

RESEARCH

Open Access



Isolation and characterization of a roseophage representing a novel genus in the N4-like *Rhodovirinae* subfamily distributed in estuarine waters

Xingyu Huang¹, Chen Yu² and Longfei Lu^{3*}

Abstract

Background *Roseobacteraceae*, often referred to as the marine roseobacter clade (MRC), are pivotal constituents of bacterial communities in coastal and pelagic marine environments. During the past two decades, 75 roseophages that infect various *Roseobacteraceae* lineages have been isolated. The N4-like roseophage clade, which encompasses 15 members, represents the largest clade among these roseophages. N4-like phages form a monophyletic group, classified as family *Schitoviridae*. And all N4-like roseophages form a unique clade within *Schitoviridae* and has been classified as the *Rhodovirinae* subfamily.

Results In this study, we isolated a novel roseophage, vB_DshP-R7L, that infects *Dinoroseobacter shibae* DFL12 from Xiamen Bay in the East China Sea. Conserved genes of *Schitoviridae* have been identified in the genome of vB_DshP-R7L, and following phylogenetic analysis suggests that the newly isolated phage is a member of the *Rhodovirinae* subfamily and represents the sole member of a novel genus, *Gonggongvirus*. The genome of vB_DshP-R7L harbors six auxiliary metabolic genes (AMGs), most of which potentially enhance DNA de novo synthesis. Additionally, a gene encoding ribosomal protein was identified. Comparative genomic analysis of AMG content among *Rhodovirinae* indicates a distinct evolutionary history characterized by independent ancient horizontal gene transfer events. Read-mapping analysis reveals the prevalence of vB_DshP-R7L and other *Rhodovirinae* roseophages in estuarine waters.

Conclusions Our work illustrates the genomic features of a novel roseophage clade among the subfamily *Rhodovirinae*. The AMG content of vB_DshP-R7L is under severe purification selection, which reveals their possible ecological importance. We also demonstrated that vB_DshP-R7L and other *Rhodovirinae* roseophages are only detected in estuaries. Our isolation and characterization of this novel phage expands the understanding of the phylogeny, gene transfer history, and biogeography of *Rhodovirinae* infecting marine *Roseobacteraceae*.

Keywords Bacteriophage, Roseophage, Isolation, Phylogenetic analysis, Auxiliary metabolism gene, Biogeography

*Correspondence:

Longfei Lu

lulongfei@4io.org.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Phages are viruses that can infect bacteria, and play critical roles in biogeochemical and ecological functions. In the oceans, phages outnumber their hosts by an order of magnitude [1, 2], influence microbial community dynamics by modulating host mortality, and facilitate horizontal gene transfer [1, 3]. Marine phages frequently contain auxiliary metabolic genes (AMGs). AMGs are viral genes originally acquired from host genomes through horizontal gene transfer. They enable viruses to manipulate host cellular machinery during infection, alleviate metabolic bottlenecks, and augment host metabolism to maximize viral production [4–7].

The marine roseobacter clade (MRC) represents one of the most abundant heterotrophic bacterial taxa in coastal ecosystems, comprising approximately 5% and 20% of bacterial communities in open oceans and nearshore ecosystems [8, 9]. All members of the MRC form a single monophyletic group, classified as family *Roseobacteraceae*, which occupies diverse ecological niches [9, 10]. The *Roseobacteraceae* are pivotal in marine biogeochemical cycles due to their roles in both aerobic and anaerobic photosynthesis, and they are the primary reducers of dimethylsulfoniopropionate (DMSP) produced by dinoflagellates [10–12].

To date, 75 roseophages that infect members of the *Roseobacteraceae* have been identified [13–37] (Table 1). Of these, two belong to *Monodnaviria*, with the others classified as *Caudoviricetes*. Comparative genomic analyses have delineated five major clades among the *Caudoviricetes* roseophages, namely *Schitoviridae*, *Cobavirus*, *Autographiviridae*, Cbk-like, and Chi-like clades [14, 18, 38, 39]. Roseophages from the *Cobavirus*, *Autographiviridae*, and *Naomviridae* clades have shown a global distribution [13, 14, 18].

Roseophages from the *Schitoviridae* clade constitute the largest group of isolated roseophages. The family *Schitoviridae*, previously recognized as the N4-like phage clade, shares several conserved genes forming a universal backbone, notably three DNA-dependent RNA polymerase genes [40, 41]. The genomes of *Schitoviridae* are transcribed in three programmed stages during infection, with each RNA polymerase functioning at a specific stage [41]. All roseophages within the *Schitoviridae* form a unique monophyletic group classified as the subfamily *Rhodovirinae*, which is further divided into seven genera: *Aorunvirus*, *Raunefjordvirus*, *Aoquinvirus*, *Pomeroyivirus*, *Sanyabayvirus*, *Plymouthvirus*, and *Baltimorevirus* [40]. Comparative genomic analysis has revealed diverse AMGs in *Rhodovirinae* genomes, which highlights their potential ecological significance [39]. These AMGs are predominantly involved in the DNA de novo synthesis pathway, including *trx*, *grx*, *rnr*, *thyX* and *mazG* [39].

Despite detailed genomic characterizations, the distribution of *Rhodovirinae* in the marine environment remains poorly described.

In this study, we report a novel roseophage, vB_DshP-R7L, infecting *Dinoroseobacter shibae* DFL12 isolated from the coastal waters of Xiamen. Phylogenetic analysis suggests that vB_DshP-R7L represents a previously unrecognized genus within the *Rhodovirinae* subfamily. Six AMGs were identified in its genome, and their evolutionary history was comprehensively analyzed. The tRNA content in the vB_DshP-R7L genome may enhance the expression of its AMGs. Furthermore, metagenomic analysis revealed the prevalence of vB_DshP-R7L and its *Rhodovirinae* relatives in estuarine waters.

Methods

Isolation and purification of vB_DshP-R7L

vB_DshP-R7L was isolated from the surface water of Xiamen Bay station S03 (118.03N, 24.43E) during the Xiamen nearshore cruise in November 2013. Seawater samples were filtered through a 0.22 μm membrane and stored at 4°C in darkness. For cultivation, 5 ml of *Dinoroseobacter shibae* DFL12 host culture was incubated at 20°C and agitated at 100 rpm in RO medium (1 g·L⁻¹ yeast extract, 1 g·L⁻¹ peptone, 1 g·L⁻¹ sodium acetate, 1 L artificial seawater, with a pH maintained between 7.4 and 7.8) until reaching the exponential growth phase. 1 ml of the preserved seawater filtrate was then introduced to this culture medium and incubated for 7 days. The phage was isolated and purified through five successive rounds using the double-layer agar plate method.

After purification, phage plaques were collected with a pipette tip and expanded in 1 L of liquid RO medium containing *Dinoroseobacter shibae* DFL12 in its exponential phase. The culture was subsequently centrifuged at 12,000×g for 10 min at 4°C. The supernatant was transferred to a fresh container and the phage particles were concentrated via overnight precipitation using 100 g of PEG8000, followed by centrifugation at 10,000×g for 60 min at 4°C. The resulting precipitate was resuspended in 6 ml of SM buffer (100 mM NaCl, 8 mM MgSO₄, 50 mM Tris–HCl) and concentrated using CsCl density gradient ultracentrifugation following a previously reported method [42]. The phage-containing band was further purified using 30 kDa super-filters (UFC503096, Millipore) and adjusted to the desired concentration with SM buffer following the previously reported method to remove CsCl [35].

The purified viruses were also tested against various roseophage hosts including *Erythrobacter litoralis* DSM 8509, *Erythrobacter* sp. JLA75, *Erythrobacter* DSM 6997, and *Roseobacter denitrificans* DSM 7001, all cultured to their exponential phases. The host range was determined

Table 1 Summary of all isolated roseophages

Roseophage name	Roseophage order	Roseophage clade	Host name	Source water	Genome size (bp)	Number of ORFs	%G+C	Accession number	reference
CRP-113	Caudoviricetes	Autographiviridae	Roseobacter FZCC0023	Taizhou coast, East China Sea	39,709	48	44.89	OR420755	13
CRP-114	Caudoviricetes	Autographiviridae	Roseobacter FZCC0023	Yantai coast, Bohai Sea	40,413	50	48.81	OR420740	13
CRP-118	Caudoviricetes	Autographiviridae	Roseobacter FZCC0023	Taizhou coast, East China Sea	41,088	52	46.35	OR420741	13
CRP-125	Caudoviricetes	Autographiviridae	Roseobacter FZCC0023	Taizhou coast, East China Sea	38,917	50	48.7	OR420743	13
CRP-143	Caudoviricetes	Autographiviridae	Roseobacter FZCC0023	Ningbo coast, East China Sea	42,634	53	44.64	OR420748	13
CRP-171	Caudoviricetes	Autographiviridae	Roseobacter FZCC0023	Ningbo coast, East China Sea	39,796	48	44.89	OR420737	13
CRP-227	Caudoviricetes	Autographiviridae	Roseobacter FZCC0037	Yantai coast, Bohai Sea	39,373	49	48.31	OR420746	13
CRP-361	Caudoviricetes	Autographiviridae	Roseobacter FZCC0042	Yantai coast, Bohai Sea	39,084	47	48.48	OR420750	13
CRP-403	Caudoviricetes	Autographiviridae	Roseobacter FZCC0037	Pattaya coast, Indian Ocean	40,567	52	47.04	OR420752	13
CRP-804	Caudoviricetes	Autographiviridae	Roseobacter FZCC0196	Qingdao coast, Yellow Sea	41,620	52	50.02	OR420734	13
vB_RpoS-V10	Caudoviricetes	Cbk-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	147,480	240	56.36	MH015255	16
DSS3phi8	Caudoviricetes	Cbk-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor	146,135	234	56.35	KT870145	17
DSS3phi1	Caudoviricetes	Chi-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	59,601	84	64.13	NC_025428	17
vB_RpoS-V11	Caudoviricetes	Chi-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	59,549	84	64.02	MH015254	17
vB_RpoS-V16	Caudoviricetes	Chi-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	61,382	84	63.64	MH015258	17
vB_RpoS-V18	Caudoviricetes	Chi-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	59,111	84	64.03	MH015252	17
vB_RpoS-V7	Caudoviricetes	Chi-like	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	59,573	86	64.12	MH015249	17
CRP-4	Caudoviricetes	cobavirus	Planktomarina temperata FZCC0023	Yantai coast, Bohai sea	40,768	57	45.2	MK613346	18
CRP-5	Caudoviricetes	cobavirus	Planktomarina temperata FZCC0040	Pingtang coast, Taiwan strait	39,601	60	45.63	MK613347	18
LenP_VB1	Caudoviricetes	cobavirus	Lentibacter sp. SH36	North Sea	40,163	56	47.02	MF431617	19
LenP_VB2	Caudoviricetes	cobavirus	Lentibacter sp. SH36	North Sea	40,907	55	47.82	MF431615	19
LenP_VB3	Caudoviricetes	cobavirus	Lentibacter sp. SH36	North Sea	40,497	56	47.25	MF431616	19
MB-2001	Caudoviricetes	cobavirus	Roseobacter SIO67	Mission Bay California USA	38,252	56	44.14	FJ867914	20
OS-2001	Caudoviricetes	cobavirus	Roseobacter SIO67	Oceanside California USA	38,205	58	44.84	FJ867913	20

Table 1 (continued)

Roseophage name	Roseophage order	Roseophage clade	Host name	Source water	Genome size (bp)	Number of ORFs	%G + C	Accession number	reference
SBR SIO67-2001	Caudoviricetes	cobavirus	Roseobacter SIO67	Solana Beach California USA	38,150	59	44.15	FJ867912	20
SIO1	Caudoviricetes	cobavirus	Roseobacter SIO67	Pacific coast	39,898	35	46.16	AF189021	20
SIO1-1989	Caudoviricetes	cobavirus	Roseobacter SIO67	The pier at Scripps Institution of Oceanography USA	38,155	62	44.2	FJ867911	20
SIO1-2001	Caudoviricetes	cobavirus	Roseobacter SIO67	The pier at Scripps Institution of Oceanography USA	39,413	56	45.26	FJ867910	20
P12053L	Caudoviricetes	cobavirus	Celeribacter sp. strain IMCC12053	The coast of The Yellow Sea South Korea	38,889	58	46.1	JQ809650	21
DSS3phi22	Malgrandaviricetes	Microviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	4248	4	57.7	MF101922	16
vB_RpoMi-V15	Malgrandaviricetes	Microviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	4248	4	57.7	MH015251	16
DSS3_PM1	Caudoviricetes	Naomviridae	Ruegeria pomeroyi DSS3	Puerto Morelos, Mexico	70,044	103	47.34	MN602267	14
DSS3_VP1	Caudoviricetes	Naomviridae	Ruegeria pomeroyi DSS3	Rialto Bridge, Venice, Italy	75,087	109	47.5	MN602266	14
vB_RpoP-V12	Caudoviricetes	Schitoviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	74,704	85	47.9	FR719956.1	16
vB_RpoP-V13	Caudoviricetes	Schitoviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	74,830	81	50.77	MH015256	16
vB_RpoP-V14	Caudoviricetes	Schitoviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	74,792	89	47.91	MH015257	16
vB_RpoP-V17	Caudoviricetes	Schitoviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	74,665	84	47.93	MH015259	16
vB_RpoP-V21	Caudoviricetes	Schitoviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor, USA	74,665	84	47.91	MH015253	16
DS-1410Ws-06	Caudoviricetes	Schitoviridae	Roseobacter denitrificans OCh114	surface water of Sanya Bay, northern South China Sea	76,466	79	50.01	KU885988	22
RD1410Ws01	Caudoviricetes	Schitoviridae	Dinoroseobacter shibae DFL12T	surface water of Sanya Bay, northern South China Sea	72,674	77	49.53	KU885989	22
RD1410Ws07	Caudoviricetes	Schitoviridae	Roseobacter denitrificans OCh114	surface water of Sanya Bay, northern South China Sea	76,298	78	50.01	KU885990	22

Table 1 (continued)

Roseophage name	Roseophage order	Roseophage clade	Host name	Source water	Genome size (bp)	Number of ORFs	%G+C	Accession number	reference
DSS3phi2	Caudoviricetes	Schitoviridae	Ruegeria pomeroyi DSS3	Baltimore Inner Harbor water USA	74,611	83	47.92	FJ591093	23
EE36phi1	Caudoviricetes	Schitoviridae	Sulfitobacter sp. EE36	Baltimore Inner Harbor water USA	73,325	81	47.04	FJ591094	23
pCB2047-B	Caudoviricetes	Schitoviridae	Sulfitobacter sp. strain 2047	Mesocosm study Raunefjorden Norway	74,485	77	42.98	HQ317387	24
R1	Caudoviricetes	Schitoviridae	Dinoroseobacter shibae DFL12T	Baicheng Harbor Xiamen China	75,028	88	49.26	KJ621082	25
R2C	Caudoviricetes	Schitoviridae	Dinoroseobacter shibae DFL12T	Huangcuo station Xiamen China	74,806	86	49.17	KJ803031	26
RLP1	Caudoviricetes	Schitoviridae	Roseovarius sp. 217	Langstone Harbor Hampshire UK	74,583	92	49.02	FR682616	27
RPP1	Caudoviricetes	Schitoviridae	Roseovarius nubinihibens	L4 sampling station Plymouth UK	74,704	91	49.05	FR719956	27
vB_DshP-R7L	Caudoviricetes	Schitoviridae	Dinoroseobacter shibae DFL12T	Xiamen Costal	75,871	83	48.96	NC_070859	this study
pCB2051-A	Caudoviricetes		Loktanella sp. CB2051	Norwegian Sea,	56,958	76	54.98	NC_020853	16
CRP-1	Caudoviricetes		Planktomarina temperata FZCC0023	Osaka Bay, Japan	54,045	73	42.43	MK613343	18
CRP-2	Caudoviricetes		Planktomarina temperata FZCC0023	Pingtang coast	54,148	70	42.93	MK613344	18
CRP-3	Caudoviricetes		Planktomarina temperata FZCC0040	Yantai coast, Bohai sea	52,963	61	49.66	MK613345	18
CRP-6	Caudoviricetes		Planktomarina temperata FZCC0042	Pingtang coast, Taiwan strait	44,927	70	47.05	MK613348	18
CRP-7	Caudoviricetes		Planktomarina temperata FZCC0042	Yantai coast, Bohai sea	58,106	72	40.29	MK613349	18
pCB2047-A	Caudoviricetes		Sulfitobacter sp. strain 2047	Mesocosm study Raunefjorden Norway	40,929	72	58.8	HQ332142	23
pCB2047-C	Caudoviricetes		Sulfitobacter sp. strain 2047	Mesocosm study Raunefjorden Norway	40,931	73	58.99	HQ317384	23
CRP-207	Caudoviricetes		FZCC0040	Pattaya Beach, Thailand	54,895	76	46.2	MZ892987	28
CRP-212	Caudoviricetes		FZCC0040	Pattaya Beach, Thailand	54,748	59	48.58	MZ892988	28
CRP-235	Caudoviricetes		FZCC0040	North Sea	52,729	61	45.69	MZ892989	28
CRP-345	Caudoviricetes		FZCC0042	Pattaya Beach, Thailand	54,718	81	42.15	MZ892990	28
CRP-603	Caudoviricetes		FZCC0012	Pattaya Beach, Thailand	54,551	77	43.12	MZ892991	28
CRP-738	Caudoviricetes		FZCC0089	Pattaya Beach, Thailand	53,826	65	45.64	MZ892992	28

Table 1 (continued)

Roseophage name	Roseophage order	Roseophage clade	Host name	Source water	Genome size (bp)	Number of ORFs	%G+C	Accession number	reference
CRP-810	Caudoviricetes		CHUG strain FZCC0198	Yellow Sea	57,692	70	52.85	OR671924	29
CRP-13	Caudoviricetes		FZCC0023	North Sea	55,015	74	44.23	MW514247	30
CRP-9	Caudoviricetes		FZCC0023	Pattaya Beach, Thailand	56,157	73	41.61	MW514246	30
CRP-901	Caudoviricetes		FZCC0083	North Sea	53,013	77	45.57	OQ401623	31
CRP-902	Caudoviricetes		FZCC0083	Yellow Sea	51,954	80	45.5	OQ401624	31
NYA-2014a	Caudoviricetes		Sulfitobacter sp. strain 2047	Mesocosm study Raunefjord, Norway	42,092	71	58.54	KM233261	33
R4C	Caudoviricetes		Dinoroseobacter shibae DFL12T	Xiamen Coastal	36,268	50	66.77	NC_009952	34
R5C	Caudoviricetes		Dinoroseobacter shibae DFL12T	oligotrophic surface of SCS	77,874	123	61.55	KY606587	35
RDJLphi1	Caudoviricetes		Roseobacter denitrificans OCh114	South China Sea surface seawater	62,668	87	57.88	HM151342	36
RDJLphi2	Caudoviricetes		Roseobacter denitrificans OCh114	Wuyuan Bay Xiamen China	63,513	76	57.3	KT266805	36
vB_PeaS-P1	Caudoviricetes		Roseobacter Pelagibaca abyssi JLT2014	south eastern Pacific, 2000 m	38,868	51	63.52	KT381865	37
vB_ThpS-P1	Caudoviricetes		Roseobacter T. profunda JLT2016	south eastern Pacific, 2571 m	39,591	52	66.73	KT381864	37

using the double-layer agar plate method with each bacterial strain tested in triplicate.

Transmission electron microscopy (TEM) and phage growth curve

For electron microscopy, the purified and desalted phage particles were adsorbed onto 200-mesh carbon-coated copper grids for 10–30 min in the dark. After staining with 1% phosphotungstic acid for 30 min, the samples were examined using a JEM-2100 transmission electron microscope (JEOL, Tokyo, Japan) at an 80 kV voltage. Images were captured using an HPF-TEM system (Gatan Inc., Pleasanton, CA, USA).

A growth curve assay was conducted to evaluate the infectivity and replication capability of vB_DshP-R7L based on epifluorescence microscopy. Prior to conducting the phage growth curve analysis, 600 ml *Dinoroseobacter shibae* DFL12 has been cultured 4 h at 28 °C to its exponential phase (OD600 value 0.30), then evenly divided into three replicate samples. Phages were added to a each sample at a phage concentration ranging from 1.85 to 2.20×10^6 ml⁻¹, with an average concentration of

2.02×10^6 ml⁻¹. The culture was incubated at 28 °C with continuous shaking. Samples were collected every 30 min, and viral abundance was quantified using the epifluorescence microscopy method [43]. Growth rate was determined by dividing the average viral concentration at the end of the exponential phase of vB_DshP-R7L (ranging from 0.83 to 1.04×10^9 ml⁻¹, with an average of 9.1×10^8 ml⁻¹) by the viral concentration at the end of the latent phase (ranging from 2.18 to 3.32×10^6 ml⁻¹, with an average of 2.50×10^6 ml⁻¹).

Characterization of structural protein

Structural proteins of vB_DshP-R7L were analyzed by mixing phage particles with loading buffer and heating to 100 °C for 5 min. The heated samples were separated using 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and the gels were stained with silver. SDS-PAGE and following protein analysis were conducted using liquid chromatography–mass spectrometry at the Shanghai Applied Protein Technology [44].

DNA extraction and genome annotation

Phage DNA was extracted using the phenol–chloroform method, following the protocols described by Yang [35]. The extracted DNA was sequenced using an Illumina HiSeq 2500 sequencer with a paired-end, 2×251 bp configuration. The sequencing produced a total of 1.56 gigabases of clean reads. The CLC Genomics Workbench software was utilized to analyze the high-throughput sequencing data and to assemble these reads into a complete genome sequence. This sequencing and assembly process was performed by MAGIGENE Corporation. The complete genome sequence has been deposited in GenBank under the accession number MZ773648.1.

The genomic annotation was carried out using Prodigal v2.6.3 [45] to predict putative open reading frames (ORFs) within the vB_DshP-R7L genome. Functional annotation of these viral ORFs was performed using the DIAMOND BLASTp tool against the NCBI non-redundant protein sequence database, updated as of January 8, 2024. Additionally, InterProScan v5.52–86.0 was employed to annotate viral ORFs, referencing the Pfam-A database [46]. tRNAscan-SE was utilized to identify tRNAs within the genome, accepting only results classified as high-quality [47]. The involvement of auxiliary metabolic genes in cellular metabolic pathways was determined manually by analyzing ORF prediction and BLAST results against the KEGG pathway database and PHROGS database [48]. Beside experimental test, the genome of vB_DshP-R7L was matched against IMG environmental CRISPR spacer database and iPHoP(database version Aug_2023_pub_rw) to further check the host range of vB_DshP-R7L to search potential host of vB_DshP-R7L [49, 50].

Comparative genomics and phylogenetic analysis

The genome structure of vB_DshP-R7L and its homologous phages was compared using the EasyFig tool, which employs the TBLASTX comparison method [51]. Genomes of all isolated roseophages and *Schitoviridae* phages were retrieved from GenBank on September 1, 2024, and from a recent study [52]. Viral proteins were clustered at a 70% identity threshold using CD-hit with reference genomes to identify their homologs, adhering to a previously established method [53]. The conserved marker genes among the *Schitoviridae* phages which encoding the virion-encapsidated RNA polymerase and the major capsid protein to construct separate phylogenetic trees.

The phylogenetic tree of vB_DshP-R7L and other *Rhodovirinae* phages based on their complete genome sequences was generated using Virus Classification and Tree Building Online Resource (VICTOR, <https://ggdc.dsmz.de>). Marker genes were aligned using MAFFT

(v.7.453) under default settings and subsequently trimmed with trimAL (v.1.5) to eliminate gaps in over 10% of the genomes. The aligned sequences were used to construct a maximum likelihood phylogenetic tree with Raxml-ng [18], employing the LG+G8+F model, starting with an initial tree configuration of 10 and 100 bootstrap replicates. The resultant tree was visualized using iTOL [54]. Phylogenetic trees of AMGs in the vB_DshP-R7L and their homologs from *Rhodovirinae* hosts and other reported *Schitoviridae* phages were also constructed and visualized using the same method [52]. Additionally, amino acid sequence comparisons of the ribosomal protein encoding AMG and its homologs were visualized by AliView [55]. The genomes of isolated *Schitoviridae* phages and environmental *Schitoviridae* phages retrieved from the IMG/VR database were clustered using vContact2 [49, 56]. To investigate their potential evolutionary origins, non *Schitoviridae* roseophages and *Roseobacteraceae* bacteria genomes harboring homologs of AMGs in vB_DshP-R7L were also clustered with *Schitoviridae* genomes.

The ratio of synonymous to nonsynonymous substitutions (dN/dS) was employed to examine the genetic variation and evolutionary patterns of vB_DshP-R7L and its homologs. The substitution rates were calculated using the branch-site unrestricted statistical test for episodic diversification available in the HyPhy package [57]. In this analysis, genes from vB_DshP-R7L served as the test group, and their homologs formed the background group.

The tRNA Adaptation Index (tAI) was used to measure the tRNA usage preferences of each gene. This index calculates the tRNA copy number and the wobble interaction between the codon and anticodon using stAI-calc [58]. The tAI value of a gene, ranging from 0 to 1, is determined by the geometric mean of the normalized codon usage values. A higher tAI value indicates greater availability and translatability from the tRNA pool, signifying a gene's efficient translation. The codon usage of vB_DshP-R7L genes were compared against four distinct tRNA gene copy number pools. The first pool consists of the viral tRNA genes, encoding one tRNA for proline and one for isoleucine. The second pool includes the host tRNA gene repertoire. The third pool comprises the host's synonymous tRNA genes, specifically three for proline and two for isoleucine. The fourth pool encompasses all other tRNAs in the host genome.

Distribution of the *Rhodovirinae* roseophages in the global ocean

A comprehensive assessment of the relative abundance of isolated *Rhodovirinae* roseophages was conducted using 294 marine metagenomic datasets (Table S1). These

datasets were from Global Ocean Viromes 2.0 [59], South China Sea [60, 61], East China Sea [62], Pearl River estuary virome [63], Chesapeake Bay [64], Delaware Bay [64], Goseong Bay [65] and Yangshan Harbour [66] (Table S1). Marine metagenomic reads were mapped against the isolated *Rhodovirinae* roseophages genomes using CoverM v0.7.0 with a previously reported method (–min-read-percent-identity 95 –min-read-aligned-length 50) [29]. The relative abundances of the phages were normalized to mapped read counts per kilobase pair of genome per million read counts (RPKM). Phages were considered present at a site if their genome coverage exceeded 30% in the dataset. The distribution of *Rhodovirinae* roseophages across global oceans was visualized using the R ggmap package.

Results

Virion shape, host range, and growth curve of vB_DshP-R7L

vB_DshP-R7L was isolated from the surface waters of Xiamen Bay at the Xiangan coastal station S03 (118.03N, 24.43E) in November 2013. Electron microscopy revealed that the vB_DshP-R7L virions possess an isometric head approximately 70 nm in diameter and a short tail about 40 nm long (Fig. 1a). Host range assessments were conducted on other hosts of *Rhodovirinae* roseophages including *Erythrobacter litoralis* DSM 8509, *Erythrobacter* sp. JL475, *Erythrobacter* DSM 6997, *Roseovarius nubinhibens*, *Roseovarius* sp. 217, *Ruegeria pomeroyi* and *Roseobacter denitrificans* DSM 7001. The results demonstrated that vB_DshP-R7L has a limited host range, failing to infect reported hosts of *Rhodovirinae*. Further genomic comparisons of vB_DshP-R7L against CRISPR spacers within the *Roseobacteraceae*, the IMG

environmental CRISPR spacer database, and iPHoP CRISPR spacer database. indicated exclusive matches with spacers from *Dinoroseobacter shibae* DFL12. One-step growth curve tests of vB_DshP-R7L were failed do to the small plaque size. The following growth curve analysis based on epifluorescence microscopy revealed a latency period of approximately 1 h, followed by an exponential lysis phase lasting 5 h and achieving a peak viral density of 10^9 /ml (Fig. 1b).

Genome content and phage structural proteins

vB_DshP-R7L is a linear dsDNA phage with a genome length of 75.87 kb and a G+C content of 48.96%. A 99 bp repeat sequence was identified at the 3' end of the genome. Analysis of the genome revealed 84 open reading frames (ORFs), 57 of which encode hypothetical proteins, while the other 27 ORFs were predicted with known functions (Table 2). The genome also encodes two tRNAs for proline and isoleucine.

Structural protein analysis through gradient SDS-PAGE electrophoresis and mass spectrometry identified conserved proteins in the virion of vB_DshP-R7L (Fig. 1c). R13LH_gp66 (398.28 kDa), was predicted to function as an N4-like viral RNA polymerase, and has the largest molecular weight among all structural proteins. The most abundant structural protein, based on peptide counts from mass spectrometry, was vB_DshP-R7L_gp72(51.40 kDa), predicted as the N4-like major capsid protein. vB_DshP-R7L_gp67(69.18 kDa), the second most abundant, was predicted as a cell wall hydrolase. Peptides of vB_DshP-R7L_gp70(49.64 kDa) and vB_DshP-R7L_gp75(88.01 kDa) were also identified, which were homologs of the N4 gp54-like structural protein and the N4-like viral portal protein,

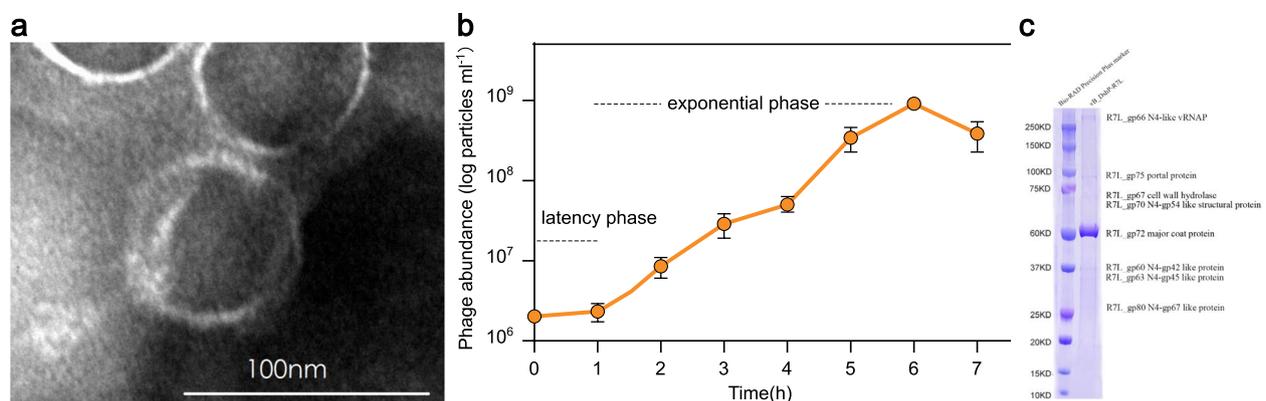


Fig. 1 Characterization of vB_DshP-R7L Phage. **a** Transmission electron micrograph of vB_DshP-R7L particles. **b** Growth curve of phage vB_DshP-R7L. The latent period and final concentration of vB_DshP-R7L phage were 1 h and 10^9 particles per milliliter, respectively. **c** Identification of structural proteins of vB_DshP-R7L phage. Purified virus particles were subjected to 8%–18% protein SDS–polyacrylamide gel electrophoresis and bands were annotated based on protein molecular weight and mass spectrometry results

Table 2 Genome annotation of vB_DshP-R7L

Gene id	Strand	Start(bp)	Stop(bp)	Length(aa)	MW(kd)	Annotated feature	Putative function	Conserved gene
vB_DshP-R7L_gp1	+	2	349	116	12.62	hypothetical protein		
vB_DshP-R7L_gp2	+	375	1358	328	36.53	hypothetical protein		
vB_DshP-R7L_gp3	+	1358	1675	106	11.52	hypothetical protein		
vB_DshP-R7L_gp4	+	1672	2040	123	13.86	hypothetical protein		
vB_DshP-R7L_gp5	+	2037	2471	145	16.78	hypothetical protein		
vB_DshP-R7L_gp6	+	2481	2714	78	8.9	hypothetical protein		
vB_DshP-R7L_gp7	-	3075	3293	73	7.94	hypothetical protein		
vB_DshP-R7L_gp8	+	3351	3605	85	8.96	hypothetical protein		
vB_DshP-R7L_gp9	+	3788	4072	95	11.05	hypothetical protein		
vB_DshP-R7L_gp10	+	4069	4209	47	5.21	hypothetical protein		
vB_DshP-R7L_gp11	+	4206	4517	104	11.81	hypothetical protein		
vB_DshP-R7L_gp12	+	4510	4740	77	8.35	hypothetical protein		
vB_DshP-R7L_gp13	+	4862	5272	137	14.81	hypothetical protein		
vB_DshP-R7L_gp14	+	5589	6374	262	30.16	nucleotide metabolism	RNAP1	
vB_DshP-R7L_gp15	+	6367	6633	89	10.15	Auxiliary metabolic gene	Ribosomal protein	
vB_DshP-R7L_gp16	+	6630	6836	69	7.8	hypothetical protein		
vB_DshP-R7L_gp17	+	7076	7321	82	9.28	hypothetical protein		
vB_DshP-R7L_gp18	+	7321	7470	50	6.17	hypothetical protein		
vB_DshP-R7L_gp19	-	7425	7646	74	8.18	hypothetical protein		
vB_DshP-R7L_gp20	+	7747	7947	67	7.8	hypothetical protein		
vB_DshP-R7L_gp21	+	7941	9140	400	45.44	conserved gene	RNAP2	N4 gp16-like
vB_DshP-R7L_gp22	+	9196	9381	62	7.34	hypothetical protein		
vB_DshP-R7L_gp23	+	9426	10,610	395	43.38	conserved gene	putative AAA superfamily ATPase	N4 gp24-like
vB_DshP-R7L_gp24	+	10,582	11,154	191	21.34	hypothetical protein		
vB_DshP-R7L_gp25	+	11,126	12,391	422	48.23	hypothetical protein		
vB_DshP-R7L_gp26	+	12,357	12,797	147	17.06	hypothetical protein		
vB_DshP-R7L_gp27	+	12,778	13,047	90	10.26	hypothetical protein		
vB_DshP-R7L_gp28	+	13,044	13,220	59	6.85	hypothetical protein		
vB_DshP-R7L_gp29	+	13,217	13,648	144	16.2	Auxiliary metabolic gene	Cytidine and deoxycytidylate deaminase	
vB_DshP-R7L_gp30	+	13,648	14,070	141	16.19	hypothetical protein		
vB_DshP-R7L_gp31	+	14,067	14,498	144	15.83	hypothetical protein		
vB_DshP-R7L_gp32	+	14,482	15,411	310	35.64	Auxiliary metabolic gene	Thymidylate synthase	
vB_DshP-R7L_gp33	+	15,457	16,614	386	43.92	hypothetical protein		
vB_DshP-R7L_gp34	+	16,607	16,867	87	9.9	hypothetical protein		
vB_DshP-R7L_gp35	+	16,881	17,000	40	4.43	hypothetical protein		
	+	17,069	17,144			tRNA	tRNA-Pro	
	+	17,147	17,222			tRNA	tRNA-Ile	
vB_DshP-R7L_gp36	+	17,306	17,626	107	11.86	hypothetical protein		
vB_DshP-R7L_gp37	+	17,623	18,123	167	17.86	hypothetical protein		
vB_DshP-R7L_gp38	+	18,125	18,631	169	17.05	hypothetical protein		
vB_DshP-R7L_gp39	+	18,647	20,680	678	68.55	hypothetical protein		
vB_DshP-R7L_gp40	+	20,680	23,052	791	86.38	Auxiliary metabolic gene	Sialate O-acetyltransferase	
vB_DshP-R7L_gp41	+	23,049	23,279	77	9.15	hypothetical protein		
vB_DshP-R7L_gp42	+	23,328	23,870	181	20.27	hypothetical protein		
vB_DshP-R7L_gp43	+	23,863	24,180	106	12.15	Auxiliary metabolic gene	Thioredoxin	
vB_DshP-R7L_gp44	+	24,254	26,821	856	98.65	hypothetical protein		
vB_DshP-R7L_gp45	+	26,841	28,271	477	51.92	lysis inhibitor	RIB-like protein	
vB_DshP-R7L_gp46	+	28,428	28,739	104	11.75	nucleotide metabolism	homing endonuclease	
vB_DshP-R7L_gp47	+	28,720	29,160	147	16.17	hypothetical protein		

Table 2 (continued)

Gene id	Strand	Start(bp)	Stop(bp)	Length(aa)	MW(kd)	Annotated feature	Putative function	Conserved gene
vB_DshP-R7L_gp48	+	29,151	29,522	124	14.11	hypothetical protein		
vB_DshP-R7L_gp49	+	29,519	29,908	130	14.35	hypothetical protein		
vB_DshP-R7L_gp50	+	29,908	30,201	98	10.76	hypothetical protein		
vB_DshP-R7L_gp51	+	30,198	30,341	48	5.37	hypothetical protein		
vB_DshP-R7L_gp52	+	30,338	30,487	50	5.49	hypothetical protein		
vB_DshP-R7L_gp53	+	30,536	31,078	181	19.78	hypothetical protein		
vB_DshP-R7L_gp54	+	31,075	31,422	116	13.28	hypothetical protein		
vB_DshP-R7L_gp55	+	31,496	33,877	794	89.36	Auxiliary metabolic gene	Ribonucleotide reductase, large subunit	
vB_DshP-R7L_gp56	+	33,999	34,364	122	14.57	hypothetical protein		
vB_DshP-R7L_gp57	+	34,352	35,626	425	48.72	hypothetical protein		
vB_DshP-R7L_gp58	+	35,623	35,952	110	12.3	hypothetical protein		
vB_DshP-R7L_gp59	+	35,961	38,579	873	98.82	conserved gene	DNA polymerase-like protein	N4 gp39-like
vB_DshP-R7L_gp60	+	38,576	39,571	332	38	conserved gene	exonuclease	N4 gp42-like
vB_DshP-R7L_gp61	+	39,564	41,759	732	83.21	conserved gene	Primase	N4 gp43-like
vB_DshP-R7L_gp62	+	41,769	42,506	246	28.37	conserved gene	Sak4-like ssDNA annealing protein	N4 gp44-like
vB_DshP-R7L_gp63	+	42,563	43,384	274	29.91	conserved gene	single strand DNA binding protein	N4 gp45-like
vB_DshP-R7L_gp64	+	43,417	44,322	302	32.88	hypothetical protein		
vB_DshP-R7L_gp65	+	44,312	44,515	68	7.17	hypothetical protein		
vB_DshP-R7L_gp66	-	44,698	55,548	3617	398.28	nucleotide metabolism	vRNAP	
vB_DshP-R7L_gp67	-	55,605	57,539	645	69.2	structural protein	Cell wall hydrolase	
vB_DshP-R7L_gp68	-	57,550	58,011	154	16.19	hypothetical protein		
vB_DshP-R7L_gp69	-	58,014	60,764	917	100.23	hypothetical protein		
vB_DshP-R7L_gp70	-	60,776	62,128	451	49.65	structural protein	virion structural protein	
vB_DshP-R7L_gp71	-	62,168	62,890	241	26.88	conserved gene		N4 gp55-like
vB_DshP-R7L_gp72	-	63,006	64,454	483	51.41	conserved gene	major head protein	N4 gp56-like
vB_DshP-R7L_gp73	-	64,484	65,845	454	48.71	conserved gene	tail length tape measure protein	N4 gp57-like
vB_DshP-R7L_gp74	-	65,890	66,276	129	14.31	hypothetical protein		
vB_DshP-R7L_gp75	-	66,330	68,717	796	88.02	conserved gene	portal protein	N4 gp59-like
vB_DshP-R7L_gp76	-	68,900	69,232	111	12.12	hypothetical protein		
vB_DshP-R7L_gp77	-	69,229	69,876	216	24.27	hypothetical protein		
vB_DshP-R7L_gp78	-	69,873	70,067	65	7.13	hypothetical protein		
vB_DshP-R7L_gp79	-	70,069	70,410	114	11.75	hypothetical protein		
vB_DshP-R7L_gp80	-	70,410	71,078	223	25.11	structural protein	virion structural protein	
vB_DshP-R7L_gp81	-	71,089	72,702	538	61.52	nucleotide metabolism	Terminase large unit	
vB_DshP-R7L_gp82	-	72,695	73,384	230	25.81	conserved gene	Terminase small unit	N4 gp69-like
vB_DshP-R7L_gp83	+	73,459	73,779	107	11.71	hypothetical protein		
vB_DshP-R7L_gp84	+	73,779	75,869	697	76.71	structural protein	Roseovarius sp. 217 phage like structural protein	

respectively. vB_DshP-R7L_gp60(38.00 kDa), vB_DshP-R7L_gp63(29.91 kDa) and vB_DshP-R7L_gp80(25.10 kDa) were identified through gradient SDS-PAGE electrophoresis, and were homologs of N4 gp42-like, N4 gp45-like and N4 gp67-like structural proteins, respectively.

Phylogenetic analysis of vB_DshP-R7L

A comparative analysis of the full genome of vB_DshP-R7L with all other isolated roseophages (Table 1) conducted using the Virus Classification and Tree Building Online Resource(VICTOR) revealed that vB_DshP-R7L forms a separate branch (Fig. 2a) and represents a novel genus among *Rhodovirinae* phages (Figure S1). Clustering

of all *Schitoviridae* phages by their shared genes using vContact2.0 marked vB_DshP-R7L as a unique genus-level viral cluster (VC), further substantiating its status as a distinct genus (Table S2). To prevent classification errors of vB_DshP-R7L caused by horizontal gene transfer (HGT) events, additional evidence supporting this classification was obtained by constructing phylogenetic trees based on two marker genes of *Schitoviridae*, which are homologous to *Enterobacteria* phage N4 gp16 and gp56, which encodes vRNAP2 (Virion-encapsidated RNA polymerase) and MCP (Major capsid protein) (Fig. 2bc) [52]. The closest clades to vB_DshP-R7L in the single-gene phylogenetic trees are *Baltimorevirus* and *Plymouthvirus* in single-gene phylogenetic trees, with genetic distances of 0.143 and 0.02, respectively. Which exceed the distance between *Aorunvirus* and *Plymouth* within *Rhodovirinae* (0.08 and 0.01). All these results suggest that vB_DshP-R7L represents a novel genus within subfamily *Rhodovirinae*, family *Schitoviridae*. Comparative genomic analysis of vB_DshP-R7L with its closely related genera, *Baltimorevirus* and *Plymouthvirus*, along with the representative phage of *Schitoviridae*, *Enterobacteria* phage N4, revealed a conserved gene organization (Fig. 3).

Biogeography of vB_DshP-R7L and the *Rhodovirinae* roseophages

To investigate the ecological distribution of vB_DshP-R7L and other *Rhodovirinae* roseophages, we conducted a read-mapping analysis of 294 marine datasets from both open oceans and estuaries, as detailed in the Methods section. Our analysis revealed that all *Rhodovirinae* roseophages were isolated exclusively from estuarine environments. Specifically, we detected phages R1, R2C, and vB_DshP-R7L in metagenomic samples from the Pearl River Estuary, whereas vB_RpoP-V12 was found in metagenomic samples from the Delaware Bay (Fig. 4).

Function of phage auxiliary metabolic genes

We identified six auxiliary metabolic genes (AMGs) in the genome of vB_DshP-R7L, these include vB_DshP-R7L_gp29 (*dcd*) encoding dCMP deaminase, vB_DshP-R7L_gp32 (*thyX*) encoding thymidylate synthase, vB_DshP-R7L_gp43 (*trx*) encoding thioredoxin,

vB_DshP-R7L_gp55 (*rnr*) encoding ribonucleotide reductase, vB_DshP-R7L_gp15 encoding a ribosomal protein, and vB_DshP-R7L_gp40 (*nanS*) encoding sialate O-acetyltransferase.

In *Rhodovirinae*, the *trx* and *rnr* genes are conserved across all genomes, while the *dcd* and *thyX* genes are also prevalent [39]. Thioredoxin reductase, encoded by the *trx* gene, functions as a hydrogen donor, utilizing NADPH to transfer electrons to the ribonucleotide reductase encoded by the *rnr* gene, thereby reducing ribonucleoside diphosphate (rNDP) to deoxyribonucleoside diphosphate (dNDP) [67, 68]. DCMP deaminase, encoded by the *dcd* gene, deaminates deoxycytidylic acid (dCMP) to deoxyuridine monophosphate (dUMP). Thymidylate synthase, encoded by the *thyX* gene, is an essential enzyme for pyrimidine synthesis, converting dUMP to deoxythymidine monophosphate (dTMP) [69]. These four AMGs are related to the DNA de novo synthesis pathway and may enhance the production of precursors required for DNA synthesis, thus supporting DNA replication and modifications during infection (Fig. 5). vB_DshP-R7L_gp15 encodes a ribosomal protein bL12, which forms the prominent stalk-like structure on the large ribosomal subunit in bacteria and is the most common ribosomal protein encoded by viruses from aquatic environments [70]. vB_DshP-R7L_gp40 is predicted to be the *nanS* gene, which encodes sialate O-acetyltransferase. This enzyme catalyzes the hydrolysis of N-acetylneuraminic acid, a compound commonly found in the genomes of pathogenic bacteria [71].

Origin and evolution of phage vB_DshP-R7L AMGs

To elucidate the evolutionary history of the AMGs in phage vB_DshP-R7L, we identified homologs for each AMG from *Schitoviridae* phages and hosts of *Rhodovirinae* roseophages. A gene-sharing network constructed using vConTACT2 to highlight potential horizontal gene transfer (HGT) events revealed that all AMGs in vB_DshP-R7L share homologs with *Schitovirinae* phages or *Roseobacteraceae* hosts (Fig. 6). Phylogenetic trees for each AMG of vB_DshP-R7L were generated with their homologs from relatives and hosts of *Rhodovirinae* roseophages to further explore their origins (Figure S2). Additionally, the ratio of non-synonymous to synonymous

(See figure on next page.)

Fig. 2 Phylogenetic relationships of vB_DshP-R7L and related phages. Nodes with bootstrap support values over 75 were marked with a black dot. The background color of each leaf and labels shows its subfamily among the *Schitoviridae*. The phage vB_DshP-R7L, isolated in this study, is highlighted with an asterisk. **a** Phylogenetic analyses based on full genome similarity among all isolated roseophages using the Virus Classification and Tree Building Online Resource (VICTOR). **b** Phylogenetic tree based on amino acid sequences alignment of vRNAP2 (virion-encapsidated RNA polymerase) using the maximum likelihood method. **c** Phylogenetic tree based on amino acid sequences alignment of MCP (major capsid protein) using the maximum likelihood method

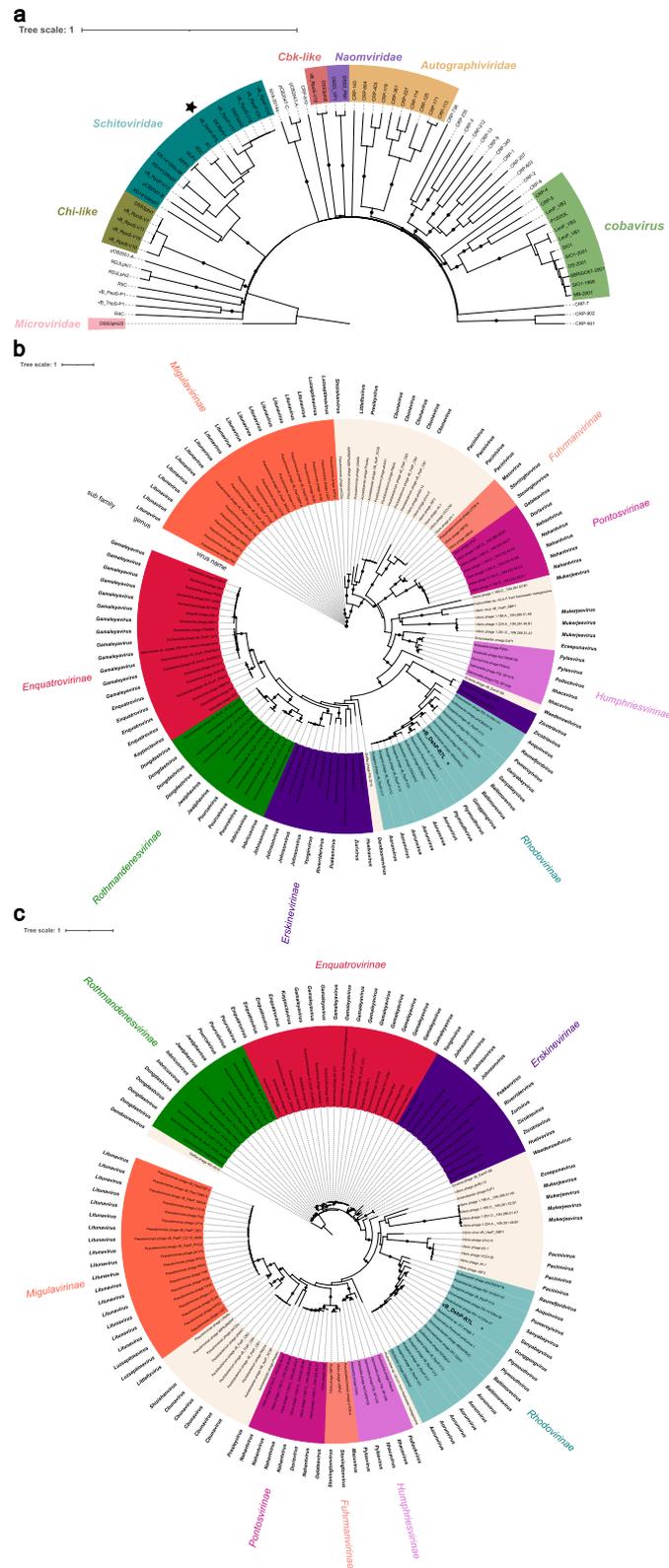


Fig. 2 (See legend on previous page.)

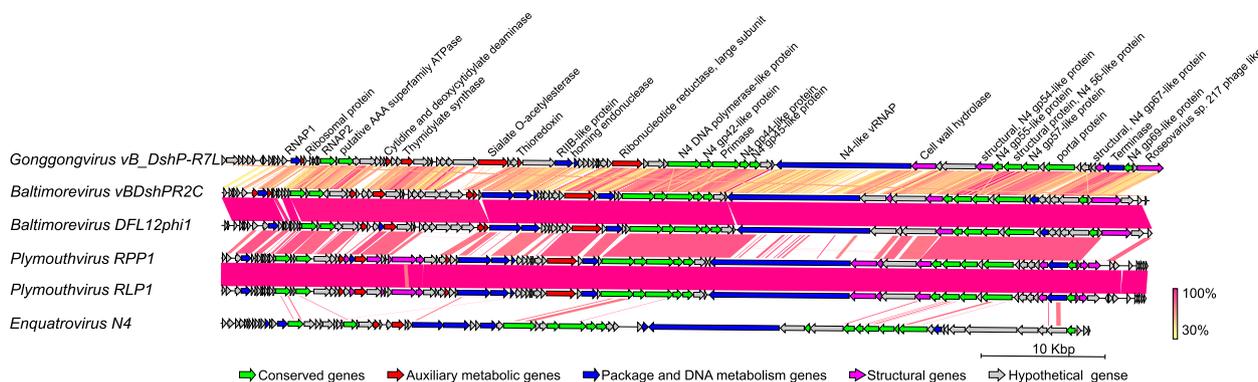


Fig. 3 Genome comparison of the vB_DshP-R7L with its closest relatives and *Enquatrovirus* N4 using EasyFig. The open reading frames are indicated by leftward or rightward arrows depending on the direction of transcription. The colors of the arrows represent the predicted gene function, and homologs are linked according to the similarity values obtained by TBLASTX comparison. The color gradient representing similarity levels progresses from yellow to red, as detailed in the legend

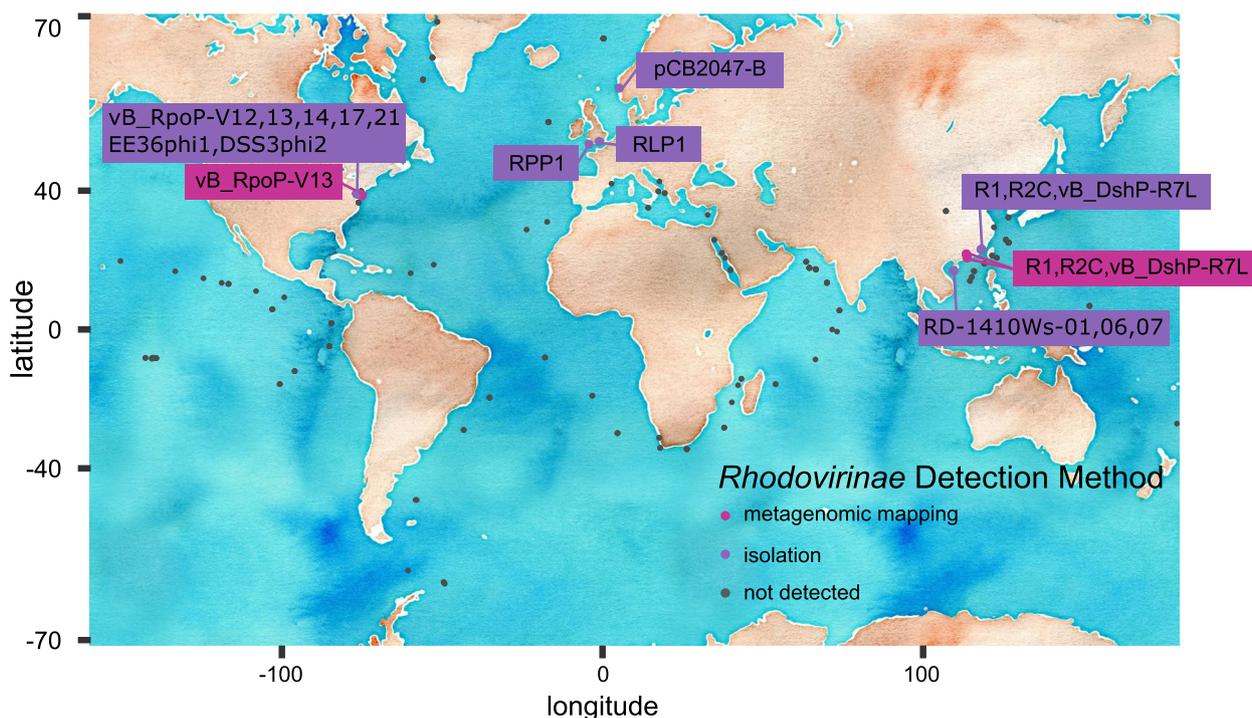


Fig. 4 Global distribution of *Rhodovirinae* roseophages in marine waters. Stations where *Rhodovirinae* roseophages have been isolated are marked with purple circles, and those where they have been detected in metagenomes are marked with red circles. The names of the *Rhodovirinae* roseophages are labeled on the map near the stations where they were detected. Each name is colored according to the method of detection

mutation rates (dN/dS) was calculated for each AMG and conserved gene to assess selective pressures (Table S3).

The *rnr* and *trx* AMGs, conserved within *Rhodovirinae* roseophages, form unique clusters distinct from their homologs. The *rnr* genes of *Rhodovirinae* roseophages grouped with homologs from their *Roseobacteraceae* hosts, suggesting a unique evolutionary path distinct from other *Schitoviridae* subfamilies. Conversely, the

trx genes of *Rhodovirinae* roseophages clustered with homologs of their hosts and the *Rothmandenesvirinae* subfamily, indicating a likely HGT event across different subfamilies of *Schitoviridae* phages. The dN/dS ratios for *rnr* and *trx* in *Rhodovirinae* were 0.12 and 0.14, respectively.

The *thyX* and *dcd* AMGs appear exclusively in four *Rhodovirinae* genera: *Plymouthvirus*, *Baltimorevirus*,

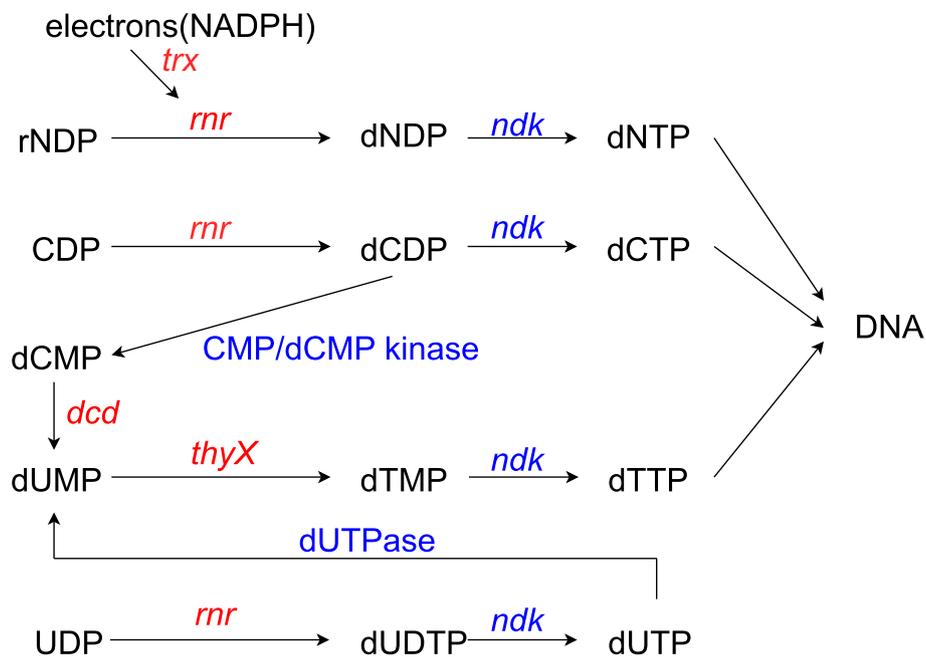


Fig. 5 Putative functions of auxiliary metabolic genes in vB_DshP-R7L to promote DNA synthesis in host cells. Phage AMGs are marked in red and host functional are marked in blue

Gonggongvirus, and *Aorunvirus*. Phylogenetic analyses based on full length genomes suggest these four genera form a monophyletic clade within *Rhodovirinae* subfamily (Fig. 2a). The *thyX* genes of *Rhodovirinae* cluster with their homologs of *Rhizobium azibense* and its infecting roseophages (Figure S2). The *dcd* genes of *Rhodovirinae* homologs cluster with homologs from *Cohaesibacter haloalkalitolerans*, *Bradyrhizobium* sp. CCBAU_21365, and phages of the *Erskinevirinae* subfamily. The origins of *thyX* and *dcd* in *Rhodovirinae* likely trace back to an ancient HGT event occurred in the common ancestor of these four genera, with dN/dS ratios of 0.18 and 0.23, respectively.

In *Rhodovirinae*, the distribution of *nanS* genes is limited to RD-1410W1-01 and vB_DshP-R7L, while ribosomal protein genes are found only in vB_DshP-R1 and vB_DshP-R7L. The *nanS* gene sequences are highly similar and cluster with homologs from plasmids of *Roseobacter denitrificans* and its infecting roseophages (Figure S2). Given RD-1410W1-01's broad host range, including *Roseobacter denitrificans* and *Dinoroseobacter shibae*, the *nanS* gene in vB_DshP-R7L likely originated from a recent HGT event facilitated by RD-1410W1-01 as they both infect *Dinoroseobacter shibae* cell. The ratio of synonymous and nonsynonymous substitutions (dN/dS ratio) for *nanS* among vB_DshP-R7L, RD-1410W1-01, and their host homologs is 0.27. The ribosomal protein genes of *Rhodovirinae* roseophages are genetically similar

to their host homologs and significantly different from previously reported viral homologs, indicating they may originate from recent HGT events from their hosts (Figure S3).

We utilized the tRNA adaptation index (tAI) to assess the adaptation of codon usage of vB_DshP-R7L genes to the phage's own tRNA content. The tAI values of vB_DshP-R7L genes ranged from 0.961 to 0.968 when aligned against the viral tRNA content (one gene for proline and one for isoleucine), and from 0.456 to 0.533 against the *Dinoroseobacter shibae* host tRNA pool (Table S4). There are 3 tRNA gene copies for proline and 2 tRNA gene copies for isoleucine in the genome of *Dinoroseobacter shibae*, and the tAI values of vB_DshP-R7L genes against host tRNA content for proline and isoleucine ranged from 0.689 to 0.769. This analysis indicates that although the phage can utilize the full complement of host tRNAs, the genes of vB_DshP-R7L are highly adapted to the tRNA content and ratio within the phage genome, reflecting evolutionary adjustments to optimize translation efficiency.

Discussion

The virion morphology and structural protein profiles of the newly isolated vB_DshP-R7L closely match those reported in previous studies on other *Rhodovirinae* roseophages [23, 25, 27, 72]. Mass spectrometry and electrophoresis analysis identified 8 proteins in

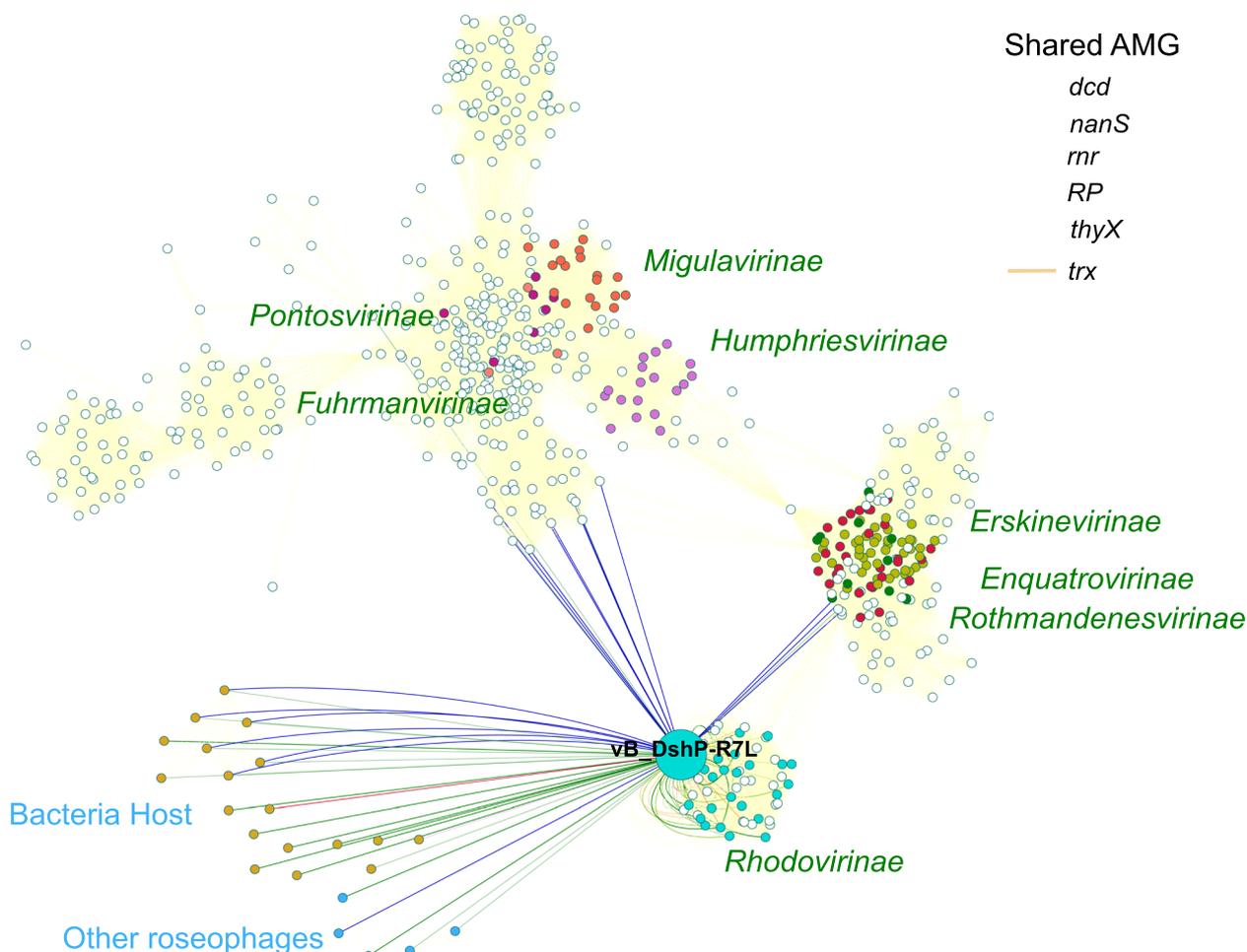


Fig. 6 Horizontal gene transfer of auxiliary metabolic genes (AMGs) between genomes of vB_DshP-R7L, *Schitoviridae* phages, non *Schitoviridae* roseophages and *Rhodovirinae* bacteria. The color of each dot represents the taxonomic status of the genome, and the distance between the dots represents whole genome-wide distance between *Schitoviridae* phage genomes calculated by vContact2. Colored lines connect AMGs in the vB_DshP-R7L genome to their homologs in other genomes. Other genomes that share at least one non AMG gene are linked by yellow lines

the virion of vB_DshP-R7L. However, other conserved structural proteins of the *Schitoviridae*, present in the genome of vB_DshP-R7L, were not detected in electrophoretic or mass spectrometry analyses. This discrepancy is likely due to their low abundance, which is overshadowed by the high concentration of detected structural proteins, particularly the major capsid protein. Similar challenges in identifying structural proteins have been reported for other *Schitoviridae* phages [27, 73]. Despite sharing virion structure and host specificity with other *Rhodovirinae* roseophages, vB_DshP-R7L exhibits a unique small plaque sizes observed, which are too small to measure accurately. Previous studies have indicated that most *Rhodovirinae* roseophages, except for those in the genus *Sanyabayvirus*, can only infect one specific host [23, 27, 74]. Our host range analyses suggest that vB_DshP-R7L exhibits a

similar host-specificity, infecting only *Dinoroseobacter shibae* DFL12.

Due to the absence of universally conserved genes in viruses, the International Committee on Taxonomy of Viruses (ICTV) recommends using multiple independent methods for phage classification. These include assessing nucleotide sequence similarity, phylogenetic analysis of conserved genes with appropriate outgroups, and the ratio of homologous genes within a viral family [75]. Given that vB_DshP-R7L and its close relatives share a conserved gene organization, we employed two independent phylogenetic analyses to determine its taxonomic classification: genome-wide sequence similarities among all isolated roseophages, and single marker gene phylogenetic analyses based on two conserved genes. All results consistently support the classification of vB_DshP-R7L as a novel genus within the subfamily

Rhodovirinae. The International Committee on Taxonomy of Viruses (ICTV) has formally classified vB_DshP-R7L into a novel genus, named *Gonggongvirus*. It is currently the only known member of this genus.

A comprehensive global distribution analysis of *Rhodovirinae* based on accumulated marine viromes has not yet been conducted. Chan et al. conducted distribution analysis using tblastx method of *Rhodovirinae* conserved genes against marine databases collected from 2003, 2007 and 2012 [27]. Zhan et al. identified *Rhodovirinae* roseophages in cold bays using pol gene PCR amplification [76], and a previous study reported extremely low relative abundances of *Schitoviridae* phages in marine metagenomic libraries [18]. A recent study highlighted their predominance in estuarine rather than open ocean environments in the Chesapeake Bay and Delaware Bay [77]. Our research explores the distribution of *Rhodovirinae* roseophages with the newly isolated vB_DshP-R7L. We employ a viral recruitment approach utilizing full-length viral genomes to analyze global marine viromes collected in recent years. This includes the Global Ocean Viromes 2.0 and smaller marine datasets from diverse oceans and estuaries. The results revealed that all detected *Rhodovirinae* roseophages were confined to estuarine ecosystem, specifically within metagenomic samples from the Pearl River Estuary and the Delaware Bay. This limited distribution may be due to their restricted host range, potentially hindering their expansion across open oceans.

AMGs are prevalent in marine phages, including those of the *Rhodovirinae* roseophages, where they hijack host metabolism to enhance phage production [39]. In vB_DshP-R7L, AMGs are predominantly located at the 3' end of the genome, consistent with patterns observed in its close relatives. Six AMGs have been identified in this genome; four are involved in the DNA de novo synthesis pathway and may enhance the production of DNA precursors, thus facilitating DNA replication and modification during infection. The remaining two AMGs in vB_DshP-R7L include a gene encoding the ribosomal protein bL12, and another encoding sialate O-acetyltransferase (*nanS*). Studies indicate that viral ribosomal protein bL12 genes are typically shorter than their host counterparts, and experiments in vitro have shown that viral encoded bL12 can integrate into the host 70S ribosomes in *E. coli* [70]. Thus, the bL12 gene in vB_DshP-R7L may enhance viral protein synthesis during infection. Previous research has identified *nanS* genes in the genomes of Stx2a phages infecting *E. coli*, suggesting that the presence of *nanS* may support phage proliferation in the gastrointestinal microbiome [78, 79]. Given that vB_DshP-R7L was isolated near a populated city and detected in metagenomic samples from the Pearl River Estuary, its *nanS* gene could potentially acquired

in estuary environments influenced by gastrointestinal microbiome.

Our analyses indicate that the AMGs of vB_DshP-R7L originate from distinct horizontal gene transfer (HGT) events, each following unique evolutionary trajectories. Phylogenetic analyses of the four DNA synthesis-related AMGs, alongside homologs from other *Rhodovirinae* roseophages, show similar tree topologies to those based on concatenated conserved genes. These genes cluster with homologs from their *Roseobacteraceae* hosts, suggesting ancient HGT events from these hosts to a common ancestor of modern *Rhodovirinae* roseophages, with stable transmission across generations.

Unlike most *Rhodovirinae* roseophages, RD-1410W1-01 has a broad host range. The *nanS* genes from RD-1410W1-01, vB_DshP-R7L form a distinct cluster with the homolog gene from *Roseobacter denitrificans*, which is one reported host of RD-1410W1-01 [22]. The *nanS* AMG is present in the genome of the phage vB_DshP-R7L, but is absent from both the genome and plasmids of its host, as well as from its closely related phages. Therefore, the origin of the *nanS* AMG in vB_DshP-R7L is unlikely to be the result of HGT from its host or vertical gene transfer from the common ancestors of its close relatives. This implying RD-1410W1-01 may have facilitated the transfer of the *nanS* gene from *Roseobacter denitrificans* to vB_DshP-R7L, given their overlapping host ranges. The ribosomal protein genes in *Rhodovirinae* roseophages exhibit genetic similarity to their host homologs, and are distinct from other viral homologs. We hypothesize that they have independent evolutionary paths and originate through recent HGT events from their hosts.

Viral AMGs may influence host metabolism during infection, and their ecological roles result in evolutionary pressures, which can be quantified using the ratio of nonsynonymous/synonymous substitution (dN/dS ratio). A gene with dN/dS ratio less than 1 indicates purifying selection, suggesting that the gene is actively eliminating harmful mutations throughout its evolutionary history and could play a crucial role during infection [80]. A gene with dN/dS ratio close to 1 suggests that the gene is predominantly affected by neutral mutations and is likely less critical for viral proliferation. Our results suggest that all AMGs of vB_DshP-R7L exhibit dN/dS ratios significantly below 1, indicating strong purifying selection and implying a vital role in infection processes (Table S3). The average dN/dS ratio for conserved genes in vB_DshP-R7L is 0.15, and dN/dS ratio of *trx* and *rnr* genes similar to those of conserved structural genes. AMGs involved in the DNA de novo synthesis pathway have experienced the highest selection pressures and may play significant roles during infection, as evidenced by their significantly

low dN/dS ratios. Other AMGs including *nanS* and ribosomal protein encoding gene, display relatively higher dN/dS values, provide another evidence that they originate from more recent HGT events (Table S3).

Viruses typically depend on the host's tRNA pool for translation, exerting evolutionary pressure on their genomes to adapt codon usage to align with the host's translational machinery [81]. However, certain bacteriophages employ an alternative strategy by encoding specialized tRNAs within their genomes to enhance the translation efficiency of stable viral genes [82, 83]. In our study, we identified two tRNA-encoding genes in the genome of vB_DshP-R7L, each present in a single copy. Intriguingly, the codon usage of all genes in vB_DshP-R7L shows a high preference for the phage encoded tRNA content, highlighting a sophisticated adaptation mechanism of phage genes and its tRNA repertoire. Despite AMGs being products of HGT, the AMGs in vB_DshP-R7L exhibit a tRNA adaptation index (tAI) closely matching that of conserved genes (Table S4). This similarity emphasizes the substantial selective pressure shaping AMG codon usage to match phage tRNA content, underlining their significant ecological potential.

Conclusions

In this study, we isolated and characterized a novel roseophage, vB_DshP-R7L, which infects *Roseobacteraceae*, a significant component of coastal bacterial communities. Phylogenetic results reveal that vB_DshP-R7L representing a new genus within the *Rhodovirinae* subfamily of roseophages. This phage was exclusively detected in metagenomic samples from the Pearl River Estuary, consistent with the environmental specificity observed for all *Rhodovirinae* roseophages in estuarine ecosystems. We identified six AMGs within the genome of vB_DshP-R7L, four are related to de novo nucleotide synthesis, while the others may involved in protein production and carbon metabolism. Comparative genomic analyses suggest these AMGs originated from distinct HGT events, and are all subject to strong evolutionary pressure, indicating their stable incorporation into the genome and functional potential in estuarine ecosystems of the South China Sea. Our comprehensive analysis enhances the understanding of the phylogeny and AMG content within the *Rhodovirinae* roseophage clade, offering new insights into the diversity, evolution, and biogeography of *Rhodovirinae* roseophages in marine environments. These findings not only identify a new phage genus but also broaden our knowledge of the phylogenetic diversity and evolutionary dynamics of marine phages.

Abbreviations

AMG	Auxiliary metabolic gene
dN/dS ratio	Ratio of synonymous and nonsynonymous substitutions

HGT	Horizontal gene transfer
ICTV	International Committee on Taxonomy of Viruses
MRC	Marine roseobacter clade
NCBI	National Centre for Biotechnology Information
ORFs	Open reading frames
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
tAI	tRNA adaptation index
TEM	Transmission electron microscopy

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11463-7>.

Supplementary Material 1: Fig. S1. Phylogenetic assignment of vB_DshP-R7L based on full genomic analysis with relatives using the auto classification method of the Virus Classification and Tree Building Online Resource (VICTOR). Phage vB_DshP-R7L isolated in this study is highlighted with an asterisk

Supplementary Material 2: Fig. S2. Phylogenetic trees of each vB_DshP-R7L AMG with homologous from Schitoviridae phages and Roseobacteraceae hosts. Nodes with bootstrap support values over 75 were marked with a black dot. The background color of each leaf and labels show its subfamily among the Schitoviridae. Phage vB_DshP-R7L isolated in this study is highlighted with an asterisk

Supplementary Material 3: Fig. S3. Comparison of ribosomal protein genes from genomes of Rhodovirinae roseophages and their hosts

Supplementary Material 4: Table S1. Marine metagenomic samples used in this study and coverage of Rhodovirinae phages in each site

Supplementary Material 5: Table S2. Clusters assigned by vContact2 for Schitoviridae viruses and other genomes sharing auxiliary metabolic genes (AMGs) with vB_DshP-R7L

Supplementary Material 6: Table S3. Summary of on non-synonymous/synonymous ratio values of AMGs and conserved genes in vB_DshP-R7L comparing with homologs

Supplementary Material 7: Table S4. Summary of tAI (tRNA Adaptation Index) values for each vB_DshP-R7L gene, evaluated against both phage-encoded and host-encoded tRNA content

Acknowledgements

Not applicable.

Authors' contributions

Xingyu Huang wrote the main text of the manuscript and conducted the analysis. Chen Yu was responsible for virus purification, while Longfei Lu was responsible for isolation and provided the basic ideas.

Funding

This research was funded by the Young Scientists Fund of the National Natural Science Foundation of China, 42306268 and Beihai Science and Technology Program, 201995076.

Data availability

The complete genome and protein sequences have been deposited in GenBank under accession number MZ773648.1 (<https://www.ncbi.nlm.nih.gov/nuccore/MZ773648.1>). Sequencing data for this study are also available in the China National Center for Bioinformation repository under accession number CRA018945 (<https://ngdc.cnbc.ac.cn/gsa/browse/CRA018945>). The genome and protein sequences of vB_DshP-R7L are included in a GenBank format supplementary material.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Tibetan Plateau Research, Chinese Academy of Sciences, No. 16 Lincui Road, Chaoyang District, Beijing 100101, People's Republic of China.

²State Key Laboratory of Marine Environmental Science, College of Ocean & Earth Sciences, Fujian Key Laboratory of Marine Carbon Sequestration, Xiamen University, Xiamen 361102, People's Republic of China. ³Key Laboratory of Tropical Marine Ecosystem and Bioresource, Ministry of Natural Resources, Fourth Institute of Oceanography, Beihai 536000, China.

Received: 6 September 2024 Accepted: 10 March 2025

Published online: 25 March 2025

References

- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol.* 2007;5(10):801–12.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol.* 2018;3(7):754–66.
- Obeng N, Pratama AA, van Elsland JD. The significance of mutualistic phages for bacterial ecology and evolution. *Trends Microbiol.* 2016;24(6):440–9.
- Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Shedding new light on viral photosynthesis. *Photosynth Res.* 2015;126(1):71–97.
- Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, Waldbauer JR, Coleman ML. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol.* 2020;18(1):21–34.
- Kieft K, Zhou Z, Anderson RE, Buchan A, Campbell BJ, Hallam SJ, Hess M, Sullivan MB, Walsh DA, Roux S, et al. Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat Commun.* 2021;12(1):3503.
- Jacobson TB, Callaghan MM, Amador-Noguez D. Hostile takeover: how viruses reprogram prokaryotic metabolism. *Annu Rev Microbiol.* 2021;75:515–39.
- Luo H, Moran MA. Evolutionary ecology of the marine Roseobacter clade. *Microbiol Mol Biol Rev.* 2014;78(4):573–87.
- Liang KYH, Orata FD, Boucher YF, Case RJ. Roseobacters in a sea of poly- and paraphyly: whole genome-based taxonomy of the family rhodobacteraceae and the proposal for the split of the “Roseobacter clade” into a novel family roseobacteraceae fam nov. *Front Microbiol.* 2021;12:683109.
- Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, Klenk H-P, Schomburg D, Petersen J, Göker M. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 2017;11(6):1483–99.
- Zubkov MV, Fuchs BM, Archer SD, Kiene RP, Amann R, Burkil PH. Linking the composition of bacterioplankton to rapid turnover of dissolved dimethylsulphoniopropionate in an algal bloom in the North Sea. *Environ Microbiol.* 2001;3(5):304–11.
- Brock NL, Menke M, Klapschinski TA, Dickschat JS. Marine bacteria from the Roseobacter clade produce sulfur volatiles via amino acid and dimethylsulphoniopropionate catabolism. *Org Biomol Chem.* 2014;12(25):4318–23.
- Du S, Wu Y, Ying H, Wu Z, Yang M, Chen F, Shao J, Liu H, Zhang Z, Zhao Y. Genome sequences of the first Autographiviridae phages infecting marine Roseobacter. *Microbial Genomics.* 2024;10(4): 001240.
- Rihtman B, Puxty RJ, Hapeshi A, Lee Y-J, Zhan Y, Michniewski S, Waterfield NR, Chen F, Weigle P, Millard AD. A new family of globally distributed lytic roseophages with unusual deoxythymidine to deoxyuridine substitution. *Curr Biology.* 2021;31(14):3199–206 e3194.
- Zhan Y, Huang S, Voget S, Simon M, Chen F. A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agents. *Sci Rep.* 2016;6: 30372.
- Zhan Y, Chen F. Bacteriophages that infect marine roseobacters: genomics and ecology. *Environ Microbiol.* 2019;21(6):1885–95.
- Zhan Y, Huang S, Chen F. Genome sequences of five bacteriophages infecting the marine Roseobacter bacterium *ruegeria pomeroyi* DSS-3. *Microbiol Resource Annot.* 2018;7(12):10.
- Zhang Z, Chen F, Chu X, Zhang H, Luo H, Qin F, Zhai Z, Yang M, Sun J, Zhao Y. Diverse, abundant, and novel viruses infecting the marine Roseobacter RCA lineage. *mSystems.* 2019;4(6):e00494-19.
- Bischoff V, Bunk B, Meier-Kolthoff JP, Spröer C, Poehlein A, Dogs M, Nguyen M, Petersen J, Daniel R, Overmann J, et al. Cobaviruses – a new globally distributed phage group infecting Rhodobacteraceae in marine ecosystems. *ISME J.* 2019;13(6):1404–21.
- Angly F, Youle M, Nosrat B, Srinagesh S, Rodriguez-Brito B, McNairnie P, Deyanat-Yazdi G, Breitbart M, Rohwer F. Genomic analysis of multiple Roseophage SIO1 strains. *Environ Microbiol.* 2009;11(11):2863–73.
- Kang I, Jang H, Oh H-M, Cho J-C. Complete genome sequence of Celeribacter bacteriophage P12053L. *J Virol.* 2012;86(15):8339–40.
- Li B, Zhang S, Long L, Huang S. Characterization and complete genome sequences of three N4-Like Roseobacter phages isolated from the South China Sea. *Curr Microbiol.* 2016;73(3):409–18.
- Zhao Y, Wang K, Jiao N, Chen F. Genome sequences of two novel phages infecting marine roseobacters. *Environ Microbiol.* 2009;11(8):2055–64.
- Ankrah NYD, Budinoff CR, Wilson WH, Wilhelm SW, Buchan A. Genome sequence of the Sulfitobacter sp. strain 2047-infecting lytic phage ΦCB2047-B. *Genome Announcements.* 2014;2(1):e00945-00913.
- Ji J, Zhang R, Jiao N. Complete genome sequence of Roseophage vB_DshP-R1, which infects Dinoroseobacter shibae DFL12. *Standards in genomic sciences.* 2015;10(6):1–8.
- Cai L, Yang Y, Jiao N, Zhang R. Complete genome sequence of vB_DshP-R2C, a N4-like lytic roseophage. *Mar Genomics.* 2015;22:15–7.
- Chan JZ, Millard AD, Mann NH, Schäfer H. Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like Roseophage specific genes. *Front Microbiol.* 2014;5:506.
- Qin F, Du S, Zhang Z, Ying H, Wu Y, Zhao G, Yang M, Zhao Y. Newly identified HMO-2011-type phages reveal genomic diversity and biogeographic distributions of this marine viral group. *ISME J.* 2022;16(5):1363–75.
- Wu Z, Guo L, Wu Y, Yang M, Du S, Shao J, Zhang Z, Zhao Y. Novel phage infecting the Roseobacter CHUG lineage reveals a diverse and globally distributed phage family. *Mosphere.* 2024;9:e00458-00424.
- Zhai Z, Zhang Z, Zhao G, Liu X, Qin F, Zhao Y. Genomic characterization of two novel RCA phages reveals new insights into the diversity and evolution of marine viruses. *Microbiology Spectrum.* 2021;9(2):e01239–e01221.
- Zhang Z, Wu Z, Liu H, Yang M, Wang R, Zhao Y, Chen F. Genomic analysis and characterization of phages infecting the marine Roseobacter CHAB-1-5 lineage reveal a globally distributed and abundant phage genus. *Front Microbiol.* 2023;14: 1164101.
- Ankrah NYD, Budinoff CR, Wilson WH, Wilhelm SW, Buchan A. Genome sequences of two temperate phages, ΦCB2047-A and ΦCB2047-C, infecting Sulfitobacter sp. strain 2047. *Genome Annot.* 2014;2(3):e00108–00114.
- Ankrah NYD. Elucidating the impact of Roseophage on Roseobacter metabolism and marine nutrient cycles. 2015.
- Cai L, Ma R, Chen H, Yang Y, Jiao N, Zhang R. A newly isolated roseophage represents a distinct member of Siphoviridae family. *Virology Journal.* 2019;16(1):128.
- Yang Y, Cai L, Ma R, Xu Y, Tong Y, Huang Y, Jiao N, Zhang R. A novel roseosiphophage isolated from the oligotrophic South China Sea. *Viruses.* 2017;9(5): 109.
- Zhang Y, Jiao N. Roseophage RDJLΦ1, infecting the aerobic anoxygenic phototrophic bacterium *Roseobacter denitrificans* OCh114. *Applied and environmental microbiology.* 2009;75(6):1745–9.
- Tang K, Lin D, Zheng Q, Liu K, Yang Y, Han Y, Jiao N. Genomic, proteomic and bioinformatic analysis of two temperate phages in Roseobacter clade bacteria isolated from the deep-sea water. *BMC Genomics.* 2017;18(1):485.
- Forcone K, Coutinho FH, Cavalcanti GS, Silveira CB. Prophage genomics and ecology in the family Rhodobacteraceae. *Microorganisms.* 2021;9(6):1115.
- Huang X, Jiao N, Zhang R. The genomic content and context of auxiliary metabolic genes in roseophages. *Environ Microbiol.* 2021;23(7):3743–57.
- Wittmann J, Turner D, Millard AD, Mahadevan P, Kropinski AM, Adriaenssens EM. From orphan phage to a proposed new family—the diversity of N4-like viruses. *Antibiotics (Basel).* 2020;9(10):663.

41. Zheng K, Liang Y, Paez-Espino D, Zou X, Gao C, Shao H, Sung Yeong Y, Mok Wen J, Wong Li L, Zhang Y-Z et al. Identification of hidden N4-like viruses and their interactions with hosts. *mSystems*. 2023;8(5):e00197–00123.
42. Zhang Z, Tian C, Zhao J, Chen X, Wei X, Li H, Lin W, Feng R, Jiang A, Yang W. Characterization of tail sheath protein of N4-Like Phage phiAxp-3. *Front Microbiol*. 2018;9:450.
43. Liang Y, Li L, Luo T, Zhang Y, Zhang R, Jiao N. Horizontal and vertical distribution of marine viroplankton: a basin scale investigation based on a global cruise. *PLoS One*. 2014;9(11): e111634.
44. Zhou Z-d, Liu W-y, Li M-q. Enhanced thiosulfate-silver staining of proteins in gels using Coomassie Blue and chromium. *Biotech Lett*. 2002;24:1561–7.
45. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):1–11.
46. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40.
47. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49(16):9077–96.
48. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Tousseint A, Petit MA, Enault F. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics*. 2021;3(3):lqab067.
49. Camargo AP, Nayfach S, Chen I-MA, Palaniappan K, Ratner A, Chu K, Ritter Stephan J, Reddy TBK, Mukherjee S, Schulz F, et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Research*. 2022;51(D1):D733–43.
50. Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol*. 2023;21(4): e3002083.
51. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics*. 2011;27(7):1009–10.
52. Zheng K, Liang Y, Paez-Espino D, Zou X, Gao C, Shao H, Sung Yeong Y, Mok Wen J, Wong Li L, Zhang YZ, et al. Identification of hidden N4-like viruses and their interactions with hosts. *mSystems*. 2023;8(5):e00197–00123.
53. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26:680.
54. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
55. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30(22):3276–8.
56. Bin Jang H, Bolduc B, Zabolocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol*. 2019;37(6):632–9.
57. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, Shank SD, Magalis BR, Bouvier D, Nekrutenko A, et al. HyPhy 2.5-a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol*. 2020;37(1):295–9.
58. Sabi R, Daniel RV, Tuller T. stAl(calc): tRNA adaptation index calculator based on species-specific weights. *Bioinformatics*. 2017;33(4):589–91.
59. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell*. 2019;177(5):1109–1123. e1114.
60. Liang Y, Wang L, Wang Z, Zhao J, Yang Q, Wang M, Yang K, Zhang L, Jiao N, Zhang Y. Metagenomic analysis of the diversity of DNA viruses in the surface and deep sea of the South China Sea. *Frontiers in microbiology*. 2019;10:1951.
61. Xie Z-X, He Y-B, Wang M-H, Zhang S-F, Kong L-F, Lin L, Liu S-Q, Wang D-Z. Dissecting microbial community structure and metabolic activities at an oceanic deep chlorophyll maximum layer by size-fractionated metaproteomics. *Prog Oceanogr*. 2020;188: 102439.
62. Tada Y, Marumoto K, Takeuchi A. Nitrospina-like bacteria are potential mercury methylators in the mesopelagic zone in the East China Sea. *Front Microbiol*. 2020;11: 1369.
63. Xu B, Li F, Cai L, Zhang R, Fan L, Zhang C. A holistic genome dataset of bacteria, archaea and viruses of the Pearl River estuary. *Scientific Data*. 2022;9(1):49.
64. Sun M, Zhan Y, Marsan D, Páez-Espino D, Cai L, Chen F. Uncultivated viral populations dominate estuarine viromes on the spatiotemporal scale. *mSystems*. 2021;6(2). <https://doi.org/10.1128/msystems.01020-01020>.
65. Hwang J, Park SY, Park M, Lee S, Jo Y, Cho WK, Lee T-K. Metagenomic characterization of viral communities in Goseong Bay, Korea. *Ocean Science Journal*. 2016;51(4):599–612.
66. Wolf YI, Silas S, Wang Y, Wu S, Bocek M, Kazlauskas D, Krupovic M, Fire A, Dolja VV, Koonin EV. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol*. 2020;5(10):1262–70.
67. Schmidt EE. DNA replication in animal systems lacking thioredoxin reductase I. In: *DNA replication-current advances*. Rijeka: IntechOpen; 2011. <https://www.intechopen.com/chapters/16952>.
68. Arner ES, Holmgren A. Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem*. 2000;267(20):6102–9.
69. Tai N, Schmitz JC, Liu J, Lin X, Bailly M, Chen TM, Chu E. Translational autoregulation of thymidylate synthase and dihydrofolate reductase. *Front Biosci*. 2004;9:2521–6.
70. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, Gillet R, Forterre P, Krupovic M. Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun*. 2019;10(1):752.
71. Visser EA, Moons SJ, Timmermans SB, de Jong H, Boltje TJ, Büll C. Sialic acid O-acetylation: from biosynthesis to roles in health and disease. *J Biol Chem*. 2021;297(2):100906.
72. McPartland J, Kaganman I, Bowman VD, Rothman-Denes LB, Rossmann MG. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol*. 2008;378(3):726–36.
73. Fouts DE, Klumpp J, Bishoplilly KA, Rajavel M, Willner KM, Butani A, Henry M, Biswas B, Li M, Albert MJ. Whole genome sequencing and comparative genomic analyses of two *Vibrio cholerae* O139 Bengal-specific Podoviruses to other N4-like phages reveal extensive genetic diversity. *Virology Journal*. 2013;10(1):165.
74. Ceysens PJ, Brabban A, Rogge L, Lewis MS, Pickard D, Goulding D, Dougan G, Noben JP, Kropinski A, Kutter E, et al. Molecular and physiological analysis of three *Pseudomonas aeruginosa* phages belonging to the “N4-like viruses.” *Virology*. 2010;405(1):26–30.
75. Simmonds P, Aiewsakun P. Virus classification—where do you draw the line? *Adv Virol*. 2018;163(8):2037–46.
76. Zhan Y, Buchan A, Chen F. Novel N4 bacteriophages prevail in the cold biosphere. *Appl Environ Microbiol*. 2015;81(15):5196–202.
77. Sun M, Chen F. Distribution of rare N4-like viruses in temperate estuaries unveiled by viromics. *Environ Microbiol*. 2022;24(12):6100–11.
78. Robinson LS, Lewis WG, Lewis AL. The sialate O-acetyltransferase EstA from gut Bacteroidetes species enables sialidase-mediated cross-species foraging of 9-O-acetylated sialoglycans. *J Biol Chem*. 2017;292(28):11861–72.
79. Rangel A, Steenbergen SM, Vimr ER. Unexpected diversity of *Escherichia coli* sialate O-acetyl esterase NanS. *J Bacteriol*. 2016;198(20):2803–9.
80. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *PLoS Genet*. 2008;4(12): e1000304.
81. Burns CC, Shaw J, Campagnoli R, Jorja J, Vincent A, Quay J, Kew O. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J Virol*. 2006;80(7):3259–72.
82. Limor-Waisberg K, Carmi A, Scherz A, Pilpel Y, Furman I. Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies. *Nucleic Acids Res*. 2011;39(14):6016–28.
83. Mahler M, Malone LM, van den Berg DF, Smith LM, Brouns SJ, Fineran PC. An OmpW-dependent T4-like phage infects *Serratia* sp. ATCC 39006. *Microbial genomics*. 2023;9(3):000968.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.