



A Novel and Ubiquitous Marine Methylophage Provides Insights into Viral-Host Coevolution and Possible Host-Range Expansion in Streamlined Marine Heterotrophic Bacteria

Holger H. Buchholz,^a Luis M. Bolaños,^a Ashley G. Bell,^a Michelle L. Michelsen,^a Michael J. Allen,^{a,b}  Ben Temperton^a

^aUniversity of Exeter, School of Biosciences, Exeter, UK

^bPlymouth Marine Laboratory, Plymouth, UK

ABSTRACT The methylophage 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 between marine primary production and remineralization of carbon back to the atmosphere. Bacteriophages shape microbial communities and are major drivers of 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 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 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 are the foundations on which the mechanistic understanding of virus-host interactions and parameterization of bioinformatic tools for viral ecology are based. This study isolated and characterized the first myophage known to infect the OM43 clade, expanding our knowledge of this understudied group of microbes. The nearly identical genomes of four strains of Melnitz isolated from different marine provinces and the 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 may not be necessarily

Editor Laura Villanueva, Royal Netherlands Institute for Sea Research

Copyright © 2022 Buchholz et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ben Temperton, b.temperton@exeter.ac.uk.

The authors declare no conflict of interest.

Received 9 February 2022

Accepted 10 February 2022

Published 21 March 2022

restricted to phylogenetically related hosts but is a function of shared physical and biochemical properties of the cell.

KEYWORDS OM43, bacteriophages, cultivation, marine microbiology, virus-host systems

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 contributes to nutrient recycling by releasing cell-bound organic compounds into the environment (4, 5). Viral infection can alter host metabolism through metabolic hijacking (6, 7), which has been shown to reprogram resource acquisition and central carbon and energy metabolism (8, 9), influencing oceanic nutrient cycles. The selective pressure of the predator-prey relationship of bacteria and phages is also a main driver of microbial evolution (10), where a constant arms race requires phages to evolve and overcome host defense mechanisms (11). 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 (12–15). However, many of these sequences lack representation in viral culture collections (16), limiting experimental determination of parameters of infection such as the 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 (16–19). However, 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 of ~1.3 Mbp) type I methylotrophs (20) of the class *Gammaproteobacteria* (20) (previously *Betaproteobacteria* [21]). The catabolism of methanol and other volatile organic compounds (VOCs) is an important link between primary production and remineralization of carbon back to atmospheric CO₂ (22–24). In the surface ocean, the peak abundance of OM43 coincides with phytoplankton blooms which provide their main carbon source (24). OM43 are particularly abundant in coastal ecosystems where they comprise up to 5% of the microbial community (25). Members of the OM43 clade are somewhat challenging to grow in the laboratory. Increased levels of auxotrophy compared to copiotrophs and largely constitutive metabolism renders them sensitive to media composition (20). As a result, only two OM43 phages have been reported, and their influence on OM43 populations 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 (26). Based on metagenomic read recruitment against reference genomes, Venkman was the third most abundant phage in the WEC sample, indicating that phages of methylotrophs are a major component of this coastal ecosystem (16). 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 (26). 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 (ANI) were isolated independently on three separate occasions: twice at station L4 in the temperate coastal WEC (April and June 2019) and from the station BATS in the Sargasso Sea (June 2019) 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 (27), and we show here that it is also a feature of their associated viruses. Like previously reported SAR11 myophages (15, 16), Melnitz also encoded the pore proteins of a curli operon that are absent in the host. Structural analysis suggests a putative reconfiguration may allow 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 toward a recent shared ancestor indicative of host transitioning between OM43 and SAR11.

RESULTS AND DISCUSSION

Phage Melnitz infecting *Methylophilales* sp. H5P1 shares viral gene 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 (16). 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. Comparison to publicly available phage genomes with CheckV (28) 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 (see Fig. S1 in the supplemental material). Capsid size is not reported due to the presence of sample preparation artifacts, which may affect accuracy of capsid dimension measurements (29).

All four phage genomes shared 99.95 to 100% average nucleotide identity across their full genomes therefore all four phages should be regarded as the same viral species (30). We named this species “Melnitz” after a character in the popular Ghostbusters franchise, continuing the theme of another previously isolated phage on OM43 (Venkman) (16). 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 summarized in Table 1. Shared gene network analysis (VConTACT2 [31]) using assembled contigs from Global Ocean Viromes (GOV2 [13]) and RefSeq (V88 with ICTV and NCBI taxonomy [32]) was performed to evaluate Melnitz phylogeny. All four Melnitz variants were assigned to a subcluster with joint membership of two clusters: Cluster_2 and Cluster_3 (Fig. 1). Cluster_2 contained an additional 15 viral contigs from nine different GOV2 viromes. Cluster_3 contained three virome contigs shared with Cluster_2, as well as 25 pelagimyophage genomes assembled from metagenomes (PMP-MAVGs) (33). The subcluster containing Melnitz also contained *Pelagibacter* myophages HTVC008M and Mosig (17, 34), suggesting they belong to the same family. A total of 18 contigs clustering with Melnitz isolates were identified from GOV2 and WEC viromes (13, 35). Genome alignments (BLASTn) of all contigs sharing the VConTACT2 cluster with Melnitz are provided in Fig. S2 in the supplemental material. Further phylogenetic analysis showed that only two environmental contigs consistently shared a branch with Melnitz (L4_2016_09_28_HYBRID_000000000048 and Station137-DCM-ALL-assembly-NODE1-length-137507-cov-14.215202), based on single shared genes encoding: tail sheath protein (see Fig. S3), a terminase large subunit (see Fig. S4), and scaffolding proteins (see Fig. S5), as well as four concatenated structural genes (Fig. 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 (33).

Metagenomic analysis shows cosmopolitan nature of Melnitz-like phages. To establish the distribution patterns of Melnitz, reads of 131 GOV2 viromes were mapped

TABLE 1 General features of OM43 strain H5P1 and four Melnitz phages isolated in this study

Phage	Culture ID	Host	Phage group	Morphotype	Genome size (bp)	G+C (%)	No. of:		Source water	Sampling date	Latitude	Longitude
							ORFs	tRNAs				
Melnitz	EXVC044M	H5P1	Melnitz	Myovirus	141,548	37.6	224	4	BATS	1 June 2019	N31°40'	W64°10'
Melnitz variant 1	EXVC043M	H5P1	Melnitz	Myovirus	141,552	37.6	226	4	BATS	1 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	1 April 2019	N50°15'	W04°13'
		H5P1	<i>Methylophilales</i> sp.	<i>Gammaproteobacteria</i>	1,336,408	34.4	1382	38	WEC	18 September 2018	N50°15'	W04°13'

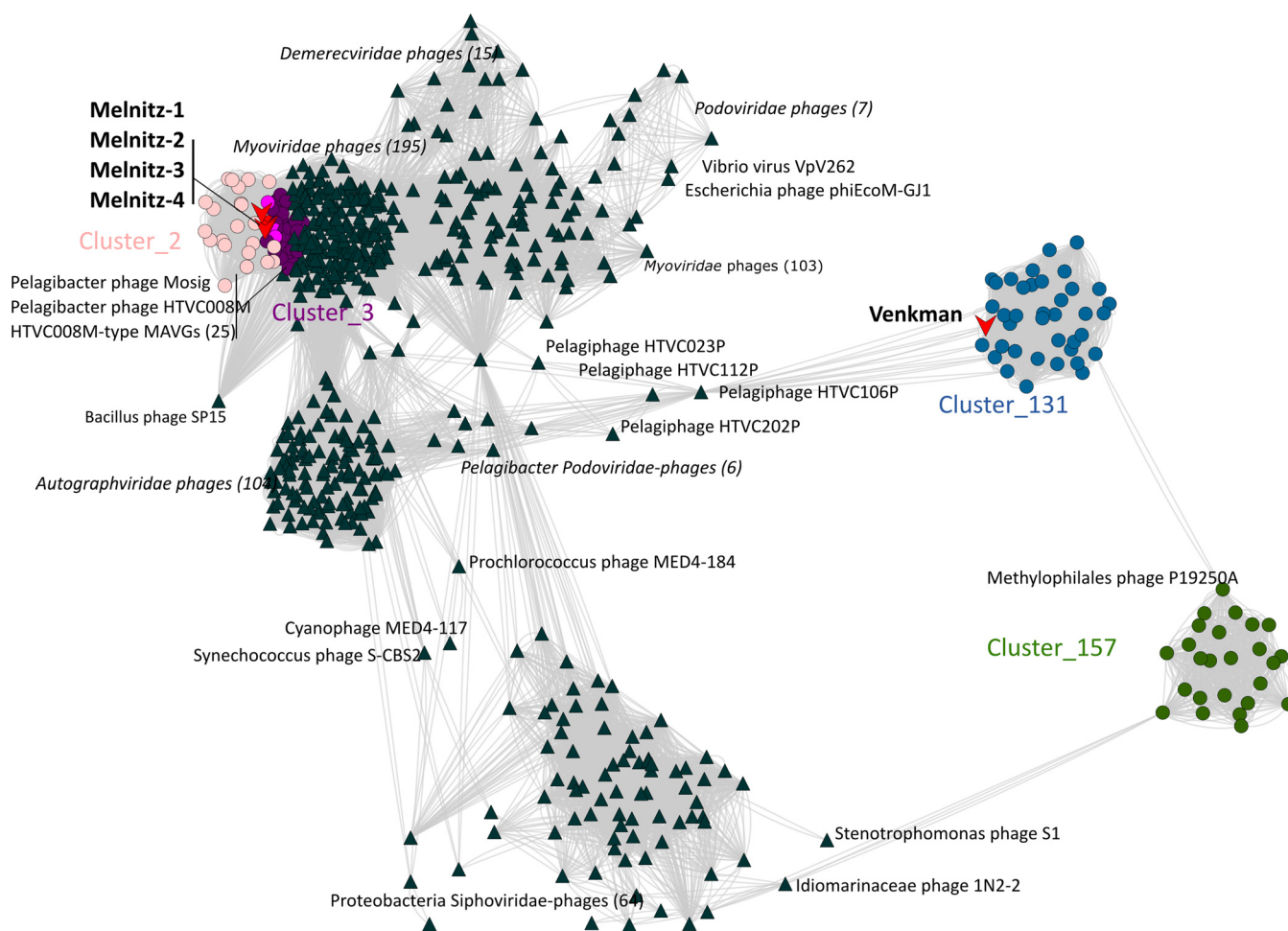


FIG 1 Viral shared gene content network of OM43 phages, related bacteriophages from the NCBI, and related sequences from the Global Ocean Virome (GOV2.0). Nodes represent viral genomes; edges represent the similarity between phages based on shared gene content. NCBI reference genomes that were greater than more than two neighboring edges from contigs of interest were excluded for clarity. Phage isolates are indicated by red arrows. Colored 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.

against phage genomes ($\geq 95\%$ nucleotide identity over $\geq 90\%$ read length; genome coverage of $>40\%$ was required to register a phage as “present” to avoid false positives [36, 37]) (see Fig. S6). The relative abundance of phages was calculated based on the number of reads mapped to a contig, normalized by contig length and sequencing depth (mapped reads per kilobase pair of genome per million reads [RPKM]). Linear regression of relative abundance of all three phages known to infect OM43 (Melnitz from this study, plus Venkman and MEP301) as well as the *Pelagibacter* phages HTVC008M and Mosig (genomes related to Melnitz) did not show significant correlation to depth (see Fig. S7). In water samples from the epipelagic zone (max. 200m depth) abundance of OM43 phages Melnitz, Venkman and MEP301 as a function of temperature showed a significant and negative relationship (Melnitz: $P = 2 \times 10^{-6}$, $R^2 = 0.277$; Venkman: $P = 9 \times 10^{-7}$, $R^2 = 0.406$; MEP301: $P = 0.008$, $R^2 = 0.154$). Relative abundance was positively correlated with absolute latitude (Melnitz: $P = 2 \times 10^{-5}$, $R^2 = 0.226$; Venkman: $P = 6 \times 10^{-7}$, $R^2 = 0.416$; MEP301: $P = 0.015$, $R^2 = 0.13$). However, for MEP301, temperature and latitude correlation to relative abundance is weak and likely a poor explanation for its distribution. Venkman and Melnitz correlation analyses suggest that these viruses are most abundant in colder regions. In contrast, relative abundances of Melnitz-related *Pelagibacter* phages HTVC008M and Mosig were not significantly correlated with temperature (Mosig: $P = 0.032$, $R^2 = 0.032$; HTVC008M: $P = 0.096$, $R^2 = 0.019$) or latitude (Mosig: $P = 0.096$, $R^2 = 0.018$; HTVC008M: $P = 0.048$, $R^2 = 0.03$).

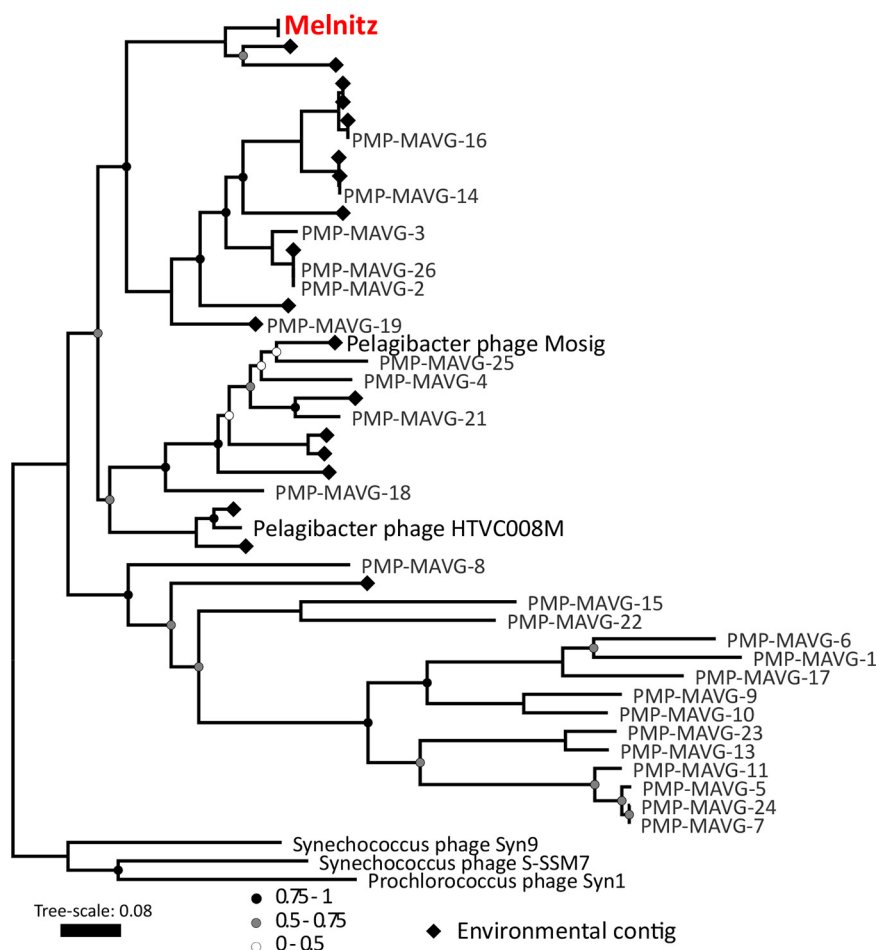


FIG 2 Phylogenetic tree of metagenomic contigs and marine Melnitz-type myophages. Neighbor-joining maximum-likelihood tree (500 bootstraps) based on four individually aligned and concatenated structural genes (capsid assembly, major capsid, sheath subtilisin, terminase large subunit) of genomes and contigs 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) data set (13), which were omitted for clarity. A fully labeled tree is available in Fig. S3 in the supplemental material.

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. While 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. To further evaluate the abundance of isolated methylophages over depth and time, 382 metagenomic samples from the ALOHA station (located in the North Pacific Subtropical Gyre [NPSG]) (38) were randomly subsampled to five million reads (without replacement) and mapped against phage genomes using the same thresholds. None of the known OM43 phage isolates were identified as present in the samples using a minimum cutoff 40% genome coverage (data not shown). Related SAR11 myoviruses (Mosig and HTVC008M) were also absent, with the exception of 462 RPKM recruited by HTVC008M in a single sample (HSD20-02a-277-S2C009-0200-170214). The low representation of isolated myoviruses for either SAR11 or OM43 in this region might indicate that unknown local viral strains (with at least 60% difference across the genomes) dominate, or that these viruses are rare in the NPSG at all depths. It is also worth noting that while GOV2 and WEC viromes were produced using iron chloride flocculation to concentrate viral particles, those from ALOHA were produced by capturing viral particles on 0.02- μ m filters.

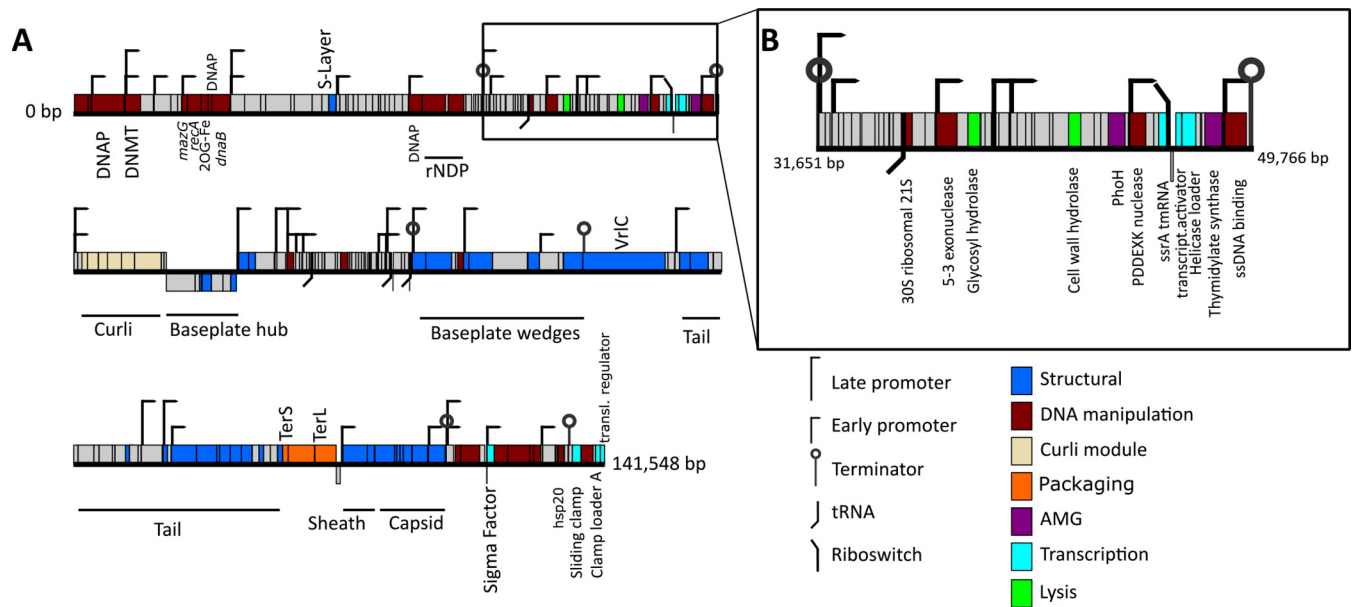


FIG 3 Gene map showing identified genomic features of the OM43 phage Melnitz. (A) Gene map of the 141,548-bp genome of Melnitz contains 143 hypothetical ORFs (62%) without known function (indicated gray). (B) Section of the Melnitz genome between two terminator sequences that contains the *ssrA* gene, the glutamine riboswitch, and the transcription coactivator gene for transcription-translation regulation.

Therefore, it is possible (although unlikely) that the absence of Melnitz-like genomes in the NPSG could result from methodological differences.

It has been demonstrated that the marine biosphere maintains persistent bacterial “seed banks” (39), meaning there is a high probability that any given marine bacteria can be found in any marine ecosystem, albeit in extremely low abundance, awaiting favorable conditions for growth. Similarly, many viruses infecting globally distributed bacteria such as SAR11 are found in viromes from all oceans (19), suggesting the majority of viral species are shared between oceans (40). 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 (~5,000-km distance between sites) suggests either low population-level variance or that the isolation conditions used favor 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 favorable conditions select them for propagation (41, 42). 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 and 141,552 bp in length, all with a G+C content of 37.6% (Table 1), similar to that of their hosts (34.4%). For each, 224 open reading frames (ORFs) were predicted (using five different gene callers and manual curation [43]), except Melnitz-2, which had two additional small ORFs (NCBI accession numbers [QZI94509.1](#) and [QZI94511.1](#)) of unknown function (226 ORFs total). Four additional tRNA sequences for all four genomes were identified (genes 59, 125, 147, and 153) (tRNAscan [44] v2.0, ARAGORN v1.2.38 [45]), and functional annotation (using BLASTp [46] phmmer v.241.1 against Pfam [47] and Swiss-Prot [48]) of ORFs suggest that all Melnitz strains encoded the same set of genes. Out of 224 ORFs, 143 (~63%) had unknown function (Fig. 3A). The tail-assembly associated region in the Melnitz-4 genome between bp 99465 and 105981 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 [49]) 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 (50, 51), all four Melnitz variants had identical host ranges when screened against a panel of other OM43 isolates from the WEC (*Methylophilales* sp. strains C6P1, D12P1, and H5P1) (15; data not shown). Melnitz possessed a set of structural genes typically associated with T4-type myophages (52), including T4-like base-plate, tail tubes, base plate wedges, tail fibers, virus neck, tail sheath stabilization, pro-head core and capsid proteins. Melnitz encodes orthologs of the auxiliary metabolic genes (AMGs) *mazG* (gene 10) and *phoH* (gene 81), which are involved in cellular phosphate starvation induced stress responses and are a common feature of phages from P-limited marine environments (53–55). In addition, *hsp20* (gene 223) was identified, which, together with *mazG* and *phoH*, is considered part of the core genes in T4-like cyanomyophages (53, 56). Both Melnitz and its host, *Methylophilales* sp. H5P1, encode a type II DNA methylase to potentially protect DNA against cleavage by restriction endonucleases (57). In T-even phages, DAM methylase (similar to type II DNA methylase) protects phage DNA from restriction endonucleases through competitive inhibition (58); thus, DNA methylase in Melnitz (gene 5) could also be involved in protection from host restriction endonucleases during infection.

For DNA replication and metabolism, Melnitz encodes *nrdA/nrdE* (KEGG Orthology KO00525) and *nrdB/nrdF* (KEGG Orthology KO00526) genes (genes 39 and 41) that together form active ribonucleotide reductases for catalyzing the synthesis of deoxyribonucleoside triphosphates required for DNA synthesis (59). Additional DNA replication- and manipulation-associated genes were *polB* (DNA polymerase beta, gene 4), *recA* (DNA recombinase, gene 11), and *dnaB* (helicase-like, gene 14), as well as two DNA primases (gene 13 and 38) and a 2OG-Fe oxygenase (gene 12). The *rpsU* gene (gene 60, encoding the bacterial 30S ribosomal subunit 21S) was found downstream adjacent to gene 59—a tRNA-Arg (*tct*). Virally encoded 21S genes were previously identified in the *Pelagibacter* phages HTVC008M (17) and Mosig (34), as well as the OM43 phage Venkman (16). It is thought that the 21S subunit in phages might be required for initiating polypeptide synthesis and mediation of mRNA binding (60); the proximate tRNA sequence may support a translational and protein synthesis role of the viral ribosomal gene. Within the same region, Melnitz's Gene_23 was annotated as a putative surface layer-type (S-layer) protein, previously associated with a strong protection mechanism against superinfection in bacteriophages of *Bacillus* spp. and *Pseudomonas aeruginosa* (61–63). Though protection against superinfection is sometimes considered to be more important in temperate phages, the lytic *Escherichia coli* T4-like phage Spackle also possesses structural defense mechanisms against superinfection from closely related phages (64). Therefore, 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 organize into early, middle and late stages (65). 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 (52, 66). Like other marine T4-like phages, such as those infecting streamlined cyanobacteria (67), Melnitz lacked identifiable middle promoters. Reduced numbers of regulatory elements are a common feature of genomically streamlined bacteria such as SAR11 and OM43 (68). 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 an 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 (gene 215), gp45-like sliding clamp C terminal (gene 226), clamp loader A subunit (gene 91), transcription coactivator (gene 90), and a translational regulator protein (gene 1). 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 σ -factor protein and host RNA polymerase (69). To activate it, the sliding clamp is loaded by a clamp-loader DNA polymerase protein complex, allowing the complex to move along DNA strands (70, 71). 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. Stalling of ribosomes can occur when mRNAs lack stop codons (nonstop mRNA), which may cause sequestration of ribosomes and production of defective polypeptides. *ssrA* encodes 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. An additional role of tmRNA is to add a hydrophobic peptide tag onto nonstop mRNA, instigating proteolysis of the polypeptide (72). Permuted two-piece tmRNA are transcribed as a single precursor RNA but split into two RNA molecules (comprising a tRNA-like domain and an mRNA-like domain) (73). Two-piece tmRNA are found in most *Alphaproteobacteria*, and some groups of *Gammaproteobacteria* and *Cyanobacteria* (74), whereas other bacteria encode one-piece tmRNA. In Melnitz, the *ssrA* gene (gene 86) encoding a two-piece tmRNA was situated on the same operon and directly upstream of the transcription coactivator (Fig. 3B). Host-encoded *ssrA* can be used as sites of integration for prophages in deep-sea *Shewanella* isolates (62), and fragments of *ssrA* have previously been identified in prophage genomes (63), 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 possible gene exchange with a temperate prophage may explain the presence of a complete *ssrA* gene in Melnitz. A possible role for a complete viral tmRNA is to help maintain the hijacked bacterial machinery. Alternatively, the phage might use tmRNA to selectively tag and degrade host proteins to recycle amino acids for viral protein synthesis. Since the tmRNA is located immediately downstream of an early promoter, we postulate that this occurs early during infection to minimize degradation of viral proteins. We evaluated the frequency of *ssrA* genes found in 18,146 publicly available genomes in the phage genome database at millardlab.org (last updated on 21 January 2021), which includes all RefSeq genomes (75). We also evaluated whether those identified were either one-piece or two-piece tmRNAs. 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 (45), of which 50 were encoded by phages isolated from marine or aquatic samples (see Table S1). In contrast, only 2.5% of one-piece tmRNAs were marine, suggesting that two-piece tmRNA is a feature more prevalent in marine and aquatic phages compared to phages from other environments. Of the 47 contigs (27 of which were classed as complete by CheckV) that were clustered with Melnitz based on shared gene content (Fig. 1), 17 had permuted tmRNA sequences, 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 (76). Riboswitches are a common regulatory mechanism in streamlined marine bacteria due to their low metabolic maintenance cost compared to protein-encoded promoters and repressors (68). Like tmRNAs, riboswitches are a rare feature in phages. A phage-encoded riboswitch putatively controlling regulation of *psbA* was previously identified in a cyanophage (71). Of the 133 isolate phage genomes encoding tmRNAs only two phages other than Melnitz possessed a riboswitch (both glutamine): *Pelagibacter* phage Mosig (34) and *Prochlorococcus* phage AG-345-P14 (77). Additional riboswitches were identified in metagenomic assembled virome contigs. In 10 metagenomically assembled viral genomes of pelagimyophages thought to infect SAR11 (PMP-MAVGs) and five GOV2 contigs related to

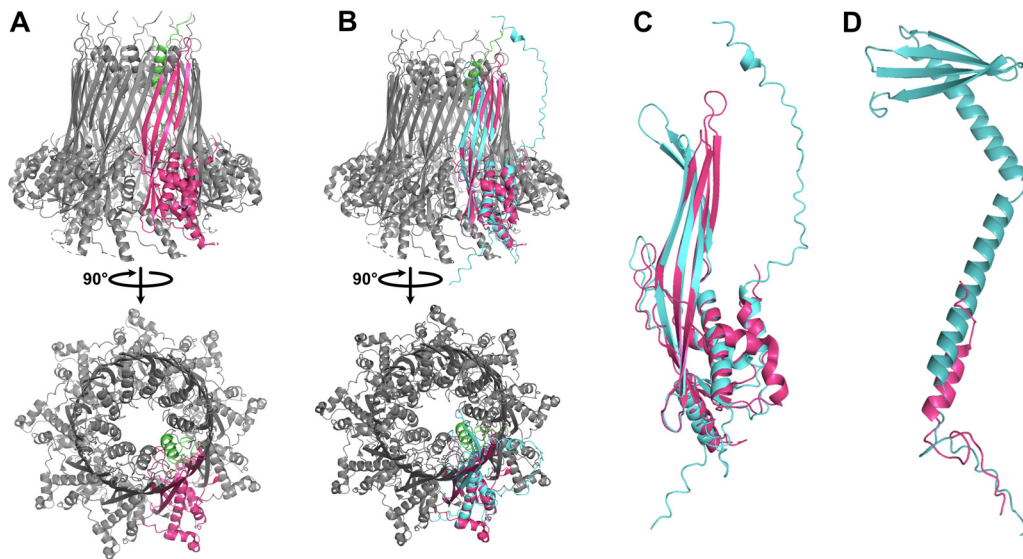


FIG 4 Structural prediction of CsgGF complex encoded by Melnitz. (A) Predicted structure of CsgGF complex in *E. coli* (PDB model 6L7A) comprises a hetero-18-mer with 1:1 stoichiometry of CsgG (pink) and CsgF (green), forming a pore in the outer membrane with two constrictions: one provided by CsgG at 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.

Melnitz, only one of these contigs had both tmRNA and a riboswitch; 11 PMP-MAVGs encoded neither. This suggests that (i) glutamine riboswitches are a common but not defining feature in Melnitz-like marine phages and that (ii) 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, and MBRSH7, as well as *Methylophilum planktonicum*, *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 (78). Phages infecting *Synechococcus* have been shown to use extracellular nitrogen for phage protein synthesis (79), 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 fibers that are part of biofilm formation. 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 fibers (80). 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 ~ 12.9 Å in diameter. CsgF forms a secondary channel ~ 14.8 Å in diameter at the neck of the beta barrel (Fig. 4A; see also Fig. S8A and B) and assists in excretion of the amyloid fiber. Like previously described pelagimyophages (33), Melnitz possesses *csgF* and *csgG*, but lacks the genes for amyloid fiber production: *csgA* and *csgB* (Fig. 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 fibers for “sibling capture” of proximate host cells (33). However, we saw no evidence of “clumping” as a result of sibling capture, neither in cytograms nor TEMs, under the culturing conditions used in this study. In addition, 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 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. Since streamlined genomes would be expected to lose residual or otherwise unnecessary genes encoding incomplete machinery (27), it is unlikely that curli-associated phage encoded genes supplement a partial operon in the host. The presence of curli genes across multiple phage species suggests that their presence provides the virus with a competitive advantage that has yet to be identified.

Structural homology modeling of the Melnitz-encoded CsgGF pore with Swiss-Model (81) yielded low structural similarity to known CsgGF structures (GMQE < 0.5, with Q-MEAN scores < -3 and z-scores $\gg 2$ outside normalized Q-mean score distributions from a nonredundant set of structures from PDB). Structural similarity was greatest in regions 238 to 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 (82) 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: (i) a narrowing of the barrel at the neck and (ii) an extension of the α -helix outside the barrel (Fig. 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 (Fig. 4D). Similar modeling of CsgG and CsgF encoded by pelagiophage HTVC008M also contained these features (see Fig. S9A 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 14.7 Å in diameter, matching that provided by CsgF in *E. coli* (see Fig. S8B 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 ~ 15 Å in diameter in the inner membrane, resulting in rapid membrane depolarization and subsequent activation of membrane-bound lysins (83).

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 (Fig. 4C) might interact to stabilize the elongated hinged α -helices of CsgF. The C-terminal β -sheets of CsgF could then form a second channel beneath the CsgG barrel (see Fig. S10A). 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) (84). 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 predicted to be positively charged and narrower (14.7 Å compared to 20.6 Å 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. Using cryo-electron tomography could resolve the cell surface and curli protein structure or provide evidence of extracellular amyloid fibers and/or curli pores.

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 coactivator genes (Fig. 3). Overexpression of thymidylate synthase in the D29 phage infecting *Mycobacterium tuberculosis* results in delayed lysis and higher phage yields (85). In phage T4, postponing lysis is used to delay the release of viral progeny until conditions are favorable, thereby maximizing virion production before viral release and successful replication after (86). In the T4-like Melnitz, *thyX* could therefore be involved in postponing lysis as well. Lytic control may putatively be related to the upstream glutamine riboswitch. Host 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 (87). 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 pore, which may result in rapid depolarization of the membrane, triggering activation of lysins and rapid cell lysis.

Structural homology modeling (Swiss-Model) 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 localized gaps in the peptidoglycan layer for assembly of type IV secretion systems (88). 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 (see Fig. S11) (89). Outside this active site, structural similarity to T4 endolysin (PDB model 2561.1) was low (13% identity over 51% coverage; GMQE = 0.18). Melnitz lacked any other lysin-like genes. Like the T4-encoded endolysin *e* (90), gp67 is under the control of an early transcription promoter (Fig. 3B). We therefore propose that the gp67 gene in Melnitz potentially serves two functions: (i) as an endolysin for degradation of cellular peptidoglycans during lysis or (ii) as an autolysin to enable assembly of the CsgGF or CsgG-only pore protein. Phages infecting *Streptococcus pneumoniae* upregulate host-encoded autolysins alongside phage-encoded lysins during late-stage infection to accelerate cell lysis (91). 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 fibers that enable the phage to adsorb and inject its genetic material. Mutation in phage proteins, either through point mutations or recombination during coinfection can result in host range expansion or transition (92). 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 fibers (93). 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

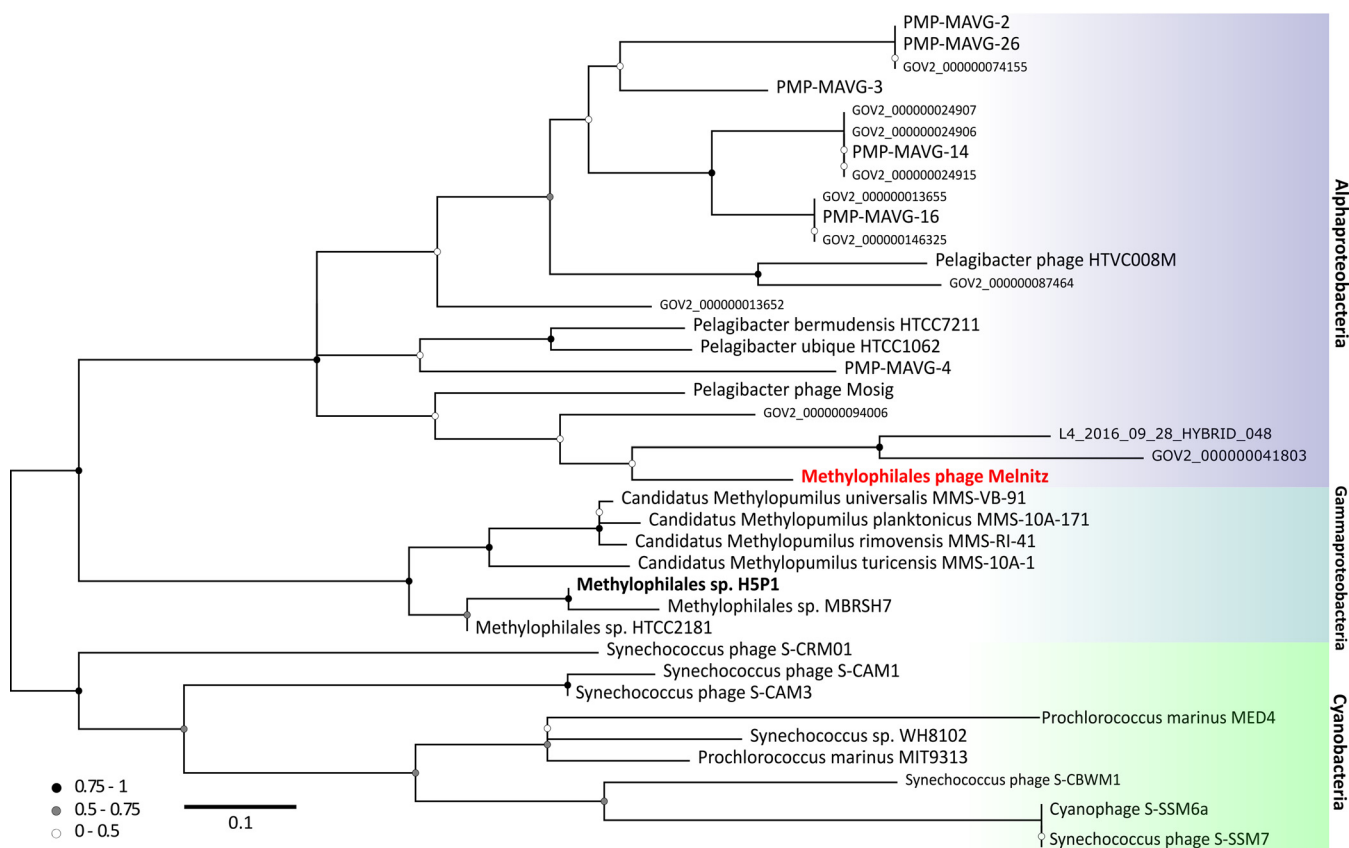


FIG 5 Phylogeny of tmRNA genes in major marine lineages. Neighbor-joining maximum-likelihood tree (100 bootstraps) of tmRNA genes found in marine phages and host lineages (not exhaustive) suggests that three the three known major lineages between *Cyanobacteria*, *Gammaproteobacteria*, and *Alphaproteobacteria* 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.

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* [21]), and *Cyanobacteria* (73, 74). 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* (Fig. 5). Second, two of the four tRNAs encoded by Melnitz were more closely related to tRNAs encoded by pelagiomyophages 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 (94). Phage-encoded tRNAs also enable phages to optimize 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 (95). Indeed, tRNAs are often used for computational host prediction of phages due to high sequence conservation of the gene between host and viral forms (96, 97). 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 associated 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 (Fig. 6). A list of tRNAs used is provided in Table S2. Sequences encoding tRNAs were aligned using all-versus-all BLASTN with a minimum expect-value of 1×10^{-5} . Two of four tRNAs in

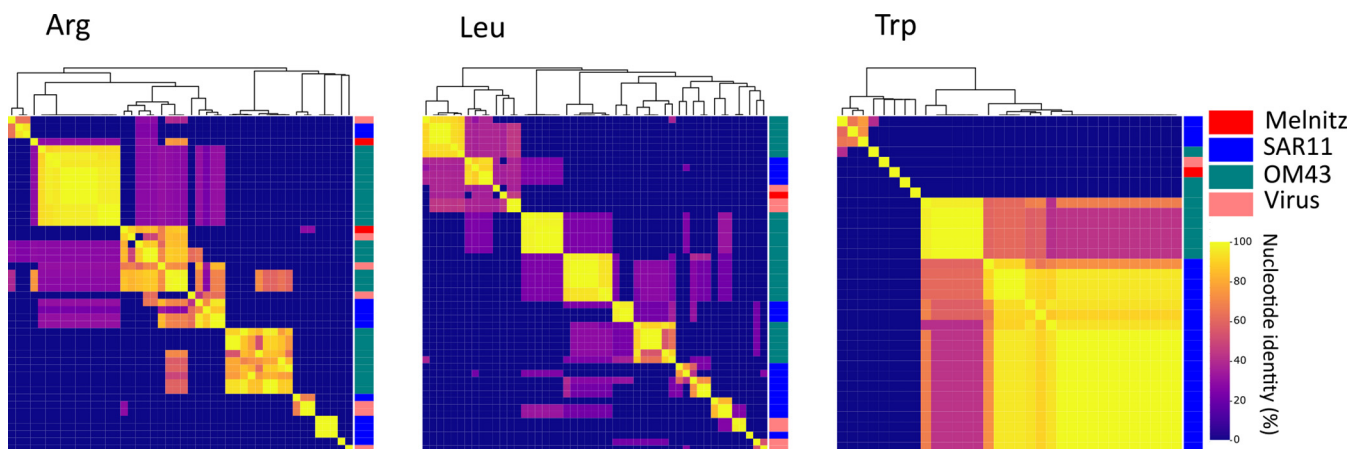


FIG 6 Alignment of tRNA genes found in Melnitz, SAR11 and OM43 lineages. Heatmaps and dendrograms (Euclidian similarity matrices) were prepared based on similarity between alignments for arginine (Arg), leucine (Leu), and tryptophan (Trp) tRNA genes found in OM43 phage Melnitz, OM43, and SAR11, as well as tRNA derived from isolated SAR11 and OM43 phages.

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 with PMP-MAVG-17, previously classified as *Pelagibacter* phage (33) but only had 45% nucleotide identity. The fourth tRNA-Trp (CCA) found in Melnitz did not align (E values $> 1 \times 10^{-5}$) with any tRNA found in OM43, SAR11, or any of their respective phages. Using tRNA as host 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. Full-length genome alignments of Melnitz and HTVC008M (see Fig. S12) support this hypothesis, since rare genetic features such as the curli operon, as well as gene synteny, are conserved in both species. We speculate that a recently shared host which led to the acquisition of these genes followed by viral speciation is more likely than two separate events of horizontal gene transfer of the same genetic module.

To assess the breadth of host ranges, we challenged the SAR11 strains HTCC7211 and HTCC1062 and OM43 strains H5P1, D12P1, and C6P1 (16) with the phages Melnitz and Mosig (34), but found no evidence that either phage is able to replicate in cells beyond host class boundaries as observed via cell lysis (see Fig. S13). 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 noncanonical 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 coinfecting phages (92, 98). 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 fiber tips in gene 37, yielding variants that could infect both hosts and others that could only infect *Y. pseudotuberculosis* (93). Host-prediction of viral contigs from metagenomes has identified rare phages (115 of 3,687) possibly able to infect multiple classes (99). We propose that two properties of SAR11 and OM43 increase the likelihood of such an event in natural communities. (i) 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. (ii) 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 maximize replication on minimum resources in nutrient-limited marine environments (20, 100). 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 (16). 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 coinfecting with a methylophage that enabled a host-transition event through homologous recombination. Though homologous recombination for SAR11 and/or OM43 virus-host systems have not been shown before, similar examples exist for lambdoid *E. coli* phages (101) and in marine cyanophages high rates of recombination have been reported (102). 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 coevolution 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 among viruses within gene-sharing clusters (31, 103).

Conclusion. Here, we provide evidence that supports a putative interclass host transition event between two important clades of streamlined marine heterotrophs and expands our knowledge about the dynamics and characteristics in genomically streamlined heterotrophic virus-host systems. We isolated four nearly 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 data sets 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 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 AND 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 (16). Continuous cultures were grown using artificial seawater-based artificial seawater medium (ASM1) (104) amended with 1 mM NH_4Cl , 10 μM KH_2PO_4 , 1 μM FeCl_3 , 100 μM pyruvate, 25 μM glycine, and 25 μM methionine, as well as 1 nM (each) 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP), pantothenate, biotin, pyrroloquinoline quinone (PQQ), and B_{12} . Additional 1 mM methanol and 5 μL of an amino acid mix (MEM amino acids [50 \times] solution; Sigma-Aldrich) were added per 100 mL of medium. Bacterial stocks were grown in 50 mL of acid-washed

polycarbonate flasks at 15°C without shaking using 12-h light-dark cycles. All culturing was performed in a Sanyo Versatile environmental test chamber (model MLR-350HT).

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 5 m 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 μ m pore size), a 142 mm, 0.2 μ m-pore polycarbonate filter (Merck Millipore) and a 142 mm, 0.1 μ m-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, Gottingen, Germany) and used as inoculum in a multistep viral enrichment experiment followed by dilution-to-extinction purification as described previously (16). Briefly, viral inoculum (10% [vol/vol]) was added to 96-well Teflon plates (Radleys, UK) with exponentially growing host cultures. After a 1- to 2-week incubation period, cells and cellular debris were removed with 0.1 μ m syringe polyvinylidene difluoride filters. The filtrate was added as a 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 (16). Phages were purified by dilution-to-extinction methods (our detailed protocol is available here <https://dx.doi.org/10.17504/protocols.io.bb73irqn>).

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 μ L of an amino acid mix (MEM amino acids [50 \times] solution). Wells were inoculated to 1×10^6 cells mL⁻¹ with *Pelagibacter* hosts HTCC1062 or HTCC7211 or with OM43 host 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 μ L of viral culture. The plate was then incubated at 15°C and monitored daily using flow cytometry for ~2 weeks. Successful infections were identified by 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 of acid-washed, polycarbonate flasks and infected at a cell density of about 1×10^6 to 5×10^6 cells mL⁻¹ with 10% (vol/vol) viral inoculum. Infected cultures were incubated at 15°C for 7 to 14 days, and then the lysate was transferred to 50 mL Falcon tubes. Larger cellular debris was removed by centrifugation (GSA rotor; Thermo Scientific, catalog no. 75007588) at 8,500 rpm/10,015 $\times g$ for 1 h. Supernatant was filtered through 0.1 μ m-pore-size polyvinylidene difluoride syringe filter membranes to remove any remaining smaller cellular debris. Phage particles were precipitated using a PEG8000/NaCl flocculation approach (105). Briefly, 50 mL of lysate was amended with 5 g of PEG8000 and 3.3 g of NaCl (Sigma), shaken until dissolved, and incubated on ice overnight. Precipitated phages were pelleted by centrifugation at 8,500 rpm/10,015 $\times g$ for 1 h at 4°C. After discarding the supernatant, we resuspended phage particles by rinsing the bottom of the tubes twice with 1 mL of SM buffer (100 mM NaCl, 8 mM MgSO₄·7H₂O, 50 mM Tris-Cl). DNA from the suspended phages was extracted with the Wizard DNA Clean-Up system (Promega) according to the manufacturer's instructions, using preheated 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 \times 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 BMAP v38.22 (106; available at <https://sourceforge.net/projects/bbmap/>). Contigs were assembled using SPAdes v.3.13 with default settings (107). Viral contigs were confirmed with VirSorter v1.05 (108) (categories 1 or 2, >15 kbp). The quality and completeness of the contigs, as well as the terminal repeats, were evaluated using CheckV v0.4.0 with standard settings (109). Open Reading Frames (ORFs) were identified with phanotate v2019.08.09 (110) and imported into DNA Master for manual curation (43) using additional ORF predictions made by GenMarkS2 (111), GenMark.heuristic (112), Prodigal v2.6.3 (113) and Prokka v1.14.6 (114). ORFs were functionally annotated using BLASTp against NCBI's nonredundant protein sequences (46), phmmer v2.41.1 against Pfam (47), and Swiss-Prot (48). All genes called were listed and compared using a scoring system evaluating length and overlap of ORFs, as well as the quality of annotation (43). tRNA and tmRNA were identified with tRNAScan-SE v2.0 (44) and ARAGORN v1.2.38 (45). Genomes were scanned for riboswitches using the web application of Riboswitch Scanner (115, 116). FindTerm (energy score < -11) and BPROM (LDF > 2.75) from the Fgenesb_annotator pipeline (117) 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' (52, 71) and middle-promoter MotA box (TGCTTtA)-dependent middle promoters were used as a query for a BLASTN search over the whole genome. Promoters and/or terminators were excluded if they were not intergenic or not within a 10-bp overlap of the start/end of ORFs.

Host DNA preparation, genome sequencing, and annotation. Host cultures were grown in 50 mL of ASM1 medium amended with 1 mM methanol in 250 mL polycarbonate flasks. Upon reaching maximum cell density, genomic DNA was extracted using a Qiagen DNeasy PowerWater kit (14900-50-NF) from biomass retained on 0.1 μ m pore-size PC filters according to the manufacturer's protocol, with minor modifications to increase the yield. The bead-beating step was lengthened from 5 to 10 min. DNA elution was performed with a 2 min 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 \times 250 bp], targeting 30-fold coverage). In addition, 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 BMAP v38.22 (106; available at <https://sourceforge.net/projects/bbmap/>). Contigs were assembled using SPAdes v.3.13 with default settings

(107). Gene calls were made with PROKKA v1.14.6 (114) and submitted to BlastKOALA (118) for further annotation and prediction of KEGG pathways using the *Methylophilales* strain HTCC2181 as a reference.

Phylogeny and network analysis. All contigs from the Global Ocean Virome (GOV2) data set and a WEC virome (13, 35), *Methylophilales* phage Venkman (16) and LD28 phage P10250A (119), all isolated *Pelagibacter* phages (16–19, 34) and *Pelagibacter*-like MAGs (33), 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 (113) and imported into the Cyverse Discovery Environment 2.0 (<https://de.cyverse.org/>), where vContact-Genome 1.1.0 was used to prepare protein sequences before protein clustering using VConTACT2 (v.0.9.8) with default settings (31) to assess relatedness via shared gene networks. The gene sharing network was visualized using Cytoscape v3.7.1 (120). Virome contigs, isolate genomes, and MAVGs clustering with Melnitz genomes were aligned using EasyFig v2.2.2 using the following parameters: “–ann_height 150 –blast_height 200 –f2 2000 –f arrow 86 180 233 –e 0.00005 –l 30 –blastn” (121). 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) (122). Genes were aligned within the Phylogeny.fr online server (123) opting for MUSCLE alignment (124) and built-in curation function (125) with default settings, removing positions with gaps for calculating phylogenetic trees. Maximum-likelihood trees were calculated with PhyML (126, 127) using 100 bootstraps, unless specified otherwise. The trees were visualized using FigTree (v1.4.4; <http://tree.bio.ed.ac.uk/software/figtree/>). Figures were edited with Inkscape (www.inkscape.org) for aesthetics.

Metagenomic reads recruitment. Marine virome data sets were used to assess the relative abundance of phage contigs (including a single virome from the Western English Channel), 131 samples from the Global Ocean Virome data set (GOV2) (13, 35), and 382 samples of the ALOHA 2.0 database from the North Pacific Subtropical Gyre (13, 35, 38). Metagenomic reads were subsampled to 5 million reads using the `reformats.sh` command within the `bbmap` suite (128). Bowtie2 (128) indexes of dereplicated contigs were created for all known pelagiphage isolate genome (16–19, 34, 119), a selection of cyanophages, a selection of abundant *Roseobacter* phages (129), and enterobacterial phages T4 and T7 (as negative controls), as well as one genome from the viral population isolated in the present study (Melnitz). Metagenomic reads from each virome were mapped against all contigs with `bowtie2 –seed 42 –non-deterministic`. To calculate coverage and 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`”.

Search and alignment of tRNA. tRNAs were identified from bacteria and virus genomes using ARAGORN v1.2.38 (45) using the “–t –gcbact –c –d –fons” flags and tRNAscan-SE v2.0.7 using flags “–B –fasta” (44). tRNAs were deduplicated using `seqkit v0.15.0` (130) with “`rmdup –by-seq –ignore-case`”. A BLAST database of tRNA genes was made in BLAST v2.5.0+ (122) and used for sequence alignment with BLASTN with the flags “–outfmt ‘6 std qlen slen’ –evaluate 1e–05 –task blastn-short”. The percent identity was calculated as follows: (alignment length/query length) × 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 Swiss-Model server (81) using BLAST and HHBlits. Putative models were downselected based on suitable quaternary structure properties and $QME > 0.7$. Autolysin SagA from *Brucella abortus* (PDB model 7NDP.1) was selected as the best hit and used for subsequent modeling of the structure of gp67. Models and associated figures were visualized 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 Swiss-Model server using BLAST and HHBlits. The best hit was determined by predicted quaternary structure properties. Global QME scores were < 0.5 , and Q -MEAN scores identified the best-predicted structures as unreliable (range, -3.28 to -5.11), although localized regions had Q -MEAN scores of > 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/1LVPSOF4L502F21RWBmYJJYLDIOU2NNTL>). Structures were determined both with or without postprediction relaxation (`use_amber`) and use of `MMSeqs2` templates (`use_templates`). Since 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, visualized and aligned to *E. coli* CsgGF (7NDP) in PyMOL v.2.5.1. CsgGF from pelagimyophage HTVC008M was similarly analyzed 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 (<https://github.com/HBuchholz/Genomic-evidence-for-inter-class-host-transition-between-streamlined-heterotrophs>). Electrostatic potential of protein surfaces was calculated and visualized 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 3×5 min washes in droplets of deionized water, structures were contrasted on droplets of 2% (wt/vol) uranyl acetate in 2% (wt/vol) methyl cellulose (ratio, 1:9) on ice for 10 min, the grids were picked up in a wire loop, and excess contrasting medium was 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 entries under NCBI accession numbers [MZ577095](https://www.ncbi.nlm.nih.gov/nuccore/MZ577095) to [MZ577098](https://www.ncbi.nlm.nih.gov/nuccore/MZ577098) of BioProject [PRJNA625644](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA625644); the reference genome used for the analysis was deposited under [MZ577097](https://www.ncbi.nlm.nih.gov/nuccore/MZ577097). Sequencing data for all phages sequenced in this study

can be found on the SRA data bank under accession numbers [SAMN18926670](https://www.ncbi.nlm.nih.gov/SRA/acc/acc.cgi?acc=SAMN18926670) to [SAMN18926674](https://www.ncbi.nlm.nih.gov/SRA/acc/acc.cgi?acc=SAMN18926674). Reads for *Methylophilales* bacterial host H5P1 are available under [SAMN20856461](https://www.ncbi.nlm.nih.gov/SRA/acc/acc.cgi?acc=SAMN20856461).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 7.4 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.02 MB.

SUPPLEMENTAL FILE 3, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS

We thank Christian Hacker and the Bioimaging Centre of the University of Exeter for performing the TE microscopy and imaging. We also 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 also 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 facility in carrying out this work. Genome sequencing was provided by the Exeter Sequencing Service.

This project utilized 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 H.H.B. were funded by the Natural Environment Research Council (NERC) GW4+ Doctoral Training program. L.M.B., M.L.M., and B.T. 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 analyzed 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

- Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–361. <https://doi.org/10.1038/nature04160>.
- Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, Haldal 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. <https://doi.org/10.1038/ismej.2008.31>.
- Wommack E, Colwell K, Rita R. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* 64:69–114. <https://doi.org/10.1128/MMBR.64.1.69-114.2000>.
- Weitz JS, Wilhelm SW. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* 4:17. <https://doi.org/10.3410/B4-17>.
- Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, Follows MJ, Fuhrman JA, Jover LF, Lennon JT, Middelboe M, Sonderregger 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>.
- Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology* 16:15. <https://doi.org/10.1186/s12985-019-1120-1>.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nat Microbiol* 3:754–766. <https://doi.org/10.1038/s41564-018-0166-y>.
- 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* 14:881–895. <https://doi.org/10.1038/s41396-019-0580-z>.
- Forterre P. 2011. Manipulation of cellular syntheses and the nature of viruses: the virocell concept. *C R Chim* 14:392–399. <https://doi.org/10.1016/j.crci.2010.06.007>.
- Martiny JBH, Riemann L, Marston MF, Middelboe M. 2014. Antagonistic coevolution of marine planktonic viruses and their hosts. *Annu Rev Mar Sci* 6:393–414. <https://doi.org/10.1146/annurev-marine-010213-135108>.
- Avrani S, Schwartz DA, Lindell D. 2012. Virus-host swinging party in the oceans: incorporating biological complexity into paradigms of antagonistic coexistence. *Mob Genet Elements* 2:88–95. <https://doi.org/10.4161/mge.20031>.
- 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, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB, Tara Oceans Coordinators. 2015. Ocean plankton: patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. <https://doi.org/10.1126/science.1261498>.
- Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Dominguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay J-É, Vik D, Babin M, Bowler C, Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB, Tara Oceans Coordinators. 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177:1109–1123. <https://doi.org/10.1016/j.cell.2019.03.040>.
- Martinez-Hernandez F, Fornas O, Lluésma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, Anton J, Gasol JM, Rosselli R, Rodríguez-Valera F, Sullivan MB, Acinas SG, Martínez-García M. 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 8:15892. <https://doi.org/10.1038/ncomms15892>.
- Martinez-Hernandez F, Fornas O, Lluésma Gomez M, García-Heredia I, Maestre-Carballa L, López-Pérez M, Haro-Moreno JM, Rodríguez-Valera F, Martínez-García M. 2019. Single-cell genomics uncover *Pelagibacter* as the putative host of the extremely abundant uncultured 37-F6 viral

- population in the ocean. *ISME J* 13:232–236. <https://doi.org/10.1038/s41396-018-0278-7>.
16. Buchholz HH, Michelsen ML, Bolaños LM, Browne E, Allen MJ, Temperton 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>.
 17. 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>.
 18. 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>.
 19. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, Du S, Tian Z, Zhao Y. 2021. Culturing novel and abundant pelagiphages in the ocean. *Environ Microbiol* 23:1145–1161. <https://doi.org/10.1111/1462-2920.15272>.
 20. Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C, Oh H-M, Kitner JB, Vergin KL, Rappé MS. 2008. The small genome of an abundant coastal ocean methylotroph. *Environ Microbiol* 10:1771–1782. <https://doi.org/10.1111/j.1462-2920.2008.01598.x>.
 21. Parks DH, Chuvpochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36: 996–1004. <https://doi.org/10.1038/nbt.4229>.
 22. Thrash JC, Temperton B, Swan BK, Landry ZC, Woyke T, DeLong EF, Stepanauskas R, Giovannoni SJ. 2014. Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J* 8:1440–1451. <https://doi.org/10.1038/ismej.2013.243>.
 23. Beale R, Dixon JL, Smyth TJ, Nightingale PD. 2015. Annual study of oxygenated volatile organic compounds in UK shelf waters. *Mar Chem* 171: 96–106. <https://doi.org/10.1016/j.marchem.2015.02.013>.
 24. Halsey KH, Carter AE, Giovannoni SJ. 2012. Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* 14:630–640. <https://doi.org/10.1111/j.1462-2920.2011.02605.x>.
 25. Rappé MS, Kemp PF, Giovannoni SJ. 1997. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnol Oceanogr* 42:811–826. <https://doi.org/10.4319/lo.1997.42.5.0811>.
 26. Yang M, Xia Q, Du S, Zhang Z, Qin F, Zhao Y. 2021. Genomic characterization and distribution pattern of a novel marine OM43 phage. *Front Microbiol* 12:657.
 27. Giovannoni S, Thrash C, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME J* 8:1553–1565. <https://doi.org/10.1038/ismej.2014.60>.
 28. Nayfach S, Camargo AP, Eloe-Fadrosh E, Roux S. 2020. CheckV: assessing the quality of metagenome-assembled viral genomes. *bioRxiv* <https://doi.org/10.1101/2020.05.06.081778>.
 29. Ackermann HW. 2009. Basic phage electron microscopy. In Clokie MR, Kropinski AM (ed), *Bacteriophages: methods in molecular biology*. Humana Press, New York, NY. https://doi.org/10.1007/978-1-60327-164-6_12.
 30. Turner D, Kropinski AM, Adriaenssens EM. 2021. A roadmap for genome-based phage taxonomy. *Viruses* 13:506. <https://doi.org/10.3390/v13030506>.
 31. Bolduc B, Jang HB, Doucier 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>.
 32. Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated nonredundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 33:D501–4. <https://doi.org/10.1093/nar/gki025>.
 33. 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>.
 34. 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>.
 35. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, Sullivan MB, 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>.
 36. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817. <https://doi.org/10.7717/peerj.3817>.
 37. Tithi SS, Aylward FO, Jensen RV, Zhang L. 2018. FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. *PeerJ* 6:e4227. <https://doi.org/10.7717/peerj.4227>.
 38. Luo E, Eppley JM, Romano AE, Mende DR, DeLong EF. 2020. Double-stranded DNA viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J* 14:1304–1315. <https://doi.org/10.1038/s41396-020-0604-8>.
 39. Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA. 2013. Evidence for a persistent microbial seed bank throughout the global ocean. *Proc Natl Acad Sci U S A* 110:4651–4655. <https://doi.org/10.1073/pnas.1217767110>.
 40. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F. 2006. The marine viromes of four oceanic regions. *PLoS Biol* 4:e368. <https://doi.org/10.1371/journal.pbio.0040368>.
 41. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* 13:278–284. <https://doi.org/10.1016/j.tim.2005.04.003>.
 42. Brum JR, Ignacio-Espinoza JC, Kim E-H, Trubl G, Jones RM, Roux S, VerBerkmoes NC, Rich VI, Sullivan MB. 2016. Illuminating structural proteins in viral “dark matter” with metaproteomics. *Proc Natl Acad Sci U S A* 113:2436–2441. <https://doi.org/10.1073/pnas.1525139113>.
 43. Salisburry 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>.
 44. Lowe TM, Chan PP. 2016. tRNAscan-SE On-Line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44: W54–W57. <https://doi.org/10.1093/nar/gkw413>.
 45. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
 46. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated nonredundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:D61–D65. <https://doi.org/10.1093/nar/gkl842>.
 47. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger 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>.
 48. 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>.
 49. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction, and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>.
 50. Mahichi F, Synnott AJ, 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 Microbiol Lett* 295:211–217. <https://doi.org/10.1111/j.1574-6968.2009.01588.x>.
 51. 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. *Front Microbiol* 6:1107. <https://doi.org/10.3389/fmicb.2015.01107>.
 52. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rügner W. 2003. Bacteriophage T4 genome. *Microbiol Mol Biol Rev* 67:86–156. <https://doi.org/10.1128/MMBR.67.1.86-156.2003>.
 53. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigle PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12: 3035–3056. <https://doi.org/10.1111/j.1462-2920.2010.02280.x>.
 54. Bryan MJ, Burroughs NJ, Spence EM, Clokie MRJ, Mann NH, Bryan SJ. 2008. Evidence for the intense exchange of MazG in marine

- cyanophages by horizontal gene transfer. *PLoS One* 3:e2048. <https://doi.org/10.1371/journal.pone.0002048>.
55. Morris RM, Longnecker K, Giovannoni SJ. 2006. Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* 8:1361–1370. <https://doi.org/10.1111/j.1462-2920.2006.01029.x>.
 56. Rihtman B, Bowman-Grahl S, Millard A, Corrigan RM, Clokie MRJ, Scanlan DJ. 2019. Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environ Microbiol Rep* 11:448–455. <https://doi.org/10.1111/1758-2229.12741>.
 57. Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D. 2013. Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Appl Environ Microbiol* 79:7547–7555. <https://doi.org/10.1128/AEM.02229-13>.
 58. Kossykh VG, Schlagman SL, Hattman S. 1995. Phage T4 DNA N-adenine-6-methyltransferase: overexpression, purification and characterization. *J Biol Chem* 270:14389–14393. <https://doi.org/10.1074/jbc.270.24.14389>.
 59. Jordan A, Pontis E, Atta M, Krook M, Gibert I, Barbé J, Reichard P. 1994. A second class I ribonucleotide reductase in *Enterobacteriaceae*: characterization of the *Salmonella* Typhimurium enzyme. *Proc Natl Acad Sci U S A* 91:12892–12896. <https://doi.org/10.1073/pnas.91.26.12892>.
 60. 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. <https://doi.org/10.1038/s41467-019-08672-6>.
 61. Lewis LO, Yousten AA. 1988. Bacteriophage attachment to the S-layer proteins of the mosquito-pathogenic strains of *Bacillus sphaericus*. *Curr Microbiol* 17:55–60. <https://doi.org/10.1007/BF01568820>.
 62. 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 *spoA* and the sporulation genes *spo0A*, *spo0B*, and *spo0F* in Phage AP50c infection of *Bacillus anthracis*. *J Bacteriol* 196:1143–1154. <https://doi.org/10.1128/JB.00739-13>.
 63. 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. <https://doi.org/10.1038/ismej.2016.79>.
 64. Shi K, Oakland JT, Kurniawan F, Moeller NH, Banerjee S, Aihara H. 2020. Structural basis of superinfection exclusion by bacteriophage T4 Spackle. *Commun Biol* 3:691. <https://doi.org/10.1038/s42003-020-01412-3>.
 65. 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. <https://doi.org/10.4161/bioe.32110>.
 66. Hinton DM. 2010. Transcriptional control in the prereplicative phase of T4 development. *Virology* 407:289. <https://doi.org/10.1016/j.virol.2010.07.029>.
 67. Doron S, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, Feingersh 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. <https://doi.org/10.1038/ismej.2015.210>.
 68. 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. <https://doi.org/10.1126/science.1114057>.
 69. 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. <https://doi.org/10.1073/pnas.1113328108>.
 70. 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. <https://doi.org/10.1073/pnas.0408028101>.
 71. Geiduschek EP, Kassavetis GA. 2010. Transcription of the T4 late genes. *Virology* 407:288. <https://doi.org/10.1016/j.virol.2010.07.028>.
 72. Withey JH, Friedman DI. 2003. A salvage pathway for protein structures: tmRNA and trans-translation. *Annu Rev Microbiol* 57:101–123. <https://doi.org/10.1146/annurev.micro.57.030502.090945>.
 73. Keiler KC, Shapiro L, Williams KP. 2000. tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: a two-piece tmRNA functions in *Caulobacter*. *Proc Natl Acad Sci U S A* 97:7778–7783. <https://doi.org/10.1073/pnas.97.14.7778>.
 74. Sharkady SM, Williams KP. 2004. A third lineage with two-piece tmRNA. *Nucleic Acids Res* 32:4531–4538. <https://doi.org/10.1093/nar/gkh795>.
 75. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stelkel DJ, Hobman J, Jones MA, Millard A. 2021. Infrastructure for a PHAGE Reference Database: identification of large-scale biases in the current collection of phage genomes. *PHAGE* 2:214–223. <https://doi.org/10.1089/phage.2021.0007>.
 76. 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. <https://doi.org/10.1186/gb-2010-11-3-r31>.
 77. 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, Sintès 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. <https://doi.org/10.1038/sdata.2018.154>.
 78. 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. <https://doi.org/10.1093/nar/gky709>.
 79. 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. <https://doi.org/10.1073/pnas.1901856116>.
 80. Barnhart MM, Chapman MR. 2006. Curli biogenesis and function. *Annu Rev Microbiol* 60:131–147. <https://doi.org/10.1146/annurev.micro.60.080805.142106>.
 81. 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 modeling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303. <https://doi.org/10.1093/nar/gky427>.
 82. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek 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* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
 83. 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. <https://doi.org/10.1073/pnas.0907941106>.
 84. Conners R, Mc Laren M, Łapińska U, Sanders K, Rhia LM, Blaskovich MAT, Pagliara S, Daum B, Rakonjac J, Gold VAM. 2021. CryoEM structure of the outer membrane secretin channel pIV from the f1 filamentous bacteriophage. *Nat Commun* 12:6316. <https://doi.org/10.1038/s41467-021-26610-3>.
 85. Ghosh S, Shaw R, Sarkar A, Gupta SKD. 2020. Evidence of positive regulation of mycobacteriophage D29 early gene expression obtained from an investigation using a temperature-sensitive mutant of the phage. *FEMS Microbiol Lett* 367:fnaa176. <https://doi.org/10.1093/femsle/fnaa176>.
 86. Catalão MJ, Gil F, Moniz-Pereira J, São-José C, Pimentel M. 2013. Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiol Rev* 37:554–571. <https://doi.org/10.1111/1574-6976.12006>.
 87. Rajaure M, Berry J, Kongari R, Cahill J, Young R. 2015. Membrane fusion during phage lysis. *Proc Natl Acad Sci U S A* 112:5497–5502. <https://doi.org/10.1073/pnas.1420588112>.
 88. Del Giudice MG, Ugalde JE, Czubener C. 2013. A lysozyme-like protein in *Brucella abortus* is involved in the early stages of intracellular replication. *Infect Immun* 81:956–964. <https://doi.org/10.1128/IAI.01158-12>.
 89. 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*. *Mol Cells* 44:517–528. <https://doi.org/10.14348/molcells.2021.0011>.
 90. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczepinski B, Fraser CM. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol* 185:5220–5233. <https://doi.org/10.1128/JB.185.17.5220-5233.2003>.
 91. Frias MJ, Melo-Cristino J, Ramirez M. 2009. The autolysin LytA contributes to efficient bacteriophage progeny release in *Streptococcus pneumoniae*. *J Bacteriol* 191:5428–5440. <https://doi.org/10.1128/JB.00477-09>.
 92. Burrows BH, Molineux IJ, Fralick JA. 2019. Directed in vitro evolution of therapeutic bacteriophages: the Appelmans Protocol. *Viruses* 11:241. <https://doi.org/10.3390/v11030241>.

93. Tétart F, Repoila F, Monod C, Krisch HM. 1996. Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. *J Mol Biol* 258:726–731. <https://doi.org/10.1006/jmbi.1996.0281>.
94. 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* 12:771–779. <https://doi.org/10.1016/j.cels.2021.05.019>.
95. Santos SB, Kropinski AM, Ceyskens 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* 85:11265–11273. <https://doi.org/10.1128/JVI.01769-10>.
96. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev* 40:258–272. <https://doi.org/10.1093/femsre/fuv048>.
97. Coutinho FH, Zaragoza-Solas A, López-Pérez M, Barylski J, Zieleszinski A, Dutilh BE, Edwards R, Rodríguez-Valera F. 2021. RaFAH: host prediction for viruses of *Bacteria* and *Archaea* based on protein content. *Patterns* 2: 100274. <https://doi.org/10.1016/j.patter.2021.100274>.
98. 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. <https://doi.org/10.1007/BF02428046>.
99. de Jonge PA, Nobrega FL, Brouns SJJ, Dutilh BE. 2019. Molecular and evolutionary determinants of bacteriophage host range. *Trends Microbiol* 27:51–63. <https://doi.org/10.1016/j.tim.2018.08.006>.
100. Giovannoni SJ. 2017. SAR11 bacteria: the most abundant plankton in the oceans. *Annu Rev Mar Sci* 9:231–255. <https://doi.org/10.1146/annurev-marine-010814-015934>.
101. De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, Petit M-A. 2014. Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. *PLoS Genet* 10:e1004181. <https://doi.org/10.1371/journal.pgen.1004181>.
102. Kupczok A, Dagan T. 2019. Rates of molecular evolution in a marine *Synechococcus* phage lineage. *Viruses* 11:720. <https://doi.org/10.3390/v11080720>.
103. 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. <https://doi.org/10.1038/s41587-019-0100-8>.
104. 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. <https://doi.org/10.1038/ismej.2012.122>.
105. Solonenko N. 2016. Isolation of DNA from phage lysate. *Protocols.io* <https://doi.org/10.17504/protocols.io.c36yrd>.
106. 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>.
107. 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>.
108. 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>.
109. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol* 39:578–585. <https://doi.org/10.1038/s41587-020-00774-7>.
110. 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. <https://doi.org/10.1093/bioinformatics/btz265>.
111. 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>.
112. Besemer J, Borodovsky M. 1999. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res* 27:3911–3920. <https://doi.org/10.1093/nar/27.19.3911>.
113. 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>.
114. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
115. Singh P, Bandyopadhyay P, Bhattacharya S, Krishnamachari A, Sengupta S. 2009. Riboswitch detection using profile hidden Markov models. *BMC Bioinformatics* 10:325. <https://doi.org/10.1186/1471-2105-10-325>.
116. Mukherjee S, Sengupta S. 2016. Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in genomic sequences. *Bioinformatics* 32:776–778. <https://doi.org/10.1093/bioinformatics/btv640>.
117. Solovyev V, Salamov A. 2011. Automatic annotation of microbial genomes and metagenomic sequences, p 61–78. In Li RW (ed), *Metagenomics and its application in agriculture, biomedicine and environmental studies*. Nova Science Publishers, Hauppauge, NY.
118. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
119. 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. *Environ Microbiol* 19:4714–4727. <https://doi.org/10.1111/1462-2920.13936>.
120. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>.
121. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.
122. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
123. 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:W465–9. <https://doi.org/10.1093/nar/gkn180>.
124. 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>.
125. 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>.
126. Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* 55: 539–552. <https://doi.org/10.1080/10635150600755453>.
127. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704. <https://doi.org/10.1080/10635150390235520>.
128. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
129. 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:e00494-19. <https://doi.org/10.1128/mSystems.00494-19>.
130. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. <https://doi.org/10.1371/journal.pone.0163962>.