in a mesocosm experiment. *Atmospheric Chemistry and Physics*, *7*(3), 739–755. https://doi.org/10.5194/acp-7-739-2007

Skewes-Cox, P., Sharpton, T. J., Pollard, K. S., & DeRisi, J. L. (2014). Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PloS One*, *9*(8), e105067. http://dx.doi.org/10.1371/journal.pone.0105067

Slagstad, D., Wassmann, P. F. J., & Ellingsen, I. (2015). Physical constrains and productivity in the future Arctic Ocean. *Frontiers in Marine Science*, *2*, 85. http://dx.doi.org/10.3389/fmars.2015.00085

Smeal, S. W., Schmitt, M. A., Pereira, R. R., Prasad, A., & Fisk, J. D. (2017a). Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. *Virology*, *500*, 259–274. http://dx.doi.org/10.1016/j.virol.2016.08.017

Smeal, S. W., Schmitt, M. A., Pereira, R. R., Prasad, A., & Fisk, J. D. (2017b). Simulation of the M13 life cycle II: Investigation of the control mechanisms of M13 infection and establishment of the carrier state. *Virology*, *500*, 275–284. http://dx.doi.org/10.1016/j.virol.2016.08.015

Solonenko, S. A., César Ignacio-Espinoza, J, Alberti, A., Cruaud, C., Hallam, S., Konstantinidis, K., Tyson, G., Wincker, P., & Sullivan, M. B. (2013) Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics, 14*, 320, https://doi.org/10.1186/1471-2164-14-320

Solonenko, N. (2016). Isolation of DNA from phage lysate. *Protocols.io*. https://doi.org/10.17504/protocols.io.c36yrd

Solovyev, V., & Salamov, A. (2011). AUTOMATIC ANNOTATION OF MICROBIAL GENOMES AND METAGENOMIC SEQUENCES. In R. W. Li (Ed.), *Metagenomics and its Application in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers. https://www.researchgate.net/publication/259450599

Song, J., Oh, H.-M., & Cho, J.-C. (2009). Improved culturability of SAR11 strains in dilution-to-extinction culturing from the East Sea, West Pacific Ocean. *FEMS Microbiology Letters*, *295*(2), 141–147. http://dx.doi.org/10.1111/j.1574-6968.2009.01623.x

Sosa, O. A., Gifford, S. M., Repeta, D. J., & DeLong, E. F. (2015). High molecular weight dissolved organic matter enrichment selects for methylotrophs in dilution to extinction cultures. *The ISME Journal*, *9*(12),

2725–2739. http://dx.doi.org/10.1038/ismej.2015.68

Sowell, S. M., Abraham, P. E., Shah, M., Verberkmoes, N. C., Smith, D. P., Barofsky, D. F., & Giovannoni, S. J. (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *The ISME Journal*, *5*(5), 856–865. http://dx.doi.org/10.1038/ismej.2010.168

Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., Carlson, C. A., Smith, R. D., & Giovanonni, S. J. (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *The ISME Journal*, *3*(1), 93–105. http://dx.doi.org/10.1038/ismej.2008.83

Spencer, S. J., Tamminen, M. V., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., A Weitz, D., Pitkänen, L. K., Vigneault, F., Juhani Virta, M. P., & Alm, E. J. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal*, *10*(2), 427–436. http://dx.doi.org/10.1038/ismej.2015.124

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, *39*(1), 321–346. http://dx.doi.org/10.1146/annurev.mi.39.100185.001541

Steen, A. D., Crits-Christoph, A., Carini, P., DeAngelis, K. M., Fierer, N., Lloyd, K. G., & Cameron Thrash, J. (2019). High proportions of bacteria and archaea across most biomes remain uncultured [Review of *High proportions of bacteria and archaea across most biomes remain uncultured*]. *The ISME Journal*, *13*(12), 3126–3130. http://dx.doi.org/10.1038/s41396-019-0484-y

Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F., & Giovannoni, S. J. (2011). Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PloS One*, *6*(5), e19725. http://dx.doi.org/10.1371/journal.pone.0019725

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology*, *178*(3), 591–599. http://dx.doi.org/10.1128/jb.178.3.591-599.1996

Steward, G. F., Culley, A. I., Mueller, J. A., Wood-Charlson, E. M., Belcaid, M., & Poisson, G. (2013). Are we missing half of the viruses in the ocean? *The ISME Journal, 7*(3), 672-679, http://dx.doi.org/10.1038/ismej.2012.121

Still, P. C., Johnson, T. A., Theodore, C. M., Loveridge, S. T., & Crews, P. (2014). Scrutinizing the scaffolds of marine biosynthetics from different source organisms: Gram-negative cultured bacterial products enter center stage. *Journal of Natural Products*, *77*(3), 690–702. http://dx.doi.org/10.1021/np500041x

Stingl, U., Tripp, H. J., & Giovannoni, S. J. (2007). Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *The ISME Journal*, *1*(4), 361–371. http://dx.doi.org/10.1038/ismej.2007.49

Stoddard, L. I., Martiny, J. B. H., & Marston, M. F. (2007). Selection and characterization of cyanophage resistance in marine Synechococcus strains. *Applied and Environmental Microbiology*, *73*(17), 5516–5522. http://dx.doi.org/10.1128/AEM.00356-07

Sulcius, S., Staniulis, J., & Paškauskas, R. (2011). Morphology and distribution of phage-like particles in a eutrophic boreal lagoon. *Oceanologia*, *53*(2), 587–603. http://dx.doi.org/10.5697/oc.53-2.587

Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., & Chisholm, S. W. (2005). Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *PLoS Biology*, *3*(5), e144. http://dx.doi.org/10.1371/journal.pbio.0030144

Sullivan, M. B., Krastins, B., Hughes, J. L., Kelly, L., Chase, M., Sarracino, D., & Chisholm, S. W. (2009). The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environmental Microbiology*, *11*(11), 2935-2951, http://dx.doi.org/10.1111/j.1462-2920.2009.02081.x

Sullivan, M. B., Huang, K. H., Ignacio-Espinoza, J. C., Berlin, A. M., Kelly, L., Weigele, P. R., DeFrancesco, A. S., Kern, S. E., Thompson, L. R., Young, S., Yandava, C., Fu, R., Krastins, B., Chase, M., Sarracino, D., Osburne, M. S., Henn, M. R., & Chisholm, S. W. (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology*, *12*(11), 3035–3056. http://dx.doi.org/10.1111/j.1462-2920.2010.02280.x

Sullivan, M. B., Waterbury, J. B., & Chisholm, S. W. (2003). Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature*, *424*(6952), 1047–1051. http://dx.doi.org/10.1038/nature01929

Sun, J., Steindler, L., Thrash, J. C., Halsey, K. H., Smith, D. P., Carter, A. E.,

Landry, Z. C., & Giovannoni, S. J. (2011). One carbon metabolism in SAR11 pelagic marine bacteria. *PloS One*, *6*(8), e23973. http://dx.doi.org/10.1371/journal.pone.0023973

Suttle, C. A. (1993). Enumeration and isolation of viruses. In P. F. Kemp, B. Sherr, E. Sherr, & J. J. Cole (Eds.), *Handbook of methods in aquatic microbial ecology* (pp. 121–134). Lewis Publishers. https://www.researchgate.net/publication/312982272

Suttle, C. A. (1994). The Significance of Viruses to Mortality in Aquatic Microbial Communities. *Microbial Ecology*, *28*, 237–243. https://doi.org/10.1007/BF00166813

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356–361. http://dx.doi.org/10.1038/nature04160

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews. Microbiology*, *5*(10), 801–812. http://dx.doi.org/10.1038/nrmicro1750

Suttle, C. A., Chan, A. M., & Cottrell, M. T. (1991). Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Applied and Environmental Microbiology*, *57*(3), 721–726. http://dx.doi.org/10.1128/aem.57.3.721-726.1991

Székely, A.J. & Breitbart, M. (2016). Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiology Letters, 363*(6), fnw027, http://dx.doi.org/10.1093/femsle/fnw027

Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Pulcini, C., Kahlmeter, G., Kluytmans, J., Carmeli, Y., Ouellette, M., Outterson, K., Patel, J., Cavaleri, M., Cox, E. M., Houchens, C. R., Grayson, M. L., Hansen, P., Singh, N., … WHO Pathogens Priority List Working Group. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, *18*(3), 318–327. http://dx.doi.org/10.1016/S1473-3099(17)30753-3

Tal, N., Morehouse, B. R., Millman, A.,Stokar-Avihail, A., Avraham, C., Fedorenko, T., Yirmiya, E., Herbst, E., Brandis, A., Mehlman, T., Oppenheimer-Shaanan, Y., Keszei, A. F. A., Shao, S., Amitai, G., Kranzusch, P. J. & Sorek, R. (2021) Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell*, *184*(23), 5728-5739.e16, http://dx.doi.org/10.1016/j.cell.2021.09.031

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564–577. http://dx.doi.org/10.1080/10635150701472164

Tam, W., Pell, L. G., Bona, D., Tsai, A., Dai, X. X., Edwards, A. M., Hendrix, R. W., Maxwell, K. L., & Davidson, A. R. (2013). Tail tip proteins related to bacteriophage λ gpL coordinate an iron-sulfur cluster. *Journal of Molecular Biology*, *425*(14), 2450–2462. http://dx.doi.org/10.1016/j.jmb.2013.03.032

Taubert, M., Grob, C., Howat, A. M., Burns, O. J., Dixon, J. L., Chen, Y., & Murrell, J. C. (2015). XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments. *Environmental Microbiology*, *17*(10), 3937–3948. http://dx.doi.org/10.1111/1462-2920.12896

Terhaar, J., Lauerwald, R., Regnier, P., Gruber, N., & Bopp, L. (2021). Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion. *Nature Communications*, *12*(1), 169. http://dx.doi.org/10.1038/s41467-020-20470-z

Tesson, F., Herve, A., Touchon, M., d'Humieres, C., Cury, J., Bernheim, A. (2021) Systematic and quantitative view of the antiviral arsenal of prokaryotes. *bioRxiv*, https://doi.org/10.1101/2021.09.02.458658

Tétart, F., Repoila, F., Monod, C., & Krisch, H. M. (1996). Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *Journal of Molecular Biology*, *258*(5), 726–731. http://dx.doi.org/10.1006/jmbi.1996.0281

Thompson, L. R., Zeng, Q., Kelly, L., Huang, K. H., Singer, A. U., Stubbe, J., & Chisholm, S. W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), E757–E764. http://dx.doi.org/10.1073/pnas.1102164108

Thrash, J. C., Temperton, B., Swan, B. K., Landry, Z. C., Woyke, T., DeLong, E. F., Stepanauskas, R., & Giovannoni, S. J. (2014). Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *The ISME Journal*, *8*(7), 1440–1451. http://dx.doi.org/10.1038/ismej.2013.243

Thrash, C.J., Boyd, A., Huggett, M. J., Grote, J., Carini, P., Yoder, R. J., Robbertse, B., Spatafora, J. W., Rappe, M. S., & Giovannoni, S. J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Scientific Reports*, *1*(13), https://doi.org/10.1038/srep00013

Thyrhaug, R., Larsen, A., Thingstad, T. F., & Bratbak, G. (2003). Stable coexistence in marine algal host-virus systems. *Marine Ecology Progress Series*, *254*, 27–35. http://dx.doi.org/10.3354/meps254027

Tithi, S.S., Aylward, F.O., Jensen, R.V. & Zhang, L. (2018). FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. *PeerJ*, 6, e4227 http://dx.doi.org/10.7717/peerj.4227

Torrella, F., & Morita, R. Y. (1979). Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, oregon: ecological and taxonomical implications. *Applied and Environmental Microbiology*, *37*(4), 774–778. http://dx.doi.org/10.1128/aem.37.4.774-778.1979

Tripp, H. J., Kitner, J. B., Schwalbach, M. S., Dacey, J. W. H., Wilhelm, L. J., & Giovannoni, S. J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature*, *452*(7188), 741–744. http://dx.doi.org/10.1038/nature06776

Tripp, H. J., Schwalbach, M. S., Meyer, M. M., Kitner, J. B., Breaker, R. R., & Giovannoni, S. J. (2009). Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environmental Microbiology*, *11*(1), 230–238. http://dx.doi.org/10.1111/j.1462-2920.2008.01598.x

Trubl, G., Solonenko, N., Chittick, L., Solonenko, S. A., Rich, V. I., & Sullivan, M. B. (2019). Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils. *PeerJ*, 4, e1999, http://dx.doi.org/10.7717/peerj.1999

Tsementzi, D., Wu, J., Deutsch, S., Nath, S., Rodriguez-R, L. M., Burns, A. S., Ranjan, P., Sarode, N., Malmstrom, R. R., Padilla, C. C., Stone, B. K., Bristow, L. A., Larsen, M., Glass, J. B., Thamdrup, B., Woyke, T., Konstantinidis, K. T., & Stewart, F. J. (2016). SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature*, *536*(7615), 179–183. http://dx.doi.org/10.1038/nature19068

Tucker, K. P., Parsons, R., Symonds, E. M., & Breitbart, M. (2011). Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *The ISME Journal*, *5*(5), 822-830, http://dx.doi.org/10.1038/ismej.2010.188

Turner, D., Kropinski, A. M., & Adriaenssens, E. M. (2021). A Roadmap for Genome-Based Phage Taxonomy. *Viruses*, *13*(506). https://doi.org/10.3390/v13030506

Twist, K.-A. F., Campbell, E. A., Deighan, P., Nechaev, S., Jain, V., Geiduschek, E. P., Hochschild, A., & Darst, S. A. (2011). Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the β-subunit flap domain of Escherichia coli RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 19961–19966. http://dx.doi.org/10.1073/pnas.1113328108

Underwood, G. J. C., Michel, C., Meisterhans, G., Niemi, A., Belzile, C., Witt, M., Dumbrell, A. J., & Koch, B. P. (2019). Organic matter from Arctic sea-ice loss alters bacterial community structure and function. *Nature Climate Change*, *9*(2), 170–176. http://dx.doi.org/10.1038/s41558-018-0391-7

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. http://dx.doi.org/10.1093/nar/gky1049

Våge, S., Storesund, J. E., Giske, J., & Thingstad, T. F. (2014). Optimal defense strategies in an idealized microbial food web under trade-off between competition and defense. *PloS One*, *9*(7), e101415. http://dx.doi.org/10.1371/journal.pone.0101415

Våge, S., Storesund, J. E., & Thingstad, T. F. (2013). SAR11 viruses and defensive host strains [Review of *SAR11 viruses and defensive host strains*]. *Nature*, *499*(7459), E3–E4. http://dx.doi.org/10.1038/nature12387

Van Duin, J., & Wijnands, R. (1981). The function of ribosomal protein S21 in protein synthesis. *European Journal of Biochemistry / FEBS*, *118*(3), 615–619. http://dx.doi.org/10.1111/j.1432-1033.1981.tb05563.x

Van Etten, J. L., Lane, L. C., & Dunigan, D. D. (2010). DNA viruses: the really big ones (giruses). *Annual Review of Microbiology*, *64*, 83–99. http://dx.doi.org/10.1146/annurev.micro.112408.134338

Van Gerven, N., Klein, R. D., Hultgren, S. J., & Remaut, H. (2015). Bacterial amyloid formation: structural insights into curli biogensis. *Trends in Microbiology*, *23*(11), 693–706. http://dx.doi.org/10.1016/j.tim.2015.07.010

van Tol, H. M., Amin, S. A., & Armbrust, E. V. (2017). Ubiquitous marine bacterium inhibits diatom cell division. *The ISME Journal*, *11*(1), 31–42. http://dx.doi.org/10.1038/ismej.2016.112

Vega-Thurber, R., Haynes, M., Breitbart, M., Wegley, L., & Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. *Nature Protocols*, *4*(4), 470-483, http://dx.doi.org/10.1038/nprot.2009.10

Vergin, K. L., Beszteri, B., Monier, A., Thrash, J. C., Temperton, B., Treusch, A.

H., Kilpert, F., Worden, A. Z., & Giovannoni, S. J. (2013). High-resolution SAR11 ecotype dynamics at the Bermuda Atlantic Time-series Study site by phylogenetic placement of pyrosequences. *The ISME Journal*, *7*(7), 1322–1332. http://dx.doi.org/10.1038/ismej.2013.32

Vergin, K. L., Tripp, H. J., Wilhelm, L. J., Denver, D. R., Rappé, M. S., & Giovannoni, S. J. (2007). High intraspecific recombination rate in a native population of Candidatus pelagibacter ubique (SAR11). *Environmental Microbiology*, *9*(10), 2430–2440. http://dx.doi.org/10.1111/j.1462-2920.2007.01361.x

Vidaver, A. K., Koski, R. K., & Van Etten, J. L. (1973). Bacteriophage phi6: a Lipid-Containing Virus of *Pseudomonas phaseolicola*. *Journal of Virology*, *11*(5), 799-805, http://dx.doi.org/10.1128/JVI.11.5.799-805.1973

Vieira-Silva, S., & Rocha, E. P. C. (2010). The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genetics*, *6*(1), e1000808. http://dx.doi.org/10.1371/journal.pgen.1000808

Waldbauer, J. R., Coleman, M. L., Rizzo, A. I., Campbell, K. L., Lotus, J., & Zhang, L. (2019). Nitrogen sourcing during viral infection of marine cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(31), 15590–15595. http://dx.doi.org/10.1073/pnas.1901856116

Wambaugh, M. A., Shakya, V. P. S., Lewis, A. J., Mulvey, M.A., & Brown, J. C. S. (2017). High-throughput identification and rational design of synergistic small-molecule pairs for combating and bypassing antibiotic resistance. *PLoS Biology*, *15*(6), e2001644, http://dx.doi.org/10.1371/journal.pbio.2001644

Wang, M., Gao, C., Jiang, T., You, S., Jiang, Y., Guo, C., He, H., Liu, Y., Zhang, X., Shao, H., Liu, H., Liang, Y., Wang, M., & McMinn, A. (2020). Genomic analysis of Synechococcus phage S-B43 and its adaption to the coastal environment. *Virus Research*, *289*, 198155. http://dx.doi.org/10.1016/j.virusres.2020.198155

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

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A. C., Allen, M. J., Sullivan, M. B., & Temperton, B. (2019). Long-read viral metagenomics captures abundant and microdiverse viral populations and

their niche-defining genomic islands. *PeerJ*, *7*, e6800, https://doi.org/10.7717/peerj.6800

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. http://dx.doi.org/10.1093/nar/gky427

Weigel, C., & Seitz, H. (2006). Bacteriophage replication modules. *FEMS Microbiology Reviews*, *30*(3), 321–381.DOI: 10.1111/j.1574-6976.2006.00015.x

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, *28*(2), 127–181. http://dx.doi.org/10.1016/j.femsre.2003.08.001

Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H., & Breaker, R. R. (2010). Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology*, *11*(3), R31. http://dx.doi.org/10.1186/gb-2010-11-3-r31

Weitz, J. S., Hartman, H., & Levin, S. A. (2005). Coevolutionary arms races between bacteria and bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9535–9540. http://dx.doi.org/10.1073/pnas.0504062102

Weitz, J. S., Li, G., Gulbudak, H., Cortez, M. H., & Whitaker, R. J. (2019). Viral invasion fitness across a continuum from lysis to latency. *Virus Evolution*, *5*(1), vez006. http://dx.doi.org/10.1093/ve/vez006

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. http://dx.doi.org/10.1038/ismej.2014.220

Weitz, J. S., & Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports*, *4*, 17. http://dx.doi.org/10.3410/B4-17

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

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. http://dx.doi.org/10.2307/1313569

Wilson, W. H., Carr, N. G., & Mann, N. H. (1996). The effect of phosphate status on the kinetics of Cyanophage infection in the oceanic Cyanobacterium Synechococcus Sp. Wh78031. *Journal of Phycology*, *32*(4), 506–516. http://dx.doi.org/10.1111/j.0022-3646.1996.00506.x

Wilson, W. H., Turner, S., & Mann, N. H. (1998). Population Dynamics of Phytoplankton and Viruses in a Phosphate-limited Mesocosm and their Effect on DMSP and DMS Production. Estuarine, Coastal and Shelf Science, 46, 49–59. http://dx.doi.org/10.1006/ecss.1998.0333

Withey, J. H., & Friedman, D. I. (2003). A salvage pathway for protein structures: tmRNA and trans-translation. *Annual Review of Microbiology*, *57*, 101–123. http://dx.doi.org/10.1146/annurev.micro.57.030502.090945

Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., & Koonin, E. V. (2018). Origins and Evolution of the Global RNA Virome. *mBio*, *9*(6), e02329-18, http://dx.doi.org/10.1128/mBio.02329-18

Wolf, Y. I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., Krupovic, M., Fire, A., Dolja, V. V., & Koonin, E. V. (2020). Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nature Microbiology*, *5*(10), 1262-1270, http://dx.doi.org/10.1038/s41564-020-0755-4

Wommack, E, Colwell, K., & Rita. R. (2000). Virioplankton: Viruses in Aquatic Ecosystems. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, *64*(1), 69–114. https://mmbr.asm.org/content/mmbr/64/1/69.full.pdf

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*, *347*(6223), 1257594. http://dx.doi.org/10.1126/science.1257594

Xu, J., Hendrix, R. W., & Duda, R. L. (2004). Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular Cell*, *16*(1), 11–21. http://dx.doi.org/10.1016/j.molcel.2004.09.006

Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W., & Chen, X. (2014). Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. *Bioengineered*, *5*(5), 300–304. http://dx.doi.org/10.4161/bioe.32110

Yang, S.-J., Kang, I. & Cho, J.-C. (2016). Expansion of Cultured Bacterial Diversity by Large-Scale Dilution-to-Extinction Culturing from a Single Seawater Sample. *Microbial Ecology*, *71*(1), 29-43, http://dx.doi.org/10.1007/s00248-015-0695-3

Yang, J. Y., Fang, W., Miranda-Sanchez, F., Brown, J. M., Kauffman, K. M., Acevero, C. M., Bartel, D. P., Polz, M. F., & Kelly, L. (2021). Degradation of host translational machinery drives tRNA acquisition in viruses. *Cell Systems*, *12*(8), 771-779, https://doi.org/10.1016/j.cels.2021.05.019

Yang, M., Xia, Q., Du, S., Zhang, Z., Qin, F., & Zhao, Y. (2021). Genomic Characterization and Distribution Pattern of a Novel Marine OM43 Phage. *Frontiers in Microbiology*, *12*, 657. http://dx.doi.org/10.1099/mgen.0.000596

Yan, Z., Yin, M., Chen, J., & Li, X. (2020). Assembly and substrate recognition of curli biogenesis system. *Nature Communications*, *11*(1), 241. http://dx.doi.org/10.1038/s41467-019-14145-7

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R., & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews. Microbiology*, *12*(9), 635–645. http://dx.doi.org/10.1038/nrmicro3330

Yooseph, S., Nealson, K. H., Rusch, D. B., McCrow, J. P., Dupont, C. L., Kim, M., Johnson, J., Montgomery, R., Ferriera, S., Beeson, K., Williamson, S. J., Tovchigrechko, A., Allen, A. E., Zeigler, L. A., Sutton, G., Eisenstadt, E., Rogers, Y.-H., Friedman, R., Frazier, M., & Venter, J. C. (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature*, *468*(7320), 60–66. http://dx.doi.org/10.1038/nature09530

Yuan, Y., & Gao, M. (2016). Proteomic Analysis of a Novel Bacillus Jumbo Phage Revealing Glycoside Hydrolase As Structural Component. *Frontiers in Microbiology*, *7*, 745. http://dx.doi.org/10.3389/fmicb.2016.00745

Yuan, Y., & Gao, M. (2017). Jumbo Bacteriophages: An Overview. *Frontiers in Microbiology*, *8*, 403. http://dx.doi.org/10.3389/fmicb.2017.00403

Zaragoza-Solas, A., Rodriguez-Valera, F., & López-Pérez, M. (2020). Metagenome Mining Reveals Hidden Genomic Diversity of Pelagimyophages in Aquatic Environments. *mSystems*, *5*(1), e00905–e00919. http://dx.doi.org/10.1128/mSystems.00905-19

Zayed, A. A., Wainaina, J. M., Dominguez-Huerta, G., Pelletier, E., Guo, J., Mohssen, M., Tian, F., Pratama, A. A., Bolduc, B., Zablocki, O., Cronin, D., Solden, L., Delage, E., Alberti, A., Aury, J.-M., Carradec, Q., da Silva, C.,

Labadie, K., Poulain, J., Ruscheweyh, H.-J., Salazar, G., Shatoff, E., {Tara Oceans Coordinators}, Bundschuh, R., Fredrick, K., Kubatko, L. S. Chaffron, S., Culley, A. I., Sunagawa, S., Kuhn, Jens H., Wincker, P., & Sullivan, M. B. (2022) Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science*, *376*(6589), 156-162, http://dx.doi.org/10.1126/science.abm5847

Zborowsky, S., & Lindell, D. (2019). Resistance in marine cyanobacteria differs against specialist and generalist cyanophages. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 16899–16908. http://dx.doi.org/10.1073/pnas.1906897116

Zhang, Z., Chen, F., Chu, X., Zhang, H., Luo, H., Qin, F., Zhai, Z., Yang, M., Sun, J., & Zhao, Y. (2019). Diverse, Abundant, and Novel Viruses Infecting the Marine Roseobacter RCA Lineage. *mSystems*, *4*(6). https://doi.org/10.1128/mSystems.00494-19

Zhang, Z., Qin, F., Chen, F., Chu, X., Luo, H., Zhang, R., Du, S., Tian, Z., & Zhao, Y. (2020). Culturing novel and abundant pelagiphages in the ocean. *Environmental Microbiology*.*23*(2), https://doi.org/10.1111/1462-2920.15272

Zhao, X., Chen, C., Jiang, X., Shen, W., Huang, G., Le, S., Lu, S., Zou, L., Ni, Q., Li, M., Zhao, Y., Wang, J., Rao, X., Hu, F., & Tan, Y. (2016). Transcriptomic and Metabolomic Analysis Revealed Multifaceted Effects of Phage Protein Gp70.1 on *Pseudomonas aeruginosa*. *Frontiers in Microbiology*, *7*, http://dx.doi.org/10.3389/fmicb.2016.01519

Zhao, X., Schwartz, C. L., Pierson, J., Giovannoni, S. J., McIntosh, J. R., & Nicastro, D. (2017). Three-Dimensional Structure of the Ultraoligotrophic Marine Bacterium “Candidatus Pelagibacter ubique.” *Applied and Environmental Microbiology*, *83*(3), https://doi.org/10.1128/AEM.02807-16

Zhao, Y., Qin, F., Zhang, R., Giovannoni, S. J., Zhang, Z., Sun, J., Du, S., & Rensing, C. (2018). Pelagiphages in the Podoviridae family integrate into host genomes. *Environmental Microbiology*, *21*(6), 1989–2001. http://dx.doi.org/10.1111/1462-2920.14487

Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., Ellisman, M., Deerinck, T., Sullivan, M. B., & Giovannoni, S. J. (2013). Abundant SAR11 viruses in the ocean. *Nature*, *494*(7437), 357–360. http://dx.doi.org/10.1038/nature11921

Zhu, W., Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research*, *38*(12), e132.

http://dx.doi.org/10.1093/nar/gkq275

Zimmerman, A. E., Howard-Varona, C., Needham, D. M., John, S. G., Worden, A. Z., Sullivan, M. B., Waldbauer, J. R., & Coleman, M. L. (2019). Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews. Microbiology*.*18*, 21-34 https://doi.org/10.1038/s41579-019-0270-x

Appendices

1. **Buchholz, H.H.**, Michelsen, M.L., Bolaños, L.M., Browne, E., Allen, M.J., Tempterton, B. (2021). Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria. *ISME J* **15,** 1585–1598. https://doi.org/10.1038/s41396-020-00872-z

2. **Buchholz, H.H.**, Michelsen, M.L., Parsons, R.J., Bates, N.R., Temperton, B. (2021). Draft Genome Sequences of Pelagimyophage Mosig EXVC030M and Pelagipodophage Lederberg EXVC029P, Isolated from Devil's Hole, Bermuda. *Microbiology Resource Announcements* 10(7). http://dx.doi.org/10.1128/MRA.01325-20

3. **Buchholz, H. H.**; Bolaños, L. M.; Bell, A. G.; Michelsen, M. L.; Allen, M. J; Temperton, B. (2021). Genomic evidence for inter-class host transition between abundant streamlined heterotrophs by a novel and ubiquitous marine Methylophage. *bioRxiv*, http://dx.doi.org/10.1101/2021.08.24.457595

4. Warwick-Dugdale, J., **Buchholz, H.H.**, Allen, M.J., Temperton, B. (2019) Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virol J* **16,** 15. https://doi.org/10.1186/s12985-019-1120-1

5. **Holger H. Buchholz**, Kimberly Halsey, Stephen Giovannoni (2021). Viral Impacts on volatile organic carbon and energy cycling in the ocean, case for support of thesuccessful fellowship proposal for Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology Award ID:879226

6. **Appendix 6. Supplementary Video 2.1** Time series of an infection of *Pelagibacter ubique* HTCC1062 showing how host cell density achieves a steady state in infected cultures that is lower than that of uninfected cultures. Associated cytograms are shown to demonstrate shifts in fluorescence associated with viral infection. (see submitted MP4 video).

7. **Appendix 7** List of alignments for Leu-, Arg- and Trp-tRNA genes found in OM43, SAR11 and associated viruses (see submitted Excel spreadsheet).

8. **Appendix 8** Matrix of pairwise average nucleotide identity (ANI) for viruses from (Zhao *et al.*, 2013, 2018) and this study as well as related contigs from (Mizuno *et al.*, 2013). (see submitted Excel spreadsheet)

Appendix 1. Buchholz, H.H., Michelsen, M.L., Bolaños, L.M., Browne, E., Allen, M.J., Tempterton, B. (2021). Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria. *ISME J* **15,** 1585–1598. https://doi.org/10.1038/s41396-020-00872-z

The ISME Journal (2021) 15:1585–1598
https://doi.org/10.1038/s41396-020-00872-z

ARTICLE

Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria

Holger H. Buchholz[1] · Michelle L. Michelsen[1] · Luis M. Bolaños[1] · Emily Browne[1] · Michael J. Allen[1,2] · Ben Temperton[1]

Received: 20 May 2020 / Revised: 1 December 2020 / Accepted: 7 December 2020 / Published online: 25 January 2021

Abstract

Microbes and their associated viruses are key drivers of biogeochemical processes in marine and soil biomes. While viruses of phototrophic cyanobacteria are well-represented in model systems, challenges of isolating marine microbial heterotrophs and their viruses have hampered experimental approaches to quantify the importance of viruses in nutrient recycling. A resurgence in cultivation efforts has improved the availability of fastidious bacteria for hypothesis testing, but this has not been matched by similar efforts to cultivate their associated bacteriophages. Here, we describe a high-throughput method for isolating important virus–host systems for fastidious heterotrophic bacteria that couples advances in culturing of hosts with sequential enrichment and isolation of associated phages. Applied to six monthly samples from the Western English Channel, we first isolated one new member of the globally dominant bacterial SAR11 clade and three new members of the methylotrophic bacterial clade OM43. We used these as bait to isolate 117 new phages, including the first known siphophage-infecting SAR11, and the first isolated phage for OM43. Genomic analyses of 13 novel viruses revealed representatives of three new viral genera, and infection assays showed that the viruses infecting SAR11 have ecotype-specific host ranges. Similar to the abundant human-associated phage ϕCrAss001, infection dynamics within the majority of isolates suggested either prevalent lysogeny or chronic infection, despite a lack of associated genes, or host phenotypic bistability with lysis putatively maintained within a susceptible subpopulation. Broader representation of important virus–host systems in culture collections and genomic databases will improve both our understanding of virus–host interactions, and accuracy of computational approaches to evaluate ecological patterns from metagenomic data.

Introduction

It is estimated that viral predation kills ~15% of bacterial cells in marine surface water each day [1] and is a major contributor to nutrient recycling via the viral shunt, where marine viruses make cell-bound nutrients available to the neighbouring microbial community through viral lysis of host cells [2, 3]. Viruses are key players in the modulation of carbon fluxes across the oceans (150 Gt/yr), increasing particle aggregation and sinking to depth [2, 4], and accounting for 89% of the variance in carbon export from surface water to the deep ocean [5]. Viruses alter host metabolism through auxiliary metabolic genes (AMGs), increasing and altering the cellular carbon intake of infected cells [6]. Virus–host interactions also increase co-evolutionary rates of both predator and prey via Red Queen dynamics [7, 8]. While recent metagenomic advances have provided major insight into global viral diversity and abundance [9–12], mechanistic understanding of virus–host interactions in ecologically important taxa is reliant on experimental co-culturing of model systems. In cyanobacteria, such systems have shown that viruses increase the duration of photosynthetic function [13] and can inhibit CO_2 fixation, providing direct evidence that viruses of abundant phototrophs play an important role in nutrient cycling and global carbon budgets [14]. Furthermore, isolation of new viruses provides complete or near-complete viral genomes, with

Supplementary information The online version of this article (https://doi.org/10.1038/s41396-020-00872-z) contains supplementary material, which is available to authorized users.

✉ Ben Temperton
b.temperton@exeter.ac.uk

[1] School of Biosciences, University of Exeter, Exeter, UK

[2] Plymouth Marine Laboratory, Plymouth, UK

concrete evidence of known hosts. Such systems are critical to the development, ground-truthing and application of computational methods to identify and classify viral genomes in metagenomic data (e.g. VirSorter [15], VirFinder [16] and MARVEL [17]), quantify boundaries for viral populations [12, 18] and genera (VConTACT2 [19]), understand the importance of AMGs in altering nutrient flux in natural communities [20] and to predict host ranges of uncultured viruses in metagenomic data (e.g. WIsH [21, 22]).

Viruses of primary producers, such as cyanophages, are both well-represented with model systems and well-studied in the laboratory. In contrast, virus–host model systems for similarly important and abundant marine heterotrophic bacteria are rare. Isolated viruses infecting heterotrophs are heavily biased towards those with fast-growing, copiotrophic hosts that grow readily on solid agar, enabling the use of plaque assays for viral isolation. Such systems are not representative of the vast majority of heterotrophs in nutrient-limited soil and aquatic environments, which are dominated by slow-growing, oligotrophic taxa with few regulatory mechanisms and complex auxotrophies that limit growth on solid media [23–26]. Advances in dilution-to-extinction culturing of ecologically important hosts have enabled the cultivation of many fastidious bacterial taxa that are not amenable to growth on solid media from soil [27], marine [28, 29] and freshwater environments [30]. Without plaque assays to facilitate isolation and purification of viral isolates, cultivation of viruses infecting fastidious taxa in liquid media is challenging, and further exacerbated by the slow growth rates and complex nutrient requirements of their hosts. The paucity of such model systems introduces significant bias in our understanding of viral influence on global carbon biogeochemical cycles. Therefore, it is important that the efforts to isolate heterotrophic bacterial taxa for experimentation and synthetic ecology are matched by efforts to isolate their associated viruses.

Here, we adapted recent advances in dilution-to-extinction culturing of hosts [29], and protocols to isolate viruses from liquid media [31] to improve the efficiency of cultivating novel virus–host systems for fastidious taxa: First, we selected the ecologically significant SAR11 and OM43 heterotrophic marine clades as models for viral isolation; second, we used sequential enrichment of viruses from natural communities on target hosts to improve the rates of viral isolation [32]; third, we replaced the requirement for time-intensive epifluorescent microscopy with identification of putative viral infection by comparing infected and uninfected hosts by flow cytometry, followed by confirmation using transmission electron microscopy (TEM). These clades are abundant and important to global carbon biogeochemistry [23, 33–37], but little is known about their associated viruses. In the case of viruses infecting SAR11, two challenges limit our ability to evaluate host–virus ecology in natural communities: (1) Assembly of abundant and microdiverse genomes from viral metagenomes presents a challenge to short-read assembly methods, resulting in underrepresentation in subsequent datasets [12, 18]. This was demonstrated in the successful isolation of the first known pelagiphages by culturing, including the globally dominant HTVC010P, which, prior to its isolation, was entirely missed in marine viromes [38]; (2) poor representation of viral taxa in databases limits our capacity to accurately train machine-learning approaches for in silico host prediction [12, 39, 40]. This results in either a lack of host information for abundant viral contigs, or worse, incorrect assignment of host to viral contigs, confounding ecological interpretation of data. Some important taxa, such as the OM43 clade, which plays an important role in oxidation of volatile carbon associated with phytoplankton blooms [33, 35, 41], lack any isolated viruses with experimentally confirmed hosts. In addition, both SAR11 and OM43 represent model organisms for genome streamlining as a result of nutrient-limited selection [42]. The effect of genome minimalism on viral infection dynamics is poorly understood, but critical to evaluating the impact of predator-prey dynamics on global marine carbon budgets. In this study, novel SAR11 and OM43 representatives from the Western English Channel were isolated and used as bait to isolate associated viruses. We increased the initial concentration of viruses in natural seawater samples by tangential flow filtration, followed by inoculation of cultures and one to three rounds of sequential enrichment on target hosts in 96-well plates. This yielded 117 viral isolates from 218 inoculated cultures from seven monthly water samples (September 2018–July 2019). A subsample of putative viral isolates for both clades was sequenced, providing 13 novel viral genomes, including the first known siphovirus to infect SAR11 and the first known virus–host model for OM43.

Results and discussion

Isolation of a novel SAR11 strain and three new OM43 strains from the Western English Channel to use as bait for phage isolation

Dilution-to-extinction culturing for host taxa using natural seawater-based medium was performed from a water sample collected in September 2017 from the Western English Channel and yielded the first SAR11 strain (named H2P3α) and the first three OM43 strains (named C6P1, D12P1 and H5P1) from this region. The full-length 16S rRNA gene of *Pelagibacter sp*. H2P3α was 100%

identical to that of the warm-water SAR11 ecotype *Pelagibacter bermudensis* HTCC7211 (subclade 1a.3) and was considered to be a local variant [43] (Supplementary Fig. 1A). All three novel OM43 isolates were most closely related to *Methylophilales sp*. HTCC2181, a streamlined member of the OM43 clade with a 1.3-Mbp genome, isolated from surface water of the North Eastern Pacific [44] (C6P1 96.17%, D12P1 96.62% and H5P1 97.79% nucleotide identity across the full 16S rRNA gene) (Supplementary Fig. 1B). The average nucleotide identity of the 16S rRNA gene of isolates CP61, D12P1 and H5P1 to each other was ~98.46% (Supplementary Table 1), suggesting that they are representatives of the same genus [45].

An efficient, low-cost method of isolating new viruses yielded 117 new viral isolates for SAR11 and OM43 taxa

Using the four new hosts from above and established SAR11 isolates *Pelagibacter ubique* HTCC1062 (subclade 1a.1) and HTCC7211 (subclade 1a.3), we developed an optimised viral isolation pipeline (Fig. 1) and applied it to 6 monthly water samples from the Western English Channel, taken between September 2018 and April 2019 (Supplementary Table 2). Each month, we concentrated a natural viral community with tangential flow filtration and used it to inoculate one to two 96-well Teflon plates containing host cultures at ~10^6 cells mL^{-1}. Plates were monitored by flow cytometry and growth of putatively infected cultures was compared to those of unamended controls over the course of ~2 weeks to account for the slow growth rates of SAR11 and OM43 [35, 46]. An additional sample was taken in July 2019 to attempt viral isolation on OM43 strains D12P1 and H5P1. Out of a total of 218 cultures amended with concentrated viral populations, 117 viruses were isolated, purified and still infective after at least three passages, with repeatable differences observed in cytograms of infected and control cultures (Supplementary Figs. 2–5). This represents an overall isolation efficiency of 53% and an average yield of 18 viruses per environmental sample (Fig. 2). For 90% of inoculated SAR11 cultures (94 out of 105), we observed evidence of viral infection, and the putative viral isolate could be propagated and purified, fulfilling Koch's postulates for confirming a pathogenic agent. For OM43, 23 out of 113 (20%) inoculated cultures yielded positive infections that could be similarly propagated. All viral isolations required between one and three rounds of virus enrichment (Supplementary Table 3) before changes in host growth curves between infected and uninfected cultures could be observed. This suggests that putatively rare viruses can be enriched within three rounds to a level at which infection can be observed on a flow cytometer.

New viruses represent novel viral populations and support established ANI cut-offs for ecologically discrete viral ecotypes

Due to the rate-limiting step of culturing sufficient biomass for extraction of viral DNA, we subselected 16 viral isolates based on availability in November 2018 across four different hosts (HTCC1062, HTCC7211, H2P3α and H5P1) for Illumina sequencing to >30-fold coverage (Table 1). Three out of 16 sequenced samples (~19%, two from host HTCC7211, one from OM43 host H5P1) failed to assemble into single viral contigs, in line with previously reported failure rates of 18–39% for phages of *Escherichia coli* and *Salmonella spp*. [47]. For 11 of the remaining 13 samples (12 from SAR11 hosts and one from OM43), each individual sequence assembly was identified as a complete viral genome by VirSorter [15] and 95–100% complete using CheckV [48]. All assemblies yielded a single viral contig (categories 1 or 2, >15 kbp) per sample, indicating that our purification process was effective in recovering pure viral isolates. Interestingly, all viral isolate genomes were classified by VirSorter as Category 2 ("likely virus") rather than Category 1 ("most confident"), despite being complete—indicating either a lack of viral hallmark genes or a lack of enrichment of viral genes on the contigs. This finding matches our own observations for other isolated viruses infecting SAR11 (data not shown) and suggests that VirSorter classification of pelagiphages is conservative. Viral isolates were named in accordance with current ICTV guidelines for prokaryotic virus taxonomy [49], with names selected from Norse mythology and folklore, and contemporary culture (Table 1).

Viral populations are defined as discrete ecological and evolutionary units that form non-overlapping clouds of sequence space, and previous work in cyanophage populations has shown that viral populations can be delineated into populations using an average nucleotide identity (ANI) cutoff of 95% [50]. Pairwise ANI was calculated between the thirteen successfully sequenced viral genomes from this study and 85 other known or putative pelagiphages [39, 51, 52]. Pairwise ANI ranged between 77.5 and 100%, with a discrete distribution between 96.4 and 100.0% (Supplementary Fig. 6, Supplementary Table 4). This is in agreement with previous work in cyanophages [50, 53] and supports the broad use of proposed boundary cut-offs to define viral populations within viral metagenomic assemblies [10, 11]. At the proposed ANI cut-offs of 95% over 85% length [18], our 13 new viruses clustered into six viral populations, ranging from singletons to a viral population with four members (Table 1). Phages within the same populations were all isolated from the same environmental sample and on the same host, in agreement with their classification as discrete ecological and evolutionary units. All viruses isolated in this study formed their own

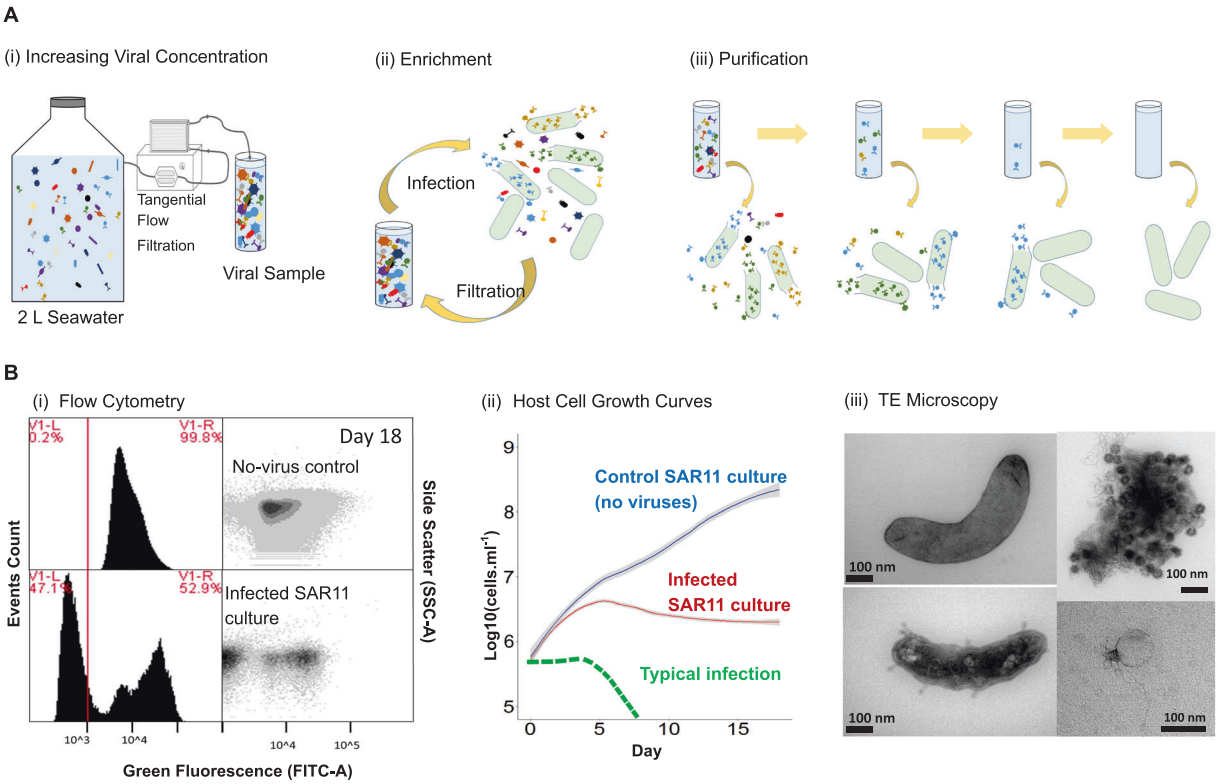


Fig. 1 Schematic of the isolation workflow. A Isolating and purifying viral cultures by: (i) Increasing concentration of viruses in water samples by tangential flow filtration (TFF); (ii) Initial infection of host cultures to enrich the sample for specific viruses; (iii) Purification of viral isolates through three rounds of dilution-to-extinction. **B** Screening cultures for viral infections using: (i). Flow cytometry, by comparing populations of no-virus controls and infected cultures; (ii) Comparing growth curves of no-virus control culture (HTCC1062) against infected SAR11 cultures; (iii) Confirming the presence of viruses in infected SAR11 cultures using transmission electron microscopy: Top left: HTCC1062 no-virus control, bottom left: infected HTCC1062, top right: aggregated cellular debris and viruses, bottom right: virus found in infected HTCC1062 culture.

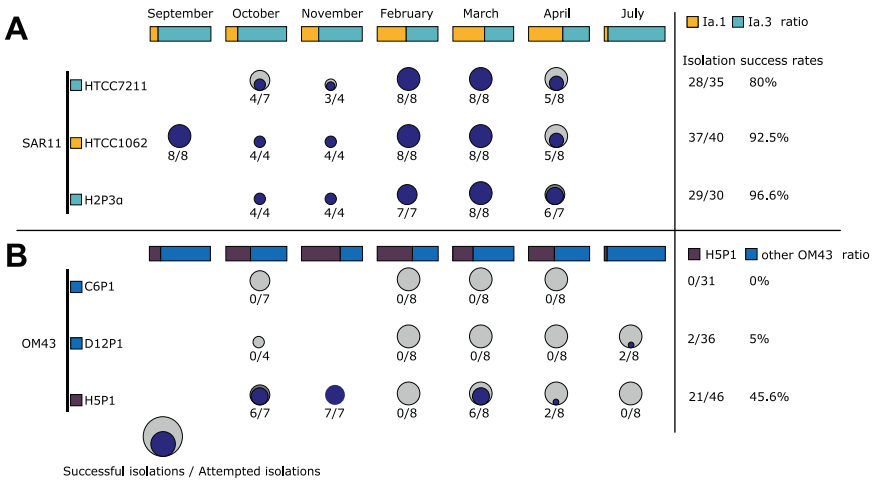


Fig. 2 Summary of success rates of viral isolations per bacterial host. A Isolation rates for phages on three SAR11 strains HTCC1062, HTCC7211 and H2P3α; and **B** Isolation rates for phages on three OM43 strains C6P1, D12P1 and H5P1. Circles and associated numbers represent the number of successful isolations (i.e. successful passage through three rounds of dilution-to-extinction), over the total number of attempts made. Bars at the top of each panel represent the relative abundance (SAR11 subclades 1a.1 and 1a.3, and OM43 strains for **A** and **B**, respectively) of strains in the microbial community sampled at the time of collection (based on 16S rRNA community profiling). Percentages on the right represent total successes for all attempts per host strain.

Table 1 Summary of phages isolated and sequenced in this study.

Viral collection number	Virus	Host	Viral population representative	Hypergeometric cluster	Genome	Length (bp)	G + C %	Taxon	Accession number	Simplified phonetic spelling	Meaning and origin of the names
EXVC025P	Eistla	HTCC1062	Eistla	Cluster A	Circular	39638	32.7	Podoviridae	MT375521	aɪs:tlaː	"Stormy one", Giantess in the poetic Edda
EXVC018P	Eyrgjafa	HTCC1062	Eyrgjafa	Cluster B	Circular	38005	32.6	Podoviridae	MT375523	eːɪrgjaːfaː	"Scar donor", Giantess in the poetic Edda
EXVC020P	Gjalp	HTCC1062	Eyrgjafa	Cluster B	Circular	37857	32.5	Podoviridae	MT375524	gjaːlp	"Roaring one", Giantess in the poetic Edda
EXVC021P	Greip	HTCC1062	Greip	Cluster D	Circular	34916	31.5	Podoviridae	MT375525	graip	"Grasp", Giantess in the poetic Edda
EXVC014P	Ran	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375530	ran	"Plundering", Goddess of the Sea
EXVC013S	Aegir	H2P3α	Kolga	Singleton	Fragment	18297	31.1	Siphoviridae	MT375519	æːgir	"Sea", God of the Sea
EXVC012P	Jörmungand	H2P3α	Ran	Cluster A	Circular	41529	34.1	Podoviridae	MT375528	jɔrmunˌgand	"Huge monster", Giant sea serpent
EXVC016S	Kólga	H2P3α	Kolga	Singleton	Circular	48659	30.5	Siphoviridae	MT375529	kɔlgaː	"Cool-wave", Daughter of Ran and Aegir
EXVC019P	Unn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375531	ʊːn	"Wave", Daughter of Ran and Aegir
EXVC015P	Hroenn	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375527	ɦrøn	"Wave", Daughter of Ran and Aegir
EXVC010P	Bylgja	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375520	bɪːlgjaː	"Billow"-wave, Daughter of Ran and Aegir
EXVC011P	Himinglæva	H2P3α	Bylgja	Cluster A	Circular	41069	33.5	Podoviridae	MT375526	hiːmɪŋglæfa	"Transparent-on-top"-wave, Daughter of Ran and Aegir
EXVC282S	Venkman	H5P1	Venkman	Singleton	Linear	38624	34.4	Siphoviridae	MT375522	vɛŋkmæn	Bill Murray's Character in the Ghostbusters movie

exclusive viral populations, with no representatives from either known isolates [38, 52] or fosmid-derived [39] genomes from other studies, indicating that a high degree of viral population diversity remains to be discovered in the Western English Channel and beyond.

Current pelagiphage isolates can be organised into five distinct phylogenetic clades

To evaluate isolate diversity at higher taxonomic organisation, we picked one representative from each of our six viral populations and compared them to previous isolates and fosmid-derived phage sequences using three approaches: first, phylogenetic analyses were performed based on conserved genes in known pelagiphage isolates and closely related taxa from viral metagenomic surveys [11, 12]; second, raw hypergeometric probability of shared-gene content was calculated (to capture broader relationships and account for genomic mosaicism) [54]; third, genomes were organised into ICTV-recognised genera using vConTACT2. vConTACT2 initially derives viral clusters using a hypergeometric approach, with subsequent refinement with Euclidean-distance-based hierarchical clustering to split mismatched, 'lumped' viral clusters [19]. All three approaches were congruent—clustering on probability of shared-gene content organised pelagiphage genomes into four main clusters and numerous singleton genomes (Fig. 3). This was broadly supported by phylogenetic (Supplementary Fig. 7) and vConTACT2 (Supplementary Fig. 8) classification approaches. Cluster A contained 23 members (nine from fosmid-derived contigs [39]; eleven previously isolated pelagiphages [52, 55]) and Pelagibacter phages *Ran* (EXVC014P), *Bylgja* (EXVC010P) and *Eistla* (EXVC025P) from this study. Cluster B contained two previously isolated pelagiphages, one fosmid-derived contig and Pelagibacter phage *Eyrgjafa* (EXVC018P) from this study. All viruses in Clusters A and B were assigned to a single viral genus by vConTACT2 that also contained 12 previously isolated pelagiphages [38, 52]. Cluster C only contained fosmid-derived contigs from the Mediterranean [39], with no isolated representatives, marking it an important target for future isolation attempts. Cluster D contained eight fosmid-derived contigs, Pelagibacter phage HTVC010P and Pelagibacter phage *Greip* (EXVC021P) from this study. Pelagibacter phage *Kolga* (EXVC016S) from this study and the only known Pelagimyophage HTVC008M from a previous study [38] fell outside the four main clusters. VConTACT2 split Cluster D, leaving Pelagibacter phages *Greip* and *Kolga* as members of two singleton clusters, suggesting that they are the first cultured representatives of novel viral genera and distinct from the globally ubiquitous Pelagibacter phage HTVC010P.

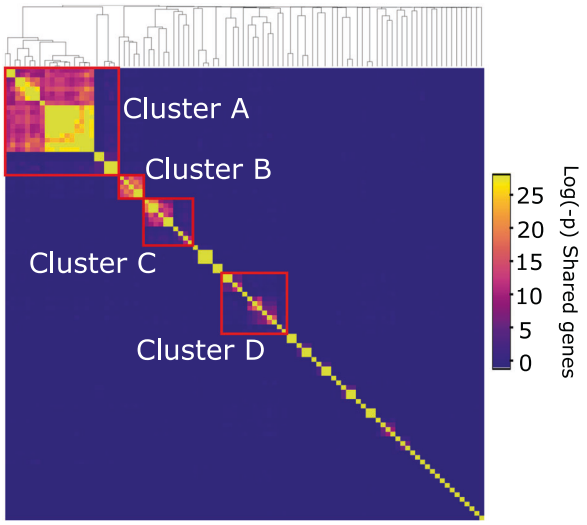


Fig. 3 Pelagibacter podophages can be grouped into four main clusters. Hypergeometric probability of shared-gene content between known pelagiphage genomes identified four main viral phylogenetic clusters (outlined with red boxes). Siphophage Kolga and myophage HTVC008M did not cluster with podophages based on shared gene content (singletons). Cluster C only contained metagenomically derived contigs. Cluster assignment for phages isolated in this study can be found in Table 1. See Supplementary Fig. 7 for a full phylogenetic tree.

Novel pelagiphages are ecotype-specific and persist in the community

Community composition analysis using 16S rRNA genes showed that during the sampling period (September 2018–July 2019), the SAR11 contribution relative to the total number of sequences ranged from a minimum of 28.0% in July 2019 to a maximum of 42.2% in September 2018 (Supplementary Fig. 9). At the clade level, SAR11 composition was relatively stable over time, except for clade II roughly doubling its relative contribution from 15.1% to about 29.6% between February and April. The SAR11 community was dominated by clade I overall throughout the sampling period. Within clade I, the ratio between SAR11 subclade Ia.1 (cold-water ecotype) and Ia.3 (warm-water ecotype) showed that the warm-water ecotype dominated from September to November as well as July (Fig. 2A). During the coldest months in the Western English Channel (February–April), the cold-water ecotype became as abundant as the warm-water ecotype with roughly a 1:1 ratio. Overall, temperature ranged from 9.3 °C to 14.1 °C from October to April, when isolations were attempted on all strains (Supplementary Table 2). HTCC1062 and HTCC7211 have specific growth rates of ~0.22 and ~0.12 divisions per day at 10 °C, respectively [46]. Our new SAR11 isolate from the

Table 2 Host infectivity of viral populations isolated and sequenced in this study.

Phage	H2P3α	HTCC7211	HTCC1062
Eistla			+
Eyrgjafa			+
Greip			+
Ran	+	+	
Kolga	+	+	
Bylgja	+	+	+

Pelagibacter sp. H2P3α and *Pelagibacter bermudensis* HTCC7211 are warm-water ecotypes of SAR11 subclade 1a; HTCC1062 is a cold-water ecotype. *Pelagibacter* phage *Bylgja* was the only virus capable of infecting both ecotypes.

Western English Channel, *Pelagibacter sp.* H2P3α, showed similar growth rates to *Pelagibacter bermudensis* HTCC7211 (Supplementary Fig. 10), with specific growth rates of ~0.10, ~0.45 and ~0.84 divisions per day at 10, 18 and 25 °C, respectively. Therefore, measured in situ temperatures during our sampling period were sufficient to support slow growth of warm-water ecotypes even during winter months, potentially providing sufficient prey to support a population of warm-water ecotype-specific phages. This result shows that slow growth of warm-water ecotypes at in situ temperatures is possible, supporting the finding of the 16S community analysis that this ecotype persists throughout the year.

An alternative explanation is that isolated viruses have a broad host range that encompasses both warm- and cold-water ecotypes of the SAR11 subclade Ia. We tested the host range (Table 2) of six pelagiphages (one of each viral population), isolated from samples in October 2018 and November 2018 (in situ water temperature of 14.8 °C and 14.2 °C, respectively) across the three SAR11 strains (Table 1). Pelagibacter phages *Eistla, Eyrgjafa* and *Greip* all infected cold-water ecotype HTCC1062 exclusively, while Pelagibacter phages *Ran* and *Kolga* only infected warm-water ecotypes HTCC7211 and H2P3α. Pelagibacter phage *Bylgia* was the only virus that could infect both warm- and cold-water ecotypes. Therefore, our new pelagiphages appear to be broadly ecotype-specific, confirming previous findings [52]. Our results suggest overall that pelagiphages persist in the water column throughout the year in sufficient densities to be isolated by our enrichment method, despite ecotypic specificity and fluctuations in warm- and cold-water ecotype community contributions of SAR11 subclade 1a. If concentration and enrichment of viruses during isolation is sufficient to successfully isolate even low-abundance phages, then a comprehensive library of representative phage isolates could be generated with relatively modest sampling effort across a few locations.

Pelagibacter phages Kolga and Aegir—the first siphovirus infecting SAR11

The 25 previously known viral isolates infecting SAR11 comprise 24 podoviruses and one myovirus [38, 52, 56]. Previous cultivation efforts for viruses of SAR11 have not isolated any siphoviruses, nor are any known from viral metagenomic studies. We isolated and sequenced the two Pelagibacter phages *Kolga* and *Aegir* using host H2P3α as bait (Fig. 4A-C). Transmission electron microscopy (TEM), which showed evidence of a long tail (Fig. 4B), suggested classification as the first reported siphovirus infecting members of the SAR11 clade.

Pelagibacter phages *Aegir* and *Kolga* were classified as members of the same population using a boundary cutoff of 95% ANI over 85% contig length; however, *Aegir* had a length of 18,297 bp compared to 48,659 bp in *Kolga*; therefore, we considered *Aegir* to be a partial genome of the same viral population. *Kolga* did not share a significant number of genes with known SAR11 podoviruses (Fig. 3), and did not cluster with other pelagiphages using hypergeometric analysis based on shared-gene content. VConTACT2 also grouped *Aegir* and *Kolga* into one cluster without any other known viruses, suggesting that it represents a novel viral genus. A number of genes found in *Kolga* were shared with other known siphoviruses (associated with different hosts) such as the Bordetella phage LK3. Screening of contigs from the Global Ocean Virome dataset [11] identified six contigs from various ecological zones, which shared a viral cluster with *Kolga*, but which belonged to different viral populations based on network analysis using vConTACT2. Phylogenetic analysis of all Pelagibacter phage TerL genes indicates that the closest known relatives to *Kolga* and *Kolga*-like contigs are members of the Myoviridae family, supporting a classification as a distinct and novel viral group (Fig. 4C, Supplementary Fig. 7). In *Kolga*, 67% of encoded genes could not be functionally annotated, and out of all hypothetical genes identified on *Kolga*, only three hypothetical genes were shared with SAR11 podoviruses. In contrast, on average, ~90% of genes without known function identified within our novel SAR11 podoviruses are shared between different SAR11 podoviruses. *Kolga* possesses a tail- tip J protein (Fig. 4A), often found in phages with long non-contractile tails such as E. coli phage ƛ, where it plays a role in DNA injection during cell entry and tail assembly [57]. *Kolga* also encodes a small S21 subunit of the 30S ribosomal gene structurally similar to the ones found in Pelagibacter phage HTVC008M, Pelagimyophage-like contigs [58] and hosts HTCC7211 and H2P3α. Encoding ribosomal genes is a feature found in numerous myoviruses and siphoviruses [59]. The S21 gene is involved in translation initiation and needed for the binding of

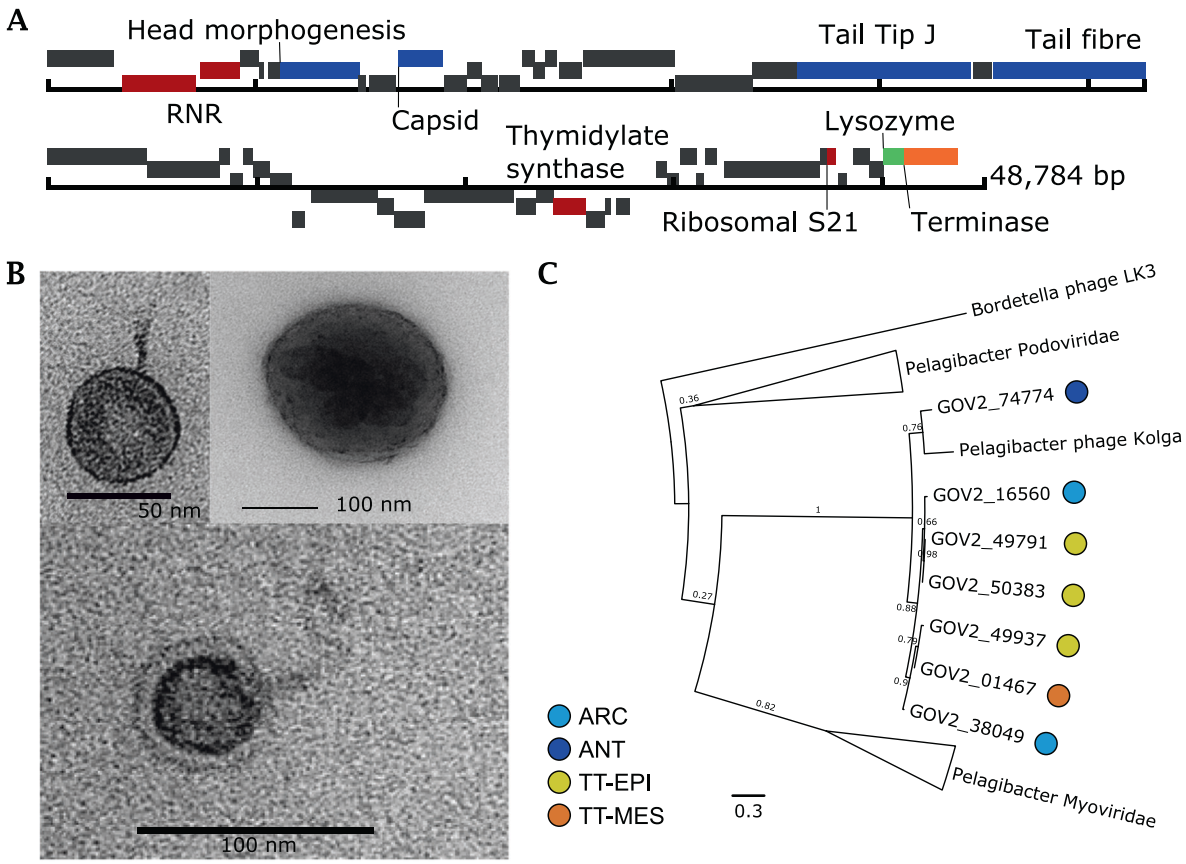


Fig. 4 The first reported siphovirus infecting SAR11 (*Pelagibacter* phage *Kolga* EXVC016S) isolated on novel host *Pelagibacter sp.* H2P3α. A Gene map of the 48,784 bp genome, which contains 80% hypothetical genes without known function. **B** Transmission electron micrographs of *Kolga* (left and bottom) and an H2P3α host cell infected with *Kolga* (top right). **C** Unrooted maximum likelihood tree of *terL* genes found in pelagiphages. Branches containing members of the Podoviridae and Myoviridae family infecting SAR11 are collapsed for clarity (full tree is available in Supplementary Fig. 7). Closely related contigs prefixed with 'GOV' represent viral contigs from the Global Ocean Virome dataset and the ecological zone from which they were assembled is marked: temperate–tropical epipelagic (TT-EPI), temperate–tropical mesopelagic (TT-MES), Antarctic (ANT), Arctic Ocean (ARC).

mRNA [60]. Virally encoded S21 genes may provide a competitive advantage for the phage as it could replace cellular S21 and assist in the translation of viral transcripts. *Kolga* may need its S21 gene for shifting the translational frame, as it has been shown that for some members of the Caudovirales, the production of tail components is dependent on programmed translational frameshifting [61]. Given the constitutive nature of gene expression in genomically streamlined bacteria [62], genes such as S21 may also provide the virus with a mechanism to manipulate host metabolism in the absence of typical promoters and repressors.

First methylophages for marine OM43 isolated

Isolation of novel phages for the OM43 clade yielded 23 positive infections, with efficiencies ranging from 0% (no viruses isolated on host C6P1) to 45% on H5P1 (Fig. 2). To the best of our knowledge, these are the first reported viruses infecting members of the OM43 clade. One explanation for the lower efficiency of isolation of phages infecting OM43 is simply one of lower host abundance concomitant with lower phage abundance in the viral community, reducing the likelihood of infective viruses coming into contact with susceptible and permissive cells. OM43 is closely associated with metabolism of extracellular substrates from phytoplankton blooms [63], but has low abundance outside of phytoplankton spring blooms [33]. Our water samples were not associated with high in situ fluorescence (used as proxy measurement for phytoplankton), and missed the April 2019 spring bloom by about two weeks (Supplementary Fig. 11). Based on 16S community analysis, the OM43 contribution to the bacterial community (0.7%) was the lowest during sampling in April 2019 and highest in October 2018 (1.9%). Relative to all other OM43, H5P1 was the most abundant OM43 in the Western English Channel, contributing more than half of all OM43 in November (Fig. 2B). This could explain the higher success rate of isolating phages on host H5P1 compared to C6P1 and D12P1.

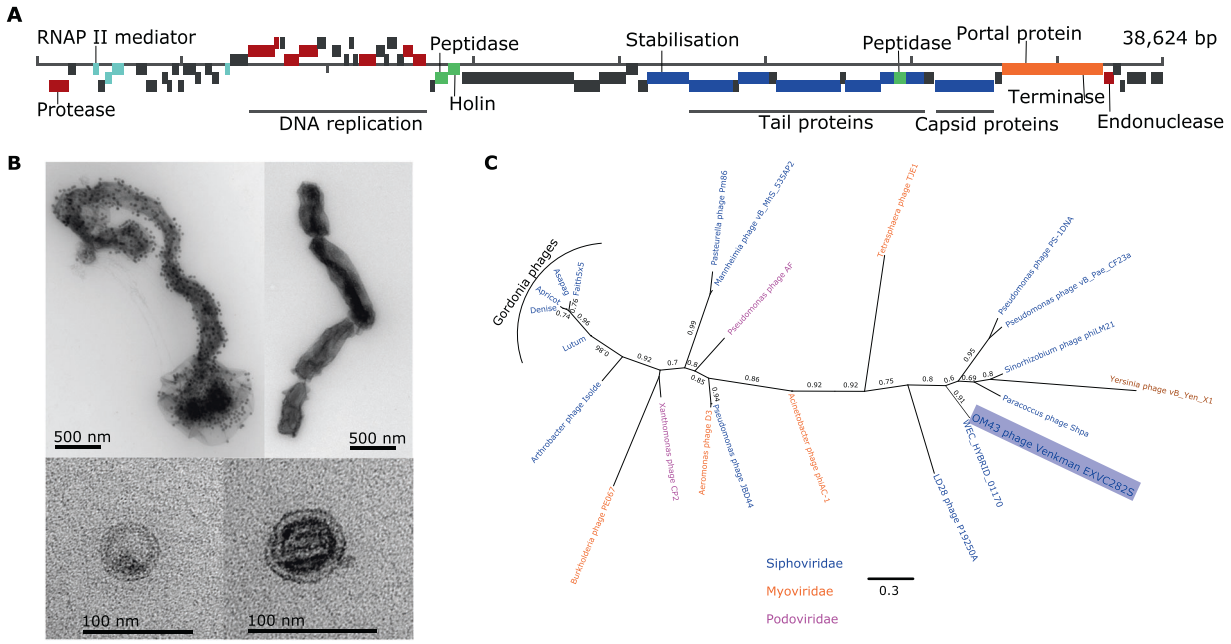


Fig. 5 The first reported cultured virus known to infect a member of the OM43 clade (Methylophilales phage Venkman EXVC282S), which was isolated on host strain H5P1 from this study. A Gene map displaying protein-coding genes. Gene function is colour-coded as follows: structural genes (blue); DNA replication (red); lysis (green); transcription (turquoise); packaging (orange); hypothetical genes (grey). **B** TEM images of infected and chaining H5P1 cells (top left), uninfected H5P1 chaining cells (top right), *Venkman* viral particles (bottom left and right). **C** Maximum likelihood tree (500 bootstraps) of concatenated viral *terL* and exonuclease genes. Host families of the phages are indicated in the figure.

Sequencing of the first OM43 phage isolate, Methylophilales phage *Venkman* (EXVC282S), returned a single genome 38,624 bp long (31.9% GC content), which was linear but complete (Supplementary Table 5). *Venkman* encodes genes (Fig. 5A) with similar synteny and function to the siphovirus P19250A (38,562 bp) that infects freshwater *Methylophilales* LD28, which is often considered a freshwater variant of OM43 [64, 65]. Unlike the siphovirus P19250A, TEM images indicated that OM43 phage *Venkman* had a short tail (Fig. 5B) similar to podoviruses, though it is possible that tail structures were lost during grid preparation. *Venkman* shared a viral population with a 23-kbp contig from the Western English Channel virome assembled from short-read data (WEC_HYBRID_01170), suggesting that this viral type does not suffer from the issues of high abundance and microdiversity that challenge assembly of some pelagiphages. Phylogenetic analysis of concatenated TerL and exonuclease genes indicated that *Venkman* is most closely related to P19250A and other siphoviruses infecting different Proteobacteria (Fig. 5C). However, branch support values were low, despite numerous attempts to refine the tree with different approaches (see Supplementary Methods). We identified a number of common phage proteins such as a capsid protein, terminase, nucleases and tail structural proteins, and the remaining 54% of genes were hypothetical. VConTACT2 assigned OM43 phage *Venkman* and LD28 phage P19250A to the same genus-level cluster; therefore, the OM43 phage *Venkman* may be a marine variant of the freshwater LD28 siphovirus P19250A.

Global abundance of novel isolates highlights niche specificity and low representation in existing datasets

To evaluate the relative global abundance of our new phage isolates, existing virome datasets from the Global Ocean Virome survey [11] and the Western English Channel [12] were randomly subsampled to 5 million reads and competitively recruited against genomes of viral population representatives from this study as well as previously isolated pelagiphages (see Supplementary methods). Overall, with the exception of Pelagibacter phage HTVC010P, pelagiphages were poorly represented in samples from temperate–tropical mesopelagic (TT-MES) and Antarctic (ANT) ecological zones, as defined in [11] (Fig. 6). Phages isolated in this study were neither globally ubiquitous, nor abundant in the single Western English Channel virome (with the exception of phage *Ran*). Phages *Bylgja, Himinglaeva, Eistla* and *Eyrgjafa* did not achieve the minimum cutoff of 40% genome coverage to be classified as present [18] in any of the viromes tested. *Ran* was the only phage isolated in this study with representation in at least two samples from temperate–tropical epipelagic (TT-EPI) zones, concordant with its host specificity of warm-water

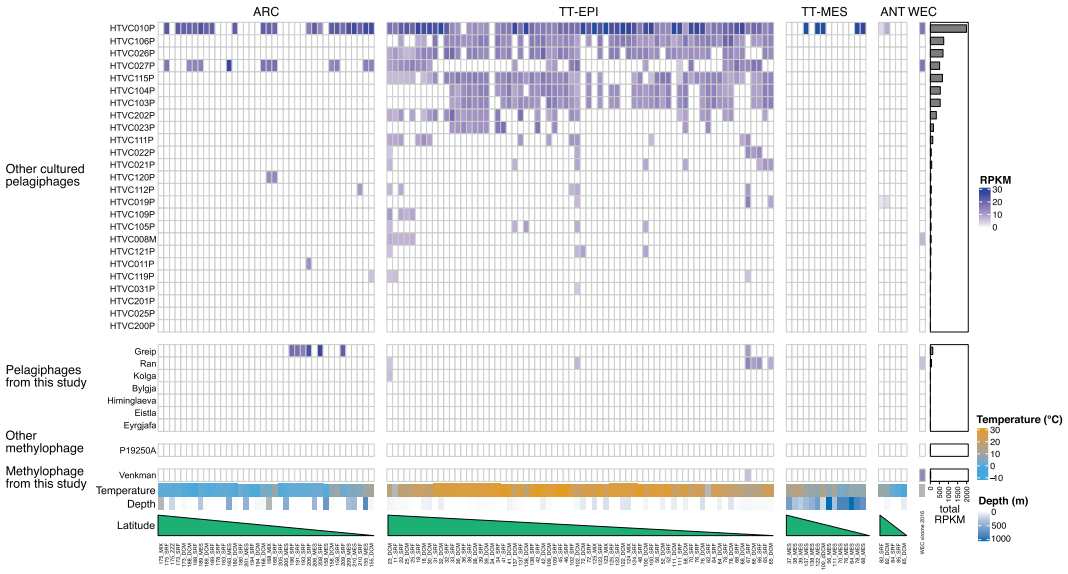


Fig. 6 Relative abundance (reads recruited per kilobase of contig per million reads (RPKM)) of known pelagiphage isolates and methylophage genomes in viromes from the Western English Channel [12] and the Global Ocean Virome [11]. Samples are organised by decreasing latitude per ecological zone: ARC Arctic, TT-EPI temperate–tropical epipelagic, TT-MES temperate–tropical mesopelagic, ANT Antarctic, WEC Western English Channel.

SAR11 ecotypes (*Pelagibacter bermudensis* HTCC7211 and H2P3α) from this study (Table 2). Similarly, the Western English Channel virome was taken in September 2016, when waters are usually highly stratified after summer heating [66] and warm-water ecotypes of SAR11 dominate the microbial community (Fig. 2A). Other viral populations from this study also isolated on warm-water SAR11 ecotypes (phages *Kolga, Bylgja* and *Himinglaeva*) were either absent or below limits of detection in this sample. This coupled with the low global abundance of these viruses suggests that pelagiphage communities comprise few highly abundant taxa and a long rare tail. Pelagibacter phage *Greip* was detected in seven samples, six of which were Arctic (ARC) samples from a discrete region of the Arctic characterised by low nutrient ratios [11]. In three of those samples (191_SRF, 193_SRF and 196_SRF), HTVC010P was not detected and in the other three (206_SRF, 208_SRF and 209_SRF), *Greip* was 1.5- to 6.6-fold more abundant than HTVC010P, identifying *Greip* as an abundant arctic pelagiphage ecotype. The fact that *Greip* was isolated on the cold-water ecotype of SAR11 (Pelagibacter ubique HTCC1062) and does not infect either warm-water ecotype HTCC7211 or H2P3 supports the hypothesis that host niche specificity shapes the phylogeography of associated viral taxa.

In the Western English Channel viral metagenomes, OM43 phage *Venkman* was the third most abundant (2,699 RPKM) after Pelagibacter phages HTVC027P (5,023 RPKM) and HTVC010P (5,026 RPKM) identifying it as an ecologically important virus in this coastal microbial community (Fig. 6). In contrast, neither *Venkman* nor LD28 phage P19250A were identified in the majority of samples from GOV2, except in one sample (67_SRF) in the South Atlantic Ocean (257 RPKM against *Venkman*). Sample 67_SRF was also classified as a coastal biome by the authors [11]. These results coincide with previous reports that OM43, and presumably their associated viruses as a consequence of host abundance, are important in some coastal regions, but largely absent in open-ocean systems [33, 41].

Unusual host–virus dynamics are prevalent in isolated phages

Pelagibacter phages *Bylgja*, *Eyrgjafa*, *Ran* and *Eistla* all encode endonucleases and exonucleases (Supplementary Fig. 12) and cluster by shared protein content with other pelagiphages (Clusters A and B), such as, e.g. HTVC011P and HTVC025P, shown previously to integrate into host genomes (Fig. 3) [52]. We therefore predict that all our phages within Clusters A and B are temperate phages. *Eyrgjafa* also encodes a tRNA-Leu that has 85% nucleotide identity over the first 34 bases of the tRNA-Leu of its host HTCC1062, suggesting a putative integration site into the host genome [67]. To date, 16 of the 29 viruses previously isolated on SAR11 strains have either been shown to be capable of lysogeny, and/or encode genes associated with a temperate infection cycle. In contrast, viruses such as the Pelagibacter phage *Greip*, and the abundant HTVC010P in Cluster D, do not possess any genes associated with lysogeny, and would therefore be classified as exclusively lytic. Viruses in Clusters A

and B (putatively temperate phage), were of much lower abundance in the environment compared with HTVC010P, which was among the most abundant pelagiphages in the Western English Channel (Fig. 6), suggesting a possible ecological difference between two groups of viruses with different infectivity strategies.

Interestingly, growth curves of hosts infected with Pelagibacter phage *Greip* and other isolated phages deviated from the expected decay in cell abundance associated with viral lysis and previously observed in isolated pelagiphages [38] (Fig. 1 Bii). The first pelagiphages HTVC010P and HTVC008M were isolated from the warm waters of the Sargasso Sea, and HTVC011P and HTVC019P were isolated from the colder waters of the Oregon Coast. All four strains were propagated on the cold-water SAR11 ecotype *P. ubique* HTCC1062. In all cases, host density was reduced from ~8×10^6 cells mL^{-1} at T_0 to $<10^6$ cells mL^{-1} over a 60–72 h period. Viruses isolated from warm waters took 17% longer than those from cold water to do so [38], suggesting that suboptimal hosts reduced the rate of infection as shown in cyanophages [68]. In contrast, infection dynamics of our isolates often resulted in host density of infected cultures growing to a steady state, but at a lower cell density than uninfected cells (Supplementary Video 1), irrespective of cluster, population assignment or evidence of genes associated with temperate lifestyles. Out of 117 viruses isolated in this study, only 16 infections reduced host abundance below their inoculum density of 10^6 cells mL^{-1}. In 53 infections, densities of infected cells increased to within an order of magnitude of uninfected cells (Supplementary Fig. 13), but demonstrated clear evidence of viral infection in cytograms, TEMs and subsequent recovery of viral genomes in selected samples.

Similar patterns of infection were recently reported in the extremely abundant bacteriophage ϕCrAss001 found in the human gut, where infection of exponentially growing cells of *B. intestinalis* 919/174 did not result in complete culture lysis, but caused a delay in stationary phase onset time and final density, despite lacking genes associated with lysogeny. As with our study, the authors observed that this only occurred in liquid culture, and isolation of the virus required numerous rounds of enrichment. They postulated that the virus may cause a successful infection in only a subset of host cells, with the remainder exhibiting alternative interactions such as pseudolysogeny or dormancy [32]. The prevalence of similar infection dynamics in the phages isolated in this study offers two intriguing possibilities: (i) many of the viruses isolated in this study are either not fully lytic, but fall somewhere on the continuum of persistence [69]. This could be controlled by genes currently lacking a known function, with lysogeny (and associated superinfection immunity) favoured at high cell density, which would support the Piggyback-the-Winner hypothesis [70, 71]; (ii) the steady state of host and virus densities observed here is an indicator of host phenotypic bistability in these streamlined heterotrophic taxa. Viral propagation occurring in only a subset of cells could explain the requirement of multiple rounds of enrichment before sufficient viral load is reached to be able to observe lytic infection on the host population. Either strategy, or a combination of both, would provide an ecological advantage of long-term stable coexistence between viruses and hosts, and offers an explanation of the paradox of stable high abundances of both predator and prey across global oceans [12, 38, 72, 73]. Infection in a subset of the population could also explain the low lytic activity observed in pelagiphages in situ, despite high host densities [74], and the small dynamic range and decoupled abundances of SAR11 and virioplankton in the Sargasso Sea [75]. Limited lysis of subpopulations of hosts such as SAR11 and OM43 that specialise in the uptake of labile carbon enriched through viral predation [76, 77] could facilitate efficient intra-population carbon recycling and explain the limited influence of SAR11 and associated viral abundances on carbon export to the deep ocean [5]. We propose the moniker the 'Soylent Green Hypothesis' for this mechanism, after the 1973 cult film in which the dead are recycled into food for the living. Further investigation leveraging our new virus–host model will provide greater insight into viral influence on ocean carbon biogeochemistry.

In conclusion, our method coupled dilution-to-extinction cultivation of hosts and associated viruses, resulting in the isolation of three new strains of OM43: a Western English Channel variant of a warm-water ecotype of SAR11, the first known methylophages for OM43 and the first siphovirus infecting SAR11, as well as eleven other viruses infecting this important marine heterotrophic clade and >100 more isolates to be sequenced and explored. The described method represents an efficient and cost-effective strategy to isolate novel virus–host systems for experimental evaluation of co-evolutionary dynamics of important fastidious taxa from marine and other biomes. Coupling these methods to existing advances in host cultivation requires minimal additional effort and will provide valuable representative genomes to improve success rates of assigning putative hosts to metagenomically derived viral contigs. Broader representation of model systems beyond cyanophages and viruses of copiotrophic, *r*-strategist hosts will reduce bias in developing methods to delineate viral population boundaries [78, 79], increasing the accuracy with which we derive ecological meaning from viral metagenomic data. We therefore hope that this method will enable viruses to be included in the current resurgence of cultivation efforts to better understand the biology and

ecology of phages, and the influence of the world's smallest predators on global biogeochemistry.

Methods summary

A complete description of the materials and methods is provided in the Supplementary Information. Four bacterial strains (*Methylophilales sp*. C6P1, D12P1 and H5P1; *Pelagibacter sp*. H2P3α) were isolated from Western English Channel station L4 seawater samples using dilution-to-extinction methods [29]. All four bacteria and two additional SAR11 strains *Pelagibacter bermudensis* HTCC7211 and *Pelagibacter ubique* HTCC1062 were used as bait to isolate phages from six monthly Western English Channel L4 seawater samples (50°15.00 N; 4° 13.00 W). Briefly, water samples were concentrated for viruses using tangential flow filtration (100 kDa Hydrosart membrane) and used as viral inoculum (10% v/v) in exponentially growing cultures of host bacteria in artificial seawater medium [80] in 96-well Teflon plates (Radleys, UK). Cells of the resulting lysate were filtered out (0.1 µm PVDF syringe filters) and the filtrate was used as viral inoculum in another round of isolation. This process was repeated until viral infection could be detected by flow cytometry—comparing cytograms and maximum density of infected cultures against uninfected cultures. Phages were purified by dilution-to-extinction methods (detailed protocol available here: 10.17504/protocols.io.c36yrd). Phage genomes were sequenced using Illumina 2 × 150 PE sequencing, assembled and manually annotated as described in [81]. Phylogenetic classification of phages was performed on concatenated shared genes using a combination of Bayesian inference trees, maximum likelihood trees and shared-gene likelihood analyses, depending on the availability of appropriate taxa. ICTV-recognised genera based on shared-gene content were assigned with VConTACT2 [19]. The relative abundance of novel phages in the Western English Channel [12] and Global Ocean viromes [11] was estimated (RPKM) by competitive read recruitment of five million randomly subsampled reads against pelagiphage and methylophage genomes from this study and others [38, 52, 64, 82].

Data availability

All reads can be found in the SRA database under BioProject number PRJNA625644 as BioSamples SAMN14604128–SAMN14604140. Annotated phage genomes are deposited as GenBank submissions under accession numbers MT375519–MT375531. Sequences used for phylogenetic analysis are deposited under https://github.com/ViralPirates/viral-dte.

Acknowledgements We would like to thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing the TE microscopy and imaging. We would also like to thank the crew of the *R/V Plymouth Quest* and our collaborators at PML for collecting water samples, and the driver Magic for delivering water samples from Plymouth to Exeter. Genome sequencing was provided by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1). This project used equipment funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). Bioinformatic analyses were conducted using the high-performance computing, ISCA, provided by the University of Exeter.

Funding The efforts of Holger Buchholz in this work were funded by the Natural Environment Research Council (NERC) GW4 + Doctoral Training program. MLM and BT were funded by NERC (NE/R010935/1) and by the Simons Foundation BIOS-SCOPE program.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Eric Wommack K, Colwell RR. Virioplankton: viruses in Aquatic Ecosystems. Microbiol Mol Biol Rev. 2000;64:69–114.
2. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, et al. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. ISME J. 2015;9:1352–64.
3. Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol Rep. 2012;4:17.
4. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. Nat Rev Microbiol. 2014;12:519–28.
5. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. Nature. 2016;532:465–70.

6. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how phages command the microbial high seas. Virol J. 2019;16:15.
7. Weitz JS, Hartman H, Levin SA. Coevolutionary arms races between bacteria and bacteriophage. Proc Natl Acad Sci USA. 2005;102:9535–40.
8. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. Long-term stability and Red Queen-like strain dynamics in marine viruses. Nat Microbiol. 2020;5:265–71.
9. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. ISME J. 2016;10:437–49.
10. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537:689–93.
11. Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Temperton B, et al. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. Cell. 2019;177:1109–23.
12. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. PeerJ. 2019;7:e6800.
13. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. Nature. 2003;424:741.
14. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Viruses Inhibit CO_2 Fixation in the Most Abundant Phototrophs on Earth. Curr Biol. 2016;26:1585–9.
15. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. 2015;3:e985.
16. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome. 2017;5:69.
17. Amgarten D, Braga LPP, da Silva AM, Setubal JC. MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. Front Genet. 2018;9:304.
18. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5:e3817.
19. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019;37:632–9.
20. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. ISME J. 2015;9:472–84.
21. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics. 2017;33:3113–4.
22. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. 2017;45:39–53.
23. Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, et al. One carbon metabolism in SAR11 pelagic marine bacteria. PLoS ONE. 2011;6:e23973.
24. Becker JW, Hogle SL, Rosendo K, Chisholm SW. Co-culture and biogeography of *Prochlorococcus* and SAR11. ISME J. 2019;13:1506–19.
25. Kraemer S, Ramachandran A, Colatriano D, Lovejoy C, Walsh DA. Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. ISME J. 2020;14:79–90.
26. Morris RM, Rappe' MS, Connon SA, Vergin KL, Siebold WA, Carlson† CA et al. SAR11 clade dominates ocean surface bacterioplankton communities. Nature. 2002;420:806–10.
27. Bartelme RP, Custer JM, Dupont CL, Espinoza JL, Torralba M, Khalili B, et al. Influence of Substrate Concentration on the Culturability of Heterotrophic Soil Microbes Isolated by High-Throughput Dilution-to-Extinction Cultivation. mSphere. 2020;5:e00024–20.
28. Connon SA, Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microbiol. 2002;68:3878–85.
29. Henson MW, Pitre DM, Weckhorst JL, Celeste Lanclos V, Webber AT, Cameron, et al. Artificial Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico. Am Soc Microbiol. 2016;1:1–9.
30. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. ISME J. 2018;12:1846–60.
31. Nagasaki K, Bratbak G. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. Manual Aqua Viral Ecol ASLO. 2010;134:92–101.
32. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. Nat Commun. 2018;9:4781.
33. Morris RM, Longnecker K, Giovannoni SJ. Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. Environ Microbiol. 2006;8:1361–70.
34. Ramachandran A, Walsh DA. Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems. FEMS Microbiol Ecol. 2015;91:fiv105.
35. Halsey KH, Carter AE, Giovannoni SJ. Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. Environ Microbiol. 2012;14:630–40.
36. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. ISME J. 2019;13:1119–32.
37. Carini P, White AE, Campbell EO, Giovannoni SJ. Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. Nat Commun. 2014;5:4346.
38. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. Nature. 2013;494:357–60.
39. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. PLoS Genet. 2013;9:e1003987.
40. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat Commun. 2017;8:15892.
41. Taubert M, Grob C, Howat AM, Burns OJ, Dixon JL, Chen Y, et al. XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments. Environ Microbiol. 2015;17:3937–48.
42. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. ISME J. 2014;8:1553–65.
43. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ. The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. Environ Microbiol. 2010;12:490–500.
44. Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, et al. The small genome of an abundant coastal ocean methylotroph. Environ Microbiol. 2008;10:1771–82.
45. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol. 2014;12:635–45.

46. Polyphasic taxonomy of Marine bacteria. *studylib.net*. https://studylib.net/doc/5577926/polyphasic-taxonomy-of-marine-bacteria. Accessed 11 May 2020.
47. Olsen NS, Hendriksen NB, Hansen LH, Kot W. A New High-Throughput Screening Method for Phages: enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential. PHAGE. 2020;1:137–48.
48. Nayfach S, Camargo AP, Eloe-Fadrosh E, Roux S, Kyrpides N. CheckV: assessing the quality of metagenome-assembled viral genomes. bioRxiv. 2020., 2020.05.06.081778
49. Adriaenssens EM, Sullivan MB, Knezevic P, van Zyl LJ, Sarkar BL, Dutilh BE, et al. Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Arch Virol. 2020;165:1253–60.
50. Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, et al. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. Nature. 2014;513:242–5.
51. Våge S, Storesund JE, Thingstad TF. SAR11 viruses and defensive host strains. Nature. 2013;499:E3–4.
52. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family integrate into host genomes. Environ Microbiol. 2018;21:1989–2001.
53. Marston MF, Martiny JBH. Genomic diversification of marine cyanophages into stable ecotypes. Environ Microbiol. 2016;18:4240–53.
54. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol. 2008;25:762–77.
55. Giovannoni S, Temperton B, Zhao YGiovannoni, et al. reply. Nature 2013;499:E4–5.
56. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Novel pelagiphages prevail in the ocean. bioRxiv. 2019; 2019.12.14.876532
57. Tam W, Pell LG, Bona D, Tsai A, Dai XX, Edwards AM, et al. Tail tip proteins related to bacteriophage λ gpL coordinate an iron-sulfur cluster. J Mol Biol. 2013;425:2450–62.
58. Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M. Metagenome Mining Reveals Hidden Genomic Diversity of Pelagimyophages in Aquatic Environments. mSystems. 2020;5:e00905–19.
59. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous cultivated and uncultivated viruses encode ribosomal proteins. Nat Commun. 2019;10:752.
60. Van Duin J, Wijnands R. The function of ribosomal protein S21 in protein synthesis. Eur J Biochem. 1981;118:615–9.
61. Xu J, Hendrix RW, Duda RL. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell. 2004;16:11–21.
62. Giovannoni SJ, Thrash JC, Temperton B. Implications of streamlining theory for microbial ecology. ISME J. 2014;8:1553–65.
63. Neufeld JD, Boden R, Moussard H, Schäfer H, Murrell JC. Substrate-specific clades of active marine methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. Appl Environ Microbiol. 2008;74:7321–8.
64. Moon K, Kang I, Kim S, Kim S-J, Cho J-C. Genome characteristics and environmental distribution of the first phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. Environ Microbiol. 2017;19:4714–27.
65. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. ISME J. 2015;9:2442–53.
66. Sargeant SL, Murrell JC, Nightingale PD, Dixon JL. Seasonal variability in microbial methanol utilisation in coastal waters of the western English Channel. Mar Ecol Prog Ser. 2016;550:53–64.
67. Bailly-Bechet M, Vergassola M, Rocha E. Causes for the intriguing presence of tRNAs in phages. Genome Res. 2007;17:1486–95.
68. Enav H, Kirzner S, Lindell D, Mandel-Gutfreund Y, Béjà O. Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. Nat Commun. 2018;9:4698.
69. Weitz JS, Li G, Gulbudak H, Cortez MH, Whitaker RJ. Viral invasion fitness across a continuum from lysis to latency. Virus Evol. 2019;5:vez006.
70. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, et al. Lytic to temperate switching of viral communities. Nature. 2016;531:466–70.
71. Silveira CB, Rohwer FL. Piggyback-the-Winner in host-associated microbial communities. NPJ Biofilms Microbiomes. 2016;2:16010.
72. Kang I, Oh HM, Kang D, Cho JC. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. Proc Natl Acad Sci USA. 2013;110:12343–8.
73. Martinez-Hernandez F, Fornas Ò, Lluesma Gomez M, Garcia-Heredia I, Maestre-Carballa L, López-Pérez M, et al. Single-cell genomics uncover *Pelagibacter* as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. ISME J. 2018;13:232–6.
74. Alonso-Sáez L, Morán XAG, Clokie MR. Low activity of lytic pelagiphages in coastal marine waters. ISME J. 2018;12:2100–2.
75. Parsons RJ, Breitbart M, Lomas MW, Carlson CA. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. ISME J. 2012;6:273–84.
76. Middelboe M, Lyck PG. Regeneration of dissolved organic matter by viral lysis in marine microbial communities. Aquat Micro Ecol. 2002;27:187–94.
77. Middelboe M, Jorgensen NOG. Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. J Mar Biol Assoc U K. 2006;86:605–12.
78. Brum JR, Sullivan MB. Rising to the challenge: accelerated pace of discovery transforms marine virology. Nat Rev Microbiol. 2015;13:147–59.
79. Gregory AC, Solonenko SA, Ignacio-Espinoza JC, LaButti K, Copeland A, Sudek S, et al. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. BMC Genom. 2016;17:930.
80. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme oligotroph '*Candidatus* Pelagibacter ubique' HTCC1062 on a defined medium. ISME J. 2013;7:592–602.
81. Salisbury A, Tsourkas PK. A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation. Int J Mol Sci. 2019;20:3391.
82. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Culturing novel and abundant pelagiphages in the ocean. Environ Microbiol. 2020. https://doi.org/10.1111/1462-2920.15272.

Appendix 2. **Buchholz, H.H.**, Michelsen, M.L., Parsons, R.J., Bates, N.R., Temperton, B. (2021). Draft Genome Sequences of Pelagimyophage Mosig EXVC030M and Pelagipodophage Lederberg EXVC029P, Isolated from Devil's Hole, Bermuda. *Microbiology Resource Announcements* 10(7). http://dx.doi.org/10.1128/MRA.01325-20

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of Pelagimyophage Mosig EXVC030M and Pelagipodophage Lederberg EXVC029P, Isolated from Devil's Hole, Bermuda

Holger H. Buchholz,[a] Michelle Michelsen,[a] Rachel J. Parsons,[b] Nicholas R. Bates,[b,c] Ben Temperton[a]

[a]School of Biosciences, University of Exeter, Exeter, United Kingdom
[b]Bermuda Institute of Ocean Sciences, St. George's, Bermuda
[c]School of Ocean and Earth Sciences, Waterfront Campus, University of Southampton, Southampton, United Kingdom

ABSTRACT We present the genomes of two isolated bacteriophages infecting *Pelagibacter ubique* HTCC1062. *Pelagibacter* phage Mosig EXVC030M (*Myoviridae*) and *Pelagibacter* phage Lederberg EXVC029P (*Podoviridae*) were isolated by dilution-to-extinction culturing from the oxygen minimum zone at Devil's Hole (Harrington Sound, Bermuda).

Viruses infecting the heterotrophic bacterial clade of *Pelagibacterales* are an important component of marine microbial communities throughout global oceans (1). Since the discovery and first isolation of four pelagiphages in 2013 (2), 38 more have been isolated and sequenced (3–5). Out of the 38 isolated pelagiphages, 36 belong to the *Podoviridae* family, with only one species each of *Myoviridae* and *Siphoviridae*. Here, we report the draft genome sequences of a novel pelagimyophage and a novel pelagipodophage, both isolated on *Pelagibacter ubique* HTCC1062.

A 2-liter water sample was taken (12 July 2019) using a hand-held Niskin bottle, fired at a 20-m depth at Devil's Hole, Bermuda, a seasonal oxygen minimum zone in Bermuda (6) (latitude 32.32421, longitude −64.71849). The water sample was taken to the Bermuda Institute of Ocean Sciences for processing, where planktonic cells were removed with 0.1-μm polyethersulfone filters. Viruses were concentrated by tangential flow filtration (50R VivaFlow 100-kDa Hydrosart filter; Sartorius Lab Instruments, Göttingen, Germany). We used previously described dilution-to-extinction-based methods (4) with HTCC1062 as a bait host (grown in artificial seawater medium ASM1 [7]) in 96-well Teflon plates, which does not rely on plaque formation, because the host does not grow on solid medium. The purification process was repeated five times; nonetheless, final sequence data contained two genomes, suggesting an impure culture.

For DNA isolation, a 50-ml HTCC1062 culture (10^6 cells/ml), amended with 5 ml of 0.1-μm-filtered lysate, was grown in ASM1 (18°C) until cell death (detected via flow cytometry). Debris was removed using 0.1-μm-pore polyvinylidene difluoride (PVDF) filters, and lysate was subjected to PEG8000/NaCl DNA isolation (modified from https://doi.org/10.17504/protocols.io.c36yrd, as described previously (4).

DNA libraries (Nextera XT) were prepared and sequenced by the Exeter Sequencing Service (Illumina paired end [2 × 250 bp], NovaSeq S Prime [SP], targeting 30-fold coverage). Raw reads (13.18 million) were trimmed, quality controlled, and error corrected using tadpole (default settings [8] within BBMap v38.22 [https://sourceforge.net/projects/bbmap/]) and assembled with SPAdes v3.13 (7). Viral contigs were confirmed and gene called with VirSorter v1.05 (9) and imported into DNA Master v5.23.3 (10) for manual curation with additional gene calls using GenMark v2.5 (11), GeneMarkS v4.28 (12), GeneMarkS-2 v1.14 (13), GeneMark.hmm v3.25 (14), Glimmer v3.02 (15), and Prodigal v2.6.3 (16). Open reading frames were annotated with NCBI's nonredundant

Citation Buchholz HH, Michelsen M, Parsons RJ, Bates NR, Temperton B. 2021. Draft genome sequences of pelagimyophage Mosig EXVC030M and pelagipodophage Lederberg EXVC029P, isolated from Devil's Hole, Bermuda. Microbiol Resour Announc 10:e01325-20. https://doi.org/10.1128/MRA.01325-20.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Ben Temperton, b.temperton@exeter.ac.uk.

Received 24 November 2020
Accepted 2 February 2021
Published 18 February 2021

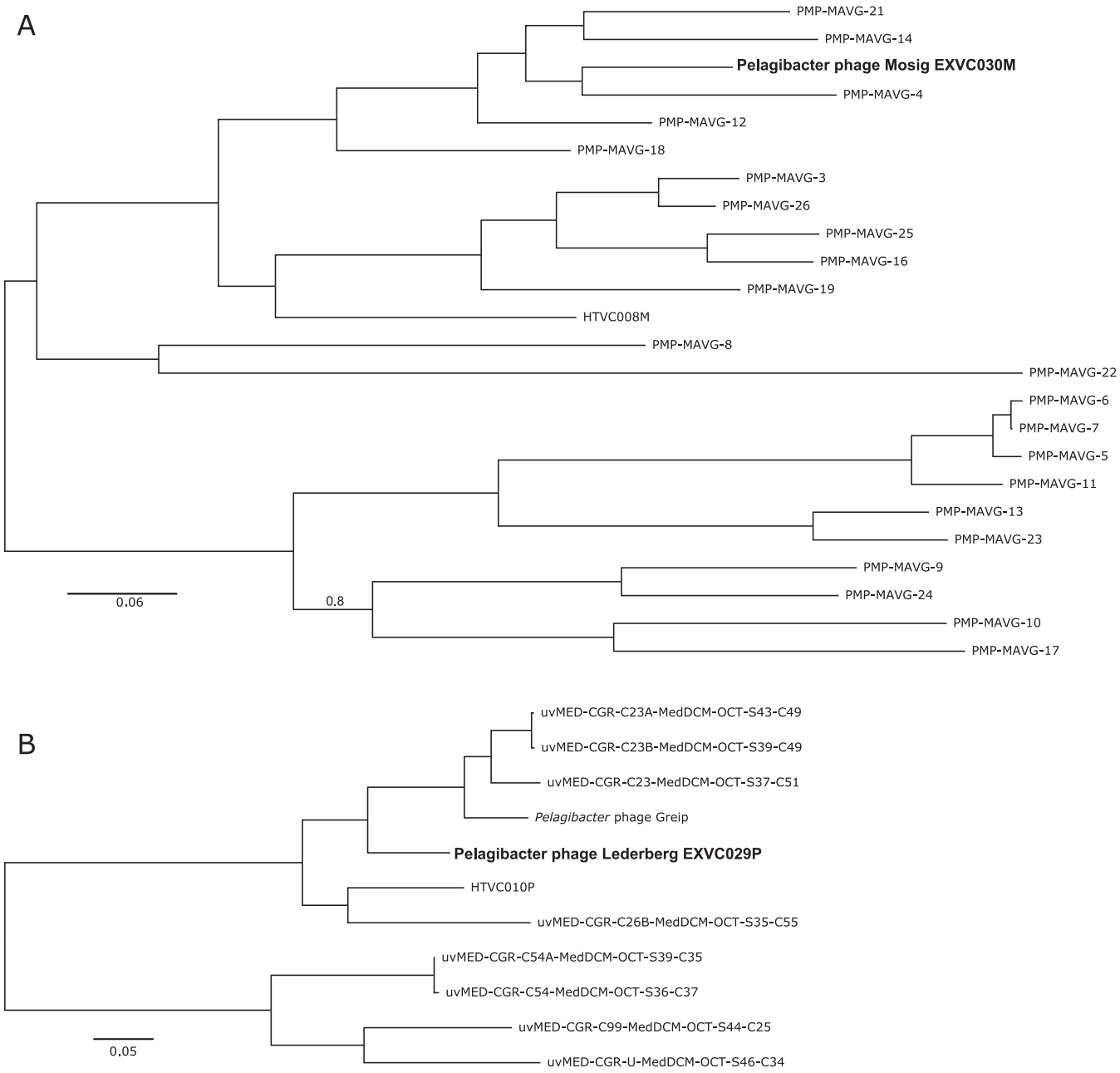


FIG 1 Bayesian inference tree from conserved genes found in pelagiphages (2, 3, 5), contigs from Mediterranean metagenomes (uvMed) (24), and putative pelagimyophages from genome-resolved metagenomics (PMP-MAVG) (25). (A) Terminase large subunit, tail sheath protein, and tail tube protein; (B) head-tail connector protein, capsid assembly protein, major capsid protein, tail tubular protein A, and putative acetyltransferase. Branch support values of 1 were omitted for clarity. The scale bar represents the estimated substitution per site.

protein database (17) and phmmer v2.41.1 (18) against the UniProtKB, uniprotrefprot (19), SWISS PROT (20), and Pfam (21) databases (accessed May 2020) and were evaluated using a previously described scoring system (10). Genome completion was verified with CheckV (22). Sequences similar to our isolates were identified with ClusterGenomes v5.1 (https://github.com/simroux/ClusterGenomes) and vConTACT 2 v0.9.19 (23) using previously isolated Pelagiphages (2, 3, 5), fosmid-derived contigs from Mediterranean metagenomes (uvMed) (24), and putative pelagimyophages from genome-resolved metagenomics (PMP-MAVG) (25). Conserved genes were identified (GET_HOMOLOGUES v09072020 [22]), aligned (MUSCLE v3.8.1551 [26]), curated (Gblocks v0.91b [27]), and concatenated manually (all with default settings). Bayesian inference trees were generated via Phylogeny.fr

(28) using MRBAYES v3.2.7 (29) (100,000 generations, sampled every 10 generations, 5,000 tree burn-in) (Fig. 1).

Pelagimyophage Mosig (named after microbiologist Gisela Mosig in recognition of her work on *Escherichia coli* phage T4) was 141,462 bp long (348× coverage; GC content, 30.01%), linear, and 75.73% complete (CheckV [22]). Out of 208 genes, 98 were putative, 3 were tRNAs, 30 were structural, and 77 were associated with DNA replication.

Pelagibacter phage Lederberg (named after microbiologist Esther Lederberg in recognition of her work on the *E. coli* phage λ) was 33,623 bp long (5,849× coverage; GC content, 33.13%) and predicted as circularly permuted/complete. Lederberg had a total of 71 genes, out of which 9 were structural, 8 were associated with DNA replication, and 54 were without known function.

Data availability. The complete genome sequences were deposited under GenBank accession numbers MT647605 (Lederberg) and MT647606 (Mosig). The corresponding read data were deposited in the Sequence Read Archive (SRA) under BioProject number PRJNA625644 and SRA accession number SRR12024324.

ACKNOWLEDGMENTS

The efforts of H.H.B. in this work were funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training Program. R.J.P., N.R.B., and B.T. were funded by the Simons Foundation BIOS-SCOPE program. M.M. and B.T. were part-funded by NERC (NE/R010935/1). This project used equipment funded by the Wellcome Trust Institutional Strategic Support Fund (WT097835MF), Wellcome Trust Multi-User Equipment Award (WT101650MA), and BBSRC LOLA award (BB/K003240/1).

We acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work.

REFERENCES

1. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, Follows MJ, Fuhrman JA, Jover LF, Lennon JT, Middelboe M, Sonderegger DL, Suttle CA, Taylor BP, Frede Thingstad T, Wilson WH, Eric Wommack K. 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. ISME J 9:1352–1364. https://doi.org/10.1038/ismej.2014.220.
2. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB, Giovannoni SJ. 2013. Abundant SAR11 viruses in the ocean. Nature 494:357–360. https://doi.org/10.1038/nature11921.
3. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, Du S, Rensing C. 2019. Pelagiphages in the Podoviridae family integrate into host genomes. Environ Microbiol 21:1989–2001. https://doi.org/10.1111/1462-2920.14487.
4. Buchholz HH, Michelsen ML, Bolaños LM, Brown E, Allen MJ, Temperton B. 2021. Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria. ISME J https://doi.org/10.1038/s41396-020-00872-z.
5. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, Du S, Tian Z, Zhao Y. 2020. Culturing novel and abundant pelagiphages in the ocean. Environ Microbiol https://doi.org/10.1111/1462-2920.15272.
6. Parsons RJ, Nelson CE, Carlson CA, Denman CC, Andersson AJ, Kledzik AL, Vergin KL, McNally SP, Treusch AH, Giovannoni SJ. 2015. Marine bacterioplankton community turnover within seasonally hypoxic waters of a subtropical sound: Devil's Hole, Bermuda. Environ Microbiol 17:3481–3499. https://doi.org/10.1111/1462-2920.12445.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
8. Bushnell B, Rood J, Singer E. 2017. BBMerge: accurate paired shotgun read merging via overlap. PLoS One 12:e0185056. https://doi.org/10.1371/journal.pone.0185056.
9. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985. https://doi.org/10.7717/peerj.985.
10. Salisbury A, Tsourkas PK. 2019. A method for improving the accuracy and efficiency of bacteriophage genome annotation. Int J Mol Sci 20:3391. https://doi.org/10.3390/ijms20143391.
11. Borodovsky M, McIninch J. 1993. GenMark: parallel gene recognition for both DNA strands. Comput Chem 17:123–133. https://doi.org/10.1016/0097-8485(93)85004-V.
12. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607.
13. Lomsadze A, Gemayel K, Tang S, Borodovsky M. 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. Genome Res 28:1079–1089. https://doi.org/10.1101/gr.230615.117.
14. Zhu W, Lomsadze A, Borodovsky M. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res 38:e132. https://doi.org/10.1093/nar/gkq275.
15. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. https://doi.org/10.1093/bioinformatics/btm009.
16. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
17. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–D65. https://doi.org/10.1093/nar/gkl842.
18. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER Web server: 2018 update. Nucleic Acids Res 46:W200–W204. https://doi.org/10.1093/nar/gky448.
19. UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049.
20. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48. https://doi.org/10.1093/nar/28.1.45.
21. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger

A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.
22. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696–7701. https://doi.org/10.1128/AEM.02411-13.
23. Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ 5:e3243. https://doi.org/10.7717/peerj.3243.
24. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. 2013. Expanding the marine virosphere using metagenomics. PLoS Genet 9:e1003987. https://doi.org/10.1371/journal.pgen.1003987.
25. Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M. 2020. Metagenome mining reveals hidden genomic diversity of pelagimyophages in aquatic environments. mSystems 5:e00905-19. https://doi.org/10.1128/mSystems.00905-19.
26. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
27. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564–577. https://doi.org/10.1080/10635150701472164.
28. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469. https://doi.org/10.1093/nar/gkn180.
29. Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755. https://doi.org/10.1093/bioinformatics/17.8.754.

Appendix 3. Buchholz, H. H.; Bolaños, L. M.; Bell, A. G.; Michelsen, M. L.; Allen, M. J; Temperton, B. (2021). Genomic evidence for inter-class host transition between abundant streamlined heterotrophs by a novel and ubiquitous marine Methylophage. *bioRxiv*, http://dx.doi.org/10.1101/2021.08.24.457595

Genomic evidence for inter-class host transition between abundant streamlined heterotrophs by a novel and ubiquitous marine Methylophage.

Holger H. Buchholz[1], Luis M. Bolaños[1], Ashley G. Bell[1], Michelle L. Michelsen[1], Michael J. Allen[1,2], Ben Temperton[1]*

[1]University of Exeter, School of Biosciences, Exeter, UK;
[2]Plymouth Marine Laboratory, Plymouth, UK

*Corresponding Author

b.temperton@exeter.ac.uk

ABSTRACT

The methylotrophic OM43 clade are Gammaproteobacteria that comprise some of the smallest free-living cells known and have highly streamlined genomes. OM43 represents an important microbial link 0between marine primary production and remineralisation of carbon back to the atmosphere. Bacteriophages shape microbial communities and are major drivers of microbial mortality and global marine biogeochemistry. Recent cultivation efforts have brought the first viruses infecting members of the OM43 clade into culture. Here we characterize a novel myophage infecting OM43 called Melnitz. Melnitz was isolated independently on three separate occasions (with isolates sharing >99.95% average nucleotide identity) from water samples from a subtropical ocean gyre (Sargasso Sea) and temperate coastal (Western English Channel) systems. Metagenomic recruitment from global ocean viromes confirmed that Melnitz is globally ubiquitous, congruent with patterns of host abundance. Bacteria with streamlined genomes such as OM43 and the globally dominant SAR11 clade use riboswitches as an efficient method to regulate metabolism. Melnitz encodes a two-piece tmRNA (*ssrA*), controlled by a glutamine riboswitch, providing evidence that riboswitch use also occurs for regulation during phage infection of streamlined heterotrophs. Virally encoded tRNAs and *ssrA* found in Melnitz were phylogenetically more closely related to those found within the alphaproteobacterial SAR11 clade and their associated myophages than those within their gammaproteobacterial hosts. This suggests the possibility of an ancestral inter-class host transition event between SAR11 and OM43. Melnitz and a related myophage that infects SAR11 were unable to infect hosts of the SAR11 and OM43, respectively, suggesting host transition rather than a broadening of host range.

IMPORTANCE

Isolation and cultivation of viruses is the foundation on which the mechanistic understanding of virus-host interactions and ground-truthing is based. This study isolated and characterised the first myophage known to infect the OM43 clade, expanding our knowledge of this understudied group of microbes. The near-identical genomes of four strains of Melnitz isolated from different marine provinces and global abundance estimations from metagenomic data suggest that this viral population is globally ubiquitous. Genome analysis revealed several unusual features in Melnitz and related genomes recovered from viromes, such as a curli operon and virally encoded tmRNA controlled by a glutamine riboswitch, neither of which are found in the host. Further phylogenetic analysis of shared genes indicates that this group of viruses infecting the gammaproteobacterial OM43 shares a recent common ancestor with viruses infecting the abundant alphaproteobacterial SAR11 clade. Host ranges are affected by compatible cell surface receptors, successful circumvention of superinfection

exclusion systems and the presence of required accessory proteins, which typically limits phages to singular narrow groups of closely related bacterial hosts. This study provides intriguing evidence that for streamlined heterotrophic bacteria, virus-host transitioning is not necessarily restricted to phylogenetically related hosts, but is a function of shared physical and biochemical properties of the cell.

Introduction

Bacteriophages are the most abundant and diverse biological entities in the oceans and are, on average, an order of magnitude more abundant than their bacterial hosts in surface water (1, 2). Viral predation kills a large proportion of bacterial cells in marine surface waters each day (3) and is a main contributor to nutrient recycling by releasing cell-bound organic compounds into the environment (4, 5). Viral infection can also alter host metabolism through metabolic hijacking (6, 7), which has been shown to reprogram resource acquisition and central carbon and energy metabolism (8), influencing oceanic nutrient cycles. The selective pressure of the predator-prey relationship of bacteria and phages is also a main driver of microbial evolution (9), where the constant arms race requires phages to evolve and improve strategies to overcome host defence mechanisms (10). Recent advances in culture-independent sequencing technology such as single-cell genomics and metagenomics have expanded our understanding of the enormous diversity of marine viruses (11–14). However, many of these sequences lack representation in viral culture collections (15), limiting experimental determination of parameters of infection such as host-range. A resurgence in bacterial cultivation efforts and improved viral isolation methods has led to the discovery of many new phages infecting abundant but fastidious marine bacteria such as SAR11. Combining genomes from viral cultures with metagenomics identified these viruses to be some of the most abundant on Earth across all marine ecosystems (15–18). Yet, many more virus-host systems occupying a range of important ecological niches such as methylotrophy remain poorly understood.

Members of the OM43 clade are small, genomically streamlined (genomes ~1.3 Mbp) Type I methylotrophs of the class Gammaproteobacteria (19). The catabolism of methanol and other volatile organic compounds (VOCs) is an important link between primary production and remineralisation of carbon back to atmospheric CO_2 (20–22). In the surface ocean, peak abundance of OM43 coincides with phytoplankton blooms which provide their main carbon source (22). OM43 are particularly abundant in coastal ecosystems where they comprise up to 5% of the microbial community (23). Members of the OM43 clade are somewhat challenging to grow in the laboratory: increased levels of auxotrophy and largely constitutive metabolism renders them sensitive to media composition (19). As a result, only two OM43 phages have been reported, therefore the influence of viral predation on OM43 is virtually unexplored. The isolation of the first viruses infecting OM43 (Venkman) from the coastal Western English Channel (WEC); and MEP301 from the Bohai Sea were both reported in 2021 (15) (24). Venkman was the third most abundant phage in the WEC sample, indicating that phages of methylotrophs are a major component of this coastal ecosystem (15). In contrast, recruiting reads from global ocean viromes against MEP301 and Venkman, indicated that their relative abundance was below detection limit in most lower-latitude pelagic viromes (24). Thus, phages infecting OM43 were thought to be predominantly found at higher latitudes in regions of high primary productivity.

Here we report the isolation and genomic analysis of Melnitz, representing a novel population of myophages infecting OM43. Four representatives of this virus that shared >99.5% average nucleotide identity were isolated independently on three separate occasions from the temperate coastal WEC and from the Sargasso Sea located within the North Atlantic subtropical gyre. This indicates that

despite low relative abundance at low latitudes, Melnitz was sufficiently abundant to be isolated through enrichment techniques. The genomic similarity between independent isolates suggests a cosmopolitan global distribution, which was supported by metagenomic read recruitment from global ocean viromes. Genome analysis of Melnitz revealed a two-piece tmRNA gene (*ssrA*), controlled by a glutamine riboswitch. Riboswitch control of regulation is a feature of streamlined organisms such as OM43 and SAR11 (25), and here we show that it is also a feature of their associated viruses. Like previously reported SAR11 myophages (15, 16), Melnitz also encoded the portal proteins of a curli operon that is absent in the host. Structural analysis suggests a putative reconfiguration allows this phage-encoded protein to serve as a novel gated secretin or pinholin, with gene synteny indicating a role in timing the release of viral progeny. Phylogenetic analysis revealed that both the *ssrA* gene and tRNA genes encoded by Melnitz were more closely related to those found within the alphaproteobacterial SAR11 host or its associated viruses, than those of its own gammaproteobacterial host. These findings point towards a recent shared ancestor indicative of host transitioning between OM43 and SAR11.

Results and Discussion

Phage Melnitz infecting Methylophilales sp. H5P1 *shares viral clusters with Pelagibacter phages.*

Two bacteriophages were isolated on the OM43 strain H5P1 from two environmental water samples taken from the Western English Channel (WEC) previously (15). Two more phages were obtained from an additional water sample taken at the Bermuda Atlantic Time Series (BATS) station in the Sargasso Sea (Table 1). Phages were successfully purified, sequenced from axenic cultures and assembled into single circular contigs. CheckV comparison to publicly available metagenomes and phage genomes suggested that the viral contigs were complete, circularly permuted genomes without terminal repeats. Transmission electron microscopy (TEM) showed straight contractile tails indicative of myophage morphology with a capsid size of 58±6 nm (Supplementary Figure 1).

All four phage genomes shared 99.95-100% average nucleotide identity across their full genomes therefore all four phages should be regarded as the same viral species (26). We named this species “Melnitz'' after a character in the popular Ghostbusters franchise, continuing the theme of another previously isolated phage on OM43 (Venkman) (15). For clarity, where individual phages within the population (Melnitz) are specified, they will subsequently be referred to with a numerical suffix (e.g. Melnitz-1). General features of the four phages are summarised in Table 1. Shared gene network analysis (VConTACT2) using assembled contigs from Global Ocean Viromes (GOV2) and RefSeq (V88 with ICTV and NCBI taxonomy) viruses assigned Melnitz to the same cluster as the *Pelagibacter* myophages HTVC008M and Mosig (16, 27), suggesting they belong to the same family (Figure 1). The four Melnitz genomes were shared across two clusters, the first contained an additional 15 viral contigs from nine different metagenomes. The second cluster contained three virome contigs shared between the two clusters, as well as 25 pelagimyophage genomes assembled from metagenomes (PMP-MAVGs) (28). *Pelagibacter* myophage isolates HTVC008M and Mosig were also placed in this second cluster. A total of 66 contigs clustering with Melnitz isolates were identified from GOV2 and WEC viromes (12, 29). Phylogenetic analysis showed that only two environmental contigs consistently shared a branch with Melnitz, based on single shared genes encoding: tail sheath protein (Supplementary Figure 2), terminase large subunit (Supplementary Figure 3) and scaffolding proteins (Supplementary Figure 4), as well as four concatenated structural genes (Figure 2). All phages within this viral group were either isolates known to infect streamlined heterotrophs, or were previously predicted to do so based on phylogenetic similarity (28).

Metagenomic analysis shows cosmopolitan nature of Melnitz-like phages

To establish the distribution patterns of Melnitz, GOV2 virome reads were mapped against phage genomes (≥95% nucleotide identity over ≥90% read length; genome coverage of <40% was required to register a phage as 'present' to avoid false positives (30)) (Supplementary Figure 5). Relative abundance of phages was calculated based on the number of reads mapped to a contig, normalised by contig length and sequencing depth (mapped reads per kilobase pair of metagenome; RPKM). Linear regression of relative abundance of all three phages known to infect OM43 (Melnitz from this study, plus Venkman and MEP301) as a function of temperature showed a significant and negative relationship (Melnitz: $p = 4.69 \times 10^{-5}$, $R^2 = 0.199$; Venkman: $p < 2.20 \times 10^{-16}$, $R^2 = 0.435$; MEP301: $p = 1.85 \times 10^{-7}$, $R^2 = 0.198$). Relative abundance was positively correlated with absolute latitude (Melnitz: $p = 1.78 \times 10^{-5}$, $R^2 = 0.094$; Venkman: $p < 2.20 \times 10^{-16}$, $R^2 = 0.545$; MEP301: $p = 4.67 \times 10^{-9}$, $R^2 = 0.247$). This suggests that these viruses are most abundant in cold, productive waters. In contrast, relative abundances of Melnitz-related *Pelagibacter* phages HTVC008M and Mosig were not correlated to temperature (Mosig: $p = 0.094$, $R^2 = 0.047$; HTVC008M: $p = 0.217$, $R^2 = 0.004$) or latitude (Mosig: $p = 0.028$, $R^2 = 0.031$; HTVC008M: $p = 0.077$, $R^2 = 0.018$). OM43 phages MEP301 and Venkman were classed as 'present' (>40% genome coverage) in 39.1% and 53.9% of 131 GOV2 viromes, respectively. Melnitz showed greater global ubiquity, being classed as 'present' in 78.8% of GOV2 viromes. Whilst there are no GOV2 samples from the Sargasso Sea, this ubiquity may explain how we were able to isolate Melnitz from both the WEC and the Sargasso Sea.

It has been demonstrated that the marine biosphere maintains persistent bacterial "seed banks" (31), meaning there is a high probability that any given marine bacteria can be found in any marine ecosystem, albeit in extremely low abundance, awaiting favourable conditions for growth. Similarly, many viruses infecting globally distributed bacteria such as SAR11 are found in viromes from all oceans (18). Studies using "viral tagging" on cyanobacteria and their associated cyanophages suggest that at single sites, viral populations in metagenomes comprise non-overlapping viral sequence space resembling "clouds", which has benchmarked metagenomic analyses and subsequent interpretations (32, 33). In that model, single strains within a viral population encode genomic variations that enable the overall population to adapt more rapidly to environmental selection pressures. In the case of Melnitz, the isolation of the same virus with up to 100% nucleotide identity across the full genome on three separate occasions in the Western English Channel and BATS station in the Sargasso Sea (~5000 km distance between sites) suggests either low population-level variance, or that the isolation conditions used favour this strain. Nonetheless, the presence of cultivable Melnitz populations at both sites supports the virus seed-bank hypothesis where viral populations are conserved and persistent in the environment, being passively transported across oceans via global currents until favourable conditions select them for propagation (34, 35). The 'environmental selection' in the case of Melnitz would likely be the enrichment culturing used for isolation, providing enough suitable hosts and nutrients for viral replication. The possibility remains that the Melnitz population at the BATS site is maintained by a resident "seed" population of OM43, though OM43 is seldom reported in microbial communities at BATS.

Genomic characterization of Melnitz-*like* Methylophilales *phages.*

Genomes of the four Melnitz-like myophage isolates were between 141,372 - 141,552 bp in length, all with G+C content of 37.6 % (Table 1), similar to that of its host (34.4%). For each, 224 open reading

frames (ORFs) were predicted, except Melnitz-2, which had two additional small ORFs of unknown function (226 ORFs total). Four additional tRNA sequences for all four genomes were identified, and functional annotation of ORFs suggest that all Melnitz strains encoded the same set of genes. Out of 224 ORFs, 143 (~63%) had unknown function (Figure 3A). The tail-assembly associated region in the Melnitz-4 genome had nine genes with altered length compared to the equivalents in the other three phages, but where annotation was possible, the genes were predicted to have the same function. Predicted protein structures (PHYRE2) had low confidence and did not allow for a meaningful structural comparison (data not shown). Though structural variations in tail and receptor genes are often considered to be important factors for defining strain-level host ranges (36, 37), all four Melnitz variants had identical host ranges when screened against a panel of other OM43 isolates from the WEC (Methylophilales sp. C6P1, D12P1 and H5P1) (15) (data not shown). Melnitz possessed a set of structural genes typically associated with *E. coli* T4-type myophages, including T4-like baseplate, tail tubes, base plate wedges, tail fibres, virus neck, tail sheath stabilization, prohead core and capsid proteins. Melnitz encodes orthologs of the auxiliary metabolic genes (AMGs) *mazG* and *phoH*, which are involved in cellular phosphate starvation induced stress responses and are a common feature of phages from P-limited marine environments (38–40). Additionally, *hsp20* was identified, which together with *mazG* and *phoH* are considered part of the core genes in T4-like cyanomyophages (38, 41). Both Melnitz and its host, Methylophilales sp. H5P1, encode for a type II DNA methylase; DNA methylase targets foreign DNA for cleavage by the restriction endonucleases (42). In T-even phages, DAM methylase (similar to type II DNA methylase) protects phage DNA from restriction endonucleases through competitive inhibition (43), thus DNA methylase in Melnitz could also be involved in protection from host restriction endonucleases during infection.

For DNA replication and metabolism, Melnitz encodes *nrdA/nrdE* and *nrdB/nrdF* genes that together form active ribonucleotide reductases for catalysing the synthesis of deoxyribonucleoside triphosphates (dNTPs) required for DNA synthesis (44). Additional DNA replication and manipulation associated genes were *polB*, *recA*, *dnaB,* as well as two DNA primases and a 2OG-Fe oxygenase. A *rpsU* gene (encoding the bacterial 30S ribosomal subunit 21S) was found downstream adjacent to a tRNA-Arg (tct). Virally encoded S21 genes were previously identified in the *Pelagibacter* phages HTVC008M (16) and Mosig (27) as well as the OM43 phage Venkman (15). It is thought that the 21S subunit in phages might be required for initiating polypeptide synthesis and mediation of mRNA binding (45); the proximate tRNA sequence may support a translational and protein synthesis role of the viral ribosomal gene. Within the same region, Melnitz also encodes for a surface layer-type (S-layer) protein, previously associated with a strong protection mechanism against superinfection in bacteriophages of *Bacillus spp*. and *Pseudomonas aeruginosa* (46–48). We postulate that phage-encoded S-layer proteins in Melnitz may be used to alter host cell surface receptors and thereby protect against superinfection from viral competitors.

Temporal regulation of T4-like phages primarily occurs at the transcriptional level and requires concomitant DNA replication and promoter/terminator sequences to organise into early, middle and late stages (49). In Melnitz, a total of 46 promoters (31 early and 15 late) and six terminator sequences were identified – considerably fewer than the >124 promoter sequences found in T4 (50, 51). Like other marine T4-like phages, such as those infecting streamlined cyanobacteria (52), Melnitz lacked identifiable middle promotors. Reduced numbers of regulatory elements are a common feature of genomically streamlined bacteria such as SAR11 and OM43 (53). Therefore, we propose that temporal regulation of Melnitz is similarly simplified as a result of host streamlining. An alternative hypothesis is that additional (middle-) promoter sequences in Melnitz are too divergent from known sequences for detection.

Like other T4-like phages, Melnitz did not encode for RNA polymerase, relying on host transcription machinery after infection for synthesis of phage proteins. Melnitz possessed the essential T4-like transcriptional genes: σ^{70}-like; gp45-like sliding clamp C terminal; clamp loader A subunit; transcription coactivator; and a translational regulator protein. In T4, the DNA sliding clamp subunit of a DNA polymerase holoenzyme coordinates genome replication in late-stage infection, forming an initiation complex with a sigma-factor protein and host RNA polymerase (54). To activate it, the sliding clamp is loaded by a clamp-loader DNA polymerase protein complex, allowing the complex to move along DNA strands (55, 56). As the same genes were found in Melnitz it is likely using them to form a T4-like protein complex for transcription and genome replication.

Marine Melnitz-like phages may use glutamine riboswitches to regulate genome expression

ssrA encodes for a tmRNA that in bacterial *trans*-translation together with *smpB* and ribosomal protein S1 is important to release ribosomes that have stalled during protein biosynthesis. These proteins are subsequently tagged and degraded (57). In Melnitz, the *ssrA* gene encoding two-piece tmRNA was situated on the same operon and directly upstream of the transcription co-activator (Figure 3B). Host-encoded *ssrA* can be used as sites of integration for prophages in deep-sea *Sherwanella* isolates (47) and fragments of *ssrA* have previously been identified in prophage genomes (48), most likely as a result of imprecise prophage excision that transfers host genetic material to the excising phage. Though we did not find integrases or other evidence that Melnitz is able to integrate into host genomes, a speculative temperate ancestral strain may explain the presence of a complete *ssrA* gene in Melnitz. A possible role for a complete viral tmRNA is to either help maintain the hijacked bacterial machinery, or the phage might use tmRNA to selectively tag and degrade host proteins to recycle amino acids for viral protein synthesis. We evaluated the frequency that *ssrA* genes were found in 18,146 publicly available genomes in the phage genome database at millardlab.org (last updated on January 21, 2021), which includes all RefSeq genomes. Only 402 phages (2.3% of all available genomes) encoded 133 unique tmRNA genetic structures. Of these, 53 were suggested to be two-piece tmRNA by ARAGORN (59), of which 50 were encoded by phages isolated from marine or aquatic samples. In contrast, only 2.5% of non-permuted tmRNAs were marine, suggesting that viral permuted tmRNA is a feature more prevalent in marine and aquatic phages compared to phages from other environments. Of the 47 contigs (27 were confirmed complete) that were clustered with Melnitz based on shared gene content (Figure 1), 17 had complete *ssrA* genes, indicating that *ssrA* encoded tmRNA is a common, but not defining feature of this viral group.

The 3' domain of the *ssrA* gene in Melnitz encodes a predicted glutamine riboswitch that resembles *glnA* RNA motifs found in cyanobacteria and other marine bacteria (58). Riboswitches are a common regulatory mechanism in streamlined marine bacteria due to their low metabolic maintenance cost compared to protein-encoded promotors and repressors (53). Like tmRNAs, riboswitches are a rare feature in phages. A phage-encoded riboswitch putatively controlling regulation of *psbA* was previously identified in a cyanophage (56). Of the 133 isolate phage genomes encoding tmRNAs only two phages other than Melnitz possessed a riboswitch (both glutamine): *Pelagibacter* phage Mosig (27) and *Prochlorococcus* phage AG-345-P14 (60). Additional riboswitches were identified in metagenomic assembled virome contigs. In ten metagenomically assembled viral genomes of pelagimyophages thought to infect SAR11 (PMP-MAVGs) and five GOV2 contigs related to Melnitz, only one of these contigs had both tmRNA and a riboswitch; eleven PMP-MAVGs encoded neither. This suggests that: (1) glutamine riboswitches are a common but not defining feature in Melnitz-like marine phages; (2) phages of that group use riboswitches or tmRNA, but rarely both, with the only two identified examples occurring in isolated phages, not metagenomically derived genomes.

Curiously, the bacterial OM43 host of Melnitz (H5P1) does not encode a glutamine riboswitch - only one cobalamin riboswitch was found located upstream of the Vitamin B12 transporter gene *btuB*. Other members of the *Methylophilaceae* (OM43 strains HTCC2181, KB13, MBRSH7 and *Methylopumilus planktonicus, M. rimovensis* and *M. turicensis*) also lack glutamine riboswitches. In cyanobacteria, the glutamine riboswitches were previously found to regulate the glutamine synthase *glnA* and are strongly associated with nitrogen limitation (61). Phages infecting *Synechococcus* have been shown to use extracellular nitrogen for phage protein synthesis (62), which might indicate a similar role for phage-encoded glutamine riboswitches in Melnitz. However, neither OM43 nor Melnitz encode the glutamine synthase *glnA* or homologous genes required for glutamine synthesis. This suggests that Melnitz-like phages use viral riboswitches to regulate their own genes rather than hijacking the cellular machinery.

Phage encoded curli operons may be involved in regulating cell lysis

Curli genes are typically associated with the bacterial production of amyloid fibres that are part of biofilm formation, where the CsgGF pore spans the outer membrane as part of a Type VIII secretion system, allowing for the secretion of CsgA and CsgB that assemble into extracellular amyloid fibres (63). In complete curli modules, CsgG and CsgF form an 18-mer heterodimer comprising nine subunits with 1:1 stoichiometry between CsgF and CsgG. The structure of the pore is dictated by CsgG, which forms a channel ~12.9Å in diameter. CsgF forms a secondary channel ~14.8Å in diameter at the neck of the beta barrel (Figure 4A, Supplementary Figure 6 A and B) and assists in excretion of the amyloid fibre. Like previously described pelagimyophages (28), Melnitz possesses *csgF* and *csgG,* but lacks the genes for amyloid fibre production - *csgA* and *csgB* (Figure 3). Zaragoza-Solas et al. speculated that phage-encoded curli pores may allow for the uptake of macromolecules, or together with unidentified homologues of missing curli genes form a complete, functional curli operon producing amyloid fibres for “sibling capture” of proximate host cells (28). However, similar to results in pelagimyophages, we did not find evidence for a complete phage encoded curli biogenesis pathway in Melnitz, nor does the bacterial H5P1 host encode for any curli associated genes, suggesting that these genes were acquired by an ancestral strain of Melnitz and pelagimyophages before host transition to OM43 and SAR11, respectively.

Structural homology modelling of the Melnitz-encoded CsgGF portal with SwissModel (64) yielded low structural similarity to known CsgGF structures (GMQE <0.5 with Q-MEAN scores < -3 and z-scores >>2 outside normalised Q-mean score distributions from a non-redundant set of structures from PDB). Structural similarity was greatest in regions 238-258 of CsgG where Q-MEAN scores were ~0.7, with conserved regions in the periplasmic-facing α-helices of CsgG. Melnitz-encoded CsgF showed poor structural similarity to any models available within Swiss-Model. Therefore, we used AlphaFold2 (65) to predict the structure of phage-encoded CsgG and CsgF *de novo* and compared predicted structures to known CsgGF structures. The predicted structure of Melnitz-encoded CsgG showed close structural similarity to known CsgG, with two notable exceptions: (1) a narrowing of the barrel at the neck; (2) an extension of the α-helix outside the barrel (Figure 4B and 4C). In contrast there was little structural similarity between Melnitz-encoded CsgF and the structure of CsgF in *E. coli*, barring a shared α-helix domain. Phage-encoded CsgF comprised two long α-helices ending in a β-sheet (Figure 4D). Similar modelling of CsgG and CsgF encoded by pelagiphage HTVC008M also contained these features (Supplementary Figure 7 A and B, respectively). Assuming the pore orientation in an infected OM43 cell is the same as that of CsgG in *E. coli*, the additional domains on phage-encoded CsgF would either place the terminal β-sheet inside the barrel of CsgG (unlikely due to steric hindrance) or the extended structure would point outwards into the extracellular milieu, in an unusual conformation with

unknown function. It is more likely that in this configuration of CsgG, phage-encoded CsgG and CsgF have diverged and evolved independently. Therefore they may no longer form a heterodimer, with CsgG retaining the function as a pore, and CsgF evolving independently to provide an alternative, unknown function. Indeed, the narrowing of the barrel neck in Melnitz-encoded CsgG creates a channel 14.7Å in diameter, matching that provided by CsgF in *E. coli* (Supplementary figure 6 B and C). One possibility is that phage-encoded CsgG is functionally analogous to pinholins, used by phage λ to regulate cell lysis. Pinholins form channels ~15Å in diameter in the inner membrane, resulting in rapid membrane depolarization and subsequent activation of membrane-bound lysins (66).

An alternative hypothesis is that in Melnitz, CsgGF forms a heterodimer in the outer membrane similar to CsgGF in *E. coli*, but that the structure is inverted so that the extended α-helices of CsgF point into the periplasm. In this conformation, the extension from Melnitz-encoded CsgG (Figure 4C) can interact to stabilise the elongated hinged α-helices of CsgF. The C-terminal β-sheets of CsgF could then form a second channel beneath the CsgG barrel (Supplementary Figure 8 A). In this conformation, CsgGF is structurally similar to a secretin, a large protein superfamily used for macromolecule transport across the outer membrane such as DNA for natural competence (e.g. PilQ within the type IV secretion system in *Vibrio cholerae*) or extrusion of filamentous phages during chronic infection (e.g. pIV in bacteriophages Ff) (67). Like CsgGF in an inverted conformation, PilQ comprises a β-barrel, with a secondary pore below, attached by two hinged α-helices. The inner surface of PilQ is negatively charged to repel the negatively charged backbone of DNA and assist in transportation across the membrane. In contrast, the inner surface of the CsgGF channel is positively charged and narrower (14.7Å compared to 20.6Å in PilQ), making a role in DNA transport across the membrane unlikely. While the function of phage-encoded CsgGF is not yet clear, the additional domains of CsgF and the lack of other curli genes within either the phage or the host suggest that it is a novel pore structure whose function has diverged from ancestral CsgGF and would be a worthy target for future structural resolution.

Whether the CsgGF complex in Melnitz acts as a pinholin or a secretin, gene synteny supports a putative role in regulation of the timing of cell lysis. First, genes *csgGF* are located immediately downstream of thymidylate synthase *thyX* and *ssrA* and transcription co-activator genes (Figure 3). Overexpression of thymidylate synthase in the D29 phage infecting *Mycobacteria tuberculosis* results in delayed lysis and higher phage yields (68). In phage T4, postponing lysis is used to delay the release of viral progeny until conditions are favourable, thereby maximising virion production before viral release and successful replication after (69). In the T4-like Melnitz, *thyX* is therefore likely to be involved in postponing lysis as well. Lytic control may putatively be related to the upstream glutamine riboswitch. H5P1 lacks glutamine synthesis pathways, but has a complete peptidoglycan biosynthesis pathway necessary to produce peptidoglycans from glutamine. Therefore, both the H5P1 host and Melnitz phage are restricted to using glutamine for protein and peptidoglycan synthesis only. In phage λ, the depletion of peptidoglycan precursors can trigger lysis through activation of a spanin complex (70). We speculate that during the course of the infection cycle, glutamine levels are kept low through incorporation into viral proteins. When protein synthesis is complete, intracellular glutamine levels increase, activating the glutamine riboswitch. This in turn could trigger an opening of the curli portal, which may result in rapid depolarization of the membrane, triggering activation of lysins and rapid cell lysis.

Structural homology modelling (SwissModel) of a Melnitz-encoded enzyme (gp67), putatively annotated as a glycosyl hydrolase, revealed structural similarity (38% sequence identity at 98% coverage, with a global model quality estimate (GMQE) of 0.82) to autolysin SagA encoded by *Brucella abortus* (PDB model 7DNP). SagA acts to generate localised gaps in the peptidoglycan layer for

assembly of type IV secretion systems (71). Therefore, it is likely that gp67 and CsgGF work in concert in Melnitz. Peptides involved in binding peptidoglycan at the active site (Glu17, Asp26, Thr31) were conserved between Melnitz gp67 and SagA (Supplementary Figure 9X) (72). Outside of this active site, structural similarity to T4 endolysin (PDB model 256l.1) was low (13% identity over 51% coverage; GMQE: 0.18). Melnitz lacked any other lysin-like genes. Like the T4-encoded endolysin *e* (73), gp67 is under control of an early transcription promotor (Figure 3B). We therefore propose that the gp67 gene in Melnitz potentially serves two functions: (1) as an endolysin for degradation of cellular peptidoglycans during lysis; or (2) as an autolysin to enable assembly of the CsgGF or CsgG-only portal protein. Phages infecting *Streptococcus pneumoniae* upregulate host-encoded autolysins alongside phage-encoded lysins during late-stage infection to accelerate cell lysis (74). The close structural match between gp67 and SagA suggests that an ancestor of Melnitz acquired a host-encoded autolysin as an alternative lysin during its evolutionary history. Whether this autolysin-derived phage lysin could be activated by membrane depolarization through CsgG is unknown.

Melnitz-encoded genes suggest a possible host transition event from SAR11 to OM43

Phage host ranges are largely determined by interactions between host receptor proteins and phage structural proteins such as tail fibres that enable the phage to adsorb and inject its genetic material. Mutation in phage proteins, either through point mutations or recombination during co-infection can result in host range expansion or transition (75). Host range expansion or transition within species boundaries is more common, but rare transition events between hosts from different genera can occur through mutations in tail fibres (76). Here, two separate lines of evidence converge to suggest that Melnitz underwent a recent host transition from SAR11 to OM43. First, *ssrA* encoded by Melnitz was identified to be more closely related to homologs within the alphaproteobacterial lineage. Two-piece circularly permuted tmRNA is common in three major lineages: Alphaproteobacteria, Betaproteobacteria (now part of Gammaproteobacteria (77)) and Cyanobacteria (78, 79). The phylogenetic evidence for alphaproteobacterial *ssrA* in Melnitz rather than *ssrA* that matches the gammaproteobacterial lineage of its OM43 host suggests that this gene was acquired from an Alphaproteobacteria (Figure 4). Second, two of the four tRNAs encoded by Melnitz were more closely related to tRNAs encoded by pelagimyophages than those of their host. Melnitz encodes two versions of tRNA-Arg(TCT) – one most closely related to that of its host H5P1, and one closely related to *Pelagibacter* phage Mosig. Phage-encoded tRNAs enable large phages to sustain translation as the host machinery is degraded to fuel phage synthesis (80). Phage-encoded tRNAs also enable phages to optimise protein synthesis in a host with different codon usage and thus serve as both a marker of increased host range and evolution through different hosts (81). Indeed, tRNAs are often used for computational host-prediction of phages due to high sequence conservation of the gene between host and viral forms (82, 83). We postulated that the host-range of Melnitz would be evident in the four tRNAs found within its genome and reflect potential hosts in the OM43 clade. Alternatively, if a recent host transition occurred, as *ssrA* phylogeny suggests, tRNAs would be similar to those found in the SAR11 bacteria and viruses. A search for tRNA genes in the genomes of isolated *Pelagibacter* phages and PMP-MAVGs (47 genomes) and isolated phages infecting OM43 (three genomes) identified tRNAs in seven *Pelagibacter* phage genomes. Melnitz was the only phage known to infect OM43 that encoded tRNAs, with four tRNAs in total (Figure 6). Sequences encoding tRNAs were aligned using all-vs-all BLASTN with a minimum expect-value of 1×10^{-5}. Two out of four tRNAs in Melnitz were tRNA-Arg(TCT), the first aligning most closely with its H5P1 host. The second tRNA-Arg best aligned with the tRNA gene found in the *Pelagibacter* phage Mosig. Its alignment to bacterial tRNA matched OM43 strains H5P1 and HTCC2181, respectively (>90% identity). The third tRNA-Leu found in Melnitz aligned

406 with PMP-MAVG-17, previously classified as *Pelagibacter* phage (28), but only had 45% nucleotide
407 identity. The fourth tRNA-Trp (CCA) found in Melnitz did not align (e-values > 1×10^{-5}) with any tRNA
408 found in OM43, SAR11 or any of their respective phages. Using tRNA as host indication in Melnitz
409 therefore reflects its OM43 host, but also indicates genetic exchange with SAR11 virus-host systems.
410 Furthermore, the related *Pelagibacter* phage Mosig possessed tRNAs matching OM43 and Melnitz as
411 well as a second tRNA aligning with its SAR11 host (93%). This may suggest either a relatively recent
412 genetic exchange between OM43 and SAR11 virus-host systems, and/or a surprisingly broad host-
413 range for *Pelagibacter* phage Mosig and *Methylophilales* phage Melnitz.

414 To assess the breadth of host ranges, we challenged the SAR11 strains HTCC7211 and HTCC1062 and
415 OM43 strains H5P1, D12P1 and C6P1 (15) with the phages Melnitz and Mosig (27), but found no
416 evidence that either phage is able to replicate in cells beyond host class boundaries as observed via
417 cell lysis (Supplementary Figure 10). Melnitz only infected D12P1 and H5P1, but not C6P1 or any
418 SAR11, whereas Mosig only caused lysis in the *Pelagibacter ubique* HTCC1062. Though the limited
419 number of available bacterial strains could have missed potential permissive hosts in a different
420 subclade, there is no evidence that non-canonical matching tRNA between SAR11/OM43 and their
421 phages is due to a speculative broad host range in these systems, and instead supports the
422 phylogenetic evidence we presented for a recent host transition event.

423 Phage host-range expansion and transition within closely related (i.e. between strains) is a common
424 feature of phage evolution, and has been shown to occur *in vitro* via homologous recombination
425 between co-infecting phages (75, 84). In contrast, expansion and transition between distantly related
426 taxa is rare but has been previously observed *in vitro*. Naturally occurring, low-abundance mutants of
427 T4 transitioned from *E. coli* to *Yersinia pseudotuberculosis* through modification of tail fibre tips in
428 gene 37, yielding variants that could infect both hosts and others that could only infect *Y.*
429 *pseudotuberculosis* (76). To our knowledge, host transition between bacteria belonging to different
430 classes, as proposed here, has not previously been observed, although host-prediction of viral contigs
431 from metagenomes has identified rare phages (115 out of 3,687) possibly able to infect multiple
432 classes (85). We propose that two properties of SAR11 and OM43 increase the likelihood of such an
433 event in natural communities: (1) Both SAR11 hosts and their associated phages possess
434 extraordinarily large, globally ubiquitous effective population sizes, making even extreme rare events
435 likely. This study has shown that phages infecting OM43 can also be abundant at higher latitudes,
436 further increasing the likelihood of rare strain variants to occur; (2) Both OM43 and SAR11 hosts share
437 highly streamlined genomes with elevated levels of auxotrophy, minimum regulation and similar G+C
438 content, shaped by selection pressure to maximise replication on minimum resources in nutrient-
439 limited marine environments (19, 86). Both fastidious hosts can be cultured on identical minimal
440 medium, as long as additional methanol is provided as a carbon source for the methylotrophic OM43
441 (15). Therefore, the phenotypic difference between OM43 and SAR11 might be smaller than suggested
442 by their taxonomic classification. Given viral selection occurs at the phenotypic level, we propose the
443 possibility of a rare event where a mutant of a T4-like phage infecting SAR11 was able to successfully
444 adsorb and inject its genome into an OM43 host, possibly co-infected with a methylophage that
445 enabled a host-transition event through homologous recombination. Once a transition event
446 occurred, the phage likely rapidly evolved specialism on the new host, losing the ability to infect the
447 original host, as demonstrated in host-range experiments. Compared to copiotrophic, *R*-strategist
448 microbial taxa such as *E. coli* and *Vibrio* spp., very little is known about the processes governing viral
449 host-range and co-evolution in *K*-strategist bacteria such as the genomically streamlined taxa that
450 dominate oligotrophic oceans. The early evidence shown here may suggest that phages can transition

to new hosts from distantly related taxa in natural communities. Such events would explain in part the diverse host ranges predicted amongst viruses within gene-sharing clusters (87, 88).

Conclusion

In this study, we provide evidence that supports a putative inter-class host transition event between two important clades of streamlined marine heterotrophs and expand our knowledge about the dynamics and characteristics in genomically streamlined heterotrophic virus-host systems. We isolated four near-identical strains of the new myophage Melnitz from subtropical and temperate marine provinces infecting the important methylotrophic OM43 clade, which we showed to be closely related to myophages infecting the abundant SAR11 clade. The analysis of metagenomic datasets provides evidence that this phage group is ubiquitous in global oceans despite relatively low overall abundance, supporting the viral seed bank hypothesis. Our genomic analysis of Melnitz revealed an incomplete curli module similar to reported curli pores in pelagimyophages, representing a rare and intriguing protein dimer that is absent in their respective host clades. We propose that these virally encoded curli pores may have been repurposed as a functional analogue for the regulation of viral lysis. We also identified an *ssrA* gene encoding for a complete viral tmRNA protein controlled by a glutamine riboswitch, showing that virus-host interactions can be regulated through riboswitches reflecting the extensive use of riboswitches in streamlined marine heterotrophs. Further phylogenetic analysis showed that the *ssrA* gene is related to the alphaproteobacterial SAR11 lineage, not the gammaproteobacterial OM43 lineage, providing evidence for host transition events in natural marine microbial communities, which was supported by the alignment of viral and bacterial tRNA genes of both lineages. These findings support the conclusion that in heterotrophic streamlined virus-host systems evolution of viral diversity is likely to be driven by host transition and expansion between closely related phages infecting hosts across broad taxonomic groups, likely increasing mosaicism and genetic exchange.

MATERIALS & METHODS

OM43 strain, media, and growth conditions

The OM43 strains *Methylophilales* H5P1 and D12P1 were isolated previously from surface water from the Western English Channel (15). Continuous cultures were grown using artificial seawater-based artificial seawater medium (ASM1) (89) amended with 1 mM NH_4Cl, 10 µM KH_2PO_4, 1 µM $FeCl_3$, 100 µM pyruvate, 25 µM glycine, 25 µM methionine as well as 1 nM each of 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP), pantothenate, biotin, pyrroloquinoline quinone (PQQ) and B12. Additional 1 mM methanol and 5 µL of an amino acid mix (MEM Amino Acids (50x) solution, Sigma-Aldrich) were added per 100 mL medium. Bacterial stocks were grown in 50 mL acid-washed polycarbonate flasks at 15 °C without shaking and 12-hour light-dark cycles.

Water sources and phage isolation

Water samples were collected from two different stations: Station L4 in the Western English Channel (WEC; 50°15'N; 4°13'W) and the Bermuda Atlantic Time Series sampling station (BATS; 31°50'N; 64°10'W) in the Sargasso Sea. For each sample we used Niskin bottles mounted on a CTD rosette to

collect 2 L of seawater at 5m depth (Table 1) into acid-washed, sterile, polycarbonate bottles. To obtain a cell-free fraction, water samples were filtered sequentially through a 142 mm Whatman GF/D filter (2.7 µm pore size), a 142 mm 0.2 µm pore polycarbonate filter (Merck Millipore) and a 142 mm 0.1 µm pore polycarbonate filter (Merck Millipore) using a peristaltic pump. The cell-free viral fraction was concentrated to about 50 mL with tangential flow filtration (50R VivaFlow, Sartorius Lab instruments, Goettingen, Germany) and used as inoculum in a multi-step viral enrichment experiment followed by Dilution-to-Extinction purification as described previously (15). Briefly, viral inoculum (10% v/v) was added to 96-well Teflon plates (Radleys, UK) with exponentially growing host cultures. After a 1 to 2 weeks incubation period, cells and cellular debris were removed with 0.1 µm syringe PVDF filters. The filtrate was added as viral inoculum to another 96-well Teflon plate with exponentially growing host culture. This process was repeated until viral infection was detected by observing cell death using flow cytometry. Phages were purified by Dilution-to-Extinction methods (Our detailed protocol is available here: 10.17504/protocols.io.c36yrd).

Assessment of viral host ranges

An acid-washed (10% hydrochloric acid) 48-well Teflon plate was prepared with 2 mL per well of ASM1 amended with 1 mM methanol and 5 µL of an amino acid mix (MEM Amino Acids (50x) solution, Sigma-Aldrich). Wells were inoculated to $1x10^6$ cells·mL^{-1} with *Pelagibacter* hosts HTCC1062 or HTCC7211, or OM43 hosts C6P1, D12P1 and H5P1 cultures in replicates of three for each of the Mosig and Melnitz phages, plus one set of no virus controls for each host. The wells marked for viral infections were infected with 200 µL of viral culture. The plate was then incubated at 15 °C and monitored daily using flow cytometry for ~2 weeks. Successful infections were identified observing cell lysis in virus amended wells compared to no-virus controls (15).

Phage DNA preparation, genome sequencing and annotation

For each viral isolate, 50 mL OM43 host cultures were grown in 250 mL acid-washed, polycarbonate flasks and infected at a cell density of about 1-5x 10^6 cells·mL^{-1} with 10% v/v viral inoculum. Infected cultures were incubated at 15 °C for 7 to 14 days after which the lysate was transferred to 50 mL Falcon tubes. Larger cellular debris was removed by centrifugation (GSA rotor, Thermo Scientific 75007588) at 8,500 rpm/ 10,015 x g for one hour. Supernatant was filtered through pore-size 0.1 µm PVDF syringe filter membranes to remove any remaining smaller cellular debris. Phage particles were precipitated using a PEG8000/NaCl flocculation approach (90). Briefly, 50 mL lysate was amended with 5g PEG8000 and 3.3 g NaCl (Sigma), shaken until dissolved and incubated on ice overnight. Precipitated phages were pelleted by centrifugation at 8,500 rpm / 10,015 x g for one hour at 4 °C. After discarding the supernatant, phage particles were resuspended by rinsing the bottom of the tubes twice with 1 mL SM buffer (100 mM NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-Cl). DNA from the suspended phages was extracted with the Wizard DNA Clean-Up system (Promega) following manufacturer's instructions, using pre-heated 60 °C PCR-grade nuclease free water for elution.

Nextera XT DNA libraries were prepared and sequenced by the Exeter Sequencing Service (Illumina paired end [2 x 250 bp], NovaSeq S Prime [SP], targeting 30-fold coverage). Reads were quality controlled, trimmed and error corrected with the tadpole function (default settings) within BBMap v38.22 ((91) available at https://sourceforge.net/projects/bbmap/). Contigs were assembled using SPAdes v.3.13 with default settings (92). Viral contigs were confirmed with VirSorter v1.05 (93) (categories 1 or 2, > 15kbp). Quality and completeness of contigs as well as terminal repeats were

evaluated using CheckV v0.4.0 with standard settings (94). Genes were called with phanotate v2019.08.09 (95) and imported into DNA Master for manual curation (96) using additional gene calls made by GenMarkS2 (97), GenMark.heuristic (98), Prodigal v2.6.3 (99) and Prokka v1.14.6 (100). ORFs were functionally annotated using BLASTp against NCBI's non-redundant protein sequences (101), phmmer v2.41.1 against Pfam (102) and SWISS PROT (103). All genes called were listed and compared using a scoring system evaluating length and overlap of ORFs as well as quality of annotation (96); tRNA and tmRNA were identified with tRNAScan-SE v2.0 (104) and ARAGORN v1.2.38 (59). Genomes were scanned for riboswitches using the web application of Riboswitch Scanner (106, 107). FindTerm (energy score < -11) and BPROM (LDF > 2.75) from the Fgenesb_annotator pipeline (108) were used to predict promoter and terminator sequences, using default parameters. The σ^{70} promoters predicted this way were considered early promoters. Known T4-like late promoter sequence 5'-TATAAAT-3' (50, 56) and middle-promoter MotA box (TGCTTtA) dependent middle promoters were used as query for a BLASTN search over the whole genome. Promoters and/or terminators were excluded if they were not intergenic or not within 10 bp overlap of the start/end of ORFs.

Host DNA preparation, genome sequencing and annotation

Host cultures were grown in 50 mL ASM1 medium amended with 1mM methanol in 250 mL polycarbonate flasks. Upon reaching maximum cell density, genomic DNA was extracted using Qiagen DNAeasy PowerWater Kit (REF 14900-50-NF) from biomass retained on 0.1 µm PC filters following the manufacturer's protocol with minor modifications to increase the yield. The bead beating step was lengthened from 5 minutes to 10 minutes. DNA elution was performed with a 2 minute incubation with elution buffer warmed to 55 °C. Nextera DNA libraries were prepared and sequenced by MicrobesNG (Birmingham, UK) for Illumina short read sequencing on the HiSeq2500 (Illumina paired end [2 x 250 bp], targeting 30-fold coverage). Additionally, long read sequencing was prepared using a MinION flow cell. Reads were quality controlled, trimmed and error corrected with the tadpole function (default settings) within BBMap v38.22 ((91) available at https://sourceforge.net/projects/bbmap/). Contigs were assembled using SPAdes v.3.13 with default settings (92). Gene calls were made with PROKKA v1.14.6 (100) and submitted to BlastKOALA (105) for further annotation and prediction of KEGG pathways using the *Methylophilales* strain HTCC2181 as reference.

Phylogeny and Network analysis

All contigs from the Global Ocean Virome (GOV2) dataset and a WEC virome (12, 29), *Methylophilales* phage Venkman (15) and LD28 phage P19250A (109), all isolated *Pelagibacter* phages (15–18, 27) and *Pelagibacter*-like MAGs (28) were screened for contigs that share a viral population with the genomes of OM43 phages (95% ANI over 80% length) using ClusterGenomes.pl v5.1 (https://github.com/simroux/ClusterGenomes). Genes of all contigs were identified by Prodigal v2.6.3 (99) and imported into the Cyverse Discovery Environment 2.0 (available at https://de.cyverse.orge/), where vContact-Gene2Genome 1.1.0 was used to prepare protein sequences before protein clustering using VConTACT2 (v.0.9.8) with default settings (87) to assess relatedness via shared gene networks. The gene sharing network was visualised using Cytoscape v3.7.1 (110). For the single-gene based phylogeny, genes of contigs that fell into the same viral protein clusters as the isolated OM43 phages from this study were aligned to selected genes in annotated OM43 phage genomes, and other respective genes of interest, with BLASTp (default parameters) (111). Genes were aligned within the

Phylogeny.fr online server (112) opting for MUSCLE alignment (113) and built-in curation function (114) with default settings, removing positions with gaps for calculating phylogenetic trees. Maximum-likelihood trees were calculated with PhyML (115, 116) using 100 bootstraps. The trees were visualised using FigTree (v1.4.4 available at http://tree.bio.ed.ac.uk/software/figtree/). Figures were edited with Inkscape (www.inkscape.org) for aesthetics.

Metagenomic reads recruitment

Marine virome datasets were used to assess the relative abundance of phage contigs including a single virome from the Western English Channel and 131 samples from the Global Ocean Virome dataset (GOV2) (12, 29). Metagenomic reads were subsampled to 5 million reads using the reformat.sh command within the bbmap suite. Bowtie2 (117) indexes of dereplicated contigs were created for all known pelagiphage isolate genome (15–18, 27), the LD28 phage P19250A (109), a selection of cyanophages, a selection of abundant *Roseobacter* phages (118), *Enterobacteria* phages T4 and T7 (as negative controls) as well as one genome from the viral population isolated in this study (Melnitz). Metagenomic reads from each virome were mapped against all contigs with bowtie2 (bowtie2 --seed 42 --non-deterministic). To calculate coverage and Reads Per Kilobase of contig per Million reads (RPKM) we used coverm (available at https://github.com/wwood/CoverM), with the following commands: “coverm contig --bam-files *.bam --min-read-percent-identity 0.9 --methods rpkm --min-covered-fraction 0.4”.

Search and alignment of tRNA

tRNAs were identified from bacteria and virus genomes using ARAGORN v.1.2.38 (59) using the ‘-t –gcbact –c –d –fons’ flags and tRNAscan-SE v2.0.7 using flags ‘-B –fasta’ (104). tRNAs were deduplicated using seqkit v0.15.0 (119) with ‘rmdup –by-seq –ignore-case’. A BLAST database of tRNA genes was made in BLAST v2.5.0+ (111) and used for sequence alignment with BLASTN with the flags `-outfmt '6 std qlen slen' -evalue 1e-05 -task blastn-short`. Percentage identity was calculated by dividing alignment-length by query-length times alignment-percentage.

Structural analysis of a putative endolysin.

The predicted amino acid sequence of gene product 67 (gp67) was used as a query to identify putative structures on the SwissModel server (64) using BLAST and HHBits. Putative models were downselected based on suitable quaternary structure properties and GQME>0.7. Autolysin SagA from *Brucella abortus* (PDB model 7dnp.1) was selected as the best-hit and used for subsequent modelling of the structure of gp67. Models and associated figures were visualised in PyMOL v. 2.5.1 (https://pymol.org/2/).

Structural analysis of CsgGF

The predicted amino acid sequences of CsgG and CsgF from Melnitz were used as a query to identify putative structures on the SwissModel server using BLAST and HHBits. The best-hit was determined by predicted quaternary structure properties. Global GQME scores were <0.5 and Q-MEAN scores identified the best predicted structures as unreliable (ranging from -3.28 to -5.11), although localised

regions had Q-MEAN scores > 0.7. Therefore, to improve structural predictions, amino acid sequences for CsgG and CsgF were independently run through AlphaFold2 using the available Colab web interface (https://colab.research.google.com/drive/1LVPSOf4L502F21RWBmYJJYYLDlOU2NTL). Structures were determined both with and without post-prediction relaxation (use_amber) and use of MMSeqs2 templates (use_templates). As no noticeable differences were observed between these runs, we selected the top scoring unrelaxed model for downstream comparison to known CsgG and CsgF structures. Predicted structures from AlphaFold2 were downloaded, visualised and aligned to *E. coli* CsgGF (7NDP) in PyMOL v. 2.5.1. CsgGF from pelagimyophage HTVC008M was similarly analysed with AlphaFold2 to confirm structural similarity to that of Melnitz. Structural prediction of the CsgGF heterodimer in Melnitz was assumed to conform to the 18-mer structure of CsgGF in *E. coli.* Scripts for generating structures in PyMOL are available here https://github.com/HBuchholz/Genomic-evidence-for-inter-class-host-transition-between-streamlined-heterotrophs. Electrostatic potential of protein surfaces was calculated and visualised using the APBS Electrostatics plugin available within PyMOL.

Transmission Electron Microscopy

For ultrastructural analysis, bacterial cells and/or phages were adhered onto pioloform-coated 100 mesh copper EM grids (Agar Scientific, Standsted, UK) by floating grids on sample droplets placed on parafilm for 3 min. After 3 x 5 min washes in droplets of deionized water, structures were contrasted on droplets of 2% (w/v) uranyl acetate in 2% (w/v) methyl cellulose (ratio 1:9) on ice for 10 min, the grids picked up in a wire loop and excess contrasting medium removed using filter paper. The grids were then air dried, removed from the wire loop and imaged using a JEOL JEM 1400 transmission electron microscope operated at 120 kV with a digital camera (Gatan, ES1000W, Abingdon, UK).

DATA AVAILABILITY

All four Melnitz-like genome were deposited as GenBank entry on NCBI under accession number MZ577095-MZ577098 of BioProject PRJNA625644; the reference genome used for the analysis was deposited under MZ577097. Sequencing data for all phages sequenced in this study can be found on the SRA databank under Accession numbers SAMN18926670-SAMN18926674. Reads for Methylophilales bacterial host H5P1 are available under SAMN20856461.

ACKNOWLEDGEMENTS

We would like to thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing the TE microscopy and imaging. We would also like to thank the crew of the R/V Plymouth Quest and our collaborators at Plymouth Marine Laboratory for collecting water samples, and the driver Magic for delivering water samples from Plymouth to Exeter. We would also like to thank the crew of the R/V Atlantic Explorer and our collaborators at the Bermuda Institute of Ocean Sciences. We acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. Genome sequencing was provided by the Exeter Sequencing Service. This project utilised equipment funded by the Wellcome Trust Institutional Strategic Support Fund

(WT097835MF), Wellcome Trust Multi User Equipment Award (WT101650MA) and BBSRC LOLA award (BB/K003240/1). The efforts of Holger H. Buchholz in this work were funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training program. LMB, MLM and BT were funded by NERC (NE/ R010935/1) and by the Simons Foundation BIOS-SCOPE program.

H.H.B. and B.T. designed experiments and wrote the manuscript, H.H.B. performed experimental research and analysed data, B.T. performed protein structure predictions, L.M.B. assisted with data analysis, A.G.B. assisted with tRNA analysis, M.L.M. assisted with laboratory work.

REFERENCES

1. Suttle CA. 2005. Viruses in the sea. Nature 437:356–361.

2. Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, Heldal M, Kimmance SA, Middelboe M, Nagasaki K, Paul JH, Schroeder DC, Suttle CA, Vaqué D, Wommack KE. 2008. Global-scale processes with a nanoscale drive: the role of marine viruses. ISME J 2:575–578.

3. Eric Wommack K, Colwell RR. 2000. Virioplankton: Viruses in Aquatic Ecosystems. MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS 64:69–114.

4. Weitz JS, Wilhelm SW. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. F1000 Biol Rep 4:17.

5. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, Follows MJ, Fuhrman JA, Jover LF, Lennon JT, Middelboe M, Sonderegger DL, Suttle CA, Taylor BP, Frede Thingstad T, Wilson WH, Eric Wommack K. 2015. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. ISME J 9:1352–1364.

6. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. Virol J 16:15.

7. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. Nat Microbiol 3:754–766.

8. Howard-Varona C, Lindback MM, Bastien GE, Solonenko N, Zayed AA, Jang H, Andreopoulos B, Brewer HM, Glavina Del Rio T, Adkins JN, Paul S, Sullivan MB, Duhaime MB. 2020. Phage-specific metabolic reprogramming of virocells. ISME J https://doi.org/10.1038/s41396-019-0580-z.

9. Martiny JBH, Riemann L, Marston MF, Middelboe M. 2014. Antagonistic Coevolution of Marine Planktonic Viruses and

691 Their Hosts. Ann Rev Mar Sci 6:393–414.

692 10. Avrani S, Schwartz DA, Lindell D. 2012. Virus-host swinging party in the oceans: Incorporating biological complexity
693 into paradigms of antagonistic coexistence. Mob Genet Elements 2:88–95.

694 11. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM,
695 Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S,
696 Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P,
697 Karsenti E, Sullivan MB. 2015. Ocean plankton. Patterns and ecological drivers of ocean viral communities. Science
698 348:1261498.

699 12. Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K,
700 Carmichael M, Cruaud C, Ferland J, Kandels S, Liu Y, Marec C, Picheral M, Pisarev S, Poulain J, Vik D, Babin M, Bowler C,
701 Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S. 2019. Marine DNA Viral Macro- and Microdiversity
702 from Pole to Pole. Cell 177:1109–1123.

703 13. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, Anton J, Gasol JM,
704 Rosselli R, Rodriguez-Valera F, Sullivan MB, Acinas SG, Martinez-Garcia M. 2017. Single-virus genomics reveals hidden
705 cosmopolitan and abundant viruses. Nat Commun 8:15892.

706 14. Martinez-Hernandez F, Fornas Ò, Lluesma Gomez M, Garcia-Heredia I, Maestre-Carballa L, López-Pérez M, Haro-
707 Moreno JM, Rodriguez-Valera F, Martinez-Garcia M. 2019. Single-cell genomics uncover Pelagibacter as the putative
708 host of the extremely abundant uncultured 37-F6 viral population in the ocean. ISME J 13:232–236.

709 15. Buchholz HH, Michelsen ML, Bolaños LM, Browne E, Allen MJ, Temperton B. 2021. Efficient dilution-to-extinction
710 isolation of novel virus–host model systems for fastidious heterotrophic bacteria. ISME J 1–14.

711 16. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB,
712 Giovannoni SJ. 2013. Abundant SAR11 viruses in the ocean. Nature 494:357–360.

713 17. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, Du S, Rensing C. 2018. Pelagiphages in the Podoviridae family
714 integrate into host genomes. Environ Microbiol 21:1989–2001.

715 18. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, Du S, Tian Z, Zhao Y. 2020. Culturing novel and abundant pelagiphages
716 in the ocean. Environ Microbiol https://doi.org/10.1111/1462-2920.15272.

717 19. Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, Oh H-M, Kitner JB, Vergin KL, Rappé MS. 2008. The
718 small genome of an abundant coastal ocean methylotroph. Environ Microbiol 10:1771–1782.

719 20. Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, Stepanauskas R, Giovannoni SJ. 2014. Single-cell
720 enabled comparative genomics of a deep ocean SAR11 bathytype. ISME J 8:1440–1451.

721 21. Beale R, Dixon JL, Smyth TJ, Nightingale PD. 2015. Annual study of oxygenated volatile organic compounds in UK shelf
722 waters. Mar Chem 171:96–106.

723 22. Halsey KH, Carter AE, Giovannoni SJ. 2012. Synergistic metabolism of a broad range of C1 compounds in the marine
724 methylotrophic bacterium HTCC2181. Environ Microbiol 14:630–640.

725 23. Rappé MS, Kemp PF, Giovannoni SJ. 1997. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes
726 cloned from the continental shelf off Cape Hatteras, North Carolina. Limnol Oceanogr 42:811–826.

727 24. Yang M, Xia Q, Du S, Zhang Z, Qin F, Zhao Y. 2021. Genomic Characterization and Distribution Pattern of a Novel
728 Marine OM43 Phage. Front Microbiol 12:657.

729 25. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. ISME
730 J 8:1553–1565.

731 26. Turner D, Kropinski AM, Adriaenssens EM. 2021. A Roadmap for Genome-Based Phage Taxonomy. Viruses 13.

732 27. Buchholz HH, Michelsen M, Parsons RJ, Bates NR, Temperton B. 2021. Draft Genome Sequences of Pelagimyophage
733 Mosig EXVC030M and Pelagipodophage Lederberg EXVC029P, Isolated from Devil's Hole, Bermuda. Microbiol Resour
734 Announc 10.

735 28. Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M. 2020. Metagenome Mining Reveals Hidden Genomic Diversity
736 of Pelagimyophages in Aquatic Environments. mSystems 5:e00905–19.

737 29. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, Sullivan MB, Temperton B. 2019. Long-
738 read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic
739 islands. PeerJ 7:e6800.

740 30. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics: an in silico evaluation of
741 metagenome-enabled estimates of viral community composition and diversity. PeerJ 5:e3817.

742 31. Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. 2013. Evidence for a persistent microbial seed bank
743 throughout the global ocean. Proc Natl Acad Sci U S A 110:4651–4655.

744 32. Deng L, Gregory A, Yilmaz S, Poulos BT, Hugenholtz P, Sullivan MB. 2012. Contrasting life strategies of viruses that
745 infect photo- and heterotrophic bacteria, as revealed by viral tagging. MBio 3.

746 33. Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, Sullivan MB. 2014. Viral tagging reveals
747 discrete populations in Synechococcus viral genome sequence space. Nature 513:242–245.

748 34. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol 13:278–284.

749 35. Brum JR, Ignacio-Espinoza JC, Kim E-H, Trubl G, Jones RM, Roux S, VerBerkmoes NC, Rich VI, Sullivan MB. 2016.
750 Illuminating structural proteins in viral "dark matter" with metaproteomics. Proc Natl Acad Sci U S A 113:2436–2441.

751 36. Mahichi F, Synnott AJ, Yamamichi K, Osada T, Tanji Y. 2009. Site-specific recombination of T2 phage using IP008 long
752 tail fiber genes provides a targeted method for expanding host range while retaining lytic activity. FEMS Microbiol Lett
753 295:211–217.

754 37. Casey A, Jordan K, Neve H, Coffey A, McAuliffe O. 2015. A tail of two phages: genomic and functional analysis of
755 Listeria monocytogenes phages vB_LmoS_188 and vB_LmoS_293 reveal the receptor-binding proteins involved in host
756 specificity. Front Microbiol 6:1107.

757 38. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR,
758 Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW. 2010. Genomic
759 analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and
760 environments. Environ Microbiol 12:3035–3056.

761 39. Bryan MJ, Burroughs NJ, Spence EM, Clokie MRJ, Mann NH, Bryan SJ. 2008. Evidence for the intense exchange of
762 MazG in marine cyanophages by horizontal gene transfer. PLoS One 3:e2048.

763 40. Morris RM, Longnecker K, Giovannoni SJ. 2006. Pirellula and OM43 are among the dominant lineages identified in an
764 Oregon coast diatom bloom. Environ Microbiol 8:1361–1370.

765 41. Rihtman B, Bowman-Grahl S, Millard A, Corrigan RM, Clokie MRJ, Scanlan DJ. 2019. Cyanophage MazG is a
766 pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. Environ Microbiol Rep 11:448–455.

767 42. Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D. 2013. Bacteriophage orphan DNA methyltransferases:
768 insights from their bacterial origin, function, and occurrence. Appl Environ Microbiol 79:7547–7555.

769 43. Kossykh VG, Schlagman SL, Hattman S. 1995. Phage T4 DNA [N]-adenine6Methyltransferase. OVEREXPRESSION,
770 PURIFICATION, AND CHARACTERIZATION ∗. J Biol Chem 270:14389–14393.

771 44. Jordan A, Pontis E, ATTAt M, Krook M, GIBERTt I, BARBtt J, Reichard P. 1994. A second class I ribonucleotide reductase
772 in Enterobacteriaceae: Characterization of the Salmonella typhimurium enzyme. Proc Natl Acad Sci U S A 91:12892–

12896.

45. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, Gillet R, Forterre P, Krupovic M. 2019. Numerous cultivated and uncultivated viruses encode ribosomal proteins. Nat Commun 10:752.

46. Lewis LO, Yousten AA. 1988. Bacteriophage attachment to the S-layer proteins of the mosquito-pathogenic strains ofBacillus sphaericus. Curr Microbiol 17:55–60.

47. Plaut RD, Beaber JW, Zemansky J, Kaur AP, George M, Biswas B, Henry M, Bishop-Lilly KA, Mokashi V, Hannah RM, Pope RK, Read TD, Stibitz S, Calendar R, Sozhamannan S. 2014. Genetic evidence for the involvement of the S-layer protein gene sap and the sporulation genes spo0A, spo0B, and spo0F in Phage AP50c infection of Bacillus anthracis. J Bacteriol 196:1143–1154.

48. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. ISME J 10:2854–2866.

49. Yang H, Ma Y, Wang Y, Yang H, Shen W, Chen X. 2014. Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. Bioengineered 5:300–304.

50. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. 2003. Bacteriophage T4 genome. Microbiol Mol Biol Rev 67:86–156, table of contents.

51. Hinton DM. 2010. Transcriptional control in the prereplicative phase of T4 development. Virol J 7:289.

52. Doron S, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, Feingersch R, Steglich C, Futschik M, Lindell D, Sorek R. 2016. Transcriptome dynamics of a broad host-range cyanophage and its hosts. ISME J 10:1437–1455.

53. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245.

54. Twist K-AF, Campbell EA, Deighan P, Nechaev S, Jain V, Geiduschek EP, Hochschild A, Darst SA. 2011. Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the β-subunit flap domain of Escherichia coli RNA polymerase. Proc Natl Acad Sci U S A 108:19961–19966.

55. Nechaev S, Kamali-Moghaddam M, André E, Léonetti J-P, Geiduschek EP. 2004. The bacteriophage T4 late-transcription coactivator gp33 binds the flap domain of Escherichia coli RNA polymerase. Proc Natl Acad Sci U S A 101:17365–17370.

56. Geiduschek EP, Kassavetis GA. 2010. Transcription of the T4 late genes. Virol J 7:288.

57. Withey JH, Friedman DI. 2003. A salvage pathway for protein structures: tmRNA and trans-translation. Annu Rev Microbiol 57:101–123.

58. Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol 11:R31.

59. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16.

60. Berube PM, Biller SJ, Hackl T, Hogle SL, Satinsky BM, Becker JW, Braakman R, Collins SB, Kelly L, Berta-Thompson J, Coe A, Bergauer K, Bouman HA, Browning TJ, De Corte D, Hassler C, Hulata Y, Jacquot JE, Maas EW, Reinthaler T, Sintes E, Yokokawa T, Lindell D, Stepanauskas R, Chisholm SW. 2018. Single cell genomes of Prochlorococcus, Synechococcus, and sympatric microbes from diverse marine environments. Sci Data 5:180154.

61. Klähn S, Bolay P, Wright PR, Atilho RM, Brewer KI, Hagemann M, Breaker RR, Hess WR. 2018. A glutamine riboswitch is a key element for the regulation of glutamine synthetase in cyanobacteria. Nucleic Acids Res 46:10082–10094.

62. Waldbauer JR, Coleman ML, Rizzo AI, Campbell KL, Lotus J, Zhang L. 2019. Nitrogen sourcing during viral infection of marine cyanobacteria. Proc Natl Acad Sci U S A 116:15590–15595.

63. Barnhart MM, Chapman MR. 2006. Curli biogenesis and function. Annu Rev Microbiol 60:131–147.

64. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303.

65. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature https://doi.org/10.1038/s41586-021-03819-2.

66. Pang T, Savva CG, Fleming KG, Struck DK, Young R. 2009. Structure of the lethal phage pinhole. Proc Natl Acad Sci U S A 106:18966–18971.

67. Conners R, Mc Laren M, Łapińska U, Sanders K, Rhia L. M, Blaskovich MAT, Pagliara S, Daum B, Rakonjac J, Gold VAM.

827 2021. CryoEM structure of the outer membrane secretin channel pIV from the f1 filamentous bacteriophage. bioRxiv
828 https://doi.org/10.1101/2021.07.20.453082.

829 68. Ghosh S, Shaw R, Sarkar A, Gupta SKD. 2020. Evidence of positive regulation of mycobacteriophage D29 early gene
830 expression obtained from an investigation using a temperature-sensitive mutant of the phage. FEMS Microbiol Lett
831 367.

832 69. Catalão MJ, Gil F, Moniz-Pereira J, São-José C, Pimentel M. 2013. Diversity in bacterial lysis systems: bacteriophages
833 show the way. FEMS Microbiol Rev 37:554–571.

834 70. Rajaure M, Berry J, Kongari R, Cahill J, Young R. 2015. Membrane fusion during phage lysis. Proc Natl Acad Sci U S A
835 112:5497–5502.

836 71. Del Giudice MG, Ugalde JE, Czibener C. 2013. A lysozyme-like protein in Brucella abortus is involved in the early stages
837 of intracellular replication. Infect Immun 81:956–964.

838 72. Hyun Y, Baek Y, Lee C, Ki N, Ahn J, Ryu S, Ha N-C. 2021. Structure and Function of the Autolysin SagA in the Type IV
839 Secretion System of Brucella abortus. Mol Cells 44:517–528.

840 73. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC,
841 Lee J, Szczypinski B, Fraser CM. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40:
842 comparative genomics of a T4-related bacteriophage. J Bacteriol 185:5220–5233.

843 74. Frias MJ, Melo-Cristino J, Ramirez M. 2009. The autolysin LytA contributes to efficient bacteriophage progeny release
844 in Streptococcus pneumoniae. Journal of Bacteriology 191:5428–5440.

845 75. Burrowes BH, Molineux IJ, Fralick JA. 2019. Directed in Vitro Evolution of Therapeutic Bacteriophages: The Appelmans
846 Protocol. Viruses 11.

847 76. Tétart F, Repoila F, Monod C, Krisch HM. 1996. Bacteriophage T4 host range is expanded by duplications of a small
848 domain of the tail fiber adhesin. J Mol Biol 258:726–731.

849 77. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized
850 bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004.

851 78. Keiler KC, Shapiro L, Williams KP. 2000. tmRNAs that encode proteolysis-inducing tags are found in all known bacterial
852 genomes: A two-piece tmRNA functions in Caulobacter. Proc Natl Acad Sci U S A 97:7778–7783.

853 79. Sharkady SM, Williams KP. 2004. A third lineage with two-piece tmRNA. Nucleic Acids Res 32:4531–4538.

80. Yang JY, Fang W, Miranda-Sanchez F, Brown JM, Kauffman KM, Acevero CM, Bartel DP, Polz MF, Kelly L. 2021. Degradation of host translational machinery drives tRNA acquisition in viruses. Cell Syst https://doi.org/10.1016/j.cels.2021.05.019.

81. Santos SB, Kropinski AM, Ceyssens P-J, Ackermann H-W, Villegas A, Lavigne R, Krylov VN, Carvalho CM, Ferreira EC, Azeredo J. 2011. Genomic and Proteomic Characterization of the Broad-Host-Range Salmonella Phage PVP-SE1: Creation of a New Phage Genus. J Virol 85:11265.

82. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. FEMS Microbiol Rev 40:258–272.

83. Coutinho FH, Zaragoza-Solas A, López-Pérez M, Barylski J, Zielezinski A, Dutilh BE, Edwards R, Rodriguez-Valera F. 2021. RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. Patterns Prejudice 100274.

84. Riede I. 1986. T-even type phages can change their host range by recombination with gene (tail fibre) or gene (head). Mol Gen Genet 205:160–163.

85. de Jonge PA, Nobrega FL, Brouns SJJ, Dutilh BE. 2019. Molecular and Evolutionary Determinants of Bacteriophage Host Range. Trends Microbiol 27:51–63.

86. Giovannoni SJ. 2017. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. Ann Rev Mar Sci 9:231–255.

87. Bolduc B, Jang HB, Doulcier G, You Z-Q, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ 5:e3243.

88. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol 37:632–639.

89. Carini P, Steindler L, Beszteri S, Giovannoni SJ. 2013. Nutrient requirements for growth of the extreme oligotroph "Candidatus Pelagibacter ubique" HTCC1062 on a defined medium. ISME J 7:592–602.

90. Solonenko N. 2016. Isolation of DNA from phage lysate. protocols.io https://doi.org/10.17504/protocols.io.c36yrd.

91. Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read merging via overlap. PLoS One 12:e0185056.

92. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD,

Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.

93. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ 3:e985.

94. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2020. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol https://doi.org/10.1038/s41587-020-00774-7.

95. McNair K, Zhou C, Dinsdale EA, Souza B, Edwards RA. 2019. PHANOTATE: a novel approach to gene identification in phage genomes. Bioinformatics 35:4537–4542.

96. Salisbury A, Tsourkas PK. 2019. A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation. Int J Mol Sci 20:3391.

97. Lomsadze A, Gemayel K, Tang S, Borodovsky M. 2018. Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. Genome Res 28:1079–1089.

98. Besemer J, Borodovsky M. 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res 27:3911–3920.

99. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

100. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

101. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–5.

102. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–30.

103. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48.

104. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res 44:W54–7.

907 105. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways,
908 diseases and drugs. Nucleic Acids Res 45:D353–D361.

909 106. Singh P, Bandyopadhyay P, Bhattacharya S, Krishnamachari A, Sengupta S. 2009. Riboswitch detection using profile
910 hidden Markov models. BMC Bioinformatics 10:325.

911 107. Mukherjee S, Sengupta S. 2016. Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in
912 genomic sequences. Bioinformatics 32:776–778.

913 108. Solovyev V, Salamov A. 2011. AUTOMATIC ANNOTATION OF MICROBIAL GENOMES AND METAGENOMIC SEQUENCES,
914 p. . *In* Li, RW (ed.), Metagenomics and its Application in Agriculture, Biomedicine and Environmental Studies. Nova
915 Science Publishers.

916 109. Moon K, Kang I, Kim S, Kim S-J, Cho J-C. 2017. Genome characteristics and environmental distribution of the first
917 phage that infects the LD28 clade, a freshwater methylotrophic bacterial group. Environ Microbiol 19:4714–4727.

918 110. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a
919 software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.

920 111. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

921 112. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, Claverie J-
922 M, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–9.

923 113. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res
924 32:1792–1797.

925 114. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks
926 from protein sequence alignments. Syst Biol 56:564–577.

927 115. Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful
928 alternative. Syst Biol 55:539–552.

929 116. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum
930 likelihood. Syst Biol 52:696–704.

931 117. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

932 118. Zhang Z, Chen F, Chu X, Zhang H, Luo H, Qin F, Zhai Z, Yang M, Sun J, Zhao Y. 2019. Diverse, Abundant, and Novel

933 Viruses Infecting the Marine Roseobacter RCA Lineage. mSystems 4.

934 119. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One
935 11:e0163962.

936
937
938
939
940
941
942
943
944
945
946
947
948
949

950 Table 1. General features of OM43 strain H5P1 and four Melnitz phages isolated in this study.

951 Figure 1. Viral shared gene-content network of OM43 phages, related bacteriophages from NCBI and related
952 sequences from the Global Ocean Virome (GOV2.0). Nodes represent viral genomes; edges represent the
953 similarity between phages based on shared gene content. NCBI reference genomes more than two neighbouring
954 edges removed were excluded for clarity. Phage isolates are indicated with red arrows. Coloured circles
955 represent genomes and virome contigs within the same cluster as OM43 phage isolate genomes. Nodes shared
956 between Cluster_2 (light pink) and Cluster_3 (purple) are highlighted in hot pink.

957 Figure 2. Phylogenetic tree of metagenomic contigs and marine Melnitz-type myophages. Neighbour-joining tree
958 (500 bootstraps) based on four individually aligned and concatenated structural genes (capsid assembly, major
959 capsid, sheath subtilisin, terminase large subunit) of genomes and contigs that were clustered with OM43 phage
960 Melnitz, with the exception of phages infecting *Synechococcus* spp. that were used to root the tree. All four
961 Melnitz-like isolates were included but the branch was collapsed for clarity. Branch support values of 1 are not
962 shown. Leaves without labels indicate contigs from the Global Ocean Virome (GOV2) dataset (12), which were
963 omitted for clarity.

964 Figure 3. Gene map showing identified genomic features of the OM43 phage Melnitz. **A:** Gene map of the
965 141,548 bp genome of Melnitz contains 143 hypothetical ORFs (62 %) without known function (indicated dark
966 grey); **B:** Section of the Melnitz genome between two terminator sequences that contains the *ssrA* gene,
967 glutamine riboswitch and transcription co-activator gene for transcription-translation regulation.

968 Figure 4. Structural prediction of CsgGF complex encoded by Melnitz. **A:** Predicted structure of CsgGF complex
969 in *E. coli* (PDB model 6l7a) comprises a hetero-18-mer with 1:1 stoichiometry of CsgG (pink) and CsgF (green),
970 forming a pore in the outer membrane with two constrictions - one provided by CsgG at the base of the barrel
971 and one provided by CsgF at the neck of the barrel. **B:** Structural prediction of Melnitz encoded CsgG using
972 AlphaFold2 (teal) showed structural conservation with CsgG from *E. coli* in the periplasmic α-helices and β-barrel
973 structure. **C:** Expanded view of the structural alignment of Melnitz CsgG with *E. coli* CsgG shows a putative

narrowing of the pore at the top of the barrel, matching the pore diameter at the top of the barrel in CsgGF in E. coli. **D:** Alignment of predicted structure of Melnitz-encoded CsgF (teal) to that of CsgF in *E. coli* (pink) showed low structural similarity, with Melnitz-encoded CsgF comprising two alpha-helices and a beta sheet. Alignment indicated that the additional structures of Melnitz-encoded CsgF extend out of the CsgG pore, with unknown function.

Figure 5. Phylogeny of tmRNA genes in major marine lineages. Neighbour-joining tree (100 bootstraps) of tmRNA genes found in marine phages and host lineages (not exhaustive) suggests that three the three known major lineages between Cyanobacteria, Gammaproteobacteria and Alphaproteobacteria are shared with their associated phages, except for OM43 phage Melnitz (infecting H5P1 on the Gammaproteobacterial branch) which has a tmRNA gene more closely related to genes found in Alphaproteobacteria and their phages.

Figure 6. Alignment of tRNA genes found in Melnitz, SAR11 and OM43 lineages. Heatmaps and dendrograms (Euclidian similarity matrices) based on similarity between alignments for arginine (Arg), leucine (Leu) and tryptophan (Trp) tRNA genes found in OM43 phage Melnitz, OM43 and SAR11 as well as tRNA derived from isolated SAR11 and OM43 phages.

Supplementary Figure 1. TEM images of phage Melnitz virions. Bar indicates 100 nm.

Supplementary Figure 2. Neighbour-joining tree (100 bootstraps) of a tail sheath encoding gene found in Melnitz related myophages. Branch support values <1 are indicated by circle colours on the tree. The cyanophage branch was used to root the tree.

Supplementary Figure 3. Neighbour-joining tree (100 bootstraps) of the *TerL* gene found in Melnitz related myophages. Branch support values <1 are indicated by circle colours on the tree. The cyanophage branch was used to root the tree.

Supplementary Figure 4. Neighbour-joining tree (100 bootstraps) of a capsid scaffolding gene found in Melnitz related myophages. Branch support values <1 are indicated by circle colours on the tree. The cyanophage branch was used to root the tree.

Supplementary Figure 5. Global abundance of Melnitz. Reads recruited per kilobase of contigs per million reads (RPKM) of GOV2 viromes against phages infecting OM43 and LD28 as well as known *Pelagibacter* myophages. Samples are organized by ecological zone: Arctic, TT-EPI temperate-tropical epipelagic, TT-MES temperate-tropical mesopelagic, ANT Antarctic.

Supplementary Figure 6. Cross sectional structure of CsgGF. **A** Clipped predicted surface model of internal structure of CsgGF in *E. coli* showing CsgG (teal) and CsgF (pink) showing the internal channel structure. This structure comprises a series of narrowing pores. **B** with CsgF creating a pore 14.8Å in diameter and CsgG creating a smallest pore 12.9Å in diameter. **C** Clipped predicted surface model of internal structure of Melnitz-encoded CsgG comprises two channels of similar size to those of the *E. coli* CsgGF complex.

Supplementary Figure 7. Structural comparison between **A** CsgG encoded by Melnitz (teal) and HTVC008M (pink); **B** CsgF encoded by Melnitz (teal) and HTVC008M (pink).

Supplementary Figure 8. **A** Predicted structure of Melnitz CsgF in an inverted orientation compared to **B** structure of secretin PilQ (6W6M) from *Vibrio cholerae*. Left to right - cartoon model; electrostatic potential (red = -ve, blue=+ve); top-down view through the barrel.

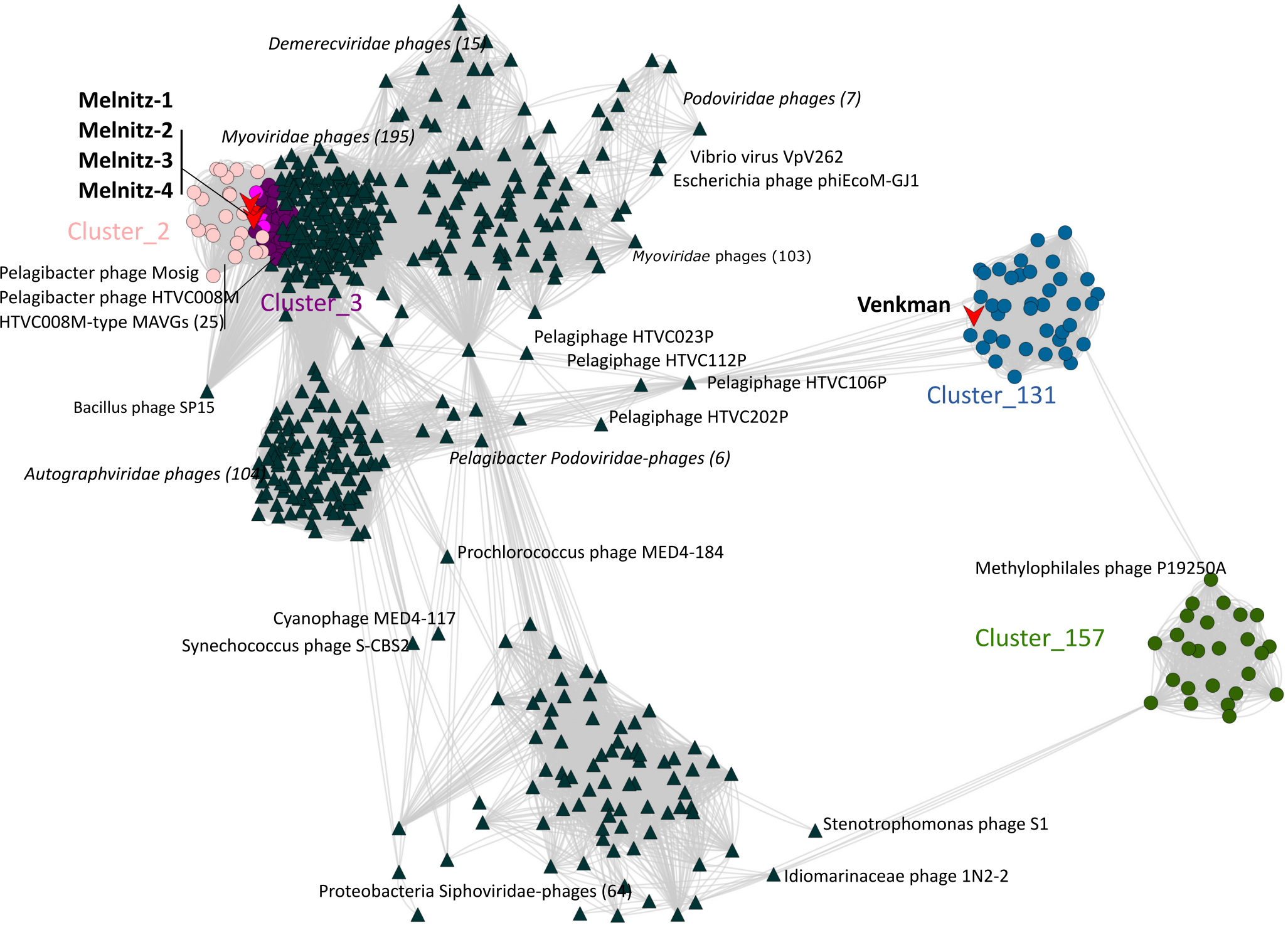
Supplementary Figure 9. SwissModel alignment of putative endolysin structure encoded by Melnitz gp67 (hot pink), with a SagA autolysin encoded by Brucella abortus (PDB model 7dnp.1, grey). The conserved peptidoglycan binding site encoded by Glu17, Asp26 and Thr31 is highlighted (teal), with the peptidoglycan substrate (yellow stick representation).

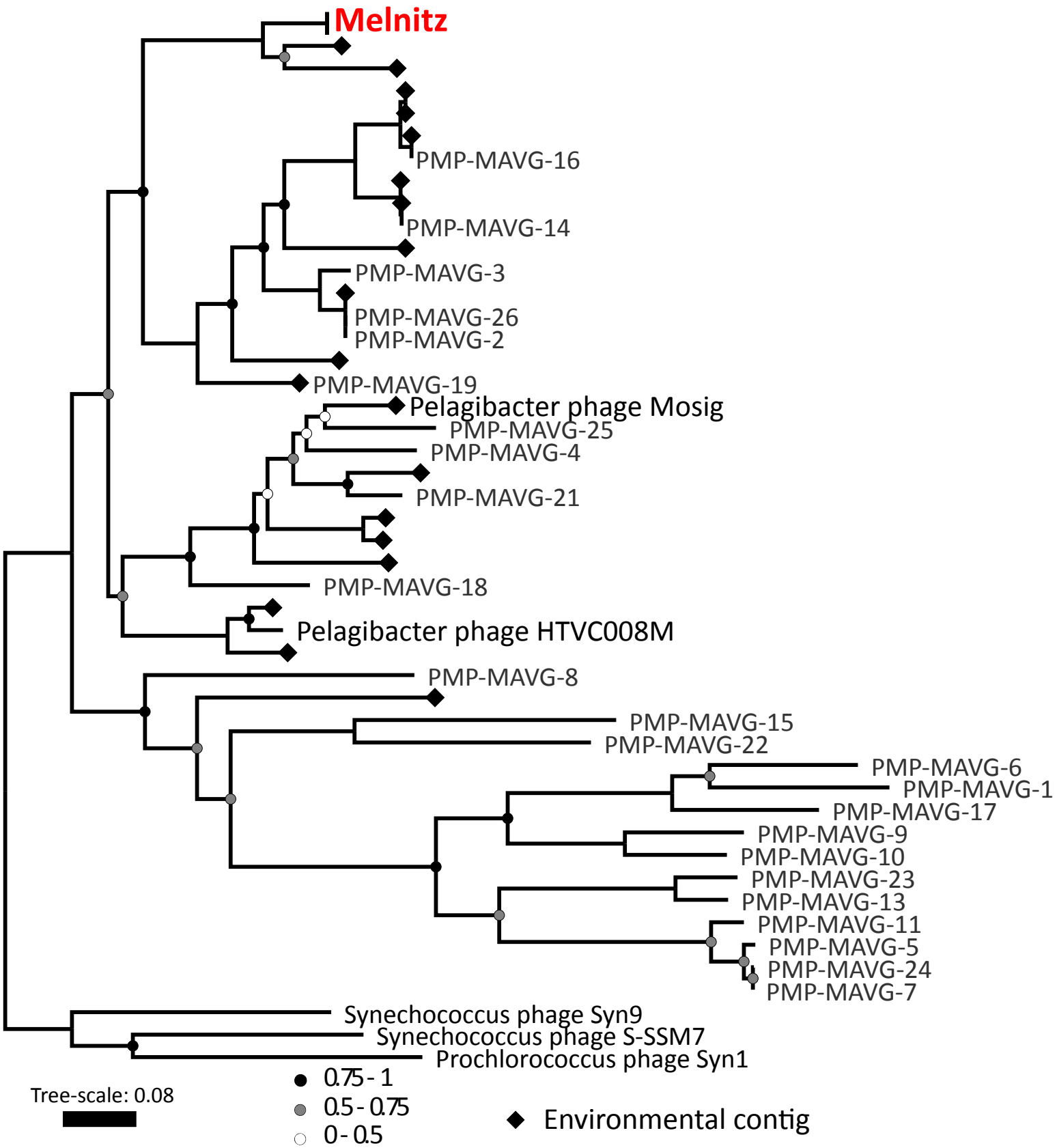
1017 Supplementary Figure 10. Host range assessment of Melnitz. **A** Growth curve of SAR11 strains HTCC1062 and
1018 HTCC7211, treated with OM43 phage Melnitz and SAR11 phage Mosig. **B** Growth curves of OM43 strains C6P1,
1019 D12P1 and H5P1 treated with OM43 phage Melnitz and SAR11 phage Mosig

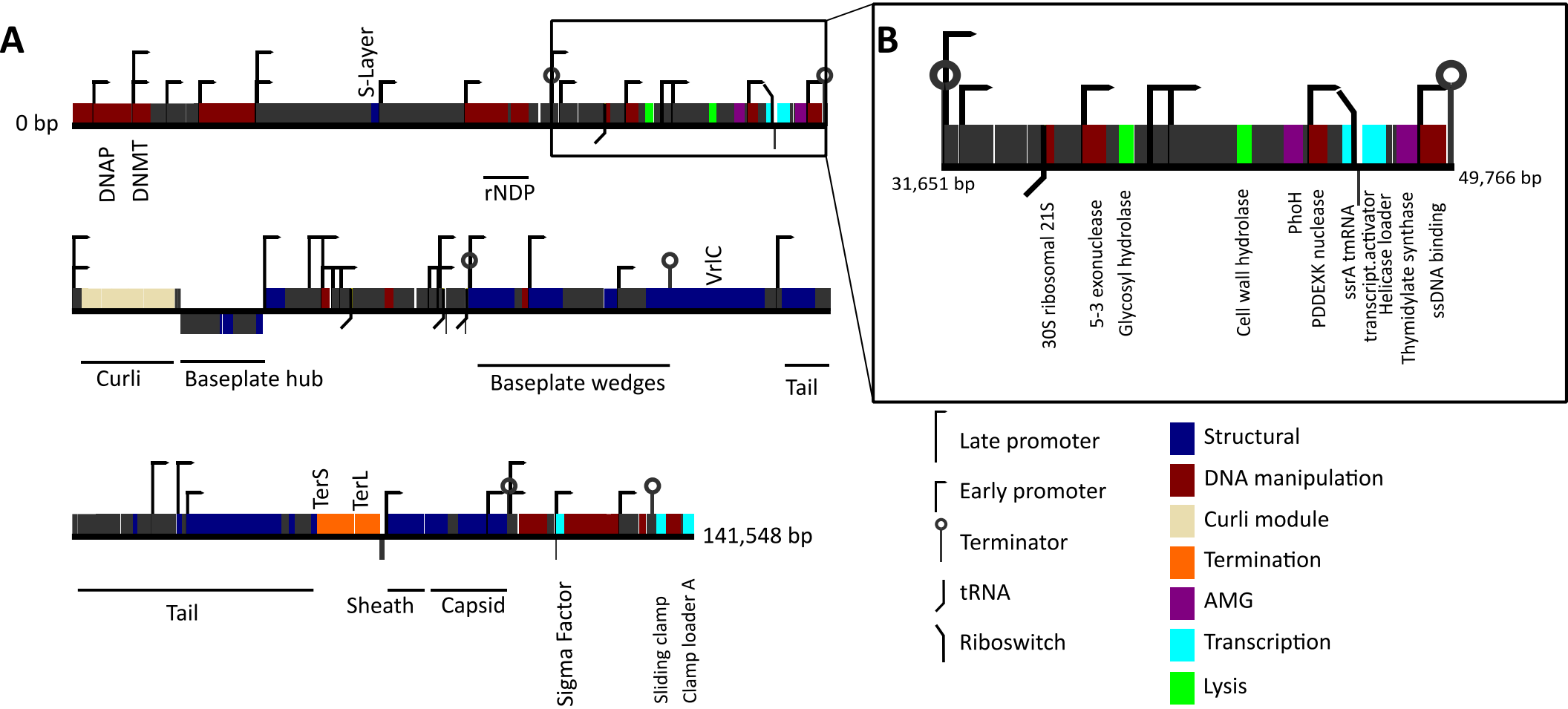
1020

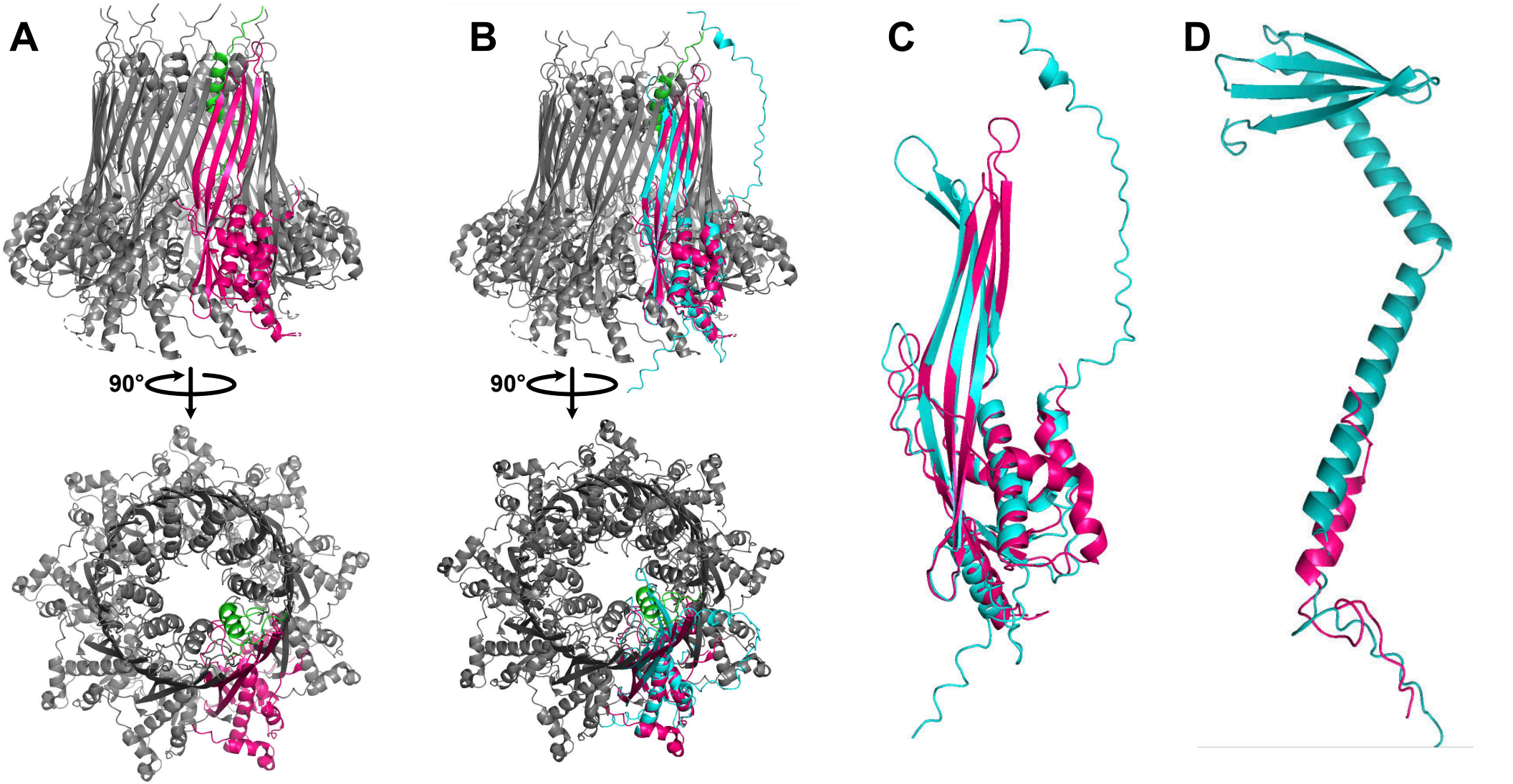
1021

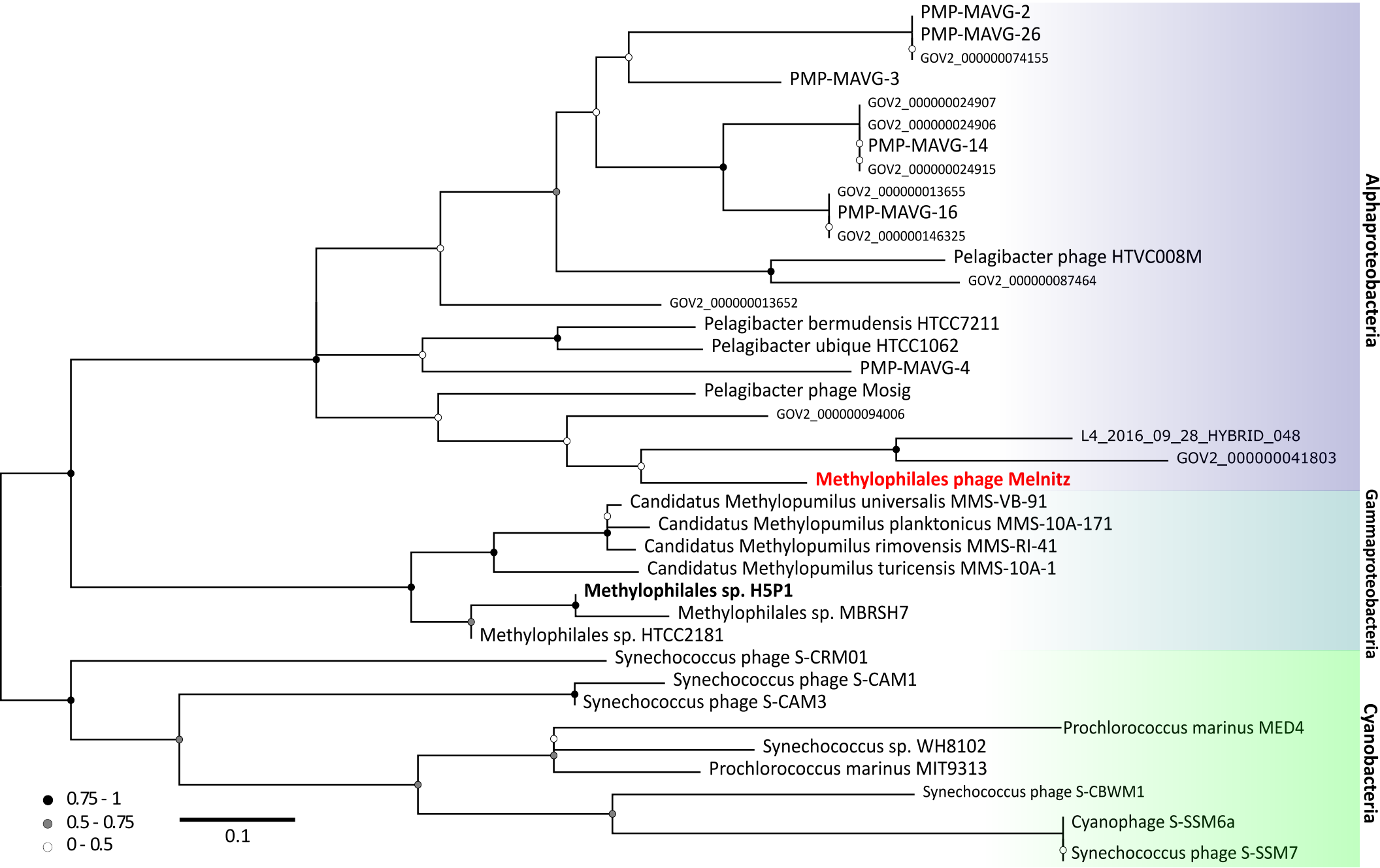
Phage	Culture ID	Host	Phage group	Morphotype	Genome size (bp)	G + C %	ORFs	tRNAs	Source Water	Sampling Date	Latitude	Longitude
Melnitz	EXVC044M	H5P1	Melnitz	Myovirus	141,548	37.6	224	4	BATS	01 June 2019	N31°40'	W64°10'
Melnitz variant 1	EXVC043M	H5P1	Melnitz	Myovirus	141,552	37.6	226	4	BATS	01 June 2019	N31°40'	W64°10'
Melnitz variant 2	EXVC040M	H5P1	Melnitz	Myovirus	141,548	37.6	224	4	WEC	22 July 2019	N50°15'	W04°13'
Melnitz variant 3	EXVC039M	H5P1	Melnitz	Myovirus	141,372	37.6	224	4	WEC	01 April 2019	N50°15'	W04°13'
		H5P1	Methylophilales sp.	Gammaproteobacteria	1,336,408	34.4	1382	38	WEC	Sep-18	N50°15'	W04°13'

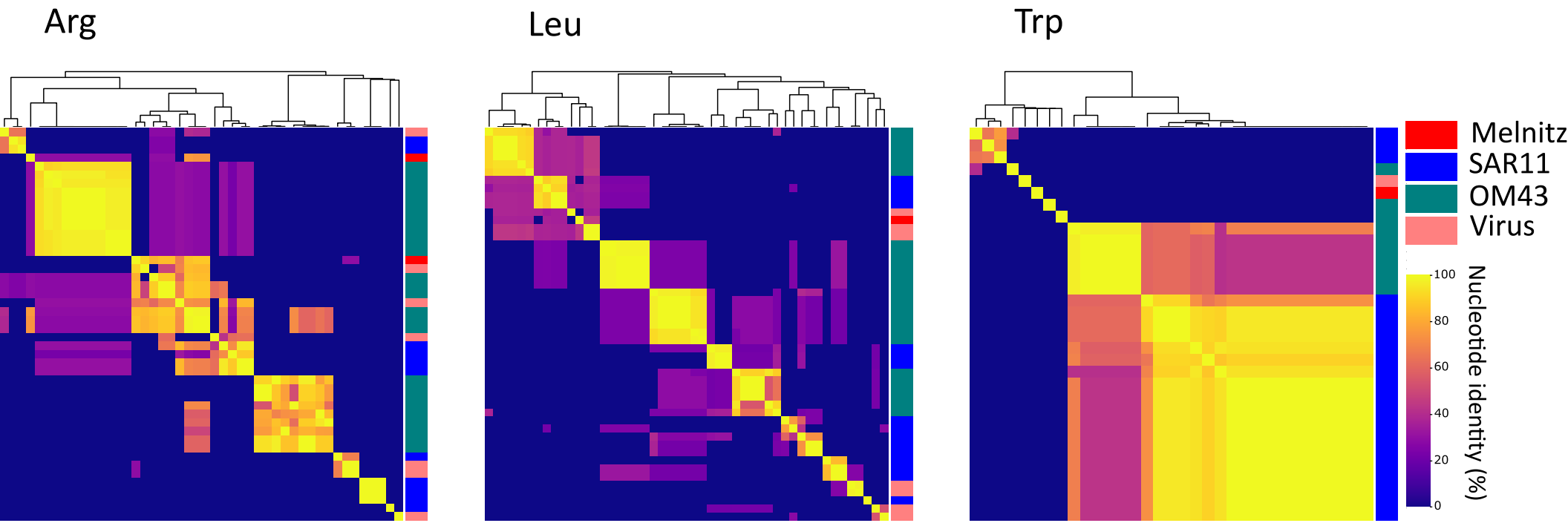

Demerecviridae phages (15)
Podoviridae phages (7)
Melnitz-1
Melnitz-2
Melnitz-3
Melnitz-4
Myoviridae phages (195)
Vibrio virus VpV262
Escherichia phage phiEcoM-GJ1
Cluster_2
Myoviridae phages (103)
Pelagibacter phage Mosig
Pelagibacter phage HTVC008M
HTVC008M-type MAVGs (25)
Cluster_3
Venkman
Pelagiphage HTVC023P
Pelagiphage HTVC112P
Pelagiphage HTVC106P
Cluster_131
Bacillus phage SP15
Pelagiphage HTVC202P
Pelagibacter Podoviridae-phages (6)
Autographviridae phages (104)
Prochlorococcus phage MED4-184
Methylophilales phage P19250A
Cyanophage MED4-117
Synechococcus phage S-CBS2
Cluster_157
Stenotrophomonas phage S1
Idiomarinaceae phage 1N2-2
Proteobacteria Siphoviridae-phages (64)


Melnitz
PMP-MAVG-16
PMP-MAVG-14
PMP-MAVG-3
PMP-MAVG-26
PMP-MAVG-2
PMP-MAVG-19
Pelagibacter phage Mosig
PMP-MAVG-25
PMP-MAVG-4
PMP-MAVG-21
PMP-MAVG-18
Pelagibacter phage HTVC008M
PMP-MAVG-8
PMP-MAVG-15
PMP-MAVG-22
PMP-MAVG-6
PMP-MAVG-1
PMP-MAVG-17
PMP-MAVG-9
PMP-MAVG-10
PMP-MAVG-23
PMP-MAVG-13
PMP-MAVG-11
PMP-MAVG-5
PMP-MAVG-24
PMP-MAVG-7
Synechococcus phage Syn9
Synechococcus phage S-SSM7
Prochlorococcus phage Syn1
Tree-scale: 0.08
0.75 - 1
0.5 - 0.75
0 - 0.5
Environmental contig


A
0 bp
S-Layer
DNAP
DNMT
rNDP
VrlC
Curli
Baseplate hub
Baseplate wedges
Tail
TerS
TerL
141,548 bp
Tail
Sheath
Capsid
Sigma Factor
Sliding clamp
Clamp loader A
B
31,651 bp
49,766 bp
30S ribosomal 21S
5-3 exonuclease
Glycosyl hydrolase
Cell wall hydrolase
PhoH
PDDEXK nuclease
ssrA tmRNA
transcript.activator
Helicase loader
Thymidylate synthase
ssDNA binding
Late promoter
Early promoter
Terminator
tRNA
Riboswitch
Structural
DNA manipulation
Curli module
Termination
AMG
Transcription
Lysis


A
90°
B
90°
C
D


PMP-MAVG-2
PMP-MAVG-26
GOV2_000000074155
PMP-MAVG-3
GOV2_000000024907
GOV2_000000024906
PMP-MAVG-14
GOV2_000000024915
GOV2_000000013655
PMP-MAVG-16
GOV2_000000146325
Pelagibacter phage HTVC008M
GOV2_000000087464
GOV2_000000013652
Pelagibacter bermudensis HTCC7211
Pelagibacter ubique HTCC1062
PMP-MAVG-4
Pelagibacter phage Mosig
GOV2_000000094006
L4_2016_09_28_HYBRID_048
GOV2_000000041803
Methylophilales phage Melnitz
Candidatus Methylopumilus universalis MMS-VB-91
Candidatus Methylopumilus planktonicus MMS-10A-171
Candidatus Methylopumilus rimovensis MMS-RI-41
Candidatus Methylopumilus turicensis MMS-10A-1
Methylophilales sp. H5P1
Methylophilales sp. MBRSH7
Methylophilales sp. HTCC2181
Synechococcus phage S-CRM01
Synechococcus phage S-CAM1
Synechococcus phage S-CAM3
Prochlorococcus marinus MED4
Synechococcus sp. WH8102
Prochlorococcus marinus MIT9313
Synechococcus phage S-CBWM1
Cyanophage S-SSM6a
Synechococcus phage S-SSM7
Alphaproteobacteria
Gammaproteobacteria
Cyanobacteria
0.75 - 1
0.5 - 0.75
0 - 0.5
0.1


Arg
Leu
Trp
Melnitz
SAR11
OM43
Virus
Nucleotide identity (%)
100
80
60
40
20
0

Appendix 4. Warwick-Dugdale, J., **Buchholz, H.H.**, Allen, M.J., Temperton, B. (2019) Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virol J* **16,** 15. https://doi.org/10.1186/s12985-019-1120-1

Warwick-Dugdale et al. Virology Journal (2019) 16:15
https://doi.org/10.1186/s12985-019-1120-1

Virology Journal

REVIEW Open Access

Host-hijacking and planktonic piracy: how phages command the microbial high seas

Joanna Warwick-Dugdale[1,2], Holger H. Buchholz[2], Michael J. Allen[1,2] and Ben Temperton[2*]

Abstract

Microbial communities living in the oceans are major drivers of global biogeochemical cycles. With nutrients limited across vast swathes of the ocean, marine microbes eke out a living under constant assault from predatory viruses. Viral concentrations exceed those of their bacterial prey by an order of magnitude in surface water, making these obligate parasites the most abundant biological entities in the ocean. Like the pirates of the 17th and 18th centuries that hounded ships plying major trade and exploration routes, viruses have evolved mechanisms to hijack microbial cells and repurpose their cargo and indeed the vessels themselves to maximise viral propagation. Phenotypic reconfiguration of the host is often achieved through Auxiliary Metabolic Genes – genes originally derived from host genomes but maintained and adapted in viral genomes to redirect energy and substrates towards viral synthesis. In this review, we critically evaluate the literature describing the mechanisms used by bacteriophages to reconfigure host metabolism and to plunder intracellular resources to optimise viral production. We also highlight the mechanisms used when, in challenging environments, a 'batten down the hatches' strategy supersedes that of 'plunder and pillage'. Here, the infecting virus increases host fitness through phenotypic augmentation in order to ride out the metaphorical storm, with a concomitant impact on host substrate uptake and metabolism, and ultimately, their interactions with their wider microbial community. Thus, the traditional view of the virus-host relationship as predator and prey does not fully characterise the variety or significance of the interactions observed. Recent advances in viral metagenomics have provided a tantalising glimpse of novel mechanisms of viral metabolic reprogramming in global oceans. Incorporation of these new findings into global biogeochemical models requires experimental evidence from model systems and major improvements in our ability to accurately predict protein function from sequence data.

Keywords: AMGs, marine, Cyanophage, Nucleotide scavenging, Biogeochemical cycling, Host-virus interactions, Lysogeny

Background

Based on their extraordinary abundance and diversity, J.B.S. Haldane once quipped that '*The Creator would appear as endowed with a passion for stars, on the one hand, and for beetles on the other*' [1]. In comparison, the Creator's zeal for viruses would make stars and beetles appear to be a side-project performed with perfunctory indifference. It is estimated that there are a million viruses in the ocean for every star in the universe [2]. Assuming an average size of 100 nm in length, placed end-to-end marine viruses would stretch to our nearest neighbour star (Proxima Centauri, 4.22 light years away) and back. The vast majority of these viruses are obligate parasites of marine bacteria - the primary drivers of global carbon biogeochemistry [3]. With 10^{28} infections occurring per day in the oceans [4], these bacteriophages are responsible for up to 50% of bacterial mortality [5] and the daily release of 12.4 μg of organic carbon per litre of seawater, or an estimated 10 billion tons globally per day [5, 6]. Release of cellular substrates through lysis have been shown to stimulate surviving members of the community and increase microbial productivity through nutrient recycling [7–9]. Thus, marine viruses are a major component of global carbon cycling.

Lysis and release of cellular material as dissolved organic matter is only the final step of a complex

* Correspondence: B.Temperton@exeter.ac.uk
[2]University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK
Full list of author information is available at the end of the article

host-virus interaction where the invading pathogen can manipulate host metabolism and alter its phenotype to favour viral replication at the expense of host function. Once a virus has overcome host defences, lytic viral infection typically involves a shut-down of host metabolism, followed by degradation of macromolecules and scavenging of intracellular resources. Virally encoded 'Auxiliary Metabolic Genes' (AMGs) are expressed during infection to augment and redirect energy and resources towards viral production [10–13]. These AMGs are often repurposed versions of host-genes picked up during ancestral infections, evolving separately to improve viral fitness [14, 15]. Conservation of AMGs across phage lineages suggests that not only are such genes critical for viral success, but that viral modulation of marine nutrient cycling via metabolic hijacking is an important, but understudied component of biogeochemical cycles. Furthermore, viral replication through lysis is only beneficial to the virus if upon release, its progeny can infect other susceptible cells. This process is governed by host availability and density-dependent selection. In resource-limited or otherwise challenging environments viral fitness may be better served by maximising host fitness until such time as the conditions are met for a lytic approach to be favourable. Single amplified genomes of bacterial cells from marine environments have shown that ~ 1/3 contain viral sequences [16, 17]. Between 28 and 71% of marine bacterial isolates contain inducible prophage, with greater occurrence of lysogeny associated with low-nutrient environments [18–21]. In extreme environments such as hot-springs, nearly all cells contained a viral signature [22]. In lysogenic cells, the viruses are not passive passengers, but actively promote the host fitness through expression of viral genes increasing metabolic flexibility, enabling antibiotic resistance, toxin production and immunity to similar viruses [23] (reviewed in [24]). The aim of this review is to illustrate the different ways in which bacteriophages alter the function of their hosts, emphasising how marine viruses modify the metabolism of environmentally relevant bacteria and in turn their impact on global carbon biogeochemistry. We also discuss how novel molecular methods are providing greater insight into the breadth and scale of viral hijacking in the global ocean.

Plunder and pillage - redirecting cellular metabolism to maximise viral production

First, we will consider the classic view of viral predation, where the virus replicates by lytic infection and redirects host resources for the purpose of viral production that ends in cell lysis and release of viral progeny (the 'Plunder and Pillage' strategy, Fig. 1a). Once a virus docks with a susceptible cell and successfully injects its

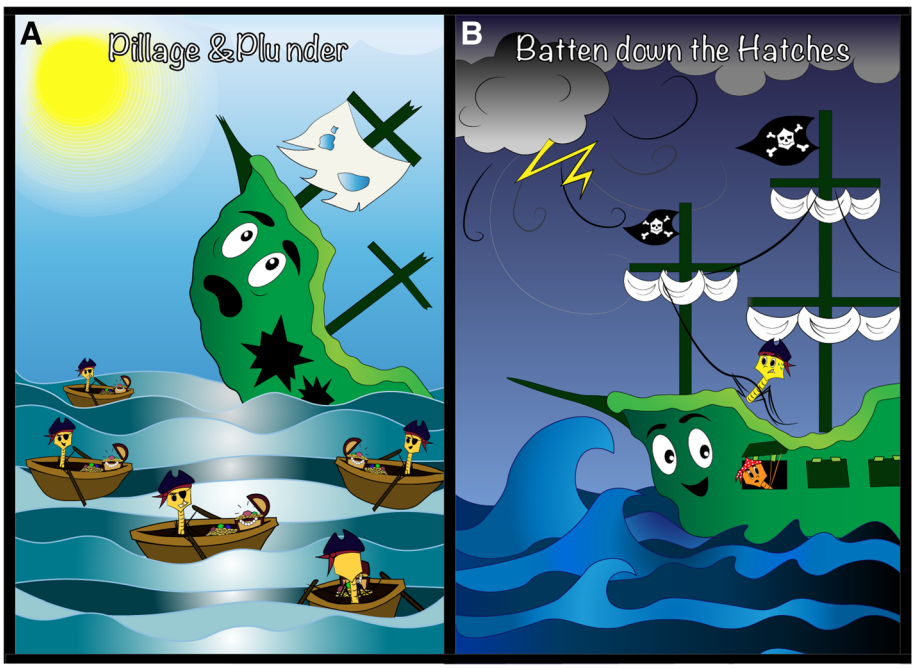


Fig. 1 A cartoonist's depiction of the two types of host-virus interactions in the oceans. Under the 'Pillage and Plunder' model (**a**), the virus infects its host and redirects energy and substrates towards viral replication before lysing the cell and releasing viral progeny for further infections. Under the 'Batten down the hatches' model (**b**), viral fitness is improved by increasing host fitness, either by augmenting metabolic flexibility through virally-encoded genes, increasing resistance against other viruses, or by curbing host metabolism to maximise host survival under nutrient limitation

genome and associated proteins into the host, hijacking of host metabolism is swift and efficient. Infection in marine T4-like cyanophages follows a similar pattern to that observed in T4 infections in *Escherichia coli*, with 3 distinct phases of transcription of viral genes: (1) defence neutralisation and host metabolic rewiring; (2) synthesis of viral components; and (3) viral assembly [25]. These stages are referred to as 'early', 'middle' and 'late' stages, respectively. Examples of expression of early stage genes include the infection of the marine Bacteroidetes *Cellulophaga baltica*. Infection with a specialist phage was initiated with expression of 15 genes putatively involved in overcoming host defences. Of these, one was an anti-restriction defence gene thought to provide viral defence against host degradation of viral DNA. The remaining 14 genes had unknown function but shared a promoter with the anti-restriction defence and were located in the same genomic neighbourhood, inferring a complementary functional role [13]. A study of the transcriptional response to infection by the cyanophage Syn9 in multiple *Synechococcus* hosts showed that within the first 30 min following infection, almost all transcription is associated with phage encoded genes and within the first hour of infection ~ 80% of cellular DNA has been scavenged for resources [25]. Reconfiguration of cellular transcription machinery is often achieved either by altering host RNA polymerases to favour viral promoters using phage factors packaged within the viral capsid and delivered during infection, or by direct delivery of a virally-encoded RNA polymerase [25–28].

Once the cellular machinery has been repurposed for viral production, the 'middle' stage genes are expressed, which include mechanisms to manipulate host metabolism. Termed "Auxiliary Metabolic Genes" (AMGs), many of these viral genes encode for central metabolic proteins and are conserved across different marine phage lineages, signifying their importance to viral success [29]. Two different classes of AMG have been described (reviewed in [30]): to be categorised as Class I, a gene must encode a protein with a central metabolic function, for example in photosynthesis, carbohydrate metabolism or amino acid metabolism, and it will appear in a Kyoto Encyclopaedia of Genes and Genomes (KEGG) metabolic pathway [31]; Class II AMGs have been defined as those which encode proteins that have only general, undefined metabolic roles, or peripheral roles (examples include assembly or membrane transport functions) and therefore do not appear in KEGG metabolic pathways [30, 32]. Table 1 provides numerous examples of metabolic pathways manipulated by virally-encoded genes in marine and non-marine systems, as identified from both experimental and metagenomic analyses.

Virus-directed augmentation and redirection of host energy and resources to increase viral progeny production has been observed in both marine heterotrophs and phototrophs, with the latter the focus of much of the seminal research in this area. The lytic cyanophage Syn9 encodes photosystem II core protein D1 (which enables transfer of electrons from water to plastoquinone), as well as genes catalysing the synthesis of deoxyribonucleotides (*nrdA/nrdB*) using NADPH as a terminal electron donor [10]. Indeed, viral encoding of genes associated with photosystem II is common in cyanophages, with 88% of cyanophage genomes containing *psbA* and 50% containing both *psbA* and *psbD* [33]. Phages encoding genes associated with photosystem I were identified in metagenomic data and postulated to increase ATP concentrations to power viral production by promoting cyclic photophosphorylation, at the expense of reducing power for CO_2 fixation [34]. Interestingly, in silico modelling of the advantage conferred by a phage encoding a *psbA* gene suggested that the increase in energy available to the virus for replication is negated by the increased cost of encoding an extra gene. However, under high-light conditions, the virus gains an advantage by replacing photosystem II machinery damaged by light stress [35, 36]. Thus, the presence of photosystem genes on viral genomes is a function of host ecotypic distribution and irradiance sensitivity. Recently a novel *Prochlorococcus*-infecting cyanophage (P-TIM68) was isolated that encodes viral genes for photosystem II as well as a cassette for photosystem I (*psaABCDEFJK*) - the first known example of a phage to manipulate both photosystems during infection. Predictions of enhanced cyclic electron flow around photosystem I were also confirmed [37]. Typically, these genes are co-expressed during early and middle infection and result in a decoupling of photosynthetic 'light reactions' from 'dark reactions', directing ATP and reducing power away from CO_2 fixation and towards nucleotide biosynthesis for phage replication [10, 37].

Re-routing of photosynthesis-derived energy to increase viral production is just one of many strategies viruses employ during infection to plunder the cellular resources of their hosts. The primary limiting nutrients in marine environments are phosphate (P), nitrate (N) and iron (Fe) [38–40]. Therefore it is perhaps unsurprising that marine viruses encode AMGs to augment host uptake of these resources during infection. Genomes of viruses isolated from P-limited environments have been found to contain more AMGs associated with P-uptake than those from P-replete environments [41]. T4-like cyanophages, S-SM1 and S-SM2, encode an alkaline phosphatase, putatively enabling cleavage of phosphate from organic phosphate sources [42]. Under P-limitation, hosts increase production of the phosphate transporter PstS to maximise uptake, and in P-limited marine environments PstS is one of the most abundant proteins identified in metaproteomic datasets [43]. Both

Table 1 Examples of studies that reported/predicted phage-mediated alteration of metabolic function in prokaryotic hosts

Host/s	Phage/s; cycle (if known)	Modification/Phenomena (molecular; physiological; phenotypic)	Observed effect (O) or Predicted effect (P) on host metabolism/ host survival	References
Vibrio cholera	VPIΦ and CTXΦ; Lysogenic	Insertion of VPIΦ results in toxin-coregulated pilus (TCP) expression; TCP-facilitated CTXIΦ insertion into host genome	(O) Expression of cholera toxin	[67, 68]
Escherichia coli	933 W; Lysogenic to lytic switch	Induction of 933 W prophages that encode for both shiga toxin (Stx) and a cleavable repressor	(O) Greatly increases *stx* gene expression, and therefore bacterial production and release of Stx.	[65]
Staphylococcus aureus	Φ13; Lysogenic	Integration of Φ13 genome with beta-toxin gene (*hlb*)	(O) Loss of beta-toxin expression (Note: beta-toxin is a sphingomyelinase)	[71]
Escherichia coli	λ; Lysogenic	λ cI protein expression; cI binds to *pckA* regulatory region preventing transcription	(O) Suppression of phosphoenolpyruvate carboxykinase production & gluconeogenesis; reduced growth rate; predation avoidance	[66]
Vibrio harveyi 645; 20; 45	VHML; Lysogenic	Integration of VHML genome via transposition	(O) Broad suppression of substrate utilization; changes in d-gluconate utilization (625); c-glutamyl transpeptidase activity (20); and sulfatase activity (45)	[69, 70]
Listonella pelagia	ΦHSIC; Pseudolysogenic	Chromosomal integration of prophage	(O) Reduction in substrate utilization	[21, 110]
Cellulophaga baltica MM#3	ΦS_M and ΦS_T; Lytic	On evolution of phage resistance: possible adaptation of amino acid transporters (likely phage receptors) in cell membrane	(O) Reduction in ability to metabolise various carbon sources, including many amino acids	[111]
Synechococcus WH8109	Cyanophage Syn9; lytic	Phage encoded carbon metabolism genes cp12, *talC, psbA, zwf, gnd*, and *nrdA/nrdB*, co-expressed in early infection; two-fold increase in NADPH/NADP ratio	(P) 'light reactions' decoupled from 'dark reactions; ATP & NADPH directed away from the Calvin Cycle. likely towards phage dNTP biosynthesis	[10]
Synechococcus WH8017	S-SM2; lytic	Phage encodes genes for photosynthesis (PSII): *psbA*; *psbD*, and carbon metabolism genes: *gnd*; *tal; zwf;* CP12	(P) Photosynthesis augmented during infection; carbon redirected from glucose and amino acid production to ribose-5P and NAPDH generation (for dNTP synthesis), via PPP-mediated glucose reduction	[42, 53]
Cyanobacteria: various Prochlorococcus and Synechococcus strains	Various: 42 cultured cyanophages	88% of cyanophage genomes include *psbA*; 50% code for both *psbA* and *psbD* (PSII genes)	(P) Boost to phototrophic metabolism during infection.	[33]
Cyanobacteria	Un-cultured cyanophages	Phage-encoded photosystem I genes *psa*A, B, C, D, E, K and JF (from environmental samples)	(P) Channelling of reducing power from respiratory chain towards PSI, possibly for ATP generation	[34]
Prochlorococcus MIT9515	P-TIM68; lytic	Phage encoded photosystem I and II proteins incorporated into host membrane	(O) Photosynthetic capacity maintained; enhanced cyclic electron flow around PSI; (P) Generation additional ATP for phage replication	[37]
Vibrios (including V. parahaemolyticus	KVP40; lytic	Phage ORFs code for: PhoH; putative pyridine nucleotide (NAD^+) salvage pathway, and hydrolysis of NADH	(P) Facilitates cross-membrane transport of NAD^+ precursors, NAD^+ synthesis, and cycling of NADH back to precursors.	[112]
Various (marine metagenomic Assemblages)	Various (marine viral metagenomes)	Most abundant putative viral-encoded enzymes: riboreductases; carboxylyases and transferases; *psbA* genes.	(P) Aids scavenging of host nucleotides (e.g. Riboreductases); supports host metabolism during the infection cycle (e.g., carboxylyases; transferases and D1 protein)	[113]
Various: from 22 'ultra-clean' viromes in POV dataset	Various; classified via protein cluster (PC) generation	35 carbon pathway AMGs, representing a near-full central carbon metabolism gene complement.	(P) In oligotrophic environments, AMGs may redirect host carbon flux into energy production and the replication of viral DNA.	[53]

Table 1 Examples of studies that reported/predicted phage-mediated alteration of metabolic function in prokaryotic hosts *(Continued)*

Host/s	Phage/s; cycle (if known)	Modification/Phenomena (molecular; physiological; phenotypic)	Observed effect (O) or Predicted effect (P) on host metabolism/ host survival	References
Various: from 32 viromes in POV dataset	Various; classified via PC generation	32 new viral AMGs (9 core; 20 photic; 3 aphotic): 9 encode Fe-S cluster proteins and genes associated with DNA replication initiation (*DnaA*), DNA repair (*dut*; *radA*) and motility augmentation (*psel*).	(P) Fe-S cluster modulation may drive phage production (in the photic zone); Genes associated with DNA replication and repair, and motility augmentation could assist high-pressure deep-sea survival.	[30]
Various: 127 SAGs from uncultivated SUP05 bacteria	Various: 69 viral contigs (from SUP05 SAGs)	4 putative AMGs (encoded by 12 viral contigs): *phoH* (on a bona fide viral contig); 2OG; 2OG-FeII oxygenase, *tctA* (protein domain only); and *dsrC*.	(P) *dsrC* likely functional in SUP05 sulfur cycling; characterisation of viral DsrC needed to elucidate roles of viruses in modulating electron transfer during viral infection.	[16]
Various: Actinobacteria, proteobacteria (α; δ; γ) Bacteroidetes, Cyanobacteria, Deferribacteres	Various, inc. members of T4 (superfamily) and T7 (genus)	243 putative AMGs (95 previously known [6]) including *dsr*C (11 genes), *sox*YZ (4 genes), both originating from T4 superfamily; *P-II* (encodes a nitrogen metabolism regulator) and *amo*C (encodes ammonia monooxygenase sub-unit)	(P) Viral roles in: Sulfur oxidation, via Dsr and Sox pathways; Nitrogen cycling (influenced by *P-II*), with potential for alternative pathways of N and NH_3 uptake during N starvation, and NH_3 oxidation (via *amo*C).	[47]
Various: 113 genomes (marine bacteria)	Various: 64 prophage-like elements (21 GTAs)	High relative incidence of transcriptional regulatory and repressor-like proteins in putative prophages (comparison: lytic phages)	(P) Suppresses non-essential host metabolic activities in unfavourable environments/periods	[21]
Listeria monocytogenes	'A118-like prophage' (reversible excision)	*comK* gene, encoding *L. monocytogenes* competence system master regulator, is activated by the excision of A118-like prophage	(O) A118-like prophage is excised only when a *L. monocytogenes* cell is engulfed by a phagosome: the host's activated competence system facilitates escape, after which prophage reintegrate with host *comK* gene, deactivating host's competence system	[74]
Anabaena spp.; Nostoc spp.	Non-infective 'prophages' (x 3; non-reversible excision)	Recombinases (prophage-encoded) act to excise prophages from 3 host genes that are involved in nitrogen fixation (*nifD*; *fdxN*; *hupL*)	(O) In low nitrogen environments, excision of prophages from host N-fixation genes enables conversion of host cell to form nitrogen-fixing heterocysts	[74]
Synechococcus elongatus	Cyanophage AS-1	Prevents normal ppGpp accumulation under nutrient limitation, and the corresponding expression of genes for starvation survival	(O) Inhibits the host's natural starvation response under nutrient limitation; (P) promotes metabolic activity otherwise undertaken only when food is plentiful, facilitating phage production in low nutrient conditions	[14, 114]

Abbreviations not used in the main text: *ORF* Open Reading Frame, *dut* Deoxyuridine triphophatase, *radA* DNA recombination protein, *psel* Pseudaminic synthase, *2OG* 2-oxoglutarate; *2OG-FeII oxygenase* Fe (II)-dependent oxygenase superfamily, *tctA* Tripartite tricarboxylate transporter, *GTA* Gene Transfer Agents

virally-encoded *pstS* and *phoA* were over-represented in viral genomes from the North Atlantic Subtropical Gyre, and occurred at similar frequencies to signature core genes. In the North Pacific Subtropical Gyre, *phoA* was much less abundant in viral genomes, providing evidence of niche-separation of AMGs corresponding to limiting nutrients [41]. Using model cyanophage/host systems, Zeng and Chisholm showed that both virally-encoded *pstS* and *phoA* AMGs are upregulated during infection of P-limited hosts [15]. The phosphate-regulon associated gene, *phoH*, is so common in viral genomes it has been proposed as a signature gene for measuring phage diversity [44]. However, it is worth noting that the function of virally-encoded *phoH* is not yet clear, and *phoH* expression in phosphate limited conditions appears to vary between hosts [45, 46].

Viral manipulation of cellular nitrogen and sulfur levels during infection is perhaps less well documented in cultured host-virus systems than phosphate regulation, but recent research is providing evidence that it is no less important. Phage genomes have been found to encode numerous proteins to manipulate concentrations of 2-oxoglutarate, which in turn, regulates cellular N-limitation response via the promotor NtcA [42]. Beyond cultured representatives, a recent global study of viral metagenomics identified 243 putative viral AMGs including genes for photosystem II, ammonium transporters (*amt*) and ammonia monooxygenases (*amoC*), as well as genes associated with sulfur reduction (*dsr*) and oxidation (*sox*) [47]. Screening of marine cellular metagenomes from the Tara Oceans cruises revealed that not only do *amoC* genes encoded by viruses infecting Thaumarchaeota form a distinct phylogenetic clade, they can also comprise up to half of the total abundance of *amoC* in some metagenomic datasets [48]. Nitrate reductase genes (*nar*) have also been found in viromes from deep-sea hydrothermal vents [49]. Evidence of virally-encoded components of dissimilatory sulfite reductase complex (Rdsr) have been found in viruses infecting the uncultivated chemoautotrophic gammaproteobacterial clade SUP05 [16, 50, 51]. SUP05 are important drivers of sulfur cycling [52] and peaks in abundance of virally-encoded *dsrC* were observed during SUP05 blooms. Investigators concluded that phage-encoded DsrC is likely to function in the sulfur cycling of infected SUP05, and may modulate vital electron transfer reactions. Evidence indicating increased viral infection of SUP05 with increasing depth, and decreasing O_2 levels, underlines the potential significance of this finding [16].

Virally-encoded genes associated with carbon uptake have also been identified in both model experimental systems and viral metagenomes, including those involved in uptake and metabolism of amino acids (*speD*, *cysK/M*, *metK*, *dapC*) and carbohydrates (*manA*, *rpiB*, *glgA*) [30, 53]. Cyanophage genomes include genes (*talC*, *zwf*, *gnd*) to divert carbon towards the pentose phosphate pathway by converting glyceraldehyde-3P to fructose-6P, which is subsequently converted to reducing power and the synthesis of dNTPs for phage replication [10]. Viral metagenomics also identified other virally encoded genes involved in glycolysis (*manA*) and a glycogen synthase (*glgA*), with the latter identified in all examined viral metagenomes [53]. In this work, Hurwitz and colleagues postulated *glgA* was used to trigger a starvation response in the host in order to push carbon through non-glycolytic pathways that promote dNTP biosynthesis. Manipulation of succinate through the glyoxylate shunt is thought to facilitate energy production at the expense of amino acid synthesis, particularly in nutrient limited, deep ocean samples. Intracellular carbon is also redirected during viral infection for energy via virally-encoded genes for the Entner-Doudoroff pathway, the TCA cycle and fatty acid metabolism [53].

So how does expression of viral AMGs during infection influence cells at the fundamental level of metabolite synthesis and utilisation? The influence of viral hijacking at the metabolite level lacks a representative study in marine systems, but De Smet and colleagues performed an exemplary study in *Pseudomonas aeruginosa* PAO1 and compared responses across host-virus infections between one host and six different phages [12]. Following infection, concentration of 92 out 375 detectable host metabolites were significantly altered, showing a major rewiring of host metabolism. Only 9 of these metabolites, all associated with nucleotide metabolism, were altered by all six tested phages. This indicates that beyond nucleotide synthesis for viral replication, the impact of viral hijacking on host metabolism is not conserved even within a single host species. Larger phage genomes encoded a larger suite of metabolic machinery for viral production, whereas smaller genomes relied more on host machinery and scavenging of resources. Phages which encoded one or more RNA polymerase shut down host transcription, and initiated their own transcription machinery. Phages with smaller genomes rewired host RNA polymerases directly to increase promoter specificity towards phage-specific promoters. Interestingly, hijacking of host metabolism appeared to favour pyrimidine production, with purine synthesis enriched only in phage YuA. Infection with YuA (genome size of 58.6 kb) also resulted in major depletion of cellular resources, recycling them for phage synthesis. In comparison, phage phiKZ (which has a 280kbp genome encoding 306 open reading frames) had almost no impact on host metabolite concentrations [12].

Efficiency in host hijacking is thus not universal, and tends to be higher in infections between specialist phages and their preferred host (which are the focus of many host-phage model systems). Generalist phages that infect multiple hosts tend to have less efficient

infections, fail to completely suppress host translation and transcription, have longer latent periods and decoupled translation and transcription [12, 13]. The 'plunder and pillage' strategy was not observed in a generalist phage infection of *Cellulophaga baltica* [13], and it is possible that utilisation of AMGs for hijacking of host metabolism is not a feature of generalist infections. Yet, members of the abundant marine Myoviridae infecting *Synechococcus* and *Prochlorococcus* include many generalists, some capable of infecting both genera [54, 55]. The fitness advantage to adopting a broad host range, at the expense of efficient infection makes ecological sense when one considers that temporal patterns in marine environments often result in transition from *Synechococcus*-dominated to *Prochlorococcus*-dominated communities as nutrient availability waxes and wanes over seasonal, depth and diel gradients [56–58]. Furthermore, recent evidence from an experimental evolution experiment suggests that infection of sub-optimal hosts increases mutation rates and diversification of phage populations compared to those infecting optimal hosts [59].

Batten down the hatches: Increased viral fitness through increased host fitness

From the perspective of viral fitness, the benefit of driving host energy and metabolites towards viral production during lytic infection only confers an advantage if viral progeny released following a lytic event can successfully infect new, actively growing hosts. Imagine a band of naïve pirates in the doldrums of a seldom-travelled ocean gyre. They seize a ship, kill the crew and strip its cargo with alarming efficiency, before scuppering the vessel and sailing off in rowboats. With no land in sight, nor new vessel to capture, their career as pirates would be short and they would soon succumb to the tropical sun, deprived of life-giving resources. In contrast, a savvy band of pirates would instead remain aboard the seized vessel, secure the mainmast during storms that may arise, ration the rum and steer her towards more profitable waters. Many marine viruses display a similar 'batten down the hatches' strategy by foregoing a lytic life cycle and instead increasing the fitness of the infected cell phenotype through metabolic manipulation (Fig. 1b). Here, instead of producing viral progeny, the infecting virus is maintained either by integration into the host genome (prophage) or as an extrachromosomal element to convert the host into a lysogen. Lysogeny includes a broad range of mechanisms including chronic infection (slow release of viral particles without killing the host), pseudolysogeny (simultaneous high viral production without host lysis) and polylysogeny (infection of multiple viruses within the same host) (reviewed in [21]). Numerous studies of the prevalence of lysogeny in marine systems all suggest that lysogeny is favoured in low productivity environments [21, 60, 61]. Chemical induction of lytic viral production in natural marine communities using mitomycin C showed the switch only occurred at cellular concentrations $> 10^6$ cells per mL [62]. A significant decrease in viral abundance has been observed at depth, corresponding to host abundance an order of magnitude lower than those at the surface [63]. These lower virus-to-microbe ratios are often interpreted as evidence of increased lysogeny. A recent meta-analysis of virus-to-microbe ratios across global oceans showed that this ratio varies from 1.4 to 160 and suggested that host-virus ratios are shaped by complex feedback mechanisms between nutrient availability, evolutionary history and selection pressures [64].

Regardless of the mechanism used to maintain the virus within the host during lysogeny, the effect of phage on host phenotype has been extensively studied in important pathogens such as *E. coli* [65, 66], *Vibrio cholerae* [67, 68], *Vibrio harveyi* [69, 70], *Staphylococcus aureus* [71] and *Listeria monocytogenes* [72] (reviewed in [73, 74]) (Table 1). However, relatively few studies into lysogeny in marine host-virus model systems have been performed. When Yu and colleagues isolated a *Pseudoalteromonas* strain from Arctic sea ice containing a filamentous phage, they noticed it had lower growth rates and cell density, and lower tolerance of NaCl and reactive oxygen species than when it was cured of the infection. Transcriptional analysis showed downregulation of succinyl-CoA synthetase and succinate dehydrogenase indicating virally-mediated suppression of central carbon metabolism. However, presence of the phage increased host motility. They postulated that the presence of the phage increased host fitness during the nutrient-limited polar winters by slowing down host metabolism and increasing its capacity to find new nutrient sources in the heterogeneous structure of the sea ice in which it lived [75]. Using metagenomic analyses, Brum and colleagues postulated that a similar mechanism increased host fitness in Antarctic bacterial communities during times of low nutrients and explained an observed seasonal prevalence of lysogeny prior to the summer blooms [76]. The underpinning mechanisms are possibly similar to those observed in the *E. coli* phage λ. This lysogen maintains integration in the host genome via a phage-encoded repressor known as *cI*. *cI* also represses the host gene *pckA*, which encodes a protein for the conversion of oxaloacetate to phosphoenolpyruvate. Thus, lysogeny results in a decoupling of central carbon metabolism from cellular synthesis, reducing host growth rate and potentially conferring a selective advantage in hosts within nutrient-poor environments [66]. Integration of the viral genome into the host genome can interrupt metabolic genes and effectively act as regulatory mechanisms,

either at the level of the individual cell by repeated integration and excision, or at the population level by non-reversible lysogeny suppressing genes in a subpopulation [74]. Although there are no known examples of the former in marine bacteria, viruses infecting the cyanobacteria *Anabena* spp. and *Nostoc* spp. appear to regulate nitrogen-fixation through active lysogeny. Here, prophages interrupting N-fixation genes (*nifD*, *fdxN* and *hupL*) are excised from the genome during N-limitation, re-activating the genes (reviewed in [74]). Many metabolic genes identified in viral metagenomes are predicted to confer a selective advantage to hosts including sulfur oxidation genes [16, 50] and genes associated with adaptation to high-pressures associated with depth [30]. One study of viral metagenomes from hydrothermal vents identified virally-encoded genes associated with pyrimidine, alanine, aspartate, glutamine, nitrogen, amino and nucleotide sugar metabolism. Pathway analysis suggested that viral AMGs allowed for branched metabolic pathways to alternative products that would provide additional metabolic flexibility to the host, thus increasing host fitness [49].

Metabolic false flags

In March 1723, the pirate Captain Low approached a Spanish merchant ship in the Bay of Honduras under the Spanish colors. Once they drew near the vessel, they:

'*hauled them down, hoisted up their black flag, fired a broadside and boarded her*' [77].

Indeed, it was commonplace for pirates to sail under false flags of different countries in order to prevent their targets from identifying them as a threat until it was too late. Similarly, viruses encode genetic tools to prevent a hijacked host from recognising an infection and taking appropriate action. As internal concentrations of cellular substrates decrease, cells can undergo a 'stringent response' where they down-regulate growth functions and, in some cases, initiate programmed cell death, regulated by the toxin-antitoxin pair MazEF. Nutrient limitation results in uncharged tRNAs, which triggers increasing concentrations of guanosine tetraphosphate (ppGpp). This in turn inhibits RNA transcription and promotes transcription of the global stress response regulator RpoS, shifting the cell into stationary phase [78]. Increasing concentrations of ppGpp also increases production of MazF, a toxin that inhibits protein synthesis (resulting in cell dormancy). In some circumstances, MazF also initiates programmable cell death, with different mechanisms used under different stressors such as antibiotic treatment or DNA damage [79, 80]. Programmable cell death is thought to sacrifice a large proportion of a population, so that a small sub-population can survive by recycling released nutrients [80]. Programmable cell death regulated by MazEF has been shown to play a role in the survival of a bacterial population against phage infection. Hazan and Engelberg-Kulka showed that 400 times more viral progeny were produced in Δ*mazEF E. coli* cells infected with phage P1 compared to wild-type cells. In addition, when lysogens were mixed with non-lysogens, the Δ*mazEF* cells were susceptible to infection and lysed, whereas the non-lysogen wild-type cells were not infected by the induced phage. They suggested that it is likely the wild-type lysogens were killed by MazEF, preventing viral replication to the benefit of the population [81]. Similar observations of the role of toxin-antitoxin mechanisms for phage resistance were observed in *Erwinia carotovora*, where expression of the toxin-antitoxin system conferred resistance to a broad range of viruses [82].

Given its dual role as both a regulator of the stringent response and of MazEF-regulated resistance and cell death, increasing cellular levels of ppGpp pose a clear threat to viral replication. Many viruses are dependent on host RNA transcription for replication, and viral replication has been shown to be drastically reduced or suspended in infected *E. coli* cells that enter stationary phase [83]. Cell death limits the spread of phages through a population by reducing host density and purges lysogens from the population. Thus, is of little surprise that in the arms-race between viruses and their hosts, the viruses have commandeered a ppGpp regulator to counter host defence mechanisms. MazG has been shown to decrease cellular concentrations of ppGpp in *E. coli* and thus repress production of MazF [79]. Thus in the host MazG serves as a switch to recover from a starvation response upon return to nutrient replete conditions and to abort programmable cell death. *mazG* is common in viral genomes from marine environments and has been reported in viruses infecting both heterotrophs [84, 85] and phototrophs [14, 86]. The oligotrophic nature of vast areas of the ocean make it likely that a starvation state for host cells is the norm, resulting in high cellular concentrations of ppGpp and minimal transcription. It is likely that phages overcome this limitation by using virally-encoded MazG to deplete cellular ppGpp concentrations and thus force the infected cell to respond as if it were nutrient-replete, enabling transcription so that it may be hijacked for viral replication [87]. In addition, suppression of programmable cell death may prevent the host population from sacrificing itself in order to limit viral infection. Recent work in has shown that phages can utilise host quorum sensing to coordinate a lytic-lysogenic switch using a phage-derived oligopeptide signal [88] . Given that programmable cell death has been shown to be orchestrated through quorum sensing [89], and that suppression of *mazEF* expression increases host resistance to phages, it is conceivable that infecting viruses can use MazG to manipulate ppGpp concentrations and thus maintain the

susceptibility of hosts within a population to subsequent infection by viral progeny. In pirate terms, viruses have evolved the capacity to sail under a false flag, disabling host alarm systems until it is too late. Further experimental evidence of the use of MazG and its effects on marine populations is required to explore whether this phenomenon is observed in nature.

Future understanding of host-virus interactions

The current understanding of how viruses hijack host metabolism during infection is the result of both culturing experiments in model systems; and culture-independent techniques such as viral metagenomics. Culture-dependant techniques have significant advantages: Model systems enable us to study the relationship between host and viruses in controlled conditions; viral replication cycles and critical parameters (e.g. host range; burst size) can be defined, and the functionality of viral genes may be determined in vitro. In culture, predator-prey interactions can be isolated from those occurring in complex microbial communities. This reduction in complexity allows for investigation into how transcriptomes, proteomes and metabolites are altered during infection and thus enable a systems-biology approach to understanding complex metabolic cascades and regulation (e.g. [12, 13, 25, 45, 90]. However, many important marine taxa have, to date, resisted efforts to culture them [91]. Consequently, our model systems of host-virus interactions in marine systems are limited to a handful of taxa, with a major focus on the cyanobacteria and a limited number of heterotrophic hosts.

For systems outside of cultured representatives, viral metagenomic studies to date have provided major insights into to viral taxonomic and functional distribution and diversity [30, 32, 47, 53]. Understanding the extent and mechanisms of metabolic hijacking by marine viruses using metagenomic data comes with its own challenges: Firstly, in any viral metagenome there is the possibility of contaminating cellular DNA or randomly packaged host DNA encapsulated in gene transfer agent particles [32, 92]. As viral genomes are assembled from short read data, there is the possibility of cellular functions being misassembled into viral contigs. In such circumstances, the function may be interpreted as a novel AMG acquired by the virus to improve fitness, rather than as an artefact of bioinformatic processing [93, 94], with a concomitant over-estimation in the degree of viral piracy that occurs in marine systems. Secondly, some of the most cosmopolitan and dominant viruses on Earth are challenging to assemble using short-read technology and are under-represented in marine viral metagenomes [95, 96]. Complementary approaches to construct viral genomes from environmental samples such as single cell genomics and the development of long-read viral sequencing can alleviate these problems to some degree. Assembly of short-read data from a genome amplified from a single cell or single virus cell vastly reduces the complexity of the De Bruijn Graph and captures taxa missing from shotgun metagenomic approaches.

This approach has successfully to identify new viruses and novel AMGs [16, 17, 96]. Long-read viral metagenomics [97] offers the potential to accurately identify putative AMGs as viral, rather than cellular contaminants, by capturing the gene neighbourhood of the AMG to reliably assess its viral origins. Capturing full length viral genomes on single reads is now technically feasible and will provide a powerful tool to explore how AMGs are distributed within viral populations and how their evolution is impacted by recombination, shown to be the dominant form of mutation in some phages [98]. Long-read metagenomics from cellular fractions will better quantify the extent of lysogeny within a population by capturing integrated viral genomes on single reads.

It is worth noting however, that no matter how sophisticated sequencing methods become, perhaps the greatest barrier to understanding how marine viruses influence cellular metabolism during infection lies in our extremely limited capacity to identify the function of viral genes, in both cultured isolates and genomes constructed from environmental DNA. Whilst machine learning approaches are rising to meet this challenge [32, 99], one must consider that the scale of the 'known unknowns' is vast, with 63–93% of protein sequence space lacking functional or taxonomic annotations [100]. < 1% of viral populations in the Pacific Ocean Virome had a closely related taxonomic representative in culture [101]. Methods to identify viral host range through computational methods such as correlative abundance [102] or nucleotide composition [103, 104], are undergoing rapid development, but must be used cautiously for inferring ecological patterns [105]. Despite these challenges, the last decade has seen a dramatic improvement in our capacity to generate and interpret viral metagenomic data, largely driven by efforts to understand marine systems (e.g. [47]). This improvement has revealed a growing body of evidence identifying viruses as important agents in global carbon biogeochemical cycles, through: 1) lysis-dependent nutrient cycling and increased community productivity [7–9]; 2) influences on host-substrate interactions through auxiliary metabolic genes, shaped through viral evolution (reviewed in [11]). Viral metagenomics allows microbial ecologists to directly ascribe such functions to viruses and provides relative quantitation of viral populations and genetic diversity in a way that is challenging from cellular metagenomic data. More recently, these methods have been applied to medical microbiology and have similarly established viruses as an equally important component

of the human microbiome alongside their cellular counterparts (reviewed in [106]). Indeed, there is a growing consensus that our view of microbial ecology must put viruses centre stage as key players in shaping community structure and function. Increasing interest in the role of viruses in microbiomes will undoubtedly catalyse a feedback loop that energises the development of novel bioinformatic and culturing methods. Such tools will ultimately overcome the technical limitations previously outlined, revealing more of the metabolic capacity for cellular piracy encoded within viral sequence space.

Conclusion

The contemporary image of pirates is typically one of swashbuckling, romantic characters of Robert Louis Stevenson's *Treasure Island* and J.M. Barrie's *Peter Pan*. Piecemeal accounts and a lack of historical records has enabled myth and legend to supersede the violent and unpleasant reality written in Charles Johnson's *A General History of the Robberies and Murders of the Most Notorious Pyrates* [107]. Similarly, the relationships between marine viruses and their hosts have been considered through the paradigm of predator and prey, with much research focused on viruses as agents of top-down control. Our understanding of viral impact on host metabolism in marine systems is derived from a few model systems, or inferred from model systems of medically relevant pathogens that have evolved in nutrient-rich environments supporting high cellular densities. It is now clear however that lysogeny is common in marine systems and has the capacity for reconfiguration of host metabolism and increasing host fitness during frequent periods of nutrient limitation. Viral metagenomics continues to offer tantalising evidence of putative mechanisms for viral piracy, but even the most advanced machine learning approaches are limited by comparison to existing model systems. Thus, if we are to better understand the impact of viral metabolic hijacking on global biogeochemical cycles, advances in computational methods must be matched with recent efforts to improve the culturing important marine taxa [108, 109] and their associated viruses, followed by in vitro determination of mechanism and impacts on host and viral fitness. Our efforts will be repaid in full as data is fed back into computational approaches to facilitate the accurate translation of experimentally observed phenotypic changes into impacts on global biogeochemical cycling in our current models.

Acknowledgements

The authors would like to thank the anonymous reviewers for their critical and constructive comments. The authors would also like to thank Kema Malki (USF) for providing the artwork for Fig. 1.

Funding

This work was funded by the BIOS-SCOPE award from Simons Foundation International, and by Natural Environment Research Council (NERC) awards NE/R010935/1 and NE/P008534/1. JWD and HB were funded by NERC GW4+ Doctoral Training Partnerships.

Authors' contributions

JWD, HB and BT wrote the manuscript; MA provided direction and edits for the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK. [2]University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter EX4 4QD, UK.

Received: 20 April 2018 Accepted: 17 January 2019
Published online: 01 February 2019

References

1. Haldane JBS. What is life? Lindsay Drummond; 1945.
2. Suttle CA. Viruses in the sea. Nature. 2005;437:356–61.
3. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. Science. 2008;320:1034–9.
4. Suttle CA. Marine viruses — major players in the global ecosystem. Nat Rev Microbiol. 2007;5:801–12.
5. Noble RT, Fuhrman JA. Rapid virus production and removal as measured with fluorescently labelled viruses as tracers. Appl Environ Microbiol. 2000; 66:3790–7.
6. Breitbart M. Marine viruses: truth or dare. Annu Rev Mar Sci. 2012;4:425–48.
7. Middelboe M, Jorgensen NOG, Kroer N. Effects of viruses on nutrient turnover and growth efficiency of noninfected marine bacterioplankton. Appl Environ Microbiol. 1996;62:1991–7.
8. Middelboe M, Jorgensen NOG. Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. J Mar Biol Assoc U K. 2006; 86:605–12.
9. Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, et al. A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. ISME J. 2015;9:1352–64.
10. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc Natl Acad Sci U S A. 2011;108:16147–8.
11. Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. Curr Opin Microbiol. 2016;31:161–8.
12. De Smet J, Zimmermann M, Kogadeeva M, Ceyssens P-J, Vermaelen W, Blasdel B, et al. High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. ISME J. 2016;10:1823–35.
13. Howard-Varona C, Hargreaves KR, Solonenko NE, Markillie LM, White RA 3rd, Brewer HM, et al. Multiple mechanisms drive phage infection efficiency in nearly identical hosts. ISME J. 2018;12:1605–18.
14. Bryan MJ, Burroughs NJ, Spence EM, Clokie MR, Mann NH, Bryan SJ. Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. PLoS One. 2008;3:e2048.
15. Zeng Q, Chisholm SW. Marine viruses exploit their Host's two-component regulatory system in response to resource limitation. Curr Biol. 2012;22:124–8.
16. Roux S, Hawley AK, Beltran MT, Scofield M, Schwientek P, Stepanauskas R, et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. elife. 2014;3:e03125.
17. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. ISME J. 2015;9(11):2386–99.
18. Jiang SC, Paul JH. Significance of Lysogeny in the marine environment: studies with isolates and a model of lysogenic phage production. Microb Ecol. 1998;35:235–43.
19. Stopar D, Cerne A, Zigman M, Poljsak-Prijatelj M, Turk V. Viral abundance and a high proportion of lysogens suggest that viruses are important

members of the microbial community in the Gulf of Trieste. Microb Ecol. 2004;47:1–8.
20. Leitet C, Riemann L, Hagström Å. Plasmids and prophages in Baltic Sea bacterioplankton isolates. J Mar Biol Assoc U K. Cambridge University Press. 2006;86:567–75.
21. Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? ISME J. 2008;2:579–89.
22. Munson-McGee JH, Peng S, Dewerff S, Stepanauskas R, Whitaker RJ, Weitz JS, et al. A virus or more in (nearly) every cell: ubiquitous networks of virus–host interactions in extreme environments. ISME J. 2018;12:1706–14.
23. Williamson SJ, McLaughlin MR, Paul JH. Interaction of the PhiHSIC virus with its host: lysogeny or pseudolysogeny? Appl Environ Microbiol. 2001;67:1682–8.
24. Howard-varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. 2017;11:1511–20.
25. Doron S, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, et al. Transcriptome dynamics of a broad host-range cyanophage and its hosts. ISME J. 2016;10:1437–55.
26. Horvitz HR. Bacteriophage T4 mutants deficient in alteration and modification of the *Escherichia coli* RNA polymerase. J Mol Biol. 1974;90:739–50.
27. Koch T, Raudonikiene A, Wilkens K, Rüger W. Over expression, purification, and characterization of the ADP-Ribosyltransferase (gpAlt) of bacteriophage T4: ADP-Ribosylation of *E. coli* RNA polymerase modulates T4 "early" transcription. Gene Expr. 1995;4:253–64.
28. De Smet J, Hendrix H, Blasdel BG, Danis-Wlodarczyk K, Lavigne R. *Pseudomonas* predators: understanding and exploiting phage-host interactions. Nat Rev Microbiol. 2017;15:517–30.
29. Breitbart M, Thompson LR, Suttle CA, Sullivan MB. Exploring the vast diversity of marine viruses. Oceanography. Oceanography Society. 2007;20:135–9.
30. Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the "core" and "flexible" Pacific Ocean Virome. ISME J. 2015;9:472–84.
31. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
32. Brum JR, Sullivan MB. Rising to the challenge: accelerated pace of discovery transforms marine virology. Nat Rev Microbiol. 2015;13:147–59.
33. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol. 2006;4:e234.
34. Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, et al. Photosystem I gene cassettes are present in marine virus genomes. Nature. 2009;461:258–62.
35. Bragg JG, Chisholm SW. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. PLoS One. 2008;3:e3550.
36. Hellweger FL. Carrying photosynthesis genes increases ecological fitness of cyanophage *in silico*. Environ Microbiol. Wiley Online Library. 2009;11:1386–94.
37. Fridman S, Flores-Uribe J, Larom S, Alalouf O, Liran O, Yacoby I, Salama F, Bailleul B, Rappaport F, Ziv T, Sharon I, Cornejo-Castillo FM, Philosof A, Dupont CL, Sánchez P, Acinas SG, Rohwer FL, Lindell D, Béjà O. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. Nat Microbiol. 2017;2(10):1350–7.
38. Tyrrell T. The relative influences of nitrogen and phosphorus on oceanic primary production. Nature. 1999;400:525–31.
39. Benitez-Nelson CR. The biogeochemical cycling of phosphorus in marine systems. Earth-Sci Rev. 2000;51:109–35.
40. Behrenfeld MJ, O'Malley RT, Siegel DA, McClain CR, Sarmiento JL, Feldman GC, et al. Climate-driven trends in contemporary ocean productivity. Nature. 2006;444:752–5.
41. Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. ISME J. 2013;7:1827–41.
42. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarrachino D, Osburne MS, Henn MR, Chisholm SW. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ Microbiol. 2010;12:3035–56.
43. Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF, Carlson CA, Smith RD, Giovannoni SJ. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. ISME J. 2009;3:93–105.
44. Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle CA, Weinbauer MG, Sandaa RA, Breitbart M. Development of *phoH* as a novel signature gene for assessing marine phage diversity. Appl Environ Microbiol. 2011;77:7730–9.
45. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR, Church GM, Chisholm SW. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature. 2007;449:83–6.
46. Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, Paulsen IT. Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus sp*. WH8102. ISME J. 2009;3:835–49.
47. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D. Tara Ocean coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537:689–93.
48. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode *amoC* nitrification genes. ISME J. 2018. https://doi.org/10.1038/s41396-018-0289-4.
49. He T, Li H, Zhang X. Deep-Sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. MBio. 2017;8. https://doi.org/10.1128/mBio.00893-17.
50. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. Sulfur oxidation genes in diverse deep-sea viruses. Science. 2014;344:757–60.
51. Chow CET, Winget DM, White RA, Hallam SJ, Suttle CA. Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. Front Microbiol. 2015;6(265). https://doi.org/10.3389/fmicb.2015.00265.
52. Wright JJ, Konwar KM, Hallam SJ. Microbial ecology of expanding oxygen minimum zones. Nat Rev Microbiol. 2012;10:381–94.
53. Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome Biol. 2013;14:R123.
54. Dekel-Bird NP, Sabehi G, Mosevitzky B, Lindell D. Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. Environ Microbiol. 2015;17:1286–99.
55. Sullivan MB, Waterbury JB, Chisholm SW. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. Nature. 2003;426:584.
56. Landry MR, Kirshtein J, Constantinou J. Abundances and distributions of picoplankton populations in the central equatorial Pacific from 12 N to 12 S, 140 W. Deep Sea Res Part 2 Top Stud Oceanogr. Elsevier. 1996;43:871–90.
57. Treusch AH, Vergin KL, Finlay LA, Donatz MG, Burton RM, Carlson CA, Giovannoni SJ. Seasonality and vertical structure of microbial communities in an ocean gyre. ISME J. 2009;3:1148–63.
58. Giovannoni SJ, Vergin KL. Seasonality in ocean microbial communities. Science. 2012;335:671–6.
59. Enav H, Kirzner S, Lindell D, Mandel-Gutfreund Y, Béjà O. Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. Nat Commun. 2018;9(1). https://doi.org/10.1038/s41467-018-07164-3.
60. Jiang SC, Paul JH. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine environment. Marine Ecol Prog Ser Oldendorf. 1994;104:163–72.
61. Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev. 2000;64:69–114.
62. Weinbauer MG, Suttle CA. Potential significance of lysogeny to bacteriophage production and bacterial mortality in coastal waters of the Gulf of Mexico. Appl Environ Microbiol. 1996;62:4374–80.
63. De Corte D, Sintes E, Yokokawa T, Reinthaler T, Herndl GJ. Links between viruses and prokaryotes throughout the water column along a North Atlantic latitudinal transect. ISME J. 2012;6:1566–77.
64. Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF, Fuhrman JA, et al. Re-examination of the relationship between marine virus and microbial cell abundances. Nat Microbiol. 2016;1:15024.
65. Tyler JS, Mills MJ, Friedman DI. The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. J Bacteriol. 2004;186:7670–9.

66. Chen Y, Golding I, Sawai S, Guo L, Cox EC. Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. PLoS Biol. 2005;3:e229.
67. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science. 1996;272:1910–4.
68. Karaolis DK, Somara S, Maneval DR Jr, Johnson JA, Kaper JB. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. Nature. 1999;399:375–9.
69. Oakey H, Cullen B, Owens L. A hypothetical model for VHML phage conversion of *Vibrio harveyi*. In: Walker P, Lester R, Bondad-Reantaso MG, editors. Diseases in Asian Aquaculture V. Manila: Fish Health Section, Asian Fisheries Society; 2005. p. 457-64.
70. Vidgen M, Carson J, Higgins M, Owens L. Changes to the phenotypic profile of *Vibrio harveyi* when infected with the *Vibrio harveyi* myovirus-like (VHML) bacteriophage. J Appl Microbiol. 2006;100:481–7.
71. Coleman D, Knights J, Russell R, Shanley D, Birkbeck TH, Dougan G, Charles I. Insertional inactivation of the *Staphylococcus aureus* β-toxin by bacteriophage φ13 occurs by site-and orientation-specific integration of the φ 13 genome. Mol Microbiol. 1991;5:933–9.
72. Rabinovich L, Sigal N, Borovok I, Nir-Paz R, Herskovits AA. Prophage excision activates *Listeria* competence genes that promote phagosomal escape and virulence. Cell. 2012;150:792–802.
73. Brüssow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev. 2004;68:560–602.
74. Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat Rev Microbiol. 2015;13:641–50.
75. Yu ZC, Chen XL, Shen QT, Zhao DL, Tang BL, Su HN, Wu ZY, Qin QL, Xie BB, Zhang XY, Yu Y, Zhou BC, Chen B, Zhang YZ. Filamentous phages prevalent in *Pseudoalteromonas spp.* confer properties advantageous to host survival in Arctic Sea ice. ISME J. 2015;9:871–81.
76. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. ISME J. 2016;10:437–49.
77. Cordingly D. Under the black flag: the romance and the reality of life among the pirates. New York: Random House Incorporated; 2006.
78. Chatterji D, Ojha AK. Revisiting the stringent response, ppGpp and starvation signaling. Curr Opin Microbiol. 2001;4:160–5.
79. Gross M, Marianovsky I, Glaser G. MazG -- a regulator of programmed cell death in *Escherichia coli*. Mol Microbiol. 2006;59:590–601.
80. Amitai S, Kolodkin-Gal I, Hananya-Meltabashi M, Sacher A, Engelberg-Kulka H. *Escherichia coli* MazF Leads to the Simultaneous Selective Synthesis of Both "Death Proteins" and "Survival Proteins.". PLoS Genet. 2009;5:e1000390.
81. Hazan R, Engelberg-Kulka H. *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol Gen Genomics. 2004;272:227–34.
82. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A. 2009;106:894–9.
83. Bryan D, El-Shibiny A, Hobbs Z, Porter J, Kutter EM. Bacteriophage T4 Infection of Stationary Phase *E. coli*: Life after Log from a Phage Perspective. Front Microbiol. 2016;7:1391.
84. Kang I, Oh HM, Kang D, Cho JC. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. Proc Natl Acad Sci U S A. 2013;110:12343–8.
85. Duhaime MB, Wichels A, Waldmann J, Teeling H, Glöckner FO. Ecogenomics and genome landscapes of marine *Pseudoalteromonas phage* H105/1. ISME J. 2011;5:107–21.
86. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. PLoS Biol. 2005;3:e144.
87. Clokie MRJ, Mann NH. Marine cyanophages and light. Environ Microbiol. 2006;8:2074–82.
88. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al. Communication between viruses guides lysis-lysogeny decisions. Nature. 2017;541:488–93.
89. Kumar S, Kolodkin-Gal I, Engelberg-Kulka H. Novel quorum-sensing peptides mediating interspecies bacterial cell death. MBio. 2013;4:e00314–3.
90. Howard-Varona C, Roux S, Dore H, Solonenko NE, Holmfeldt K, Markillie LM, Orr G, Sullivan MB. Regulation of infection efficiency in a globally abundant marine *Bacteriodetes* virus. ISME J. 2017;11(1). https://doi.org/10.1038/ismej.2016.81.
91. Rappé MS, Giovannoni SJ. The uncultured microbial majority. Annu Rev Microbiol. 2003;57:369–94.
92. Lang AS, Beatty JT. Importance of widespread gene transfer agent genes in alpha-proteobacteria. Trends Microbiol. 2007;15:54–62.
93. Roux S, Krupovic M, Debroas D, Forterre P, Enault F. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. Open Biol. 2013;3:130160.
94. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit MA. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. 2017;11:237–47.
95. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5:e3817. https://doi.org/10.7717/peerj.3817.
96. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, Anton J, Gasol JM, Rosselli R, Rodriguez-Valera F, Sullivan MB, Acinas SG, Martinez-Garcia M. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat Commun. 2017;8:15892. https://doi.org/10.1038/ncomms15892.
97. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, Sullivan MB, Temperton B. Long-read metagenomics reveals cryptic and abundant marine viruses. bioRxiv. 2018 [cited 2018 Jun 26]:345041 Available from: https://www.biorxiv.org/content/early/2018/06/12/345041.
98. Kupczok A, Neve H, Huang KD, Hoeppner MP, Heller KJ, Franz CMAP, Dagan T. Rates of mutation and recombination in Siphoviridae phage genome evolution over three decades. Mol Biol Evol. 2018;35:1147–59.
99. Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc Natl Acad Sci U S A. 2014;111:10714–9.
100. Hurwitz BL, Sullivan MB. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS One. 2013;8:e57355.
101. Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara oceans coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. Ocean plankton Patterns and ecological drivers of ocean viral communities. Science. 2015;348:1261498.
102. Needham DM, Sachdeva R, Fuhrman JA. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. ISME J. 2017;11(7):1614–29.
103. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Res. 2016;45:39–53.
104. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics. 2017;33:3113–4.
105. Coenen AR, Weitz JS. Limitations of Correlation-Based Inference in Complex Virus-Microbe Communities. mSystems. 2018;3(4). https://doi.org/10.1128/mSystems.00084-18.
106. Mirzaei MK, Maurice CF. Ménage à trois in the human gut: interactions between host, bacteria and phages. Nat Rev Microbiol. 2017;15:397–408.
107. Defoe D, Johnson C. A general history of the robberies and murders of the Most notorious Pyrates, and also their policies, discipline and government, from their first rise and settlement in the island of Providence, in 1717, to the present year 1724. 1724.
108. Henson MW, Pitre DM, Weckhorst JL, Lanclos VC, Webber AT, Thrash JC. Artificial Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico. mSphere. 2016;1(3). https://doi.org/10.1128/mSphere.00124-16.
109. Kang I, Kim S, Islam MR, Cho J-C. The first complete genome sequences of the acI lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-amplification of dilution-to-extinction cultures. Sci Rep. 2017;7:42252.
110. Williamson SJ, Paul JH. Environmental factors that influence the transition from lysogenic to lytic existence in the phiHSIC/*Listonella pelagia* marine phage-host system. Microb Ecol. 2006;52:217–25.

111. Middelboe M, Holmfeldt K, Riemann L, Nybroe O, Haaber J. Bacteriophages drive strain diversification in a marine *Flavobacterium*: implications for phage resistance and physiological properties. Environ Microbiol. 2009;11:1971–82.
112. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczypinksi B, Fraser CM. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J Bacteriol. 2003;185:5220–33.
113. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, et al. The marine viromes of four oceanic regions. PLoS Biol. 2006;4:e368.
114. Borbély G, Kaki C, Gulyás A, Farkas GL. Bacteriophage infection interferes with guanosine 3′-diphosphate-5′-diphosphate accumulation induced by energy and nitrogen starvation in the cyanobacterium *Anacystis nidulans*. J Bacteriol. 1980;144:859–64.

Appendix 5 Holger H. Buchholz, Kimberly Halsey, Stephen Giovannoni (2021). Viral Impacts on volatile organic carbon and energy cycling in the ocean, case for support of thesuccessful fellowship proposal for Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology (Award ID:879226)

Viral impacts on volatile organic carbon and energy cycling in the ocean

Volatile organic compound (VOC) cycling among plankton is a rapidly evolving research topic because: (1) new methods are making measurements of these compounds more accessible, (2) new data is showing a larger role for VOCs in carbon cycling, although there is considerable uncertainty in these estimates, and (3) VOC impacts on the atmosphere are being scrutinized because of increasing climate concerns. This proposal explores new questions in this untamed field, notably about the diversity and amount of VOCs metabolized by plankton and the impact of viral predation, which is particularly interesting because of the unusual partitioning of VOC oxidation between carbon and energy metabolism, a frequent focal point of viral control of host metabolism.

VOCs are a diverse group of chemicals that include a large range of climate-active gases such as acetaldehyde, acetone, methanol, alkenes and alkanes, methyl halides, DMS and isoprene. The oceans are a major source of atmospheric VOCs [1], and turnover of VOCs represents a significant fraction of total carbon oxidation at the ocean surface [2–4]. In the marine environment VOCs are released by phototrophic microorganisms during primary production and consumed by heterotrophic bacteria. It has been hypothesized that some VOCs are intermediates of cell metabolism that "leak" from cells due to hydrophobicity and low molecular weight of VOCs, or VOCs could be released as waste products. Overall, the role of VOCs in marine microbial systems is complex and not well understood, however, they are proposed to support specific synergistic or antagonistic interactions with other microorganisms [5,6].

Methanol is the single largest component of VOCs in the atmosphere and involved in the production of ozone in the troposphere and other climate-relevant gases. Previous work by Halsey has shown that metabolism of methanol by marine heterotrophic bacteria represents a major link between primary production and conversion of carbon back to atmospheric CO_2 [7–11]. Methylotrophy is thus an important component of global carbon budgets. Classification of marine environments as source or sinks of VOCs is a function of the balance between rates of production and consumption. Phytoplankton and algae excrete methanol with rates of 6 and 379 attomols per cell per day, respectively, peaking in late exponential growth [8,12]. Consumption of methanol is thought to be largely due to catabolism by methylotrophs, such as the globally ubiquitous OM43 bacterial clade, but also oxidation by other microbes such as the abundant SAR11 clade. In the surface ocean, peak abundance of OM43 coincides with phytoplankton blooms, which serve as their main carbon source [9]. Halsey used isotope labelled 14C-methanol to show that the OM43 clade could use methanol for both biomass (assimilation) and energy production (oxidation) at a rate of 43 attomols per cell per day. Though uncertain, it is thought that the element lanthanum might govern the biogeography of OM43, since it is a cofactor for the crucial and abundant methanol dehydrogenase protein (XoxF) in this cell [13,14]. However, metabolism of methanol and other VOCs by OM43 as well as the underlying biochemistry remains poorly understood. Understanding assimilation of VOCs by OM43 is critical to interpreting known ecological patterns of *in situ* carbon budgets in contemporary and future marine systems.

To date, the influence of viral predation on VOC cycling has not yet been explored. Viruses are major drivers of microbial mortality and global carbon biogeochemistry. Viral lysis kills ~20% of marine microbes per day, increasing carbon export to depth and releasing cell-associated biomass enriched in labile carbon into the water. This process catalyses microbial community productivity and respiration rates (known as the "viral shunt"). Methanol metabolism in global oceans is a critical link between primary production and carbon oxidation, but little is known about the influence of viral infection on this process. During infection, viruses hijack host metabolism, redirecting metabolites towards energy production to power virion synthesis [15]. Infection of the coccolithophore *Emiliania huxleyi* has been shown to decrease photochemical efficiency [16,17], and is therefore likely to influence light-dependent production of VOCs [10,18]. In heterotrophs viral hijacking has shown drastic reprogramming of resource acquisition, and central carbon and energy metabolism [19]. Hijacking of OM43 could increase methanol oxidation rates for increased energy production to power virion synthesis. Buchholz recently isolated the first phages infecting novel OM43 strains from the Western English Channel [20]. However, genomic analysis did not allow for the identification of metabolic hijacking mechanisms, and it is therefore unknown how OM43 phages impact the host's ability to metabolize methanol or other VOCs. Viral predation could play an

important role in driving marine systems towards methanol sink status by increasing consumption rates of infected OM43.

AIMS AND OBJECTIVES

We will quantify microbial VOC cycling by investigating the impact of viral infection on phototrophic VOC production, and methylotrophic VOC metabolism in the ecologically important OM43 clade.

1. We will test the hypothesis that production of VOCs by phototrophs is reduced during viral infection, as nutrients and energy are redirected towards viral production.
2. We will quantify the influence of viral predation on OM43 metabolism of phototrophic VOCs.
3. We will perform the first experimental evaluation of viral metabolic hijacking in the OM43 clade by combining the experiments in (1) and (2) with parallel cultures to quantify transcriptional responses in relation to nutrient acquisition in both phototrophs and methylotrophs.

STATE OF THE ART EXPERIMENTAL SYSTEM

The first host-virus systems for abundant marine methylotrophs: The current lack of research into the impact of viral infection on VOC cycling is due to a lack of appropriate model systems. Due to their slow growth rate and specialist substrate requirements, culturing streamlined heterotrophs required a decade of method development [21]. Dilution-to-Extinction (DtE) culturing, where one or two cells are placed in a well with appropriate medium, enabled these fastidious taxa to grow in the absence of fast-growing competitors. Similar methods were used to isolate the first representatives of OM43 [13,22]. Buchholz recently developed high-throughput methods to isolate fastidious taxa and their viruses through DtE culturing [20]. With the addition of methanol to a defined medium [23], we isolated three novel OM43 as well as the first viral isolates for the OM43 clade (Fig. 1). 16S rRNA analysis revealed OM43 isolate H5P1 as the most abundant OM43 in metagenome time series from the Western English Channel and the Sargasso Sea. Buchholz's DtE methods were also deployed to isolate the OM43 phage Melnitz from the Bermuda Atlantic Time Series (BATS) station during the AE1916 cruise of the BIOS-SCOPE project, we now have four different Melnitz strains in culture that can be reliably propagated. Metagenome mapping against Melnitz shows that this phage is ubiquitously distributed in global ocean viromes, making it a perfect model system for studying OM43 phage-host dynamics. **We will leverage the new host-virus systems for OM43 to determine the influence of viral metabolic hijacking on VOC metabolism *in vitro*.**

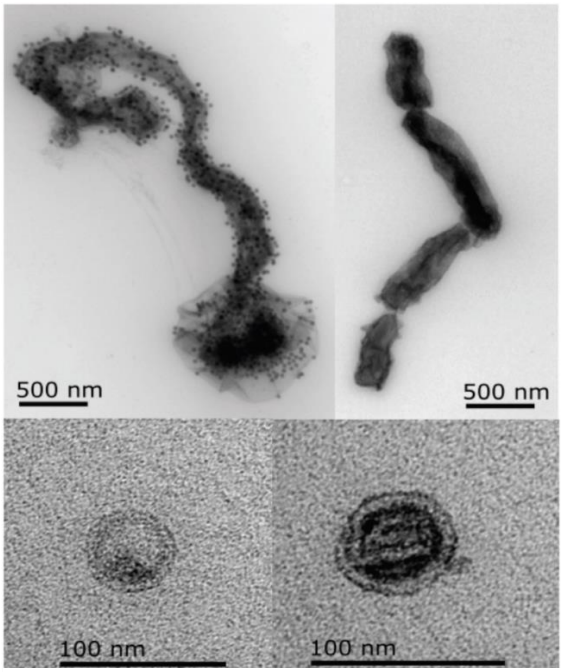


Figure 1: First viruses infecting the OM43 clade isolated from the Western English Channel. Top left: infected cells; Top Right: uninfected cells; Bottom panels: Virions of isolated methylophages.

PTR-MS enables accurate measurements of volatile carbon production: Understanding the microbial ecology of VOCs such as methanol has been challenging due to the difficulty in accurately measuring VOCs at low concentrations. Halsey coupled proton-transfer mass spectrometry (PTR-MS) with environmentally controlled dynamic stripping chambers that force liquid and gas phases into a steady state. Paired with culturing vessels, this enabled the first direct measurement of light-dependent production of methanol in algal species *Thalassiosira pseudonana* and *Dunaniella tertiolecta*, confirming phytoplankton as source of methanol in ocean surface waters [10]. **We will use PTR-MS to test the hypothesis that viral infection of phototrophs reduces VOCs production by redirecting energy away from photosynthesis. We will compare VOC profiles of infected and uninfected OM43 cultures growing on phototrophic VOC to estimate the range of VOCs metabolised by OM43, as well as the impact of viral predation on these systems.**

PROGRAM OF WORK

WP1. Measuring impact of viral infection on VOC production in *E. huxleyi:* We will experimentally determine if infection by viruses decreases VOC production rates in a model primary producer. The coccolithophore *Emiliania huxleyi* CCMP374 will be grown in f/2+Si artificial seawater [24] to mid-exponential phase in four replicate axenic cultures, then infected with virus EhV207 at a virus:host ratio of one. VOC production will be compared to four replicate controls with *E. huxleyi* treated with heat-killed viruses, by measuring VOCs in the headspace using a multiport inlet for PTR-TOF/MS as described previously [25]. "Blank" medium controls will be measured for VOCs to control for background VOCs from the medium. Supporting data (photosynthetic efficiency, chlorophyll concentrations, cell counts, virus particle counts) will be collected every 24 hours for two weeks, with additional samples taken every 6 hours after viral infection for 24 hours to assess early-, middle-, and late infection stages. ***Output: Estimated parameter of viral influence on VOC production in a model phototroph.***

WP2. Evaluating the impact viral infection on the metabolism of phototrophic VOCs by OM43: We will prepare a 20 L carboy with a *E. huxleyi* culture using the same abiotic conditions as in WP1. The VOCs from the headspace of the phototrophic culture will be vented into eight cultures with OM43 Methylophilales sp. H5P1 growing in ASM1 artificial seawater medium [26] using a peristaltic pump. Four of the replicate cultures will be treated with heat-killed myovirus Melnitz, and the other four replicate cultures will be infected in mid-exponential phase with live Melnitz phages at a virus:host ratio of one. Both treatments will be compared to additional medium "blank" controls. Rates of VOC production will be measured using the multiport system for PTR-TOF/MS as described previously[25]. The comparison between these VOC profiles will allow us to estimate the VOC consumption by OM43; the difference between infected and uninfected OM43 cultures will be used to estimate the impact of viral predation on methylotrophic VOC metabolism rates. Corresponding supporting data will be collected in the same manner as in WP1. ***Output: Estimated parameters of production and metabolism of phototrophic VOC by the OM43 clade; the first estimation of the viral impact on OM43 methylotrophy.***

WP3. Evaluating transcriptome responses to viral predation and VOC metabolism: In parallel to the VOC experiments in WP1 and WP2, we will prepare additional 100 mL cultures of OM43 and *E. huxleyi* in acid-washed polycarbonate flasks. OM43 cultures will be vented to capture VOCs produced by a single culture of *E. huxleyi* grown in a 20L carboy. We will destructively sample triplicate cultures for RNA-Seq transcriptomics in mid to late exponential phase using the protocol developed previously [27,28]. **Output: Metabolic pathways corresponding to VOC profiles from WP1 and WP2; the first mechanistic understanding of viral hijacking in the OM43 clade.**

1. Fischer, E. V. et al. Geophys. Res. Lett. 39, (2012); 2. Dachs, J. et al. Geophys. Res. Lett. 322, (2005); 3. Ruiz-Halpern, S. et al. Limnol. Oceanogr. 55, 1733–1740 (2010); 4. Hauser, E. J. et al. Limnol. Oceanogr. Methods 11, 287–297 (2013); 5. Amin, S. A. et al. Nature 522, 98–101 (2015); 6. van Tol, H. M. et al. ISME J. 11, 31–42 (2017); 7. Thrash, J. C. et al. ISME J. 8, 1440–1451 (2014); 8. Beale, R. et al. Mar. Chem. 171, 96–106 (2015); 9. Halsey, K. H. et al. Environ. Microbiol. 14, 630–640 (2012); 10. Halsey, K. H. et al. et al. Limnol. Oceanogr. 62, 2650–2661 (2017); 11. Sun, J. et al. PLoS One 6, e23973 (2011); 12. Dixon, J. L. et al. Biogeosciences 8, 2707–2716 (2011); 13. Giovannoni, S. J. et al. Environ. Microbiol. 10, 1771–1782 (2008); 14. Good, N. M. et al. Sci. Rep. 9, 4248 (2019); 15. Warwick-Dugdale, J. et al. Virol. J. 16, 15 (2019); 16. Kimmance, S. A. et al. Mar. Ecol. Prog. Ser. 495, 65–76 (2014); 17. Gilg, I. C. et al. Mar. Ecol. Prog. Ser. 555, 13–27 (2016); 18. Mincer, T. J. & Aicher, A. C. PLoS One 11, e0150820 (2016); 19. Howard-Varona, C. et al. ISME J. (2020) doi:10.1038/s41396-019-0580-z.; 20. Buchholz, H. H. et al. ISME J. 1–14 (2021); 21. Giovannoni, S. J. Ann. Rev. Mar. Sci. 9, 231–255 (2017); 22. Jimenez-Infante, F. et al. Appl. Environ. Microbiol. 82, 1215–1226 (2016); 23. Carini, P. et al. ISME J. 7, 592–602 (2013); 24. Guillard, R. R. R. & Ryther, J. H. Can. J. Microbiol. 8, 229–239 (1962); 25. Davie-Martin, C. L. et al. Frontiers in Marine Science 7, 1144 (2020); 26. Carini, P. et al. ISME J. 8, 1727–1738 (2014); 27. Smith, D. P. et al. PLoS One 5, e10487 (2010); 28. Smith, D. P. et al. MBio 4, e00133–12 (2013);