


RESEARCH

Open Access



# Arms race in a cell: genomic, transcriptomic, and proteomic insights into intracellular phage–bacteria interplay in deep-sea snail holobionts

Kun Zhou<sup>1,2</sup>, Ying Xu<sup>2,3\*</sup>, Rui Zhang<sup>4,5\*</sup> and Pei-Yuan Qian<sup>1\*</sup> 

## Abstract

**Background:** Deep-sea animals in hydrothermal vents often form endosymbioses with chemosynthetic bacteria. Endosymbionts serve essential biochemical and ecological functions, but the prokaryotic viruses (phages) that determine their fate are unknown.

**Results:** We conducted metagenomic analysis of a deep-sea vent snail. We assembled four genome bins for *Caudovirales* phages that had developed dual endosymbiosis with sulphur-oxidising bacteria (SOB) and methane-oxidising bacteria (MOB). Clustered regularly interspaced short palindromic repeat (CRISPR) spacer mapping, genome comparison, and transcriptomic profiling revealed that phages Bin1, Bin2, and Bin4 infected SOB and MOB. The observation of prophages in the snail endosymbionts and expression of the phage integrase gene suggested the presence of lysogenic infection, and the expression of phage structural protein and lysozyme genes indicated active lytic infection. Furthermore, SOB and MOB appear to employ adaptive CRISPR–Cas systems to target phage DNA. Additional expressed defence systems, such as innate restriction–modification systems and dormancy-inducing toxin–antitoxin systems, may co-function and form multiple lines for anti-viral defence. To counter host defence, phages Bin1, Bin2, and Bin3 appear to have evolved anti-restriction mechanisms and expressed methyltransferase genes that potentially counterbalance host restriction activity. In addition, the high-level expression of the auxiliary metabolic genes *narGH*, which encode nitrate reductase subunits, may promote ATP production, thereby benefiting phage DNA packaging for replication.

**Conclusions:** This study provides new insights into phage–bacteria interplay in intracellular environments of a deep-sea vent snail.

\* Correspondence: [boxuying@szu.edu.cn](mailto:boxuying@szu.edu.cn); [ruizhang@xmu.edu.cn](mailto:ruizhang@xmu.edu.cn); [boqianpy@ust.hk](mailto:boqianpy@ust.hk)

<sup>2</sup>Shenzhen University–HKUST Joint Marine Science Ph.D. Program, Shenzhen University, Shenzhen 518060, China

<sup>4</sup>State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University (Xiang'an), Xiamen, Fujian, China

<sup>1</sup>Department of Ocean Science and Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Hong Kong University of Science and Technology, Hong Kong, China

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Keywords:** Deep ocean, Hydrothermal vents, Snail holobionts, Bacterial endosymbionts, Phages, Lysogenic and lytic infections, Anti-viral defence, Counter defence, Horizontal gene transfer, Multi-omics

## Background

Deep-sea hydrothermal vent ecosystems are abundant with macrofauna, such as invertebrate animals that adapt to such extreme environments by symbioses with chemoautotrophic bacteria [1–6]. After the discovery of hydrothermal vents in 1977, numerous studies have focused on the identification and characterisation of chemosynthetic bacteria, which have been found to have remarkable functional roles, including the ability to oxidise reduced chemical compounds to fuel the ecosystems [7–9]. Intracellular chemoautotrophic bacteria in particular have received substantial attention and have been well documented [2, 4, 10–12]. For instance, gill endosymbionts in *Bathymodiolus* mussels carry hydrogenase genes and consume hydrogen as an energy source [2]. An investigation of the mussel *Bathymodiolus azoricus* discovered two groups of endosymbionts, sulphur-oxidising and methane-oxidising bacteria, which led to dual symbiosis that gave the mussels higher environmental tolerance and the ability for niche expansion [13]. Specialised intracellular bacteria play essential roles in symbiosis with marine animals, but the bacterial viruses (phages) that determine the fate of endosymbiotic bacteria are uncharacterised.

Intracellular environments are isolated niches that protect endosymbiotic bacteria from phage infection. However, phages have been found in many endosymbiotic systems of arthropods, including flour moth [14], mosquito [15, 16], cricket [17], wasp [18], and fruit fly [19]. These phages (WO) widely infect symbiotic *Wolbachia* bacteria and enter lytic lifecycles for the parasitic A and B *Wolbachia* supergroups or lysogenic lifecycles for the C and D *Wolbachia* supergroups [20]. WO phages can be vertically transmitted or horizontally transferred between *Wolbachia* cells and possibly arthropods. WO phages have a significant impact on the abundance, activity, and diversity of *Wolbachia* and then on the symbiosis between *Wolbachia* and host arthropods [20].

The detection of potentially essential phages in the endosymbiotic systems of terrestrial arthropods suggested the possible presence of phage infection of endosymbionts that inhabit deep-sea animals. Studies on deep-sea vent fields have found snails that commonly inhabit them and develop endosymbioses with bacteria [21–25]. Snails in the genus *Gigantopelta* are abundantly colonised in Longqi [26], hydrothermal vent field at 1755 m depth in the Southwest Indian Ridge [27]. *Gigantopelta aegis* harbours two phylotypes of gamma-proteobacterial endosymbionts, sulphur-oxidising (SOB)

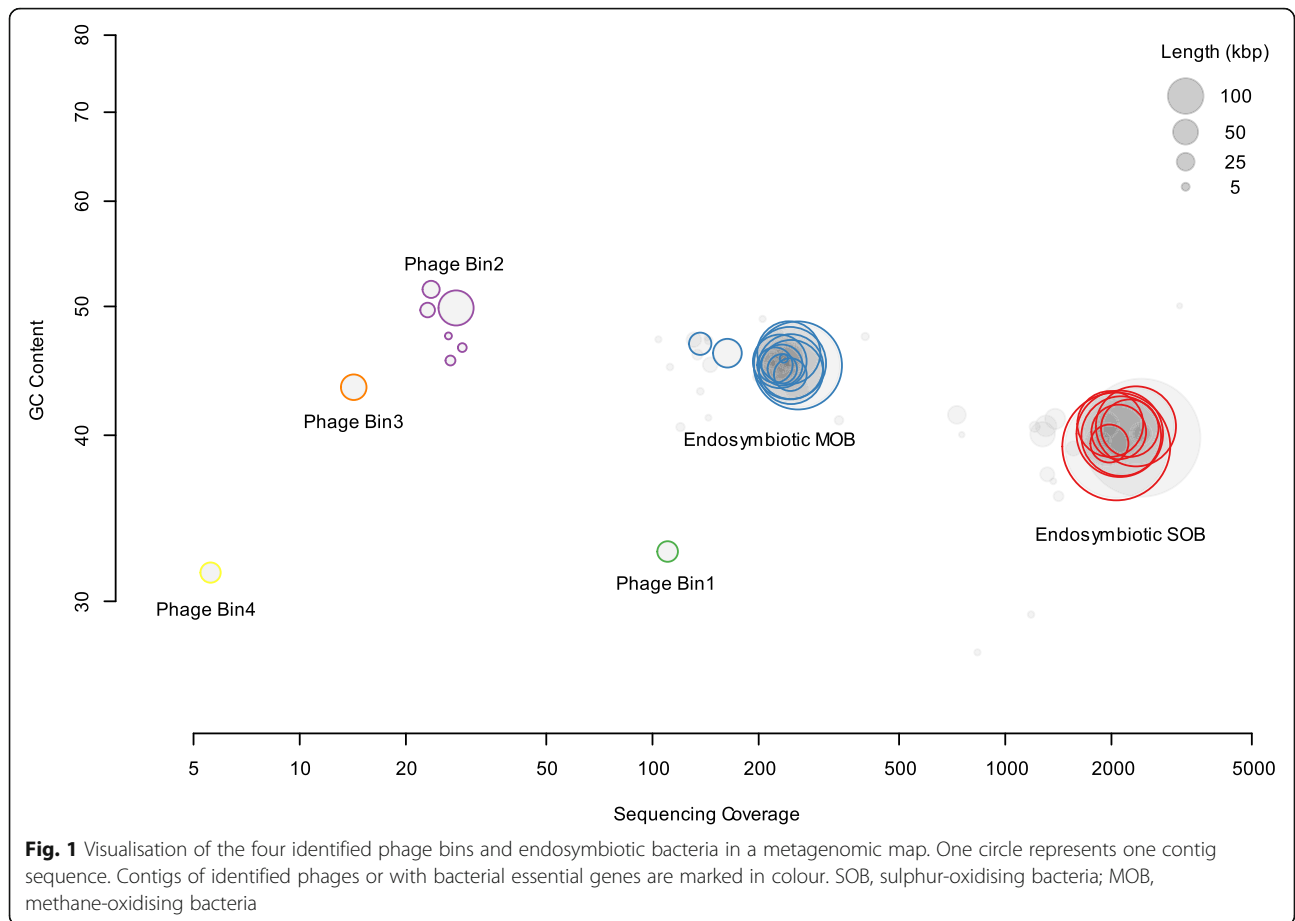
and a methane-oxidising (MOB) bacteria, in the oesophageal gland cells within the gut [24]. An omics study revealed that SOB and MOB are beneficially symbiotic to *G. aegis* snails by complementing metabolisms and providing required nutrients [24]. In the present study, we demonstrated the presence of phages that infect intracellular chemosynthetic symbionts and characterised their phylogeny and lifecycle in *G. aegis* from their genomics, transcriptomics and proteomics data. We focused on the anti-viral defence systems of bacterial endosymbionts and counter-defence mechanisms of phages to explore potential interactions between the identified phages and endosymbiotic bacteria in intracellular environments in deep-sea vent fields.

## Results

### Genomes and gene expression of phages in snail oesophageal glands

Four phage bins were recovered in the present study (Fig. 1 and Table 1), and phage hallmark genes, including genes encoding portal and head proteins, were identified in each bin (listed in Tables S1–S4, visualised in Figs. S1–S5). In each bin, the genome size ranged from 29 to 139 kb, and the GC content ranged from 31% to 49%. Two of the four bins contained circular contigs, indicating that these genomes were potentially complete. The number of predicted genes for each phage bin ranged from 17 to 142. Bin2 harboured integrase-coding genes, implying that it represented a temperate phage. Transcriptomic analysis showed that phages Bin1, Bin2 and Bin4 had reconstructed transcript support (Tables S1, S2, and S4). For example, genes 1\_10 for phage proteins GP46 and 1\_16 encoding SNF2 in phage Bin1 were transcribed into TRINITY\_DN1\_c0\_g1\_i2 and TRINITY\_DN17\_c0\_g1\_i1, respectively (Table S1). For the temperate phage Bin2, many of its genes had transcript support, including genes for integrases, recombinases, tail proteins, baseplate proteins, lysozymes, capsids and auxiliary metabolic genes (Table S2). Although phages Bin3 and Bin4 in a relatively low abundance had less/no reconstructed transcripts detected in accordance with strict criteria (95% identity and 100% coverage), their genes showed mapping-read support according to RNA-sequencing read count (Tables S3 and S4). For instance, gene 1\_12 with a length of 1173 bp for phage Bin4 capsid proteins recruited 71 reads in Sample 2 (Table S4) for mRNA sequencing.

Open reading frame-based taxonomic classification reveals that phage bins in deep-sea snails are affiliated with



the order *Caudovirales*. Network-based phylogeny showed that these phage bins were clustered with animal/human-associated phages in family *Myoviridae* (Fig. 2 and Table S5). According to the gene-sharing network, the phages recovered in the present study were classified into two groups (Group1: phages Bin1 and Bin4; Group2: phages Bin2 and Bin3). In Group1, the snail gut-associated phages were connected directly to seven Enterobacteriaceae-infecting Myoviridae phages, namely *Enterobacteria* phages SfI, SfV, and phiP27, *Salmonella* phages 118970\_sal3 and ST64B, *Shigella* phages SfII and SfIV. Notably, phages Bin1 and Bin4 were also indirectly connected to the bacteriophage APSE-2 infecting Enterobacteriaceae endosymbionts of ants, tsetse,

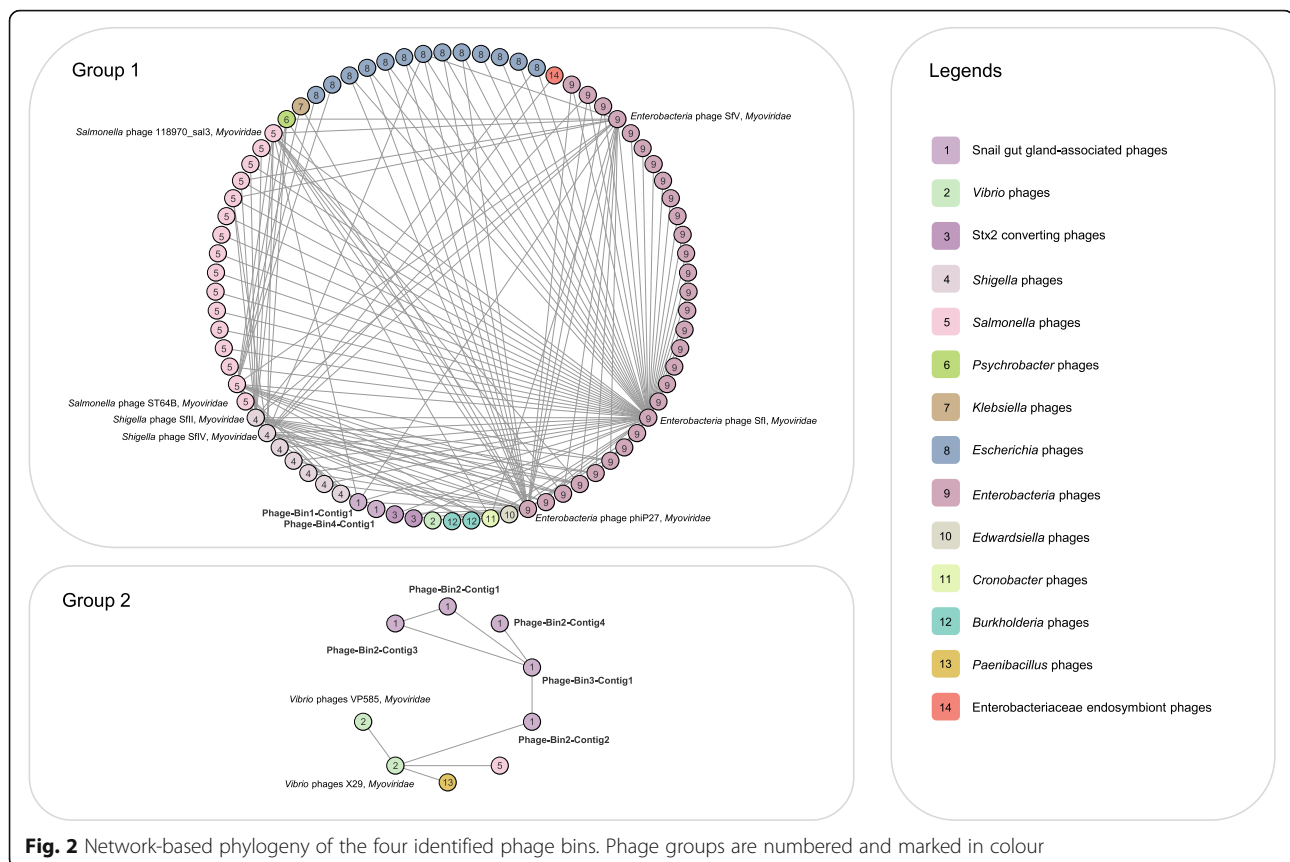
mealybugs, and aphids. Different from Group1, Group2 included fewer members, namely, phages Bin2 and Bin3, *Vibrio* phages X29 and Vp585, *Salmonella* phage Gifsy-1, and *Paenibacillus* phage Lily. Similarly, the phages associated with phages Bin2 and Bin3 were Enterobacteriaceae-infecting *Myoviridae* phages such as *Vibrio* phage X29.

**Phage–endosymbiont connections and horizontal gene transfer**

The clustered regularly interspaced short palindromic repeat (CRISPR) spacers from endosymbionts MOB and SOB were used to probe their phages by mapping to phage sequences. Phages Bin1 and Bin4 matched CRIS

**Table 1** Recovered genomes of phages from gland tissues of the snail *Gigantopelta aegis*. NO, no eukaryotic or prokaryotic lineage-specific genes were detected

	Phage Bin1	Phage Bin2	Phage Bin3	Phage Bin4
Classification	<i>Caudovirales</i>	<i>Caudovirales</i>	<i>Caudovirales</i>	<i>Caudovirales</i>
Total length (Kbp)	30.130	139.757	46.730	29.326
GC content (%)	32.70	49.69	43.47	31.53
Completeness	Potentially complete (circular)	incomplete	incomplete	Potentially complete (circular)
Contamination	NO	NO	NO	NO



**Fig. 2** Network-based phylogeny of the four identified phage bins. Phage groups are numbered and marked in colour

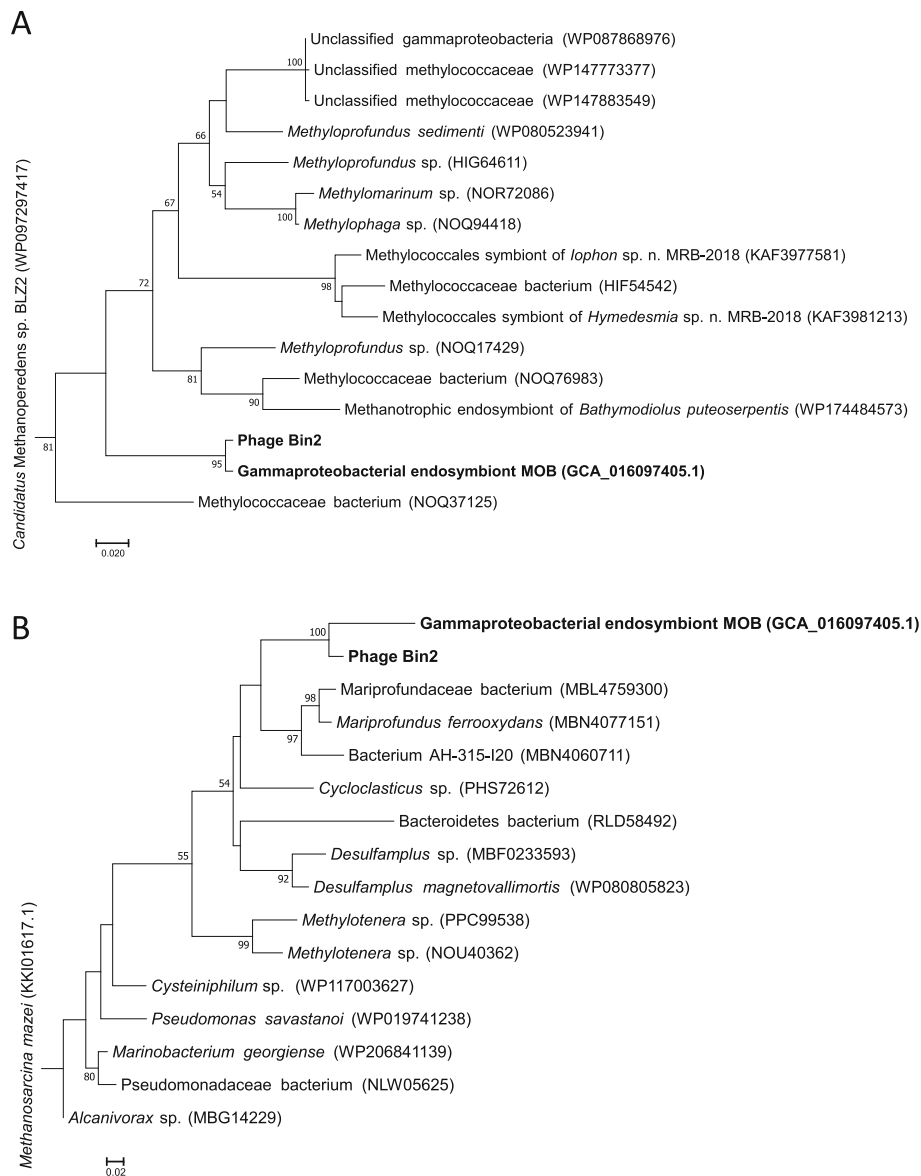
PR spacers of both endosymbiotic MOB and SOB (Table S6). Compared with MOB, SOB contained more CRISPR spacers derived from Bin1 (12 spacers) and Bin4 (4 spacers). All the mapped phage sequences of the spacers were adjacent to the bacterial genes encoding *Cas1* and *Cas2* or *Cas6f*, which are the adaptation modules for spacer insertion [28, 29] according to the genome annotations of phages and endosymbionts (Tables S6 and S7). In addition, a phage–host link was discovered between phage Bin2 and the endosymbiont MOB because of the high similarity (3.3 kb alignment length, 100% coverage, 93% identity) between phage contig6 and MOB contig2 (Table S6).

Phylogenetic analysis showed that the genes for nitrate reductase subunit  $\beta$  (*NarH*) and DNA methyltransferase were horizontally transferred between MOB and phage Bin2 (Fig. 3). The *narH*, along with the gene for nitrate reductase subunit  $\alpha$  (*narG*) located beside *narH*, is responsible for nitrate reduction in energy metabolism. The gene encoding DNA methyltransferase is related to restriction–modification systems, which are involved in innate immunity (see below). Transcriptomic analysis showed that phage *narH*, *narG*, and DNA methyltransferase genes were expressed, demonstrating that genes related to nitrate reduction and restriction–modification systems might be harnessed by phage Bin2 during the infection of MOB.

### Infection strategy of endosymbiotic phages

Thirty-eight putative prophage regions were detected in SOB and MOB using VirSorter [30], Phaster [31], and Prophage Hunter [32] (Tables S8, S9, and S10). Among them, one passed CheckV [33] examination for incomplete prophage screening based on viral-specific genes and microbial-specific genes flanked on the sides. We annotated the genes in the putative prophages against PFAM 32.0 [34] and identified phage feature genes for integrases, late control gene D protein, Gp49-like, lysozyme, tail tube protein, and tail sheath protein (Table S11). The unbinned bacterial sequences, potentially from symbionts in the metagenome, were examined by CheckV and Phaster. Two contig sequences (NODE\_11432 and NODE\_10) were identified as incomplete prophages. Transcriptomic analysis showed that the phage Bin2 integrase genes were expressed at relatively high levels (Table S2). Together, these results indicate that lysogeny is a potential infection strategy of endosymbiotic phages in the deep-sea vent snail *G. aegis*.

Conversely, genomic evidence for lytic infections was found. The putative prophages were compared with the phage genome bins identified in this study. However, no high genome similarity (identity  $\geq 95\%$ , coverage  $\geq 80\%$ ) was found between prophages and phage bins, demonstrating that the phages and potential prophages were

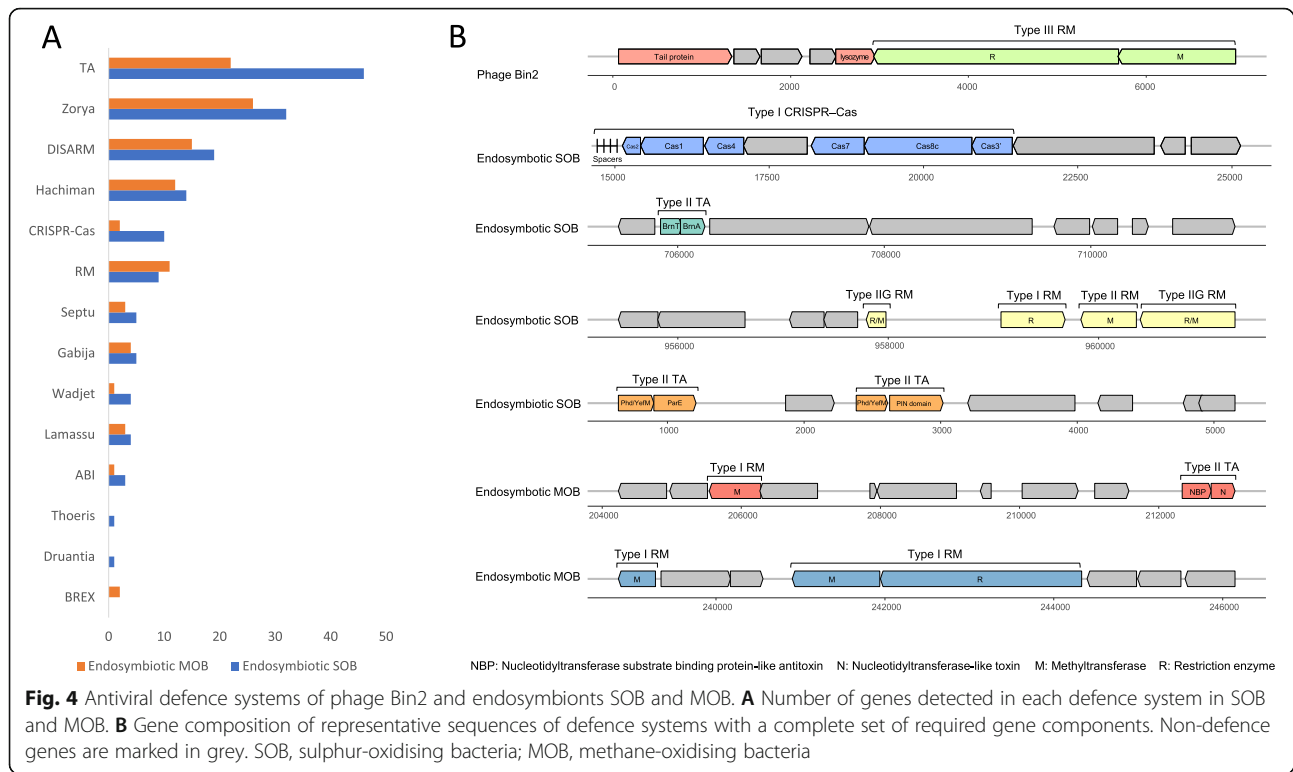


**Fig. 3** Maximum likelihood-based phylogeny of *narE* and DNA methyltransferase genes of phage Bin2 and endosymbiont MOB. **A** Phylogenetic tree of *narH* genes. **B** Phylogenetic tree of genes encoding DNA methyltransferases. Bootstrap values <50 are not shown. MOB, methane-oxidising bacteria

not from the same populations. No feature genes (e.g., integrase genes) of temperate phages were detected in phages Bin1 and Bin4, indicating they may be virulent phages with lytic lifecycles. Transcript and read mapping revealed that many genes of phages Bin1 and Bin4 for virion production had expression support, such as the genes coding portal proteins, prohead serine proteases, capsids, and head–tail joining proteins. Furthermore, the expression of lysozyme genes in phage Bin2 suggests the potential temperate phages may also have entered a lytic lifecycle and performed cell lysis to release virion progenies.

### Anti-viral defence systems in bacterial endosymbionts and their phages

A total of 153 and 102 genes of anti-viral defence systems were detected in the endosymbionts SOB and MOB, respectively (Tables S7 and S12). These genes are likely involved in diverse defence systems, including Zorya, Wadjet, toxin–antitoxin (TA, Type II), Septu, restriction–modification (RM), Hachiman, Lamassu, defence island system associated with restriction–modification (DISARM), CRISPR–Cas (Type I), abortive infection, and Gabija (Fig. 4A). Among these systems, the Type II TA, RM, and CRISPR–Cas systems were



complete with a full set of the required gene components (Fig. 4B). In SOB, Type II TA was composed of multiple toxin–antitoxin gene pairs for BrnT and BrnA, ParE and Phd/YefM, and PIN domain-containing proteins and Phd/YefM. In MOB, nucleotidyltransferase-like toxin and nucleotidyltransferase substrate-binding protein-like antitoxin constituted the two components of Type II TA. The foreign DNA-targeting RM systems were of two types: Type I RM (discovered in SOB) and Type IIG RM (detected in MOB). The Type I RM system encoded gene pairs for a methyltransferase that is used to modify nucleotides by methylation and a restriction enzyme for cleaving specific DNA sites. The detected Type IIG RM encoded a single gene for a combination of the methyltransferase and the restriction enzyme. The Type I CRISPR–Cas systems contained genes for Cas1–4, Cas6f, Cas7, and Cas8c. Adjacent to the genes for Cas2 (located in SOB) and Cas6f (located in MOB), an array of spacers was detected and the sequences were identical to sequences from phages Bin1 and Bin4 (Table S6). Transcriptomic analysis found that most of the defence genes (245 of 255) were expressed. For the most abundant SOB in the intracellular community, reconstructed transcript mapping and protein mass spectrometry showed that genes for the DEAD/DEAH box helicase of the Hachiman system, ZorB of the Zorya system, and Cas7 and Cas8c of the CRISPR–Cas system were transcribed and translated. Notably, the translation level of *cas7* was relatively high with an exponentially

modified protein abundance index (emPAI) value of 0.93, which ranked in the top 20 of the 135 sequenced peptides of SOB and MOB in one snail sample (Table S13).

In the recovered phage bins, we also identified genes related to bacterial defence systems, with 13 genes from Bin1, Bin2, and Bin3 being involved in RM, DISARM, bacteriophage exclusion system, abortive infection, Septu, or Zorya (Table S14). The sequence alignment data were used to map 12 defence genes to meta-transcriptomic reads and nine of them had transcript mappings. Over half of the expressed genes were encoded by RM systems harboured in phages Bin2 and Bin3. Notably, the Type I and Type III RM genes for methyltransferases and restriction enzymes of phage Bin2 (Table S14 and Fig. 4B) were significantly expressed with a top 50 ranking among 142 genes (Table S2).

## Discussion

Endosymbioses are common in nature and frequently refer to beneficial, intracellular symbioses [35]. The intracellular environment is usually considered as an isolated niche that protects symbiotic bacteria from phage infection. The discovery of bacteriophages that infect endosymbiont *Wolbachia* bacteria in insects led to the study of phages in intracellular habitats of invertebrate animals [20]. Our genomic, transcriptomic, and proteomic analyses demonstrated the presence of endosymbiont phages in deep-sea snails, thereby expanding phage

intracellular niches from terrestrial to marine environments.

Microbial symbionts can be horizontally transferred from the environment and/or vertically transmitted from parent hosts [36]. Accordingly, we presumed that the snail-associated phages entered the gland cells by horizontal transfer and/or vertical transmission. For horizontal transfer, the intracellular phages may be acquired by animals by uptake through the gills and guts [37], as was found in mussels, or through water flow systems, as in sponges [38]. The absorbed phages can enter animal cells, for example, through transcytosis, macropinocytosis, phagocytosis, active bacterial infection, or activation of a bacterial carrier [39]. Because symbiotic environments can select specific symbionts in horizontal transfer [36], intracellular viromes can be conserved across animal hosts, which is likely to be the reason why the snail phages are closely related to phages that infect animal-associated bacteria (Fig. 2 and Table S5). For vertical transmission, snails may acquire snail phages from their parents [20].

Previous metagenomics studies suggested the lysogenic infection strategy for phages was predominant in symbiotic systems and the model endosymbiotic WO phages were temperate phages [40, 41]. In our study of deep-sea snail endosymbiotic systems, the lysogenic lifestyles were indicated by phage integrase gene expression. We hypothesised that phage Bin2 is a temperate phage, considering the high number of transcriptomic reads identified for the integrase genes (Table S2). However, phage Bin2 did not map to any of the prophages, which suggested Bin2 may exist as a plasmid. This idea was supported by the discovery of a gene for plasmid pRiA4b ORF-3-like protein in the Bin2 genome, which was expressed with transcript support (Table S2). In Bin2, we also found relatively highly expressed genes that encoded structural proteins for virion production and lysozymes for cell lysis. Accordingly, we speculated that some of the phage Bin2 population switched back to a lytic lifecycle, whereas others maintained a lysogenic lifecycle, which is consistent with the model of spontaneous lysogenic to lytic switch [42]. Because phage Bin3 was classified along with phage Bin2 in Group2, we hypothesised that phage Bin3 is another extrachromosomal temperate phage that can be induced to enter a lytic lifecycle. However, further experimental evidence, such as plaque assays based on cultured host and virus isolates, is required to elucidate the life strategy of endosymbiotic phages in deep-sea snails.

To defend against phage infection and lysis, bacteria use their defence/immune systems as weapons to prevent population decimation by targeting invading DNA, and/or aborting viral replication, and/or conducting programmed death [43]. To target invading DNA, bacteria

depend on innate and/or adaptive immunity, such as the RM [44], DND [45], DISARM [46], bacteriophage exclusion [47], and CRISPR–Cas [29] systems. To abort viral replication, bacteria have developed TA [48] and abortive infection [49] systems. Other newly found systems with unknown mechanisms have been discovered, including Zorya, Hachiman, Gabija, Septu, Thoreris, Lamassu, Druantia, Wadjet, Kiwa, and Shedu [50]. Our exploration of endosymbiont genomes showed that over 100 genes related to 14 defence systems were harboured in SOB and MOB as arsenal against phage infection and lysis. The expression profiles of the genes in these system genes suggest they co-function to form multiple complementary defence lines [51] and generate a synergistic effect to efficiently protect hosts from phage infection [43]. For example, when phages inject their DNA into SOB or MOB cells, RM systems such as Type IIG RM activate the expression of genes for restriction endonucleases to target phage DNA and cleave specific nucleic acids. Conversely, the expression of genes for methyltransferases is used to modify the host genome to avoid self-immunisation. CRISPR–Cas and RM systems are functionally coupled. The adaptive immune CRISPR–Cas system activates genes for Cas1, Cas2, and Cas7. Then, Cas1 and Cas2 incorporate partial sequences of the invading DNA of phages, such as Bin1 and Bin4, into the CRISPR array as spacers. The spacer sequences are transcribed into CRISPR coded RNA to guide Cas7 towards the invading phage DNA to cleave it. The expression of other non-DNA-targeting systems such as TA provides another line of defence. When phages successfully inject their DNA and start replication, TA systems induce the dormancy of infected cells by inhibiting gene expression; for example, the BrnT toxins in SOB inhibit protein synthesis by functioning as ribonucleases [52]. For other systems (such as Zorya and Gabija) with partially identified required gene components, the gene expression profiles indicate that they potentially have different functional roles in the phage infection. The deployment of multiple lines of defence reflects the benefits of diverse co-functioning systems when bacterial hosts are faced with diverse heterogeneous phage predators.

In the arms race of prey and predators, phages evolved counter-defence systems to resist host defences; for example, phages that carry genes of RM systems have developed anti-restriction strategies to promote bacterial infection [53]. In our study, phages Bin1, Bin2, and Bin3 contained RM-related genes for methyltransferases and restriction enzymes mostly from Type I RM systems, which were detected in their SOB and MOB hosts. The highly similar methyltransferase genes of phage Bin2 (gene 5\_1) and MOB (gene 4\_242; Table S15) were shown to be transferred between them (Fig. 3B),

implying phages might acquire RM systems from endosymbionts. The methyltransferase-coding genes of phages were expressed at relatively high levels. These results indicate that phages may harness the expression of their genes for methyltransferases to rapidly modify the phage DNA to avoid its recognition by similar host restriction enzymes through a passive mechanism of phage evasion [53]. Moreover, the modification might protect newly synthesised phage genomes from host cleavage. The presence of RM-encoding genes may also result in a higher abundance of the phages Bin1 and Bin2 than phage Bin4, which share the same infection hosts as Bin1 and Bin2.

After successfully evading host defences, phages are faced with transcription and DNA packaging, which are energy consuming processes in the lytic replication cycle [54]. During DNA packaging, synthesised phage genomes are pumped into procapsids via ATP-dependent translocation by terminases that contain ATPase domains [54, 55]. Anaerobic respiration of prokaryotic nitrate reduction is an ATP-generating process in which the membrane-bound respiratory proteins NarG, NarH, and NarI produce a transmembrane proton motive force for ATP synthesis [56]. In the present study, Nar genes in phage Bin2, such as *narH*, were expressed by hijacking host metabolic machinery. The Nar genes worked as auxiliary metabolic genes to generate ATP along with the highly expressed host Nar genes in the endosymbiotic MOB [24]. With the high production of ATP, the synthesised DNA and proteins can be packaged efficiently for subsequent virion assembly, which contributes to a successful lytic cycle.

## Conclusions

By genomic, transcriptomic and proteomic analysis, we assembled endosymbiotic phage genomes associated with the vent snail *G. aegis* and found that the phages had infected endosymbiotic sulphur-oxidising bacteria and methane-oxidising bacteria with lysogenic or lytic lifecycles. In the arms race between endosymbiotic bacteria and phages, the bacteria encoded CRISPR–Cas systems to target phage DNA and other potentially functional defence systems to form multiple anti-viral lines. To counter these defences, the phages evolved anti-defence mechanisms and horizontally acquired auxiliary metabolic genes to benefit replication. These findings provide new insights into phage–bacteria interplay in deep-sea vent snail holobionts. According to our understanding of phage ecology in marine environments, we expected that endosymbiotic phages might be non-animal-derived factors that regulate the population size of symbiotic bacteria in deep-sea animals.

## Methods

In this study, we reanalysed the genomic, transcriptomic, and proteomic data of the deep-sea vent snail *G. aegis*. The sampling and processing of the *G. aegis* snails have been described in detail in a previous study [24]. Briefly, the samples were collected by the Jiaolong human occupied vehicle from Longqi located on the Southwest Indian Ridge and stored at  $-80^{\circ}\text{C}$  for DNA and RNA extraction. Total DNA was extracted from the oesophageal gland containing endosymbionts using a MagAttract High-Molecular-Weight DNA Kit (QIAGEN, Hilden, Netherlands) and sequenced on an Illumina NovaSeq 6000 platform. A total of  $2 \times 150$ -bp paired-end reads was obtained. In parallel, the total DNA was sequenced using a Nanopore sequencer (Oxford Nanopore MinION, UK) for scaffolding the pre-assembled contigs from the Illumina reads. The total RNA of oesophageal gland tissues of three individuals was extracted separately using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and sequenced on the Illumina NovaSeq 6000 platform. A total of  $2 \times 150$ -bp paired-end reads was obtained. Total protein was extracted from the oesophageal glands of three individuals using methanol chloroform and analysed using a Dionex Ultimate 3000 RSLCnano and Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher).

## Metagenome assembly and recovery of metagenome-assembled genomes of bacteriophages

The raw Illumina and Nanopore reads of the snail *G. aegis* oesophageal gland metagenome from the parent study [24] were retrieved from NCBI (BioProject accession: PRJNA612619). Trimmomatic (version 0.36) [57] was used with custom parameters (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:40) to trim the downloaded reads. The trimmed reads were assembled using SPAdes version 3.11.1 [58] with custom parameters (--meta, k-mer size varied from 51 to 91, with a 10-mer step size). Viral sequences with hallmark genes were identified in the assembled metagenome using the VirSorter with default parameters against the RefSeqABVir and Viromes databases [30] for subsequent binning analysis. For the genome binning of phages, we followed the approach described in [59] to manually bin phage genome sequences (contigs  $\geq 2$  kb) (see processing details in Figs. S6–S9) on the basis of GC content, sequencing depth, tetranucleotide frequency, and phage hallmark genes. The binned contigs were mapped to the Illumina reads by Bowtie2 version 2.3.4 [60] and SAMtools version 1.6 [61] to select bin-related short reads. The mapped short reads along with the long Nanopore reads were used to reassemble phage bin-related sequences using Unicycler version 0.4.7 [62] with the default



parameters. Sequences  $\geq 2$  kb were retained in the reassembled genomes.

In the resulting reassemblies, phage Bin2 contained more than one contig. Paired-end read information of each metagenomic contig was collected using a script (cytoscapeviz.pl with parameters: `-f 2 -e 500 -m 3000 -a 125 -c`) from a previous study to remove the potential contamination of eukaryotic or prokaryotic sequences in the reassembled phage Bin2 genome [59]. The eukaryotic and prokaryotic sequences in the metagenomes (after removing predicted phage contigs with hallmark genes) were identified using Autometa (parameters: `--length_cutoff 1000 --maketable --ML_recruitment`) [63]. Two unsure contig sequences in phage Bin2 were removed because (1) the two contigs had paired-end connections to the contigs from bacterial genomes (NODE\_47647 and NODE\_91930), (2) they contained no phage feature genes, such as structural protein-coding genes, and (3) the sequencing coverage of the two contigs (115 $\times$  and 65 $\times$ ) was higher than that of other contigs in phage Bin2 (approximately 25 $\times$ ) but relatively close to that of the connected bacterial contigs (NODE\_47647 was 116 $\times$ , NODE\_91930 was 213 $\times$ ). Finally, CheckM [64] with the parameter `lineage_wf` was used to identify lineage-specific marker genes of prokaryotes to assess contamination from bacterial genomes. BUSCO version 4.0.beta1 with the 'eukaryota\_odb10' database was used to check whether the viral bins lacked eukaryotic lineage-specific marker genes. The binned phage genomes were confirmed to contain no single-copy gene that was lineage specific to prokaryotes or eukaryotes. Additionally, a viral bin that contained more than one *terL* gene, indicative of a mixture of viral populations, was removed to ensure that one phage bin represented only one population. Qualified phage genome bins were retained for subsequent analysis (genome sequences were provided in Supplementary Genome sequences of Phages Bin1–4).

#### Identification of prophages

Prophage regions were separately detected from the bacterial endosymbionts SOB and MOB (NCBI GenBank: GCA\_016097415.1 and GCA\_016097405.1) with VirSorter against the RefSeqABVir and Viromes databases, PHASTER [31]. In parallel, Prophage Hunter [32] was used to identify prophages of the endosymbiont genomes. All the identified prophage sequences were functionally annotated using HMMScan in the HMMER 3.3 tool suite [65] against PFAM 32.0 [34] (E-value  $< 10^{-3}$ , bit score  $\geq 30$ ). Prophage candidates containing phage feature sequences (e.g., attachment sites [*att*] and/or genes that code integrases and/or structural proteins) or identified as intact/active prophages were retained as prophage candidates. The candidates were further aligned with identified phage sequences using BLASTn (E-value  $< 10^{-3}$ , identity  $\geq 95\%$ , coverage  $\geq 80\%$ ) to check

whether they were from the same population to infer the lifecycle of the phages [66]. CheckV [33] was used to identify viral-specific and microbial-specific genes and ensure that the candidates were from phages integrated into host genomes. Candidates containing viral-specific genes flanked by host genes were considered reliable prophage candidates. Considering that endosymbiont sequences may not have been completely recovered from the metagenome, unbinned bacterial sequences from Autometa (parameters: `--length_cutoff 1000 --maketable --ML_recruitment`) [63] also were examined by CheckV to detect prophage signals, to ensure a comprehensive investigation.

#### Phylogenetic, taxonomic, and functional annotations of phages and bacterial endosymbionts

Following [67], we use vContact2 [68] with custom parameters (`--rel-mode Diamond, --db 'ProkaryoticViral-RefSeq85-Merged', --pcs-mode MCL, --vcs-mode ClusterONE`) to cluster closely related phages. The resultant network file was imported into Cytoscape v3.8.0 [69] to visualise the network-based phylogeny of the identified phages. To analyse genome function, the genomes of the phages and endosymbiotic bacteria were imported into PRODIGAL [70] run with customised settings (`-c, -m`) to predict open reading frames (ORFs). The predicted ORFs were functionally annotated by searches against the Kyoto Encyclopedia of Genes and Genomes database [71] with KofamScan version 1.2.0 (E-value  $< 10^{-5}$ , score  $>$  predefined thresholds by KofamScan) [72]. The ORFs also were analysed with HMMScan [73] based on the PFAM database [34] with E-value  $< 10^{-3}$  and bit score  $> 30$ . The phage ORFs were compared with GOV-predicted proteins [74] and classified into eight categories: 'DNA replication, recombination, repair, nucleotide metabolism'; 'metabolism'; 'membrane transport, membrane-associated'; 'lysis'; 'structural'; 'transcription, translation, protein synthesis'; 'other'; and 'unknown' [74]. To determine the taxonomy of the phages, genome bins were annotated using CAT with custom parameter settings (`--fraction 0.8`) [5]. Taxonomy was inferred based on searches against the NCBI-nr database (<http://blast.ncbi.nlm.nih.gov/>) with BLASTp (E-value  $< 10^{-5}$  and identity  $> 30\%$ ) when no taxonomic annotation was obtained using CAT.

#### Transcriptome assembly, gene expression quantification, and protein identification by mass spectrometry

Sequenced RNA reads of metatranscriptome from the gland tissues of three snail *G. aegis* samples (SRR13131427, SRR13131416, and SRR13131407) from the parent study [24] were downloaded from the NCBI (<https://www.ncbi.nlm.nih.gov/>) Sequence Read Archive. Trimmomatic (version 0.36) was used with custom

parameters (ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:40) to trim the Illumina adapters and low-quality bases of the RNA reads. Two de novo assembled transcriptomes (total-RNA-reads-assembled version and genome-bin-mapping-reads-assembled version) were independently processed with Trinity v2.8.5 [75]. For the total-RNA-reads-assembled version, all the trimmed reads were imported into Trinity with default parameters to assemble the metatranscriptomes of three samples in parallel. For the genome-bin-mapping-reads-assembled version, Bowtie2 version 2.3.4 with default parameters was used to obtain genome bin-related reads (including SOB and MOB genomes), then Trinity was used to assemble the transcriptome of each bin (phages and endosymbiotic bacteria). Reconstructed transcript sequences were mapped to the predicted ORFs of each bin using BLASTn (E-value  $<10^{-3}$ , identity  $\geq 95\%$ , coverage = 100%). Expression levels of the ORFs of the phages, SOB, and MOB were quantified in transcripts per million using Salmon [76] with default parameters and the RNA-sequencing reads as input. In this study, we defined an expressed gene as one that had the support of a transcript or read mapping. Raw mass spectrometry data of sequenced peptides were retrieved from ProteomeXchange (<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD022852>) from the parent study [24]. A database of predicted protein sequences from the phage, SOB, and MOB genomes with their reversed protein sequences as a decoy was constructed for searching against the sequenced peptides. The sequence database and raw mass spectrometry data were input into Mascot version 2.3.0 [77] to identify and quantify the phage, SOB, and MOB proteins (confidence  $\geq 0.95$  and false discovery rate  $\leq 2.5\%$ ).

### Phage–host prediction

Phage–host associations between the bacterial endosymbionts and phage bins were identified using the following criteria: (1) Phage sequences from a bin and bacterial contigs from endosymbionts had  $\geq 70\%$  BLASTn identity (E-value  $\leq 10^{-3}$ ) and  $\geq 2.5$  kb alignment length [74]. (2) The CRISPR spacers from an endosymbiont genome identically matched the genome sequences of a phage bin [74, 78]. MetaCRT [79, 80] was used to predict CRISPR spacers and spacers  $> 6$  bp in length were matched to phage genome bins with fuzznuc [81].

### Identification of horizontal gene transfer event

Phage and bacterial contigs were aligned using BLAST to detect horizontal gene transfer (HGT) regions. According to the criteria in [74], sequences with high alignment similarity (E-value  $< 10^{-3}$ , bit score  $> 50$ , alignment length  $\geq 2.5$  kb and identity  $> 70\%$ ) were retained as HGT

candidates. Genes for RM-related DNA methyltransferases with a high similarity (93% amino acid identity, Table S15) between phage Bin2 (gene 5\_1) and endosymbiont MOB (gene 4\_242) were selected to further identify horizontally transferred genes. Putative HGT sequences were searched against the PFAM database, the Kyoto Encyclopedia of Genes and Genomes database, and the GOV classification system for functional annotation. To validate HGT events, genes for nitrate reductase subunit  $\beta$  (*narH*) and the DNA methyltransferases located in the putative HGT regions were used to infer a maximum likelihood phylogeny. The amino acid sequences of other *narH* and DNA methyltransferase genes were retrieved from the NCBI-nr database (<http://blast.ncbi.nlm.nih.gov/>) with BLASTp (E-value  $< 10^{-5}$  and identity  $> 30\%$ ) aligned with our sequences using Mafft [82] (parameters: --adjustdirectionaccurately --auto). The aligned sequences of NarH and DNA methyltransferases were then separately imported into RAxML [83] (parameters: -f a -m PROTGAMMAAUTO -N 1000) to generate maximum likelihood trees.

### Identification of prokaryotic defence system genes

BLASTp in the DIAMOND programme [84] was used to search defence system-related genes against the PADS Arsenal database [85] with custom settings (more sensitive mode, identity  $\geq 30\%$ , E-value  $< 10^{-10}$ ) to investigate the diversity of the defence systems. Bacterial genes mapped to the PADS database were checked to ensure that the identified genes contained conserved domains involved in the prokaryotic defence against phages using HMMScan in the HMMER 3.3 tool suite [65] against PFAM 32.0 [34] (E-value  $< 10^{-3}$ , bit score  $\geq 30$ ). A set of PFAM accessions for the conserved domains was retrieved from a previous study [50]. Following [86, 87], putative genes for restriction enzymes and methyltransferases were detected by querying the bacterial genes against the REBASE database [88] with BLASTp (E-value  $< 10^{-6}$ , coverage  $\geq 70\%$ ). The identified sequences were also searched against PFAM 32.0 using HMMScan in HMMER 3.3 tool suite (E-value  $< 10^{-3}$ , bit score  $\geq 30$ ). Sequences containing conserved domains of prokaryotic defence against phages were retained and incorporated into the set of RM-related genes. We detected the gene components of a system in a contig sequence or a bacterial bin as described previously to predict the completeness of the defence systems [89–92]. A system was considered complete if it included all the genes required for that system to function.

### Abbreviations

SOB: Sulphur-oxidising bacteria; MOB: Methane-oxidising bacteria; CRISPR: Clustered regularly interspaced short palindromic repeat; TA: Toxin–antitoxin system; RM: Restriction–modification system; DISARM: Defence island system associated with restriction-modification; HGT: Horizontal gene transfer

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-021-01099-6>.

**Additional file 1.** Supplementary Tables.

**Additional file 2.** Supplementary Methods and Figures.

**Additional file 3.** Supplementary Genome sequences of Phages Bin1–4.

### Acknowledgements

We thank Dr. LAN Yi (Hong Kong University of Science and Technology), Ms. WONG Wai Chuen (Hong Kong University of Science and Technology), and Mr. KWAN Yick Hang (Hong Kong University of Science and Technology) for their assistance in this project.

### Authors' contributions

PYQ conceived the project. KZ performed the data analyses and drafted the manuscript. PYQ, RZ, and YX contributed to manuscript editing. All authors read and approved the final manuscript.

### Funding

This work was supported by grants from National Natural Science Foundation of China (91951209), Major Project of Basic and Applied Basic Research of Guangdong Province (2019B030302004-04), Hong Kong Branch of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (SMSEGL20SC01), Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0409), and China Ocean Mineral Resources Research and Development Association (DY135-E2-1).

### Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Ocean Science and Hong Kong Branch of the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Hong Kong University of Science and Technology, Hong Kong, China. <sup>2</sup>Shenzhen University-HKUST Joint Marine Science Ph.D. Program, Shenzhen University, Shenzhen 518060, China. <sup>3</sup>Shenzhen Key Laboratory of Marine Bioresource and Eco-environmental Science, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, China. <sup>4</sup>State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University (Xiang'an), Xiamen, Fujian, China. <sup>5</sup>Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519080, China.

Received: 13 January 2021 Accepted: 20 May 2021

Published online: 03 September 2021

### References

- Felbeck H. Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science*. 1981;213(4505):336–8. <https://doi.org/10.1126/science.213.4505.336>.
- Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, et al. Hydrogen is an energy source for hydrothermal vent symbioses. *Nature*. 2011;476(7359):176–80. <https://doi.org/10.1038/nature10325>.
- Childress JJ, Fisher CR, Brooks JM, Kennicutt MC 2nd, Bidigare R, Anderson AE. A methanotrophic marine molluscan (bivalvia, mytilidae) symbiosis: mussels fueled by gas. *Science*. 1986;233(4770):1306–8. <https://doi.org/10.1126/science.233.4770.1306>.
- Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat Rev Microbiol*. 2019;17(5):271–83. <https://doi.org/10.1038/s41579-019-0160-2>.
- Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB. Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science*. 1981;213(4505):340–2. <https://doi.org/10.1126/science.213.4505.340>.
- Cavanaugh CM, Levering PR, Maki JS, Mitchell R, Lidstrom ME. Symbiosis of methylotrophic bacteria and deep-sea mussels. *Nature*. 1987;325(6102):346–8. <https://doi.org/10.1038/325346a0>.
- Tivey MK. Generation of seafloor hydrothermal vent fluids and associated mineral deposits. *Oceanography*. 2007;20(1):50–65. <https://doi.org/10.5670/oceanog.2007.80>.
- Fisher CR, Takai K, Le Bris N. Hydrothermal vent ecosystems. *Oceanography*. 2007;20(1):14–23. <https://doi.org/10.5670/oceanog.2007.75>.
- Canganella F. Hydrothermal vent ecosystems and representative hyperthermophilic microorganisms. *Ann Microbiol*. 2001;51(1):11–27.
- Takishita K, Takaki Y, Chikaraishi Y, Ikuta T, Ozawa G, Yoshida T, et al. Genomic evidence that methanotrophic endosymbionts likely provide deep-sea *Bathymodiolus* mussels with a sterol intermediate in cholesterol biosynthesis. *Genome Biol Evol*. 2017;9(5):1148–60. <https://doi.org/10.1093/gbe/evx082>.
- Zheng P, Wang MX, Li CL, Sun XQ, Wang XC, Sun Y, et al. Insights into deep-sea adaptations and host-symbiont interactions: a comparative transcriptome study on *Bathymodiolus* mussels and their coastal relatives. *Mol Ecol*. 2017;26(19):5133–48. <https://doi.org/10.1111/mec.14160>.
- Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, et al. Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calymptogena okutanii*. *Curr Biol*. 2007;17(10):881–6. <https://doi.org/10.1016/j.cub.2007.04.039>.
- Fiala-Médioni A, McKiness Z, Dando P, Boulegue J, Mariotti A, Alayse-Danet A, et al. Ultrastructural, biochemical, and immunological characterization of two populations of the mytilid mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge: evidence for a dual symbiosis. *Mar Biol*. 2002;141(6):1035–43.
- Fujii Y, Kubo T, Ishikawa H, Sasaki T. Isolation and characterization of the bacteriophage WO from *Wolbachia*, an arthropod endosymbiont. *Biochem Biophys Res Commun*. 2004;317(4):1183–8. <https://doi.org/10.1016/j.bbrc.2004.03.164>.
- Chauvatcharin N, Ahantari A, Baimai V, Kittayapong P. Bacteriophage WO-B and *Wolbachia* in natural mosquito hosts: infection incidence, transmission mode and relative density. *Mol Ecol*. 2006;15(9):2451–61. <https://doi.org/10.1111/j.1365-294X.2006.02947.x>.
- Sanogo YO, Dobson SL. WO bacteriophage transcription in *Wolbachia*-infected *Culex pipiens*. *Insect Biochem Mol Biol*. 2006;36(1):80–5. <https://doi.org/10.1016/j.ibmb.2005.11.001>.
- Masui S, Kuroiwa H, Sasaki T, Inui M, Kuroiwa T, Ishikawa H. Bacteriophage WO and virus-like particles in *Wolbachia*, an endosymbiont of arthropods. *Biochem Biophys Res Commun*. 2001;283(5):1099–104. <https://doi.org/10.1006/bbrc.2001.4906>.
- Bordenstein SR, Marshall ML, Fry AJ, Kim U, Wernegreen JJ. The tripartite associations between bacteriophage, *Wolbachia*, and arthropods. *PLoS Pathog*. 2006;2(5):384–93.
- Gavotte L, Vavre F, Henri H, Ravallec M, Stouthamer R, Bouletreau M. Diversity, distribution and specificity of WO phage infection in *Wolbachia* of four insect species. *Insect Mol Biol*. 2004;13(2):147–53. <https://doi.org/10.1111/j.0962-1075.2004.00471.x>.
- Kent BN, Bordenstein SR. Phage WO of *Wolbachia*: lambda of the endosymbiont world. *Trends Microbiol*. 2010;18(4):173–81. <https://doi.org/10.1016/j.tim.2009.12.011>.
- Breusing C, Johnson SB, Tunnicliffe V, Clague DA, Vrijenhoek RC, Beinart RA. Allopatric and sympatric drivers of speciation in *Alviniconcha* hydrothermal vent snails. *Mol Biol Evol*. 2020;37(12):3469–3484.
- Breusing C, Mitchell J, Delaney J, Sylva SP, Seewald JS, Gurguis PR, et al. Physiological dynamics of chemosynthetic symbionts in hydrothermal vent snails. *ISME J*. 2020;14(10):2568–79. <https://doi.org/10.1038/s41396-020-0707-2>.
- Chen C, Watanabe HK, Sasaki T. Four new deep-sea provannid snails (Gastropoda: Abyssochrysoidea) discovered from hydrocarbon seep and hydrothermal vents in Japan. *R Soc Open Sci*. 2019;6(7):190393.
- Lan Y, Sun J, Chen C, Sun YA, Zhou YD, Yang Y, et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta aegis*. *Nat Commun*. 2021;12(1):1–8.

25. Sun J, Chen C, Miyamoto N, Li RS, Sigwart JD, Xu T, et al. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nat Commun.* 2020;11(1):1657.
26. Chen C, Linse K, Roterman CN, Copley JT, Rogers AD. A new genus of large hydrothermal vent-endemic gastropod (Neomphalina: Peltospiroidae). *Zool J Linnean Soc.* 2015;175(2):319–35. <https://doi.org/10.1111/zoj.12279>.
27. Tao CH, Li HM, Jin XB, Zhou JP, Wu T, He YH, et al. Seafloor hydrothermal activity and polymetallic sulfide exploration on the southwest Indian ridge. *Chin Sci Bull.* 2014;59(19):2266–76. <https://doi.org/10.1007/s11434-014-0182-0>.
28. Dwarakanath S, Brenzinger S, Gleditsch D, Plagens A, Klingl A, Thormann K, et al. Interference activity of a minimal Type I CRISPR-Cas system from *Shewanella putrefaciens*. *Nucleic Acids Res.* 2015;43(18):8913–23. <https://doi.org/10.1093/nar/gkv882>.
29. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol.* 2015;13(11):722–36. <https://doi.org/10.1038/nrmicro3569>.
30. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ.* 2015;3:e985. <https://doi.org/10.7717/peerj.985>.
31. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 2016;44(W1):W16–21. <https://doi.org/10.1093/nar/gkw387>.
32. Song W, Sun HX, Zhang C, Cheng L, Peng Y, Deng Z, et al. Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* 2019;47(W1):W74–80. <https://doi.org/10.1093/nar/gkz380>.
33. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpidides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol.* 2020;39(5):578–85.
34. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47(D1):D427–D32. <https://doi.org/10.1093/nar/gky995>.
35. Wernegreen JJ. Endosymbiosis. *Curr Biol.* 2012;22(14):R555–61. <https://doi.org/10.1016/j.cub.2012.06.010>.
36. Bright M, Bulgheresi S. A complex journey: transmission of microbial symbionts. *Nat Rev Microbiol.* 2010;8(3):218–30. <https://doi.org/10.1038/nrmicro2262>.
37. Abad FXPR, Gajardo R, Bosch A. Viruses in mussels: public health implications and depuration. *J Food Prot.* 1997;60(6):677–81. <https://doi.org/10.4315/0362-028X-60.6.677>.
38. Hadas E, Marie D, Shpigiel M, Ilan M. Virus predation by sponges is a new nutrient-flow pathway in coral reef food webs. *Limnol Oceanogr.* 2006;51(3):1548–50. <https://doi.org/10.4319/lo.2006.51.3.1548>.
39. Bodner K, Melkonian AL, Covert MW. The enemy of my enemy: new insights regarding bacteriophage-mammalian cell interactions. *Trends Microbiol.* 2020;29(6):528–41.
40. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, et al. Lytic to temperate switching of viral communities. *Nature.* 2016;531(7595):466–70. <https://doi.org/10.1038/nature17193>.
41. Grasis JA. The intra-dependence of viruses and the holobiont. *Front Immunol.* 2017;8:1501. <https://doi.org/10.3389/fimmu.2017.01501>.
42. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* 2017;11(7):1511–20. <https://doi.org/10.1038/ismej.2017.16>.
43. Röstl JT, Marraffini L. (Ph) ighting phages: how bacteria resist their parasites. *Cell Host Microbe.* 2019;25(2):184–94. <https://doi.org/10.1016/j.chom.2019.01.009>.
44. Ershova AS, Rusinov IS, Spirin SA, Karyagina AS, Alexeevskii AV. Role of restriction-modification systems in prokaryotic evolution and ecology. *Biochemistry (Mosc).* 2015;80(10):1373–86. <https://doi.org/10.1134/S0006297915100193>.
45. Wang LR, Chen S, Vergin KL, Giovannoni SJ, Chan SW, DeMott MS, et al. DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proc Natl Acad Sci U S A.* 2011;108(7):2963–8. <https://doi.org/10.1073/pnas.1017261108>.
46. Ofir G, Melamed S, Sberro H, Mukamel Z, Silverman S, Yaakov G, et al. DISA RM is a widespread bacterial defence system with broad anti-phage activities. *Nat Microbiol.* 2018;3(1):90–8.
47. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, et al. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* 2015;34(2):169–83. <https://doi.org/10.15252/embj.2014.89455>.
48. Yamaguchi Y, Park JH, Inouye M. Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet.* 2011;45(45):61–79.
49. Chopin MC, Chopin A, Bidnenko E. Phage abortive infection in lactococci: variations on a theme. *Curr Opin Microbiol.* 2005;8(4):473–9. <https://doi.org/10.1016/j.mib.2005.06.006>.
50. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic discovery of anti-phage defense systems in the microbial pangenome. *Science.* 2018;359(6379):eaar4120.
51. Dupuis ME, Villion M, Magadan AH, Moineau S. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun.* 2013;4(1). <https://doi.org/10.1038/ncomms3087>.
52. Heaton BE, Herrou J, Blackwell AE, Wysocki VH, Crosson S. Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of *Brucella abortus*. *J Biol Chem.* 2012;287(15):12098–110. <https://doi.org/10.1074/jbc.M111.332163>.
53. Samson JE, Magadan AH, Sabri M, Moineau S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol.* 2013;11(10):675–87. <https://doi.org/10.1038/nrmicro3096>.
54. Rao VB, Feiss M. The bacteriophage DNA packaging motor. *Annu Rev Genet.* 2008;42(42):647–81.
55. Hilbert BJ, Hayes JA, Stone NP, Duffy CM, Sankaran B, Kelch BA. Structure and mechanism of the ATPase that powers viral genome packaging. *Proc Natl Acad Sci U S A.* 2015;112(29):E3792–E9. <https://doi.org/10.1073/pnas.1506951112>.
56. Moreno-Vivian C, Cabello P, Martinez-Luque M, Blasco R, Castillo F. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol.* 1999;181(21):6573–84. <https://doi.org/10.1128/JB.181.21.6573-6584.1999>.
57. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
58. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455–77. <https://doi.org/10.1089/cmb.2012.0021>.
59. Albertsen M, Hugenholz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol.* 2013;31(6):533–8. <https://doi.org/10.1038/nbt.2579>.
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
62. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
63. Miller IJ, Rees ER, Ross J, Miller I, Baxa J, Lopera J, et al. Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res.* 2019;47(10):e57–e.
64. Parks DH, Imelfort M, Skennerton CT, Hugenholz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25(7):1043–55. <https://doi.org/10.1101/gr.186072.114>.
65. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013;41(12):e121. <https://doi.org/10.1093/nar/gkt263>.
66. Gregory AC, Zayed AA, Conceicao-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell.* 2019;177(5):1109–23. <https://doi.org/10.1016/j.cell.2019.03.040>.
67. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol.* 2020;18(3):125–38. <https://doi.org/10.1038/s41579-019-0311-5>.
68. Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol.* 2019;37(6):632–9. <https://doi.org/10.1038/s41587-019-0100-8>.
69. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303>.

70. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 2010;11(1):119. <https://doi.org/10.1186/1471-2105-11-119>.
71. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
72. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics.* 2020;36(7):2251–2. <https://doi.org/10.1093/bioinformatics/btz859>.
73. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol.* 2011;7(10):e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
74. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature.* 2016;537(7622):689–93. <https://doi.org/10.1038/nature19366>.
75. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–U130. <https://doi.org/10.1038/nbt.1883>.
76. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14(4):417–9. <https://doi.org/10.1038/nmeth.4197>.
77. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* 1999;20(18):3551–67. [https://doi.org/10.1002/\(SICI\)1522-2683\(19991201\)20:18<3551::AID-ELPS3551>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2).
78. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, et al. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere.* 2017;2(2):e00359–16.
79. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinform.* 2007;8(1):209. <https://doi.org/10.1186/1471-2105-8-209>.
80. Rho M, Wu YW, Tang H, Doak TG, Ye Y. Diverse CRISPRs evolving in human microbiomes. *PLoS Genet.* 2012;8(6):e1002441. <https://doi.org/10.1371/journal.pgen.1002441>.
81. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000;16(6):276–7. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2).
82. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–66. <https://doi.org/10.1093/nar/gkf436>.
83. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
84. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
85. Zhang YD, Zhang ZW, Zhang H, Zhao YB, Zhang ZC, Xiao JF. PADS Arsenal: a database of prokaryotic defense systems related genes. *Nucleic Acids Res.* 2020;48(D1):D590–D8. <https://doi.org/10.1093/nar/gkz916>.
86. Fullmer MS, Ouellette M, Louyakis AS, Papke RT, Gogarten JP. The patchy distribution of restriction-modification system genes and the conservation of orphan methyltransferases in halobacteria. *Genes.* 2019;10(3):233. <https://doi.org/10.3390/genes10030233>.
87. Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Forster F, et al. An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Front Microbiol.* 2016;7:1751.
88. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 2015;43(D1):D298–9. <https://doi.org/10.1093/nar/gku1046>.
89. Bernheim A, Sorek R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol.* 2019;18:113–9.
90. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol.* 2017;15(3):169–82. <https://doi.org/10.1038/nrmicro.2016.184>.
91. Kamruzzaman M, Iredell J. A ParDE-family toxin antitoxin system in major resistance plasmids of Enterobacteriaceae confers antibiotic and heat tolerance. *Sci Rep.* 2019;9(1):9872. <https://doi.org/10.1038/s41598-019-46318-1>.
92. Dy RL, Przybicki R, Semeijn K, Salmond GPC, Fineran PC. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res.* 2014;42(7):4590–605. <https://doi.org/10.1093/nar/gkt1419>.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

