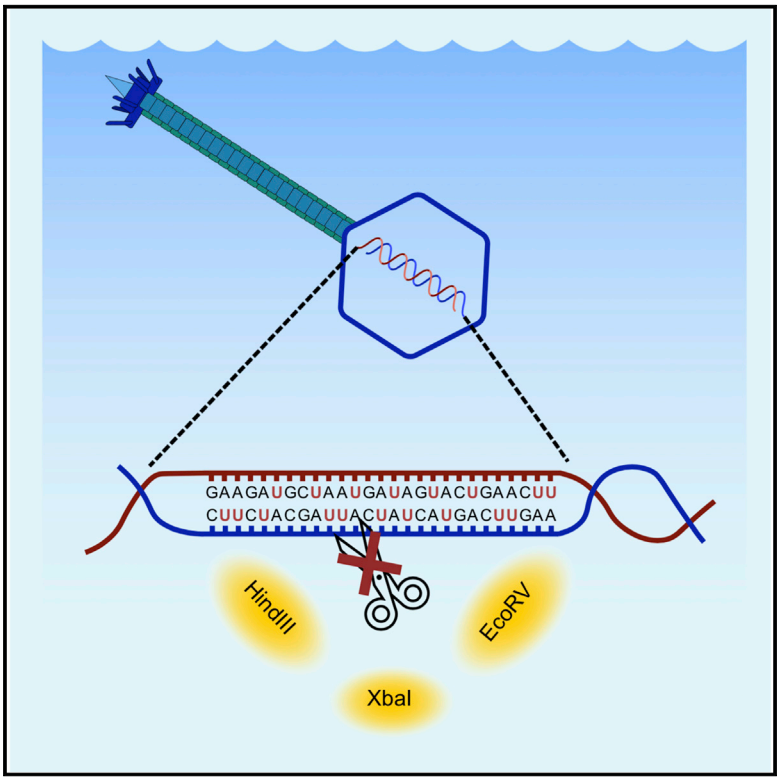
Report

Current Biology

A new family of globally distributed lytic roseophages with unusual deoxythymidine to deoxyuridine substitution

Graphical abstract



Authors

Branko Rihtman, Richard J. Puxty, Alexia Hapeshi, ..., Andrew D. Millard, David J. Scanlan, Yin Chen

Correspondence

b.rihtman.1@warwick.ac.uk (B.R.), y.chen.25@warwick.ac.uk (Y.C.)

In brief

Rihtman et al. report the discovery of two novel roseophages containing deoxyuridine (dU) instead of the canonical nucleobase deoxythymidine. Such a substitution results in resistance to a commonly used method for metagenome library preparation, suggesting that the presence of this new family of phages may have been underestimated.

Highlights

- Two new roseophages isolated from the marine environment
- They have an unusual deoxythymidine to deoxyuridine substitution in their genomes
- These dU genomes are resistant to a common method of metagenome library preparation
- These phages represent a new family and are globally distributed in the oceans

Rihtman et al., 2021, Current Biology *31*, 3199–3206
July 26, 2021
https://doi.org/10.1016/j.cub.2021.05.014

CellPress

Current Biology

CellPress
OPEN ACCESS

Report

A new family of globally distributed lytic roseophages with unusual deoxythymidine to deoxyuridine substitution

Branko Rihtman,[1,6,*] Richard J. Puxty,[1,6] Alexia Hapeshi,[2] Yan-Jiun Lee,[3] Yuanchao Zhan,[4] Slawomir Michniewski,[1] Nicholas R. Waterfield,[2] Feng Chen,[4] Peter Weigele,[3] Andrew D. Millard,[5] David J. Scanlan,[1] and Yin Chen[1,7,*]

[1]School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[2]Warwick Medical School, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[3]Research Department, New England Biolabs, 240 County Road, Ipswich, MA 01938, USA
[4]Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, 701 E. Pratt Street, Baltimore, MD 21202, USA
[5]Department of Genetics and Genome Biology, College of Life Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK
[6]These authors contributed equally
[7]Lead contact
*Correspondence: b.rihtman.1@warwick.ac.uk (B.R.), y.chen.25@warwick.ac.uk (Y.C.)
https://doi.org/10.1016/j.cub.2021.05.014

SUMMARY

Marine bacterial viruses (bacteriophages) are abundant biological entities that are vital for shaping microbial diversity, impacting marine ecosystem function, and driving host evolution.[1–3] The marine roseobacter clade (MRC) is a ubiquitous group of heterotrophic bacteria[4,5] that are important in the elemental cycling of various nitrogen, sulfur, carbon, and phosphorus compounds.[6–10] Bacteriophages infecting MRC (roseophages) have thus attracted much attention and more than 30 roseophages have been isolated,[11–13] the majority of which belong to the N4-like group (*Podoviridae* family) or the Chi-like group (*Siphoviridae* family), although ssDNA-containing roseophages are also known.[14] In our attempts to isolate lytic roseophages, we obtained two new phages (DSS3_VP1 and DSS3_PM1) infecting the model MRC strain *Ruegeria pomeroyi* DSS-3. Here, we show that not only do these phages have unusual substitution of deoxythymidine with deoxyuridine (dU) in their DNA, but they are also phylogenetically distinct from any currently known double-stranded DNA bacteriophages, supporting the establishment of a novel family ("Naomiviridae"). These dU-containing phages possess DNA that is resistant to the commonly used library preparation method for metagenome sequencing, which may have caused significant underestimation of their presence in the environment. Nevertheless, our analysis of *Tara* Ocean metagenome datasets suggests that these unusual bacteriophages are of global importance and more diverse than other well-known bacteriophages, e.g., the *Podoviridae* in the oceans, pointing to an overlooked role for these novel phages in the environment.

RESULTS AND DISCUSSION

Bacteriophages DSS3_VP1 and DSS3_PM1 have dU instead of deoxythymidine

In our attempts to isolate novel phages infecting *Ruegeria pomeroyi* DSS-3, a model organism for studying the metabolism of climate active trace gases (e.g., dimethylsulfide, methylamines), we screened seawater samples from the Grand Canal (Venice, Italy) and Puerto Morelos (Mexico) against exponentially growing DSS-3 cultures. We isolated two new phages, DSS3_PM1 and DSS3_VP1, which showed a siphovirus morphotype and had some unique characteristics (Figures 1A and 1B).

The capsids of these two phages possessed an icosahedral structure measuring 47–55 nm in diameter, with the tail measuring approximately 230–240 nm in length. Both phages produced clear, well-defined plaques on solid agar plates (Figures 1A and 1B). One-step infection experiments in liquid culture showed contrasting infection kinetics (Figures 1C and 1D) compared to previously reported double-stranded DNA (dsDNA) roseophages infecting *Ruegeria pomeroyi* DSS-3 (e.g., Phi2, Phi8).[11,15,16] Thus, infection by DSS3_VP1 and DSS3_PM1 phages caused complete lysis of the culture (Figure 1C) but with differing latent periods (110 min for DSS3_VP1; 210 min for DSS3_PM1; Figure 1D; Table S1) and burst sizes (67 PFU per cell for VP1; 38 PFU per cell for DSS3_PM1).

Attempts to sequence the genomes of DSS3_VP1 and DSS3_PM1 phages by Illumina Miseq sequencing failed to generate a library using the Nextera sample preparation kit, although a sequencing library was successfully generated using whole-genome amplified (WGA) DNA of these two phages (Figure S1). This prompted us to investigate potential DNA modifications in these two phages. We analyzed the

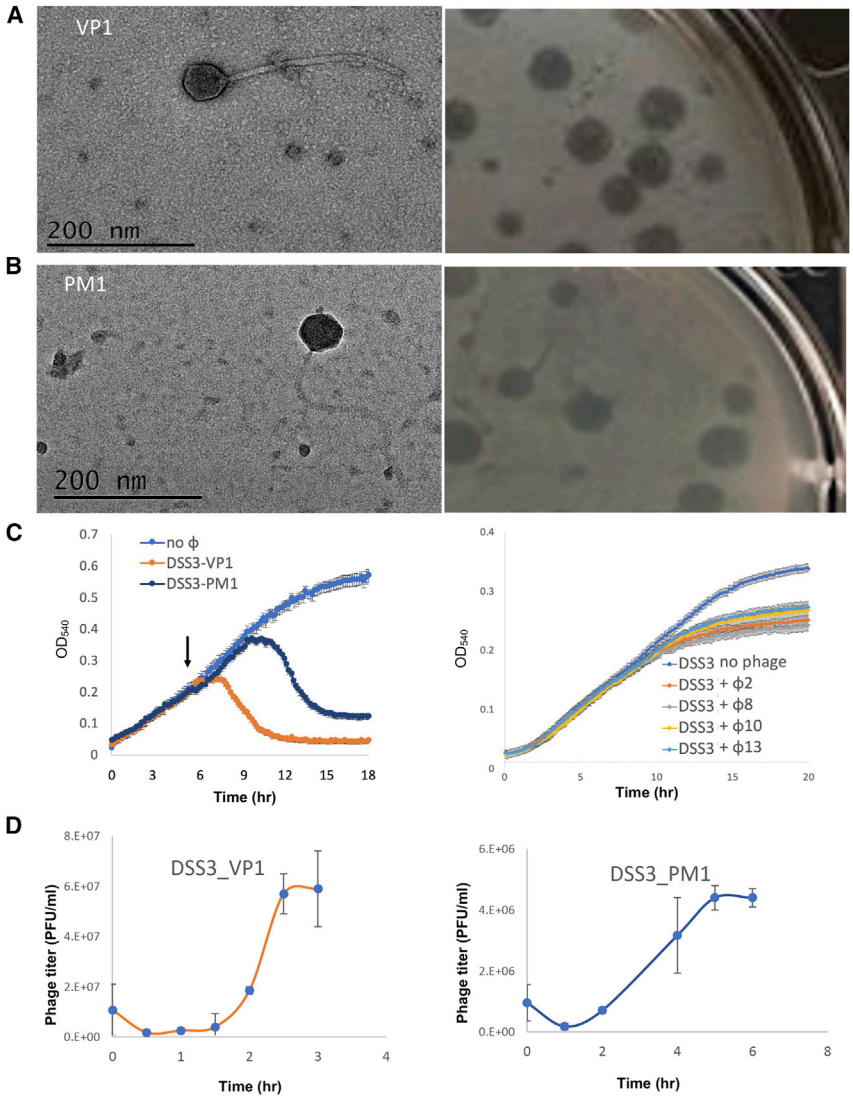


Figure 1. Isolation and characterization of novel phages DSS3_VP1 and DSS3_PM1

(A and B) Transmission electron microscopy (TEM) images of the isolated phages and plaque morphology.

(C and D) One-step infection kinetics of *R. pomeroyi* DSS-3 infected with the phages described in this study (DSS3_VP1 and DSS3_PM1) and compared with four previously isolated phages (phi2, phi8, phi10, and phi12). The arrow in the panel marks the time point at which phages were added to the exponentially growing culture and the optical density of the host was recorded continuously. Error bars represent standard error of three biological replicates. See also Table S1 for main characteristics of phages DSS3_VP1 and DSS3_PM1 and Tables S2 and S3 for predicted non-hypothetical genes.

nucleoside content obtained from enzymatic digestion of phage DSS3_VP1 genomic DNA using liquid chromatography-mass spectrometry (LC-MS).[17] Interestingly, instead of the canonical nucleoside deoxythymidine, we observed a nucleoside co-eluting with an authentic dU standard (Figure 2A). Subsequent MS analysis confirmed the identity of dU, showing the presence of an *m/z* of 251.1 species, corresponding to dU containing a sodium adduct in the positive ionization mode $[M+Na]^+$ and two species corresponding to deprotonated dU $[M-H]^-$ *m/z* 227.1 and an acetate adduct $[M+OAc]^-$ *m/z* 287.1 in the negative ionization mode (Figure S2). The normalized integration of dU and deoxyadenosine (dA) suggests that they occur in the genome at a 1:1 ratio, suggesting that all deoxythymidine positions in this phage are completely substituted by dU. Since dU modification of DNA is known to inhibit the activity of certain endonucleases such as *Eco*RV, *Xba*I, and *Hin*dIII,[18–20] we performed restriction digestion of native phage DNA, WGA phage DNA, and λ phage DNA (as a control) using these enzymes. This demonstrated unequivocally that dU in the genomes of DSS3_VP1 and DSS3_PM1 phages resisted digestion by these endonucleases (Figure 2B).

The presence of dU in phage genomes was further assessed using a uracil DNA glycosylase (UDG)-endonuclease VIII assay (Figure 2C). UDG catalyzes the removal of dU from the DNA molecule, creating an abasic site[22] that can then be cleaved by endonuclease VIII. The results showed that both the DSS3_VP1 and DSS3_PM1 genomes can be completely digested by UDG-endonuclease VIII treatment, while the products of whole-genome amplification of these genomes (which do not contain dU) remained intact. Similarly, UDG-endonuclease VIII treatment did not digest phage λ DNA. Together, our results show that dU instead of deoxythymidine is present in these phages. This unusual substitution of deoxythymidine by dU has only been found in two bacteriophages, e.g., Bacillus phage PBS1[23] and Yersinia phage phiR1-37[24] and, to the best of our knowledge, has not been reported in viruses infecting marine bacteria.

Genome sequencing reveals unique characteristics of bacteriophages DSS3_VP1 and DSS3_PM1

Whole-genome sequencing of amplified genomes (STAR Methods) of DSS3_VP1 and DSS3_PM1 phages shows that their assembled genomes are between 70 and 75 kb (Figure 2D) and highly similar to each other (Figure S3A), in spite of the large geographical distance between their isolation locations. The average nucleotide identity between the complete genomes of these two phages was 96.3% and their gene content was almost identical (Figure 2D). Both genomes contain a set of genes required for the biosynthesis and incorporation of dU into genomic DNA (Figure 2E), including a putative cytosine deaminase (DSS3_VP1_066 and DSS3_PM1_098) and a putative phosphatase/pyrophosphatase (DSS3_VP1_050 and DSS3_PM1_082) required for eliminating dTTP from the dNTP pool.[25] DSS3_VP1 has a few additional genes that are not found in the DSS3_PM1

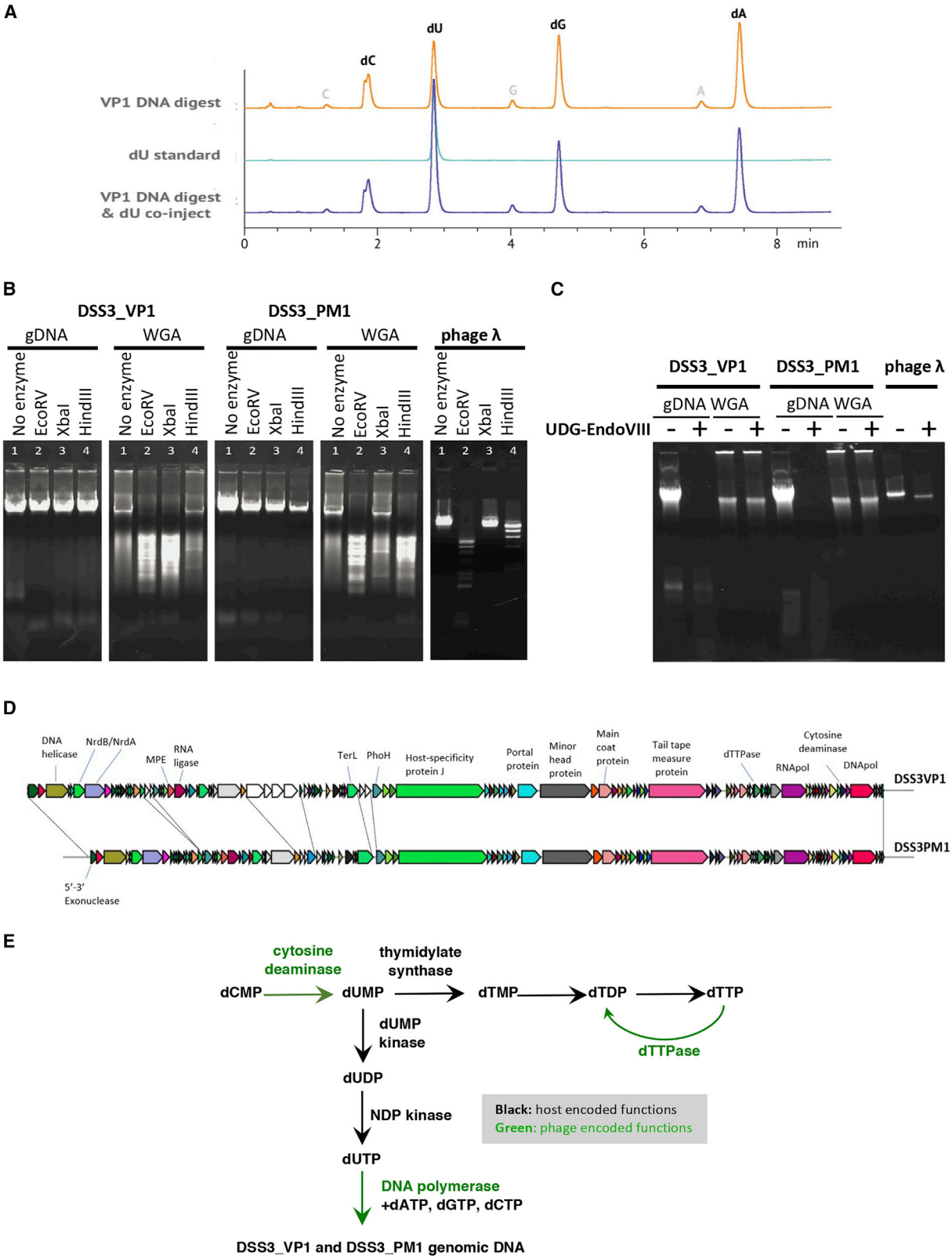


Figure 2. Characterization of dU in DSS3_VP1 and DSS3_PM1 phages and genome sequencing

(A) Liquid chromatography-mass spectrometry detection of deoxyuridine (dU) in the genome of phage DSS3_VP1. An authentic dU standard is included that co-eluted with dU from phage DSS3_VP1. See also Figure S2 for liquid chromatography-mass spectrometry detection of dU.

(legend continued on next page)

genome (Tables S2 and S3). A four-gene cluster (DSS3_VP1_00106–DSS3_VP1_00109) in DSS3_VP1 contains two hypothetical genes and two genes with predicted function, including a putative nicotinamide phosphoribosyltransferase (DSS3_VP1_00106) and a putative glutathionyl spermidine synthase (DSS3_VP1_00109). Nicotinamide phosphoribosyltransferase has been previously identified as having a potential role in NAD^+ and nucleotide metabolism.[26] Phage DSS3_VP1 seems to have acquired a type I self-splicing intron containing a Type VII homing endonuclease (DSS3_VP1_00013) in the middle of its large terminase subunit gene (Figure S3B). The large terminase subunit gene is often the site of integration of self-splicing introns and inteins in phages,[27] probably due to the high level of conservation between the gene homologs among different phages. Sometimes these self-splicing genetic elements have a regulatory function in phages (e.g., self-splicing introns within the photosynthetic *psbA* gene in cyanophage S-PM2).[28,29] However, it is not clear whether the self-splicing intron in DSS3_VP1 is under any kind of promoter control or whether the splicing occurs *in vivo* during phage gene transcription or DNA replication.

DSS3_VP1 and DSS3_PM1 represent a novel family of bacteriophages

In order to determine the phylogenetic relationship of DSS3_VP1 and DSS3_PM1 among other dsDNA phages, we used four independent approaches that are recommended by the International Committee on Taxonomy of Viruses (ICTV) to demonstrate that these phages belong to a new family. First, we carried out a uBLAST search against the curated bacteriophage genomes database (currently with ~10,000 viral genomes) extracted from GenBank,[30] using whole-genome sequences of DSS3_VP1 or DSS3_PM1 as a query and an e-value threshold of 1e−5. However, we obtained no hits with significant similarity, suggesting that these phages represent a novel clade of dsDNA phages.

To more fully establish the taxonomic position of DSS3_VP1 and DSS3_PM1, we then performed phylogenetic analysis of the DNA polymerase of all currently known roseophages.[11] This showed that DSS3_VP1 and DSS3_PM1 form a well-supported cluster distinct from other roseophages (Figure 3A). Although morphologically DSS3_VP1 and DSS3_PM1 appear to be siphoviruses, they did not cluster with known *Siphoviridae* family roseophages (e.g., the roseophage DSS3P8[15]) in the DNA polymerase phylogenetic tree. Instead, they appeared more closely related to the newly proposed *Cobaviruses* of the *Podoviridae* family.[12] Subsequent phylogenetic analyses using three other protein markers, the large terminase (TerL, phylogenetic analysis being carried out with the predicted intron sequence removed), DNA polymerase I (DNApol), and PhoH (a protein of unknown function but that is widespread in bacteriophages and has previously been used in phylogenetic analyses[31]), confirmed that DSS3_VP1 and DSS3_PM1 cluster separately from all other known described phage families (Figure S4). These phylogenetic analyses suggest that DSS3_VP1 and DSS3_PM1 represent a novel group of bacteriophages.

Given the failure of functional gene markers to phylogenetically place DSS3_PM1 and DSS3_VP1 in an established bacteriophage family, we turned to whole-genome analyses. We first tested the clustering of the DSS3_VP1 and DSS3_PM1 genomes within the *Caudovirales* order using VIPtree based on complete phage proteomes.[32] This showed that neither phage clustered with any previously isolated roseophage (Figure 3B). Indeed, these two phages appear to branch out from all known dsDNA viruses very early in evolution, before the recently established *Herelleviridae*[33] and *Ackermannviridae* families.[34]

Subsequent analysis using vConTACT 2.0 (which determines the clustering of gene families based on Blast hits against all dsDNA viruses[35]) again suggested these two newly isolated phages form a distinct cluster that is different from all currently known dsDNA viruses (Figure 3C). Together, these complementary approaches all supported that these new phages represent a unique group of bacteriophages. We therefore propose to classify these two phages with a siphovirus morphotype as a single viral species, within a novel genus (“Noahvirus”) and a new family (“Naomiviridae”).

Distribution of “Naomiviridae” in marine metagenomes and estimates of diversity

Since DSS3_VP1 and DSS3_PM1 represent the only cultured representatives of the proposed “Naomiviridae” family, we sought to expand their diversity by searching publicly available ocean metagenomes. Very few reads from the TARA Ocean metagenome mapped directly to DSS3_VP1 and DSS3_PM1 (data not shown). It is likely that dU-substituted DNA may have inhibited Tn5-based transposase insertion during library preparation, which is widely used for metagenome sequencing of environmental DNA (e.g., Illumina Nextera). Alternatively, the DNA polymerase used for metagenome library preparation might have caused stalling of the enzyme during the PCR on dU-containing templates.[36] Equally, it is also plausible that these taxa are in low abundance or form part of a diverse group with low sequence conservation. With this in mind, from the TARA dataset, we recovered 32 orthologs of the DNApol marker gene that shared a most recent common ancestor with DSS3_VP1 and DSS3_PM1 (Figure S4). From these marker genes, we recovered the resulting contigs from the assembled metagenomes. These contigs show a conserved synteny in the DNA replication module with DSS3_VP1 and DSS3_PM1 (Figure 4A), with most contigs sharing the DNApol, DNA primase-helicase,

(B) Digestion of native and whole-genome amplified (WGA) phage DNA by dU-sensitive endonucleases. DNA from phage λ, which does not contain dU, is used as a control. See also Figure S1 for NexteraXT library preparation that failed to produce a suitable library for dU-containing genome.
(C) Characterization of dU-containing DNA of phage DSS3_VP1/DSS3_PM1 and WGA DNA by uracil-DNA glycosylase (UDG) and endonuclease VIII. DNA from phage λ, which does not contain dU, was used as a control.
(D) Alignment of the genomes of phages DSS3_VP1 and DSS3_PM1. MPE, metallophosphoesterase; TerL, terminase large subunit; PhoH, phosphate stress-inducible protein; NrdB/NrdA, beta and alpha subunits of nucleotide reductase. Figure generated using Multigeneblast.[21] See also Figure S3 for pairwise genome alignment between DSS3_VP1 and DSS3_PM1 and the splicing of large terminase subunit in DSS3_VP1.
(E) The proposed pathway of deoxyuridine (dU)-containing DNA synthesis in the DSS3_VP1 and DSS3_PM1 phages. The host encoded functions are shown in black arrows and the phage encoded functions are shown in green arrows.

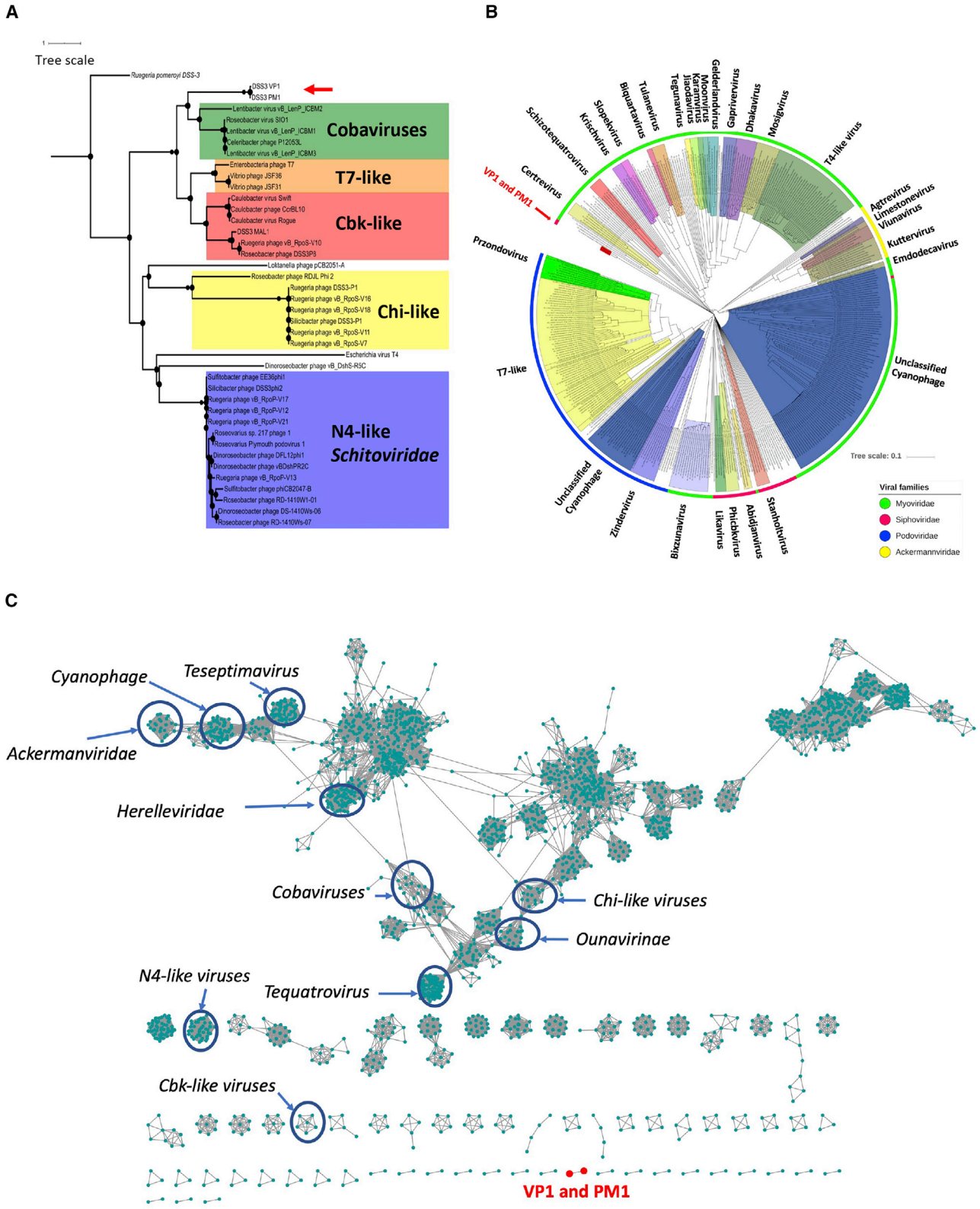


Figure 3. Phylogenetic analyses and network representation of protein content similarity of dsDNA bacteriophages generated with vConTACT v2.0 and visualized with Cytoscape

(A) Phylogenetic analysis of the DNA polymerase of double-stranded DNA viruses infecting marine *Roseobacter* clade (MRC) bacteria. The host DNA polymerase is also shown (*Ruegeria pomeroyi* DSS-3). The scale bar represents one substitution 100 amino acids.

(legend continued on next page)

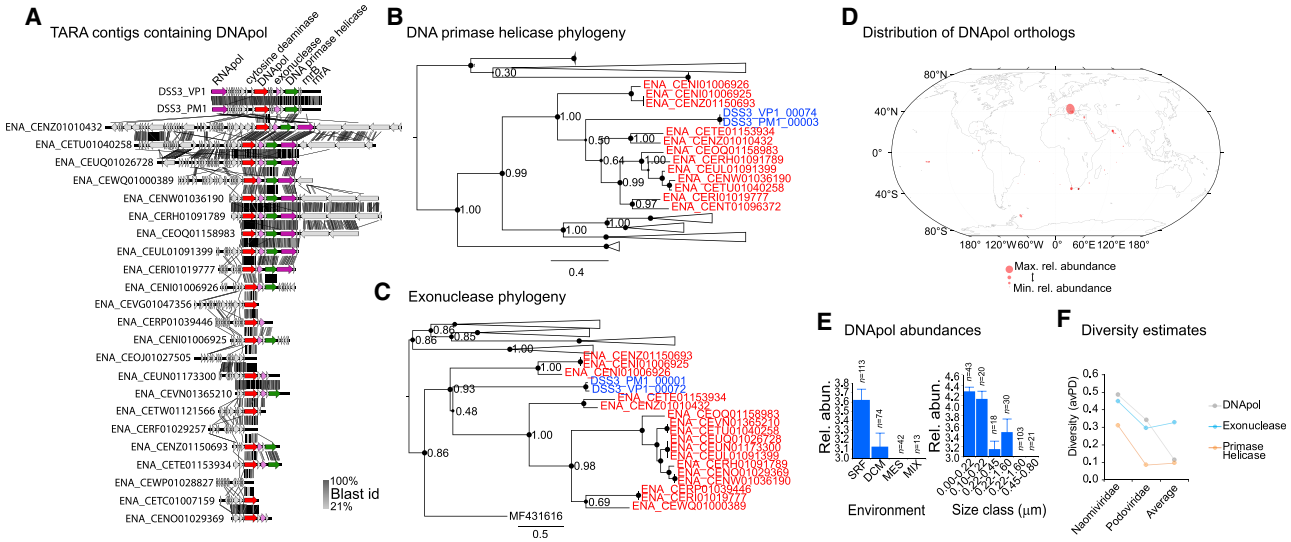


Figure 4. The relative abundance of "Naomiviridae" family phages in TARA Ocean metagenomes

(A) TARA Ocean contigs containing DNA polymerase sequences clustering with Naomiviridae.

(B and C) Phylogenies of DNA primase and exonucleases from the contigs in (A). Bootstrap support is depicted by the size of circles on nodes and is shown for major nodes. The scale bar represents substitutions per site.

(D) The relative abundance of DNA polymerase orthologs that share a most recent common ancestor with Naomiviridae from the TARA Ocean dataset.

(E) The relative abundance of the same genes in each sample and size class. Surface, SRF; surface mixed layer, MIX; deep chlorophyll maximum, DCM; mesopelagic, MES. Error bars represent standard error.

(F) Estimation of phylogenetic diversity (avPD) of "Naomiviridae." avPD was calculated using VP1 and PM1 sequences combined with all TARA sequences. For comparison, the diversity of known *Podoviridae* and the average branch length for all bacteriophage used are shown.

and an exonuclease. These genes also share a most recent common ancestor with DSS3_VP1 and DSS3_PM1 in phylogenetic analyses (Figures 4B and 4C). Furthermore, 8 out of 23 contigs also carried a putative homolog of cytosine deaminase, an enzyme likely required for replacing dTTP with dUTP (Figure 2D). The presence of this gene, both in the two isolated phages and environmental viruses represented by these contigs, further supports their monophyletic origin, both through marker gene similarity and shared potential for genomic DNA modification. Despite low species richness, we estimate that this novel clade is extremely diverse given the long branch lengths in both gene trees (Figures 4B and 4C). To better quantify this diversity, we applied a commonly used estimator of phylogenetic diversity, the average branch length (avPD[37,38]). For comparison, depending on the gene, they are between 1.4 and 3.7 times more diverse when compared to known *Podoviridae* isolates (Figure 4F). Next, we used the relative abundance of each of the 32 DNApol orthologs that clustered within the "Naomiviridae" to examine the distribution and biogeography of this clade. These analyses suggest that while "Naomiviridae" are globally distributed (being found in 93% of 243 samples) and their sequences could be recovered from all three ocean basins sampled (Pacific, Atlantic, and Indian), their abundance was very low (Figure 4D). However, we did observe increased abundance in the Adriatic and Mediterranean Sea (Figure 4D). This suggests a restricted biogeography for this clade, which explains why DSS3_VP1 was readily isolated from waters surrounding Venice. We further note that this clade is enriched in surface waters and in the sub-0.22 μm size class (Figure 4E).

Here, we report the isolation and characterization of two unique bacteriophages infecting cosmopolitan marine bacteria of the marine *Roseobacter* clade, organisms playing a crucial role in the biogeochemical cycling of C, N, and S including the metabolism of key climate-active trace gases (e.g., dimethylsulfide, methylamines). These are the first known marine bacteriophages containing noncanonical dU instead of the canonical deoxythymidine in their genomes. They represent a novel family of previously uncharacterized dsDNA viruses that are widely distributed in the global ocean. The substitution with dU adds to the expanding diversity of DNA modifications used by phages. Three recent studies found that several bacteriophages can replace adenine with noncanonical nucleobases aminoadenine, which is preferably used by phage DNA polymerase.[39–41] The ecological significance of these dU-containing marine bacteriophages remains to be determined given they have likely been underestimated in ocean surveys due to an inability of conventional metagenomic sequence libraries to detect dU-DNA. However, it is clear that phages represent a rich source of novel,

(B) Maximum likelihood proteomic tree for DSS3_PM1 and DSS3_VP1 calculated using VIPtree[32] against all double-stranded DNA viruses. The leaves labeled in red represent the position of the DSS3_VP1 and DSS3_PM1 phages. The tree shows only 413 of the closest genomes to DSS3_VP1 and DSS3_PM1 for the sake of image clarity.

(C) vConTACT analysis. Related bacteriophages are represented as blue circles with edges between them representing shared protein clusters. The genomes of DSS3_VP1 and DSS3_PM1 are represented as red circles with the connections between them also colored in red. See also Figure S4 for maximum likelihood tree of TerL, PhoH, and DNA polymerase I from roseophages.

non-canonical base modifications that together expand the genetic blueprint of life on Earth.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Growth conditions
  - Isolation of roseophages
- METHOD DETAILS
  - DNA extraction and sequencing
  - Characterization of bacteriophage-modified nucleosides
  - UDG, endonuclease VIII and restriction enzyme assays
  - Phage characterization and phage infection dynamics
  - Phylogenetic analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.05.014.

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 726116). We also acknowledge the Midlands Regional Cryo-EM Facility hosted at the Warwick Advanced Bioimaging Research Technology Platform for use of the JEOL 2100Plus, and the MRC CLIMB Infrastructure for bioinformatic analysis (MR/L015080/1). We thank Prof. Jed Fuhrman for stimulating discussions and suggestions for genome sequencing and MicrobesNG for Nextera library preparation and sequencing.

AUTHOR CONTRIBUTIONS

B.R. and Y.C. conceived the research. B.R., R.J.P., A.H., Y.-J.L., Y.Z., S.M., and A.D.M. performed research. B.R. and Y.C. wrote the manuscript and all co-authors assisted with data interpretation and manuscript editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 24, 2020
Revised: April 19, 2021
Accepted: May 10, 2021
Published: May 24, 2021

REFERENCES

1. Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature *399*, 541–548.
2. Talmy, D., Beckett, S.J., Taniguchi, D.A.A., Brussaard, C.P.D., Weitz, J.S., and Follows, M.J. (2019). An empirical model of carbon flow through marine viruses and microzooplankton grazers. Environ. Microbiol. *21*, 2171–2181.
3. Hurwitz, B.L., and U'Ren, J.M. (2016). Viral metabolic reprogramming in marine ecosystems. Curr. Opin. Microbiol. *31*, 161–168.
4. Moran, M.A., and Miller, W.L. (2007). Resourceful heterotrophs make the most of light in the coastal ocean. Nat. Rev. Microbiol. *5*, 792–800.
5. Voget, S., Wemheuer, B., Brinkhoff, T., Vollmers, J., Dietrich, S., Giebel, H.A., Beardsley, C., Sardemann, C., Bakenhus, I., Billerbeck, S., et al. (2015). Adaptation of an abundant Roseobacter RCA organism to pelagic systems revealed by genomic and transcriptomic analyses. ISME J. *9*, 371–384.
6. Wagner-Döbler, I., and Biebl, H. (2006). Environmental biology of the marine *Roseobacter* lineage. Annu. Rev. Microbiol. *60*, 255–280.
7. Buchan, A., LeCleir, G.R., Gulvik, C.A., and González, J.M. (2014). Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat. Rev. Microbiol. *12*, 686–698.
8. Curson, A.R., Todd, J.D., Sullivan, M.J., and Johnston, A.W. (2011). Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. Nat. Rev. Microbiol. *9*, 849–859.
9. Bullock, H.A., Luo, H., and Whitman, W.B. (2017). Evolution of dimethylsulfoniopropionate metabolism in marine phytoplankton and bacteria. Front. Microbiol. *8*, 637.
10. Mausz, M.A., and Chen, Y. (2019). Microbiology and ecology of methylated amine metabolism in marine ecosystems. Curr. Issues Mol. Biol. *33*, 133–148.
11. Zhan, Y., and Chen, F. (2019a). Bacteriophages that infect marine roseobacters: genomics and ecology. Environ. Microbiol. *21*, 1885–1895.
12. Bischoff, V., Bunk, B., Meier-Kolthoff, J.P., Spröer, C., Poehlein, A., Dogs, M., Nguyen, M., Petersen, J., Daniel, R., Overmann, J., et al. (2019). Cobaviruses - a new globally distributed phage group infecting Rhodobacteraceae in marine ecosystems. ISME J. *13*, 1404–1421.
13. Zhang, Z., Chen, F., Chu, X., Zhang, H., Luo, H., Qin, F., Zhai, Z., Yang, M., Sun, J., and Zhao, Y. (2019). Diverse, abundant, and novel viruses infecting the marine *Roseobacter* RCA lineage. mSystems *4*, e00494, e19.
14. Zhan, Y., and Chen, F. (2019b). The smallest ssDNA phage infecting a marine bacterium. Environ. Microbiol. *21*, 1916–1928.
15. Zhan, Y., Huang, S., Voget, S., Simon, M., and Chen, F. (2016). A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agents. Sci. Rep. *6*, 30372.
16. Zhao, Y., Wang, K., Jiao, N., and Chen, F. (2009). Genome sequences of two novel phages infecting marine roseobacters. Environ. Microbiol. *11*, 2055–2064.
17. Lee, Y.-J., Dai, N., Walsh, S.E., Müller, S., Fraser, M.E., Kauffman, K.M., Guan, C., Corrêa, I.R., Jr., and Weigele, P.R. (2018). Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. Proc. Natl. Acad. Sci. USA *115*, E3116–E3125.
18. Huang, L.H., Farnet, C.M., Ehrlich, K.C., and Ehrlich, M. (1982). Digestion of highly modified bacteriophage DNA by restriction endonucleases. Nucleic Acids Res. *10*, 1579–1591.
19. Berkner, K.L., and Folk, W.R. (1979). The effects of substituted pyrimidines in DNAs on cleavage by sequence-specific endonucleases. J. Biol. Chem. *254*, 2551–2560.
20. Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. *43*, D298–D299.
21. Medema, M.H., Takano, E., and Breitling, R. (2013). Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol. Biol. Evol. *30*, 1218–1223.
22. Schormann, N., Ricciardi, R., and Chattopadhyay, D. (2014). Uracil-DNA glycosylases-structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci. *23*, 1667–1685.
23. Takahashi, I., and Marmur, J. (1963). Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*. Nature *197*, 794–795.

24. Kiljunen, S., Hakala, K., Pinta, E., Huttunen, S., Pluta, P., Gador, A., Lönnberg, H., and Skurnik, M. (2005). Yersiniophage phiR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. Microbiology (Reading) *151*, 4093–4102.

25. Price, A.R., and Fogt, S.M. (1973). Deoxythymidylate phosphohydrolase induced by bacteriophage PBS2 during infection of *Bacillus subtilis*. J. Biol. Chem. *248*, 1372–1380.

26. Galeazzi, L., Bocci, P., Amici, A., Brunetti, L., Ruggieri, S., Romine, M., Reed, S., Osterman, A.L., Rodionov, D.A., Sorci, L., and Raffaelli, N. (2011). Identification of nicotinamide mononucleotide deamidase of the bacterial pyridine nucleotide cycle reveals a novel broadly conserved amidohydrolase family. J. Biol. Chem. *286*, 40365–40375.

27. Lavigne, R., and Vandersteegen, K. (2013). Group I introns in *Staphylococcus* bacteriophages. Future Virol. *8*, 997–1005.

28. Puxty, R.J., Evans, D.J., Millard, A.D., and Scanlan, D.J. (2018). Energy limitation of cyanophage development: implications for marine carbon cycling. ISME J. *12*, 1273–1286.

29. Millard, A.D., Gierga, G., Clokie, M.R., Evans, D.J., Hess, W.R., and Scanlan, D.J. (2010). An antisense RNA in a lytic cyanophage links *psbA* to a gene encoding a homing endonuclease. ISME J. *4*, 1121–1135.

30. Michniewski, S., Redgwell, T., Grigonyte, A., Rihtman, B., Aguilo-Ferretjans, M., Christie-Oleza, J., Jameson, E., Scanlan, D.J., and Millard, A.D. (2019). Riding the wave of genomics to investigate aquatic coliphage diversity and activity. Environ. Microbiol. *21*, 2112–2128.

31. Goldsmith, D.B., Crosti, G., Dwivedi, B., McDaniel, L.D., Varsani, A., Suttle, C.A., Weinbauer, M.G., Sandaa, R.-A., and Breitbart, M. (2011). Development of *phoH* as a novel signature gene for assessing marine phage diversity. Appl. Environ. Microbiol. *77*, 7730–7739.

32. Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017). ViPTree: the viral proteomic tree server. Bioinformatics *33*, 2379–2380.

33. Barylski, J., Enault, F., Dutilh, B.E., Schuller, M.B., Edwards, R.A., Gillis, A., Klumpp, J., Knezevic, P., Krupovic, M., Kuhn, J.H., et al. (2020). Analysis of *Spounaviruses* as a case study for the overdue reclassification of tailed phages. Syst. Biol. *69*, 110–123.

34. Adriaenssens, E.M., Wittmann, J., Kuhn, J.H., Turner, D., Sullivan, M.B., Dutilh, B.E., Jang, H.B., van Zyl, L.J., Klumpp, J., Lobocka, M., et al. (2018). Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Arch. Virol. *163*, 1125–1129.

35. Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J.H., Roux, S., Adriaenssens, E.M., Brister, J.R., Kropinski, A.M., Krupovic, M., Lavigne, R., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat. Biotechnol. *37*, 632–639.

36. Wardle, J., Burgers, P.M., Cann, I.K., Darley, K., Heslop, P., Johansson, E., Lin, L.J., McGlynn, P., Sanvoisin, J., Stith, C.M., and Connolly, B.A. (2008). Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. Nucleic Acids Res. *36*, 705–711.

37. Schweiger, O., Klotz, S., Durka, W., and Kühn, I. (2008). A comparative test of phylogenetic diversity indices. Oecologia *157*, 485–495.

38. Tucker, C.M., Cadotte, M.W., Carvalho, S.B., Davies, T.J., Ferrier, S., Fritz, S.A., Grenyer, R., Helmus, M.R., Jin, L.S., Mooers, A.O., et al. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. Biol. Rev. Camb. Philos. Soc. *92*, 698–715.

39. Zhou, Y., Xu, X., Wei, Y., Cheng, Y., Guo, Y., Khudyakov, I., Liu, F., He, P., Song, Z., Li, Z., et al. (2021). A widespread pathway for substitution of adenine by diaminopurine in phage genomes. Science *372*, 512–516.

40. Sleiman, D., Garcia, P.S., Lagune, M., Loc'h, J., Haouz, A., Taib, N., Röthlisberger, P., Gribaldo, S., Marlière, P., and Kaminski, P.A. (2021). A third purine biosynthetic pathway encoded by aminoadenine-based viral DNA genomes. Science *372*, 516–520.

41. Pezo, V., Jaziri, F., Bourguignon, P.-Y., Louis, D., Jacobs-Sera, D., Rozenski, J., Pochet, S., Herdewijn, P., Hatfull, G.F., Kaminski, P.-A., and Marliere, P. (2021). Noncanonical DNA polymerization by aminoadenine-based siphoviruses. Science *372*, 520–524.

42. Cavaluzzi, M.J., and Borer, P.N. (2004). Revised UV extinction coefficients for nucleoside-5¢-monophosphates and unpaired DNA and RNA. Nucleic Acids Res. *32*, e13.

43. Joshi, N.A., and Fass, J.N. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). https://github.com/najoshi/sickle.

44. Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling genomes and mini-metagenomes from highly chimeric reads. In Annual International Conference on Research in Computational Molecular Biology, pp. 158–170.

45. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *30*, 2068–2069.

46. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

47. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. *33*, 1870–1874.

48. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

49. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

50. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

51. Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform *23*, 205–211.

52. Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

53. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.

54. Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., Bachelerie, P., Rosnet, T., Pelletier, E., Sunagawa, S., and Hingamp, P. (2018). The Ocean Gene Atlas: exploring the biogeography of plankton genes online. Nucleic Acids Res. *46* (W1), W289–W295.

55. Rihtman, B., Meaden, S., Clokie, M.R., Koskella, B., and Millard, A.D. (2016). Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. PeerJ *4*, e2055.

56. Schindelin, J., Rueden, C.T., Hiner, M.C., and Eliceiri, K.W. (2015). The ImageJ ecosystem: an open platform for biomedical image analysis. Mol. Reprod. Dev. *82*, 518–529.

57. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47* (W1), W256–W259.

58. Sievers, F., and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. *27*, 135–145.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Ruegeria pomeroyi DSS-3	DSMZ culture collection	Strain ID, DSM 15171
Phage DSS3_VP1	Isolated near Venice, Italy. This study	N/A
Phage DSS3_PM1	Isolated near Puerto Morelos, Mexico. This study	N/A
Critical commercial assays		
TruePrime WGA kit	Expedeon	Catalog # 370025
Genomic DNA Clean & Concentrator	Zymo Research	Catalog # D4011
Qubit dsDNA BR Assay Kit	ThermoFisher	Catalog # Q32853
Uracil DNA glycosylase	NEB	Catalog # M0280S
Endonuclease VIII	NEB	Catalog # M0299S
*Xba*I	ThermoFisher	Catalog # ER0682
*Eco*RV	NEB	Catalog # R3195S
*Hin*dIII	ThermoFisher	Catalog # ER0501
Deposited data		
DSS3_VP1 genome sequence	This study	GenBank: MN602266
DSS3_PM1 genome sequence	This study	GenBank: MN602267
Software and algorithms		
Sickle	[43]	v1.33
SPAdes	[44]	v3.12.0
Prokka	[45]	v.1.12
uSearch	[46]	v.10.0.240
MEGA7	[47]	v7
IQ-tree	[48,49]	v.1.6.3
VIPtree	[32]	v1.9
vConTACT	[35]	v2.0
Cytoscape	[50]	v3.7.2
HMMER	[51]	v3.1b2
MUSCLE	[52]	v3.8.31
FastTree	[53]	2.1.8
TARA Ocean Gene Atlas	[54]	http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yin Chen (y.chen.25@warwick.ac.uk).

Materials availability

The phages isolated in this study are available from the Chen laboratory at the University of Warwick, UK.

Data and code availability

The genome sequences of DSS3_VP1 and DSS3_PM1 have been submitted to GenBank under accession numbers MN602266 and MN602267 respectively. This study did not generate new unique code.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Growth conditions

All the experiments were performed on bacterial cultures grown in Marine Broth (Difco) at 30°C, shaking at 170 rpm.

Isolation of roseophages

The two roseophages described in this study were isolated from marine coastal waters collected from geographically distant locations. DSS3_VP1 was isolated from seawater collected in the proximity of Rialto Bridge, Venice, Italy and DSS3_PM1 was isolated from surface waters near Puerto Morelos, Mexico. All of the seawater samples were filtered through a 0.22 μm pore size syringe filter and 1 mL of the filtered seawater then incubated overnight with an exponentially growing culture of *Ruegeria pomeroyi* sp. DSS-3 in Marine Broth 2216 (MB) medium (Difco) for enrichment. The infected culture was collected by centrifugation at 13,000 g for 5 min and the supernatant filtered through a 0.22 μm pore size syringe filter. Filtered lysate was then serially diluted in MB, dilutions mixed with 1 mL exponentially growing host and 4 mL 0.8% (w/v) agar in MB medium and poured into 6-well plates (Falcon). Plates were incubated overnight at 30°C and examined for plaques. Observed plaques were picked using 1ml filter pipette tips and agar plugs re-suspended in 10 mL fresh MB, incubated overnight at 4°C and again serially diluted for an additional round of plaque assays. This process was repeated 3 times in order to ensure that each plaque represented a single bacteriophage.

METHOD DETAILS

DNA extraction and sequencing

An exponentially growing *R. pomeroyi* sp. DSS-3 culture (10 ml) was infected with 100 μl phage stock solution and incubated overnight at 30°C with shaking at 170 rpm. Lysates were subsequently centrifuged at 13,000 g to remove cell debris and filtered through 0.22 μm pore size syringe filters. Filtered lysates were then used to extract viral DNA, using a modified phenol-chloroform extraction method.[55] Briefly, lysate was incubated with DNase I in order to remove bacterial DNA and submitted to phenol-chloroform-isoamyl alcohol DNA extraction. DNA was precipitated using 2:1 volume ratio of ethanol and 7.5 M ammonium acetate and purified using a Genomic DNA Clean & Concentrator Kit (Zymo Research, CA, USA). The purity of the phage DNA was examined using a Nanodrop and the amount of DNA quantified using Qubit dsDNA BR kit (Life Technologies). Purified DNA was then diluted to 0.2 ng μl^{-1} and used for Nextera library preparation, according to the manufacturer's instructions. Prepared libraries were sequenced using an Illumina MiSeq. In the case of DSS3_VP1 and DSS3_PM1, Nextera library preparations consistently failed to produce sufficient DNA for the MiSeq run. Thus, DNA from these two phages was subjected to whole genome amplification using the TruePrime WGA kit (Expedeon, CA, USA), according to the manufacturer's instructions. Amplified DNA was then used successfully in Nextera library preparations and sequenced using an Illumina MiSeq by MicrobesNG (https://microbesng.com/).

Fastq files were trimmed with Sickle v1.33, using default parameters[43] and assembled with SPAdes v3.12.0, with the careful option.[44] In each case, the assembly produced a single phage contig with > 100-fold coverage. Single contigs where then annotated with Prokka v.1.12[45] against a custom database[30] made out of previously published bacteriophage genomes.

Characterization of bacteriophage-modified nucleosides

Approximately 1-5 μg of purified PM1 and VP1 genomic DNAs were treated with 1 μl of Nucleoside Digestion Mix (New England Biolabs, Ipswich, MA) in a 20 μl reaction volume using the supplied buffer and incubated at 37°C for at least two h to overnight.[17] The product nucleosides were then subjected to the liquid chromatography (LC) and liquid chromatography–mass spectrometry (LC-MS) analyses. The LC-MS was performed on an Agilent HPLC/MS Single Quad System 1200 Series equipped with a diode array UV detector and a 6120 Quadrupole mass detector operated in electrospray ionisation mode. LC-based separation of the nucleosides upstream of the MS analysis was achieved using a reverse phase Waters Atlantis T3 C18 column (100 Å pore size, 3 μm resin particle size, and column dimensions of 4.6 × 150 mm) operated at a flow rate of 0.5 mL/min in a binary gradient from 2% solvent A (10 mM ammonium acetate, pH 4.5) to 100% B (methanol). Elution of each nucleoside species was monitored by UV absorbance at 260 nm. LC analyses were also performed on an Agilent 1290 UHPLC system equipped with a diode array detector using a reverse phase Waters XSelect HSS T3 XP column (100 Å pore size, 2.5 μm resin size, and bed dimensions of 2.1 × 100 mm) at a flow rate of 0.6 mL/min with a binary gradient from 2% solvent A (10 mM ammonium acetate, pH 4.5) to 100% B (methanol) and the effluent monitored by UV absorbance at 260 nm. The peak area of each nucleoside species resolved in the HPLC trace was calculated using the Agilent ChemStation software and divided by the corresponding nucleoside molar extinction coefficient (ε) at 260 nm to determine relative abundance. The extinction coefficients used were 7100 cm^{-1} M^{-1} for dC, 15066 cm^{-1} M^{-1} for dG, 8560 cm^{-1} M^{-1} for dT, 15060 cm^{-1} M^{-1} for dA and 9780 cm^{-1} M^{-1} for dU.[42]

UDG, endonuclease VIII and restriction enzyme assays

Genomic DNA from DSS3_VP1, DSS3_PM1, WGA products of DSS3_VP1 and DSS3_PM1 DNA and λ phage DNA (NEB) were treated with UDG (NEB) and endonuclease VIII (NEB), according to the manufacturer's instructions. Similarly, these DNA samples were treated with *Eco*RV (NEB), *Hin*dIII (ThermoFisher Scientific), and *Xba*I (ThermoFisher Scientific) by incubating them for 60 min at 37°C, followed by an inactivation step at 80°C for 1 min. The digested DNA samples were then run on a 1% agarose gel stained with GelRed Nucleic Acid Stain (Sigma-Aldrich) and visualized.

Phage characterization and phage infection dynamics

Phages were imaged using a Jeol 2100Plus transmission electron microscope. Prior to imaging, filtered phage lysates were stained with 2% (w/v) uranyl acetate and blotted onto size 200 copper mesh carbon grids (Agar Scientific). Phage images were processed using ImageJ software[56] and the size bar used to obtain the measurements of each phage particle. The reported sizes represent an average of 3 separate phage particle measurements.

Phage growth parameters were estimated using one-step infection methods. In order to estimate the length of the latent phase an exponentially growing culture of *R. pomeroyi* sp. DSS-3 was diluted to OD_{540} = 0.1 and 200 μl of this diluted culture grown in a 96-well plate in a Fluostar Omega 415-101 (BMG Labtech) plate reader at 30°C with shaking at 200 rpm until the culture reached an OD_{540}»0.4. At this point 5 μl filtered phage lysate (concentration of approx. 10^9 pfu/mL) was added to triplicate wells containing *R. pomeroyi* sp. DSS-3. The plate was then returned to the plate reader and incubated for a further 24 h, with OD_{540} measurements taken automatically every 10 min. The latent phase was estimated as the time measured between the addition of phage until the cessation of culture growth, as measured using the OD_{540}.

The burst size of each phage was also measured using one-step growth curves with 20 mL *R. pomeroyi* sp. DSS-3 culture grown in MB, using a multiplicity of infection (MOI) of 0.1, in triplicate. *R. pomeroyi* cell number was measured by flow cytometry using a Cytoflex flow cytometer (Beckman Coulter). After phage addition, the infected culture was incubated at 30°C for 30 min and then centrifuged at 4000xg in order to separate the infected cells from free phage. The pelleted cells were re-suspended in 20 mL fresh MB medium and samples taken from each infected culture for the purpose of estimating the number of infected cells at time point 0. Infected cells were serially diluted, and plaque assays performed using the soft agar overlay method. The number of plaques measured was then used as a proxy for the number of initially infected cells. Infected cultures were then sampled every 30 min for 6 h and the collected samples filtered through a 0.45 μm pore size filter into 96 well plates (Falcon) using a vacuum manifold (Promega) in order to separate the free phage from infected cells. Filtrate was subsequently used for plaque assays and the number of phages released over time estimated. Burst size was then calculated as $Burst\ size = \frac{PFUf - PFUi}{IC}$ where PFUf = the number of phages measured at the end of the latent phase, PFUi = the number of free phages measured at the first time point and IC = the number of infected cells sampled at time point zero, as described above.

Phylogenetic analyses

In order to classify these novel phages and to assign them into known taxonomic units, we performed phylogenetic analysis of protein sequences encoded by three genes found in all three phages: the large terminase subunit (TerL), DNA polymerase I and the tail tape measure protein. Amino acid sequences of each protein from all three phages were used as a query for a uBLAST search against all published bacteriophage genomes/sequences, as curated by the Millard lab at the University of Leicester.[30] uBLAST was performed using the uSEARCH package[46] with the following parameters "-ublast -evalue 1e-5." Query sequences were added to the list of orthologous proteins produced by this uBLAST search and aligned using the clustalW alignment algorithm in MEGA7 software.[47] Phylogenetic trees were then produced using the IQ-Tree script[48] using the Ultrafast Bootstrap Approximation algorithm[49] with the following parameter "-bb 1000." The chosen evolution models were VT+F+R4 (TerL), Blosum62+F+R7 (DNAPol) and VT+R6 (PhoH). Tree files were edited and further annotated using the iTOL phylogenetic trees online annotation tool.[57]

Additionally, complete phage genomes were compared against the VIPTree dsDNA viral database[32] as well as using vConTACT v2.0 using the following settings: -rel-mode Diamond,–pcs-mode MCL,–vcs-mode ClusterONE,–db ProkaryoticViralRefSeq97-Merged.[35] The structure of viral protein networks produced by vConTACT was visualized using Cytoscape[50] using an edge-weighted spring embedded model, where genomes that are closely related based on shared protein clusters are placed closer to each other.

In order to recruit environmental reads similar to the newly isolated phages, profile hidden Markov models (HMM) were created using the protein sequences of three marker genes from DSS3_VP1 and DSS3_PM1, including the phosphate stress induced protein (*phoH, PF02562*), the large subunit terminase (*terL, PF03237*) and DNA polymerase (*polA, PF00476*). HMM profiles were created from the HMMER package[51] with default parameters. The HMM profile was used to query the *TARA* Ocean Gene Atlas.[54] ClustalO[58] with default parameters was used to align orthologous sequences from the TARA Ocean Gene Atlas to protein families from DSS3_VP1 and DSS3_PM1 and those from a uBLAST search against the Millard Lab published bacteriophage genome database.[30] Alignments were manually inspected and are included in Supplemental Data. FastTree with default parameters[53] was used to reconstruct maximum likelihood phylogenetic relationships of each protein family. Those environmental orthologs that formed a well-supported monophyletic cluster with DSS3_VP1 and DSS3_PM1, were then queried for relative abundance in different sampling sites, as provided by the TARA Ocean Gene Atlas database.[54] Phylogenetic analyses of DNA primase/helicase and exonuclease sequences obtained from the *TARA* Oceans were aligned using MUSCLE.[52] Phylogenetic trees were then produced with FastTree under the gtr model.[53]

The *TARA* Ocean Gene Atlas provides an estimate of relative abundance of each orthologous gene by mapping raw reads back to a reference gene catalog (genes called from assembled metagenomes from the same samples).[54] Ocean Gene Atlas relative abundance (OGA RA) is expressed as the gene's coverage divided by the sum of all gene's coverage from the same sample. These abundances were extremely right skewed. Therefore, OGA RE estimates were log transformed (and scaled by a factor of 10^{10}). The resulting distributions were approximately normal. To recover contigs from which the DNApol orthologs were called we searched for the exact nucleotide sequence of each gene that clustered with DSS3_PM1 and DSS3_VP1 in publicly available assemblies of the TARA sequences available at https://www.ebi.ac.uk/ena/about/tara-oceans-assemblies.

QUANTIFICATION AND STATISTICAL ANALYSIS

The avPD metric was calculated according to Wardle et al.[36] and Schweiger et al.[37] For the desired node, the sum of the branch lengths within that node were divided by the number of taxa in that node. This was computed for: 1. The node containing all OGA reads clustering with VP1/PM1, 2. The node(s) containing all known *Podoviridae* sequences and 3. All taxa within the tree.