

RESEARCH

Open Access



Insights into diversity, host-range, and temporal stability of *Bacteroides* and *Phocaeicola* prophages

Nejc Stopnisek^{1,2*}, Stina Hedžet¹, Tomaž Accetto³ and Maja Rupnik^{1,2}

Abstract

Background Phages are critical components of the gut microbiome, influencing bacterial composition and function as predators, parasites, and modulators of bacterial phenotypes. Prophages, integrated forms of these phages, are prevalent in many bacterial genomes and play a role in bacterial adaptation and evolution. However, the diversity and stability of prophages within gut commensals, particularly in the genera *Bacteroides* and *Phocaeicola*, remain underexplored. This study aims to screen and characterize prophages in these genera, providing insights into their diversity, host range, and temporal dynamics in the human gut.

Results Using a combination of three bioinformatic tools—Cenote-Taker 3, Vibrant, and PHASTER—we conducted a comprehensive analysis of prophages in *Bacteroides* and *Phocaeicola*. Cenote-Taker 3 identified the most diverse set of prophages, with significant overlaps observed between the tools. After clustering high-quality prophages, we identified 22 unique viral operational taxonomic units (vOTUs). Notably, comparisons between prophages identified in isolated bacterial genomes, metaviromes, and large public gut virome databases revealed a broader host range than initially observed in single isolates. Certain prophages were consistent across time points and individuals, suggesting temporal stability. All identified prophages belonged to the *Caudoviricetes* class and contained genes related to antibiotic resistance, toxin production, and metabolic processes.

Conclusions The combined use of multiple prophage detection tools allowed for a more comprehensive assessment of prophage diversity in *Bacteroides* and *Phocaeicola*. The identified prophages were not only prevalent but also exhibited broad host ranges and temporal stability. The presence of antibiotic resistance and toxin genes suggests that these prophages may significantly influence bacterial community structure and function in the gut, with potential implications for human health. These findings highlight the importance of using diverse detection tools to accurately assess prophage diversity and dynamics.

Keywords Prophage detection, *Bacteroides*, *Phocaeicola*, Gut microbiome, Prophage diversity, Caudoviricetes, Antibiotic resistance, Phage-bacteria interactions, Temporal stability

*Correspondence:

Nejc Stopnisek
nejc.stopnisek@nlzoh.si

¹Department for Microbiological Research, Centre for Medical Microbiology, National Laboratory of Health, Environment and Food, Prvomajska 1, Maribor 2000, Slovenia

²Department of Microbiology, Faculty of Medicine, University of Maribor, Taborska ulica 8, Maribor 2000, Slovenia

³Biotechnical Faculty, University of Ljubljana, Groblje 3, Domzale 1230, Slovenia



© 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/>.

Introduction

Prophages dominate gut viromes and influence composition and function of gut bacteria as predators, parasites and modulators of bacterial phenotypes [1]. Gut virome of healthy individuals is stable, but individual specific and alterations in the gut virome impact human health [2].

Phages exhibit various life styles, from lytic to chronic infection, carrier state and lysogenic [3]. Prophages were found in 75% of more than 13,500 bacterial genomes and are particularly enriched in *Enterobacteriaceae* [4]. Dependent on environmental signals, phages can initiate lytic or lysogenic cycle after infection. Recent studies demonstrated frequent signal-dependent prophage induction in various environments [5–7]. For instance, experiments in mouse models have shown that environmental stressors can trigger prophage induction, leading to increased bacterial lysis and horizontal gene transfer [8, 9]. Similarly, human gut studies have reported active prophage induction, contributing to microbial community dynamics and potentially impacting host health [10–12]. The connection of prophage induction and associated host diseases, such as IBD, Crohn's disease and Ulcerative colitis has also been suggested [13].

Free and integrated phages are in dynamic equilibrium and therefore the insight into prophages is of importance either on metavirome level or on level of individual isolated bacterial hosts. Several bioinformatics tools are available to detect prophages in bacterial genomes [14–18], however, these predictions still lack accuracy. Individual bacterial species of human gut commensals are understudied in terms of their prophage diversity and prevalence. Temperate phages of *Faecalibacterium prausnitzii* [19], and globally present Hankyphage, propagated from *Bacteroides dorei* were characterized and found to carry diversity generating retroelements that, through reverse transcription, introduce changes, in variable regions of genes that are often involved in host recognition [20]. Some of new *Bacteroides* phages described by our group were also detected to be integrated in bacterial genomes [21]. The recent metagenomic study of infant and parent gut metagenome samples has provided extensive insight into prophages of intestinal bacteria [11].

In this study, we aimed to screen and characterize the prophages in cultivated strains of two predominant bacterial genera in the gut microbiota, *Bacteroides* and *Phocaeicola*. In populations with a western lifestyle, *Bacteroides* represent the most abundant genus in *Bacteroidota* phylum. Species of *Bacteroides* benefit to its human host, since they take part in degradation of complex polysaccharides, such as starch and xylan, produce important short-chain fatty acids (SCFA) and contribute to colonization resistance [22]. Recently some species of *Bacteroides* genus have been transferred to newly established genus *Phocaeicola* [23]. In addition to cataloguing

the prophages and their genes we also provide the insight on the stability of prophages in bacterial genomes using phage-bacterial host pairs isolated from two healthy individuals at two different time points.

Methods

Isolation of bacterial strains and genome recovery

Strains were selected from bacterial strain collection done by our group as described in Hedzet et al. [21]. In short, fecal samples, collected from two healthy donors (D1 and D2) at two temporally distant time points (S1 and S2) per each donor were stored at -80°C . For D1 the time interval between sampling was 25 Months while for D2 it was 4 Months. The complete isolation of bacterial strains and preparation of fecal suspension was carried out in an anaerobic workstation at 37°C . Homogenized fecal suspensions (20%) were plated on YBHI (brain–heart infusion media, supplemented with yeast extract (0.5%) and rumen fluid (20%)). After 72 h of incubation, approximately 100 randomly chosen colonies per sample were subcultured on YBHI plates. Identification of pure cultures was conducted by mass spectrometry (MALDI-TOF Biotyper System, Bruker Daltonik, Bremen, Germany) and by sequencing nearly complete 16 S rRNA gene. Strains identified as members of *Bacteroidaceae* ($n=74$) were selected for further whole genome sequencing (WGS).

Strains included in this study belong to genus *Bacteroides* ($n=54$), including *B. ovatus* ($n=22$), *B. uniformis* ($n=14$), *B. thetaiotaomicron* ($n=14$), *B. faecis* ($n=4$), *B. xylanisolvens* ($n=3$), and *B. stercoris* ($n=1$). Twenty isolates were from the genus *Phocaeicola*, including *P. vulgatus* ($n=18$) and *P. massiliensis* ($n=2$).

Genome sequencing

After bacterial DNA isolation with QIAamp DNA Mini Kit (Qiagen), paired-end libraries were generated using the Nextera XT Library preparation kit (Illumina) and sequenced on MiSeq (Illumina) with 600-cycle MiSeq ReagentKit v3 (2×300 bp). The quality of the raw sequencing reads was examined by FastQC tool Version 0.11.9 (Babraham Bioinformatics) [24]. Quality trimming was done by Trimmomatic Version 0.39 (USADLLAB.org) [25] with the following parameters: ILLUMINACLIP:/home/storage/DB/NexteraPE-PE.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:40. Overlapping paired-end reads were merged using FLASH software, version 1.2.11 [26]. Assembly was performed by SPAdes Assembler (meta-spades for metaviromes), version 3.14.0 [27], and the assemblies were examined using Quast version 4.0 [28]. Genome assembly statistics is provided in the Supplemental Table S1.

To identify relatedness between the obtained genomes we calculated the average nucleotide identity (ANI) using the FastANI v1.34 [29].

Prophage analysis

To identify the most diverse collection of prophages from the obtained genomes we used PHASTER [14], Vibrant v1.2.1 [16] and Cenote-Taker v3.3.0 [18]. PHASTER was accessed as a web tool in October 2021. The quality of identified prophage regions was checked using checkV [30], and only those predicted to be complete or high-quality were kept for further analyses and termed “high-quality” prophages.

Since identified high-quality prophages are from different tools, we wanted to understand how many of them are unique. For that we clustered all sequences based on 100% similarity using MMseqs2 [31, 32]. Next, we clustered prophages into viral operational taxonomic units (vOTUs) using the MIUViG [33] recommended criteria of 95% nucleotide similarity using MMseqs2 [31, 32] with the following options: `--min-seq-id 0.95 -c 0.8 --cov-mode 0`. MMseqs2 was also used to search for the presence of obtained prophages within large viral databases. The three viral databases we used were the Metagenomic Gut Virus (MGV) catalogue with 189,680 viral genomes [34], Gut Virome Database (GVD) with 33,242 unique viral populations [35], and the Gut Phage Database (GPD) with 142,809 non-redundant gut phage genomes [36]. We created MMseqs2 databases from our own prophage sequences and sequences found in these three viral databases using the command `mmseqs createdb`. The search was performed with the `mmseqs search` command using the following parameters: `--max-seqs 4 --min-seq-id 0.95 --e-profile 1.0E-06 --start-sens 1 --sens-steps 3 -s 7 --threads 24 --search-type 3`.

To assign viral taxonomy to vOTUs, we used BLAST [37], vConTACT3 [38] and PhageScope [39]. This task was performed solely for representative sequences from each prophage cluster. Initially, we assigned taxonomy using vConTACT3 with the Database v220 [40] under default settings. 13 vOTUs were successfully classified at the class level. For 9 vOTUs that vConTACT3 could not classify, we employed BLAST, considering results with an E-value greater than $1e-10$ and matching at least 60% of the query sequence. Additionally, all sequences were processed through the PhageScope online tool using the Taxonomic Annotation option. The network results from vConTACT3 were imported into Cytoscape v3.10.2 [41] for visualization purposes.

While the phage discovery tools we used provide annotations of predicted prophage regions, they rely on different databases, leading to potential discrepancies in annotations. Therefore, we annotate all prophages with vDRAM [42].

From the same donors, we obtained fecal water metaviromes [21] to determine if prophages are potentially lytic and prevalent in these samples. For read mapping, we used coverM 0.6.1 [43] with the following parameters: `-p minimap2-sr --methods tpm -t 42 --output-format dense`.

Data availability

Assembled bacterial genomes were retrieved from the NCBI under the accession numbers PRJNA843113, PRJNA636979. Code for read processing, prophage discovery and graphical outputs are available at: https://github.com/NLZOHomr/prophages_2024.

Results

Comparison of three tools for prophage detection

Altogether 597 prophages were identified in 47 out of 74 bacterial genomes, predominantly in the genus *Bacteroides* (Fig. 1A). Cenote-Taker 3 identified the most prophages ($n=232$), followed by Vibrant ($n=208$) and PHASTER ($n=157$) (Fig. 1A, B). The quality of the identified prophages varied as well. While Cenote-Taker 3 does not provide exact quality statuses, PHASTER and Vibrant do, showing that only a small number are high-quality genomes (Fig. 1A). Comparing the identified prophages, we found that 21% ($n=70$) were detected by all three tools. Interestingly, while Vibrant and PHASTER had little overlap, Cenote-Taker 3 identified a relatively high number of prophages that overlapped with both: 67 shared with Vibrant and 25 with PHASTER (Fig. 1B). Furthermore, even after filtering for high-quality or complete prophages using the CheckV tool [30] the reduction in the number of prophages was significant for all three tools (Fig. 1C). The overlap of detected phages among three tools did not change significantly.

Among the tools used, Vibrant identified the most unique prophages, while Cenote-Taker 3 identified the most diverse set, covering those identified by both Vibrant and PHASTER. This makes Cenote-Taker 3 a promising sole tool for phage discovery.

Abundance and diversity of prophages in *Bacteroides* and *Phocaeicola* strains

To identify how many unique prophages were discovered by these tools, we clustered them based on nucleotide similarity and length using MMseqs2. Out of 125 high-quality prophages, we found only 37 unique sequences at 100% similarity. At 95% similarity, 22 clusters were formed, which we refer to as viral operational taxonomic units (vOTUs) (Table 1). vOTU 4 was the only cluster containing prophages identified by all three tools (Table 1). The size of the identified prophages ranged from 27,455 bp (vOTU11) to 124,118 bp (vOTU19), encoding between 32 and 164 genes.

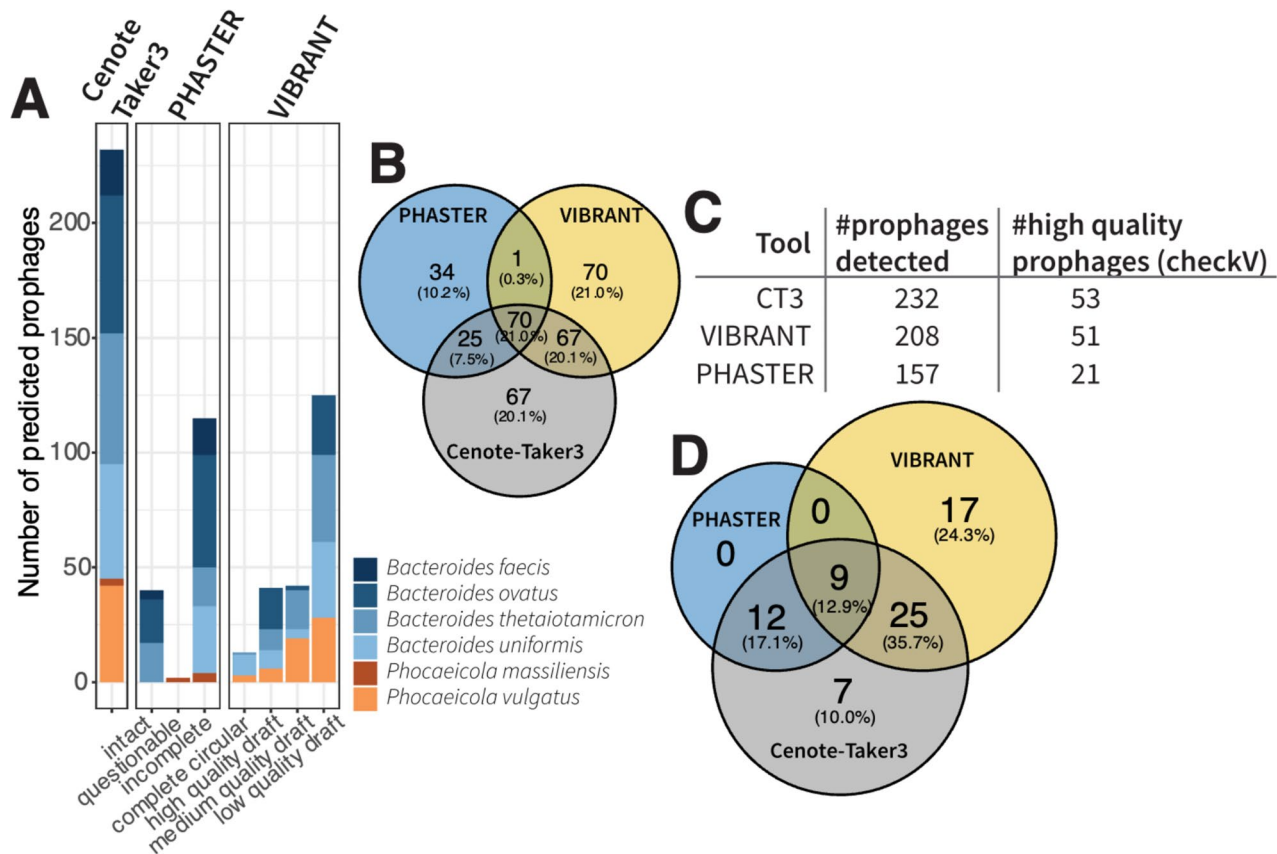


Fig. 1 Prophage discovery. The bar plot represents number of prophages identified by the three respected tools used in the study, Cenote Taker 3, Phaster and Vibrant. These are grouped by quality as predicted by each tool. Bars are colored by the bacterial species that they were found in (A). Venn diagram highlights the number of shared and unique prophages (B). Table shows the total number of identified prophages and number after filtering for only high-quality prophages as predicted by CheckV [30] (C). Corresponding Venn diagram shows the unique and shared high-quality prophages only (D). To identify shared and unique prophages we looked at the contigs where these tools predicted prophages and if these were identical we counted as shared otherwise it was unique

Per genome, we identified between 1 and 5 high-quality vOTUs, with the highest number of prophages found in two strains of *P. vulgatus* (MB20-93 and MB20-96) (Fig. 2). While bacterial genomes from a particular species show high ANI (Fig. 2A, Suppl Fig. S1), the number of high-quality prophages can vary significantly. This variation may be due to unequal integration of prophages within the population or missing genome information, as many genomes are in draft stages (56–392 contigs per genome). vOTUs 4, 10, and 16 were found in the highest number of bacterial genomes ($n=9$) (Fig. 2B). Furthermore, vOTU10 was the only vOTU identified in two different genera, *B. uniformis* and *P. vulgatus*, while all other vOTUs were highly species-specific and probably also strain-specific (Fig. 2).

Taxonomic classification revealed that all identified prophage clusters belong to the highly diverse tailed bacteriophage class *Caudoviricetes*. vConTACT3 predicted 13 vOTUs to this class and 9 were unassigned. However, using BLAST and PhageScope we were able to predict taxonomy also of the remaining 9 vOTUs which all

belong to *Caudoviricetes*. Based on vConTACT3 vOTUs make up three larger clusters, comprising of 10, 5 and 2 vOTUs, respectively and 4 vOTUs which do not have any connectivity (Fig. S2).

Prophage temporal dynamics in the gut environment of our donors

Our study allowed us to investigate not only the host range of these prophages but also their dynamics over two time periods. We found that vOTUs seemed host-specific and also highly donor-specific (Fig. 3). Only vOTU10 was found in both donors, while all others were unique to a single donor. Specifically, vOTU10 was shared between two genera only in donor 2 at sampling time 1 (Fig. 3). Additionally, when comparing the occurrences of vOTUs over time, we observed that vOTUs are temporally stable (Fig. 3). Except for vOTU12, any vOTU present at sampling time 1 (S1) also appeared at sampling time 2 (S2) within the same donor (Fig. 3). However, prophages found at sampling time 2 were not always present

Table 1 Summary of the vOTUs. Prophages that were clustered at 95% similarity formed 22 distinct vOTUs. Prophages that are within these vOTUs were identified by either one, two or once by all three tools (V– vibrant, C– cenote taker 3, P– phaster). Length of prophage genomes within each cluster and the number ORFs are also represented. Information in the brackets represents minimum and maximum values within each cluster. Taxonomy column contains viral class as predicted by either vContACT3, BLAST or PhageScope

vOTU	Tools	Length (bp)	ORF	Taxonomy
vOTU1	V, C	42160.88 (37638–48793)	51.5 (45–63)	Caudoviricetes
vOTU2	C, P	42759.29 (42,608–42961)	45	Caudoviricetes
vOTU3	C	35,454	38	Caudoviricetes
vOTU4	V, C, P	50,266 (49032–51774)	63.3 (62–65)	Caudoviricetes
vOTU5	C	29,500	38	Caudoviricetes
vOTU6	C	61,508 (59309–63707)	74 (72–76)	Caudoviricetes
vOTU7	C	43,600	54	Caudoviricetes
vOTU8	C	63,319	82	Caudoviricetes
vOTU9	V, C	73,695 (71863–75527)	109 (108–110)	Caudoviricetes
vOTU10	V, C	43789.11 (40315–45709)	46.9 (41–49)	Caudoviricetes
vOTU11	P	27,455	32	Caudoviricetes
vOTU12	P	33,914	36	Caudoviricetes
vOTU13	V, C	75645.25 (71365–81021)	109 (103–116)	Caudoviricetes
vOTU14	V, C	48,011 (44905–51117)	56 (52–60)	Caudoviricetes
vOTU15	V, C	52719.75 (50763–53372)	66.8 (66–67)	Caudoviricetes
vOTU16	V	81,403	116	Caudoviricetes
vOTU17	V	63,326 (63296–63,356)	72	Caudoviricetes
vOTU18	V	62,533	72	Caudoviricetes
vOTU19	V	120873.50 (111182–124118)	161.3 (153–164)	Caudoviricetes
vOTU20	V, C	88,336	110	Caudoviricetes
vOTU21	C	113,780 (112145–115415)	155 (154–156)	Caudoviricetes
vOTU22	V, C	69,564 (67265–71863)	100 (93–107)	Caudoviricetes

at sampling time 1, likely due to uneven bacterial culturing across different sampling periods.

Prophage host range at the level of bacterial isolates, metaviromes and global gut databases

Previously [21], we obtained three metavirome samples from the same individuals, which we used for mapping onto the reference vOTUs. Reads from all metaviromes mapped to the vOTUs, indicating that these prophages are not only lysogenic but can also exhibit a lytic cycle. Surprisingly, these prophages can become relatively abundant, as suggested by their high mapping percentages. The sample from individual 1 (D1_FW2) had the highest mapping percentage, accounting for 6% of the

total metavirome, while reads from the individual 2 (D2) mapped at 1.35% and 1.17%, for time points 1 and 2, respectively. Interestingly, reads mapped to 20 out of 22 prophage vOTUs, mainly across both donors (Fig. 4), suggesting that these prophages are prevalent in the human gut of these donors.

These results raised the question of whether vOTUs are as host- and donor-specific as our bacterial genome analysis suggests. To explore this, we examined three publicly available gut virome databases. All vOTUs identified in the present study were found in all three global gut virome databases (Supplemental Table S2). While the genus *Bacteroides* is predicted to be the sole host for all our vOTUs in the MGv database, the GPD and GVD databases predict additional hosts. For instance, vOTU5 was previously found in members of the family *Tannerellaceae*, vOTU10 in *Bacteroides dorei* and the family *Enterococcaceae*, and vOTU19 in *Bacteroides faecis*. Moreover, the GPD database includes geographical locations of the best-matched hits, showing that the studies where our prophages were discovered are spread across the globe (e.g., China, United States, Denmark, Germany, Australia, Russia, etc.) (Supplemental Table S2).

Prophage genetic content

To understand the genetic content of the identified prophages, we annotated reference vOTU genomes from 22 clusters using the vDRAM tool [42]. The gene catalogue comprises 1,659 open reading frames (ORFs), with approximately 43% of these ORFs successfully annotated (see Supplemental Table S3).

Fourteen vOTUs were predicted to carry auxiliary metabolic genes (AMGs) involved in various metabolic processes (Supplemental Table S3). Among the identified AMGs, ketopantoate hydroxymethyltransferase, an enzyme essential for vitamin B5 (pantothenate) biosynthesis, was detected. Several sugar transferases, such as glycosyl transferase, were found, which could enhance carbon metabolism and cell wall biosynthesis in the host bacteria. Notably, vOTU21 encodes a potential pullulanase, an enzyme involved in the breakdown of starches and polysaccharides, while vOTU1 encodes alpha-L-rhamnosidase, which is involved in the degradation of plant materials and microbial polysaccharides.

Several toxin genes were identified in many vOTUs, indicating potential roles in bacterial competition and survival (Supplemental Table S3). Thiol-activated cytolysin [44] was identified in vOTU8, while toxin SymE [45], part of a type I toxin-antitoxin system, was found in vOTU14. Additionally, a putative AbiEii toxin from a type IV toxin-antitoxin system [46] was present in vOTU14, and the antitoxin component MqsA was found in vOTUs 8 and 15. A type II toxin-antitoxin system HigB family

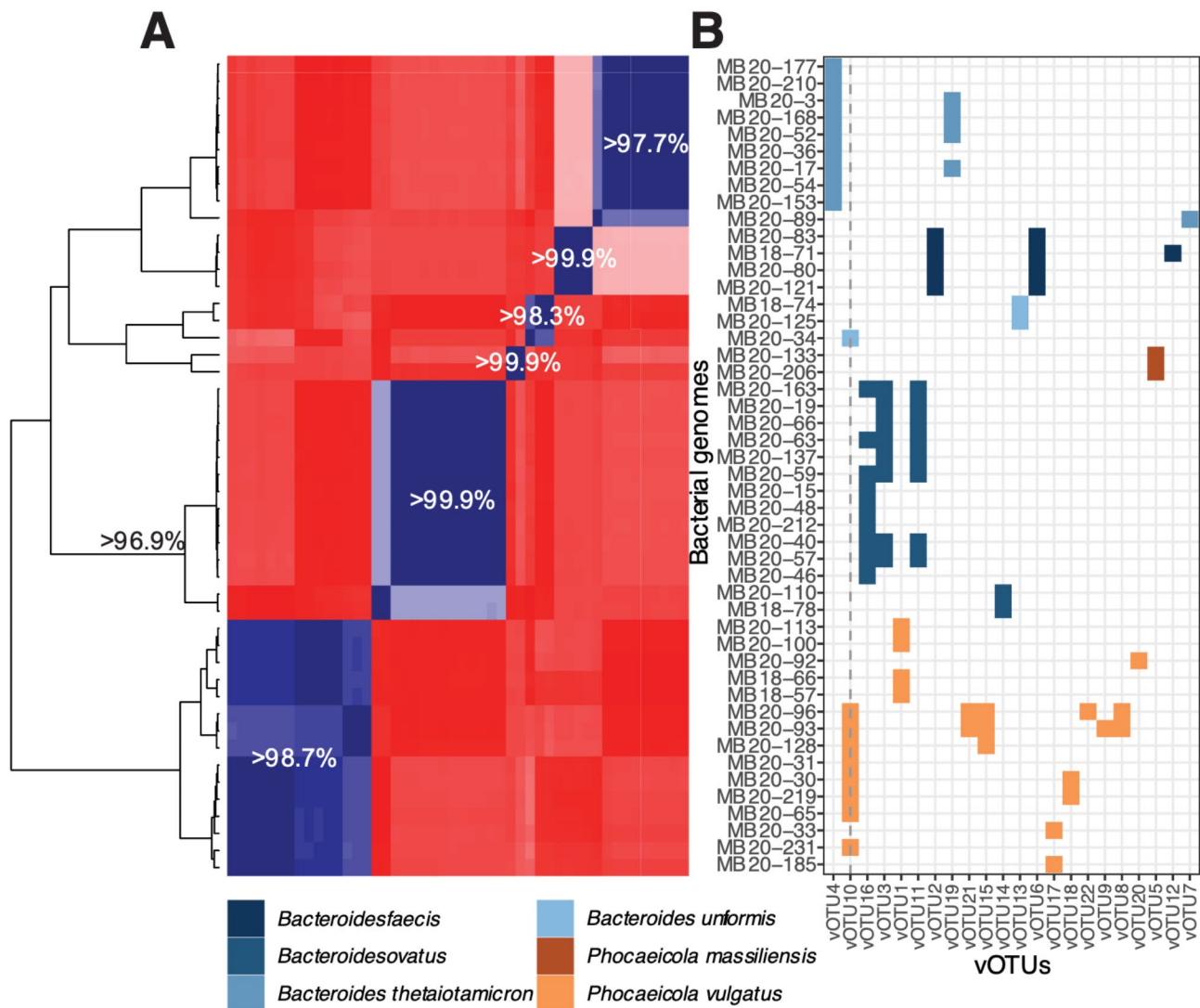


Fig. 2 vOTU host-range. Bacterial genomes where high-quality genomes were identified are represented in the ANI heatmap. The minimum ANI scores for each block are highlighted on the heatmap and are above 96.9%. Corresponding bacterial IDs are marked on right side of the heatmap. vOTUs per bacterial genome are represented as dot plot where colors represent bacterial species where vOTUs were found. vOTU10 is highlighted by break line as the only vOTU which is found in two different host species

toxin, with an associated antitoxin HigA, was identified in vOTUs 16 and 21 [47].

Furthermore, several vOTUs contained genes associated with antibiotic resistance (Supplemental Table S3). Chloramphenicol O-acetyltransferase type A, which detoxifies chloramphenicol and confers resistance, was identified in vOTUs 2 and 3. Multidrug resistance proteins of the MATE family were found in vOTUs 2 and 12. Beta-lactam resistance genes, including those encoding beta-lactamase superfamily domains, were found in

vOTUs 4, 8, and 15, while L, D-transpeptidase catalytic domains were present in vOTU16.

Discussion

In this study, we investigated the diversity, host range, and temporal dynamics of prophages within the *Bacteroidaceae* family using a combination of culture-based methods and comprehensive genomic analyses.

The use of multiple bioinformatic tools - PHASTER, Vibrant, and Cenote-Taker 3 - proved crucial in capturing the full spectrum of prophage diversity. Each tool

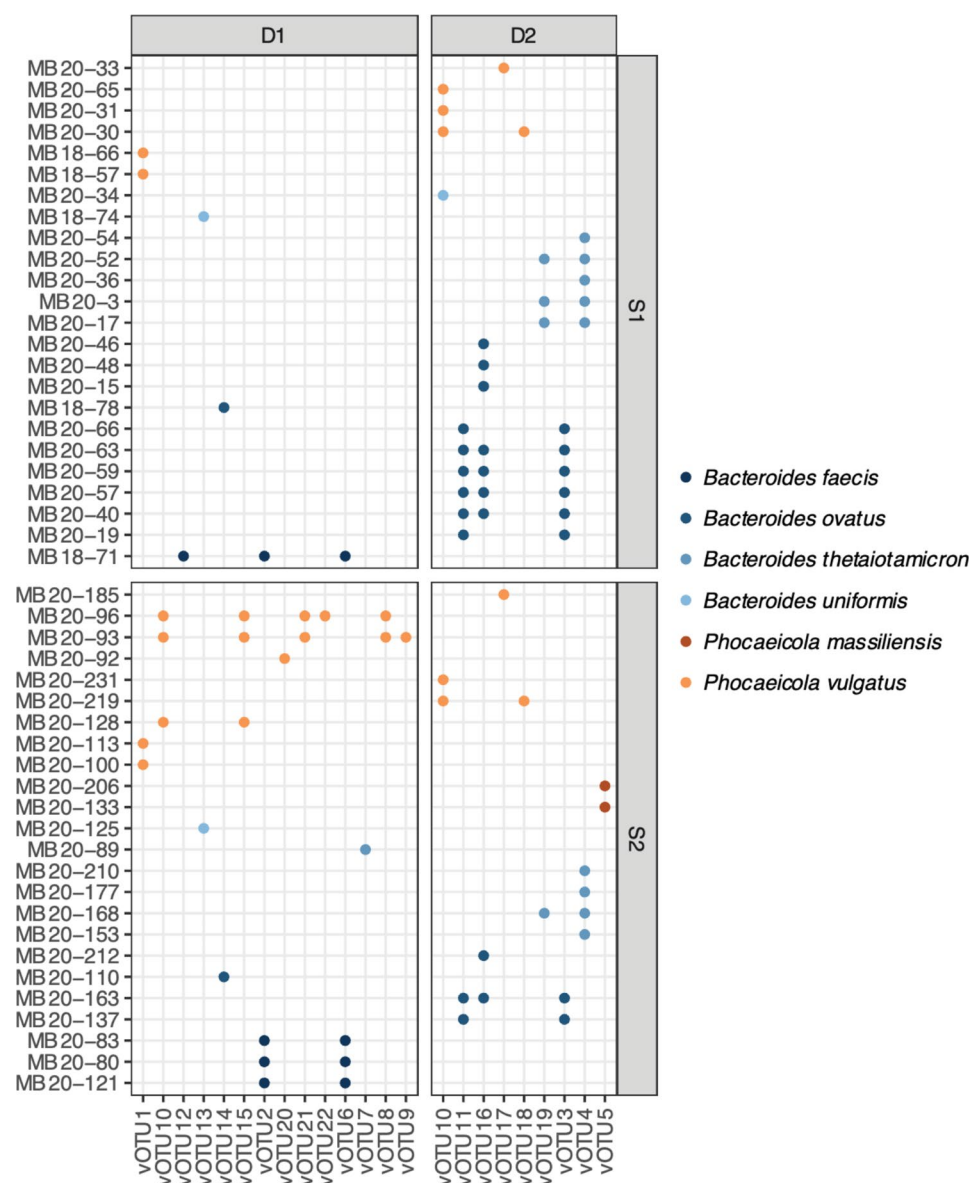


Fig. 3 vOTU over donor and time. The dot plot represents the occurrence of vOTUs in donor 1 (D1) or donor 2 (D2) at time 1 (S1) or time 2 (S2). Each dot is color coded by bacterial host

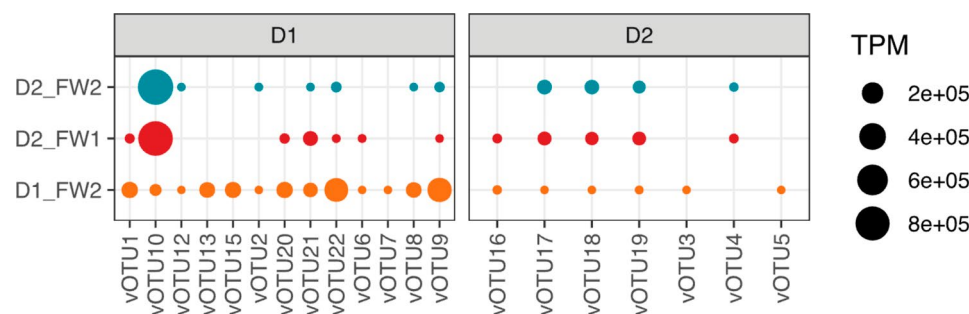


Fig. 4 Mapping summary of metavirome samples from fecal water (FW). Short reads were mapped to a database of 37 unique high-quality prophages using coverM [43] and coverage was expressed as transcripts per million (TPM), represented as dot sizes. Each dot corresponds to a prophage and is color-coded based on the metavirome sample: orange represents donor 1 fecal water at time point 2 (D1-FW2), while red and blue represent donor 2 fecal water at time points 1 (D2-FW1) and 2 (D2-FW2), respectively

employs different algorithms and reference databases, resulting in varied prophage predictions [16]. The complementary nature of these tools allowed us to identify a broader range of prophages than any single tool could achieve alone. This approach aligns with findings from other studies, which emphasize the necessity of using diverse methodologies to obtain a comprehensive view of viral diversity [10, 48, 49].

Culture-based methods, while invaluable, often present a skewed view of microbial diversity due to their selective nature. This study mitigated such biases by integrating metavirome data and leveraging large viral databases, providing a holistic perspective on prophage diversity and host range. Our initial findings from cultured bacterial strains indicated a narrow host specificity for many prophages, confined to specific species or even strains. However, metavirome analysis and database comparisons broadened this perspective, revealing that these prophages also infect a wider array of bacterial hosts. This integrative approach is crucial for understanding the true ecological roles of prophages.

All identified prophages belong to the viral class *Caudoviricetes*, a diverse group of tailed bacterial and archaeal viruses characterized by their distinctive head-tail morphology and double-stranded DNA genome. This class encompasses a wide range of bacteriophage families which differ in their tail structures and infection mechanisms but all highly prevalent in the gut microbiome [50–52]. *Caudoviricetes* play a crucial role in horizontal gene transfer, which can significantly influence bacterial evolution and adaptation [53]. The presence of *Caudoviricetes* prophages in the gut has been linked to various beneficial and detrimental effects on the human host, including modulation of immune responses and contribution to gut dysbiosis [54, 55]. Our findings reinforce the growing evidence that *Caudoviricetes* prophages are integral components of the gut microbiome, capable of shaping bacterial community structure and function.

In terms of host range and dynamics, vOTU10 stood out among the clusters examined, being consistently present in both donors across both time points. Our culture data initially identified vOTU10 in *B. uniformis* and *P. vulgatus*. Subsequent analysis using viral databases expanded its host range to include *B. dorei* and *Enterococcaceae*. BLAST results indicated a close match (99.10% identity and 94% query coverage) with members of the *Bacteroides* phage Hanky p00 (Hankyphage) group, known as temperate phages of *B. dorei* capable of lysing various *Bacteroides* species and prevalent in metaviromes [56]. Importantly, this prophage, along with vOTU1 and vOTU20, was previously identified in the same samples through meticulous manual curation of PHASTER predictions, despite these prophages not being initially predicted as high quality [21, 57]. Our integrated analysis

approach now allows us to efficiently resolve these prophages using high-throughput bioinformatics tools, providing confidence in our findings and demonstrating that we can reliably detect the same prophages without the need for labor-intensive manual curation. These findings underscore the effectiveness of combining multiple phage discovery tools to ensure comprehensive and accurate phage identification.

One of the important findings in our study is the presence of several antibiotic resistance genes (AMR) and genes that could potentially enhance and contribute to host metabolism within the prophages. This is consistent with previous research, which has documented the presence of AMR genes and auxiliary metabolic genes in prophages, emphasizing their role in bacterial adaptability and survival [58–61]. Our identification of AMR genes aligns with findings from other studies that have reported similar genes in gut-associated phages [62]. The presence of these genes in prophages is of significant concern as they can facilitate horizontal gene transfer, potentially spreading resistance traits among bacterial populations in the gut microbiome. This contributes to the growing problem of antibiotic resistance, which is a critical public health issue.

The presence of toxin genes, including thiol-activated cytolysin and various toxin-antitoxin systems, indicates potential roles in bacterial competition and survival. These toxins can help prophage-containing bacteria out-compete other microbial populations, thereby influencing the composition and dynamics of the gut microbiome [9, 44, 63, 64].

In addition to AMR and virulence genes, we discovered genes encoding proteins involved in various metabolic processes. These AMGs can enhance the metabolic capabilities of their bacterial hosts, supporting functions like vitamin B5 biosynthesis, carbon metabolism, and the degradation of complex polysaccharides [64–66].

In summary, we have used integrated approach for prophage analysis which included three different tools for phage detection, comparison of results between bacterial genomes and metaviromes and annotation of prophage genomic analysis. Our results show that identified prophages are widely found in the human gut ecosystem and are likely able infecting a broad range of bacterial species. This promiscuity may explain their stable presence over periods ranging from three months to a year. Genomic analysis revealed a variety of important functional genes including those involved in specific metabolic pathways, antibiotic resistances and toxin production.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-025-03827-6>.

Supplementary Material 1: Fig S1. Bacterial genome summary. The ANI scores are represented as heatmap where dark blue colors represent 100% identity and dark red identity below 77% as calculated by FastANI [29] (A). Bacterial species are marked on the branches of the dendrogram and donors where the bacterial cultures derived from are highlighted by color on the right side of the heatmap next to the bacterial IDs (green—donor 1 (D1) and purple—donor 2 (D2)). Number of bacterial genomes by donor, time point and phylogenetic association are represented in the bar plot (A). Fig S2. Taxonomic analysis of vOTUs. We used vConTACT3 [38] for inferring the taxonomy of viral sequences which is based on the shared protein clusters. The network was visualized and modified in Cytoscape v3.10.2 [41]

Supplementary Material 2: Supplemental Table S1. Summary of the vOTU search against the large gut virome databases. The column names represent the following: **vOTU** = vOTU ID. **DB** = reference database (see Methods for details). **DB Reference** = source of the study where vOTU was identified. **Host** = bacterial hosts as assigned from the databases. **Genome size** = genome size of the matched reference. **Different host range from our study** = with Y we highlighted those vOTUs for which bacterial host is different from ours. **Study countries** = countries where vOTUs have a match (available only for GPD database)

Supplementary Material 3: Supplemental Table S2. Combined result output from annotation with vDRAM

Supplementary Material 4

Author contributions

TA, SH, and MR designed the study. TA performed laboratory experiments. SH and TA performed WGS analyses. NS and SH performed bioinformatic analyses. NS performed data analyses and data visualization. NS and MR wrote the manuscript. NS, MR, and TA edited the manuscript. All authors read and approved the final version of the manuscript.

Funding

This work was financially supported by the Slovenian Research Agency under Grants P3-0387, P4-0097 and Slovenian Research Agency Young Investigators Grant.

Data availability

Assembled bacterial genomes were retrieved from the NCBI under the accession numbers PRJNA843113, PRJNA636979. Code for read processing, prophage discovery and graphical outputs are available at: https://github.com/NLZOHomr/prophages_2024.

Declarations

Ethics approval and consent to participate

This work utilized sequencing data from a previous study, for which written informed consent was obtained from all participants. Ethical approval was granted by the National Medical Ethics Committee of the Republic of Slovenia (reference number: 0120–142/2016–2).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 September 2024 / Accepted: 13 February 2025

Published online: 26 February 2025

References

- Sausset R, Petit MA, Gaboriau-Routhiau V, De Paepe M. New insights into intestinal phages. *Mucosal Immunol*. 2020;13:205–15.
- Avellaneda-Franco L, Dahlman S, Barr JJ. The gut virome and the relevance of temperate phages in human health. *Front Cell Infect Microbiol*. 2023;13:1–7.
- Shkoporov AN, Turkington CJ, Hill C. Mutualistic interplay between bacteriophages and bacteria in the human gut. *Nat Rev Microbiol*. 2022;20:737–49.
- López-Leal G, Camelo-Valera LC, Hurtado-Ramírez JM, Verleyen J, Castillo-Ramírez S, Reyes-Muñoz A. Mining of thousands of prokaryotic genomes reveals high abundance of Prophages with a strictly narrow host range. *mSystems*. 2022;7:e00326–22.
- Tuttle MJ, Buchan A. Lysogeny in the oceans: lessons from cultivated model systems and a reanalysis of its prevalence. *Environ Microbiol*. 2020;22:4919–33.
- Liang X, Wagner RE, Li B, Zhang N, Radosevich M. Quorum sensing signals alter in vitro soil virus abundance and Bacterial Community Composition. *Front Microbiol*. 2020;11.
- Hu J, Ye H, Wang S, Wang J, Han D. Prophage activation in the intestine: insights into functions and possible applications. *Front Microbiol*. 2021;12.
- Cornuault JK, Moncaut E, Loux V, Mathieu A, Sokol H, Petit M-A, et al. The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J*. 2020;14:771–87.
- Tyler JS, Beeri K, Reynolds JL, Alteri CJ, Skinner KG, Friedman JH, et al. Prophage induction is enhanced and required for renal disease and lethality in an EHEC Mouse Model. *PLOS Pathog*. 2013;9:e1003236.
- Redgwell TA, Thorsen J, Petit M-A, Deng L, Vestergaard GA, Russel J et al. Prophages in the infant gut are largely induced, and may be functionally relevant to their hosts. *bioRxiv*. 2021;2021.06.25.449885.
- Dikareva E, Matharu D, Lahtinen E, Kolho K-L, De Vos WM, Salonen A et al. An extended catalog of integrated prophages in the infant and adult fecal microbiome shows high prevalence of lysogeny. *Front Microbiol*. 2023;14.
- Zünd M, Ruscheweyh H-J, Field CM, Meyer N, Cuenca M, Hoces D, et al. High throughput sequencing provides exact genomic locations of inducible prophages and accurate phage-to-host ratios in gut microbial strains. *Microbiome*. 2021;9:77.
- Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, et al. Whole-virome analysis sheds light on viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe*. 2019;26:764–e7785.
- 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:W16–21.
- Reis-Cunha JL, Bartholomeu DC, Manson AL, Earl AM, Cerqueira GC. ProphET, prophage estimation tool: a stand-alone prophage sequence prediction tool with self-updating reference database. *PLoS ONE*. 2019;14:e0223364.
- Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*. 2020;8:90.
- Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*. 2021;9:37.
- Tisza MJ, Belford AK, Domínguez-Huerta G, Bolduc B, Buck CB. Cenote-Taker 2 democratizes virus discovery and sequence annotation. *Virus Evol*. 2021;7:veaa100.
- Cornuault JK, Petit M, Mariadassou M, Benevides L, Moncaut E, Langella P et al. Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. 2018;1–14.
- Wexler AG, Goodman AL. An insider's perspective: Bacteroides as a window into the microbiome. *Nat Microbiol*. 2017;2:17026.
- Hedžet S, Rupnik M, Accetto T. Broad host range may be a key to long-term persistence of bacteriophages infecting intestinal Bacteroidaceae species. *Sci Rep*. 2022;12:21098.
- Shin JH, Tillotson G, MacKenzie TN, Warren CA, Wexler HM, Goldstein EJC. Bacteroides and related species: the keystone taxa of the human gut microbiota. *Anaerobe*. 2024;85:102819.
- Al Masalma M, Raoult D, Roux V. *Phocaeicola abscessus* gen. nov., sp. nov., an anaerobic bacterium isolated from a human brain abscess sample. *Int J Syst Evol Microbiol*. 2009;59:2232–7.
- Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. 2018.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl*. 2014;30:2114–20.
- Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011;27:2957–63.
- 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;19:455–77.

28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–5.
29. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9:5114.
30. Nayfach S, Camargo AP, Schulz F, Eloë-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*. 2021;39:578–85.
31. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017;35:1026–8.
32. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. *Nat Commun*. 2018;9:2542.
33. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum information about an uncultivated virus genome (MIUViG). *Nat Biotechnol*. 2019;37:29–37.
34. Nayfach S, Pérez-Espino D, Call L, Low SJ, Sberro H, Ivanova NN, et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol*. 2021;6:960–70.
35. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The Gut Virome Database reveals age-dependent patterns of Virome Diversity in the human gut. *Cell Host Microbe*. 2020;28:724–e7408.
36. Camarillo-Guerrero LF, Almeida A, Rangel-Piner G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. *Cell*. 2021;184:1098–e11099.
37. Madden T. The BLAST Sequence Analysis Tool. 2013.
38. Bolduc B. vConTACT3 - Viral CONtig Automatic Clustering and Taxonomy. 2024.
39. Wang RH, Yang S, Liu Z, Zhang Y, Wang X, Xu Z, et al. PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. *Nucleic Acids Res*. 2024;52:D756–61.
40. Bolduc B. vConTACT3 Database v220. 2023.
41. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape automation: empowering workflow-based network analysis. *Genome Biol*. 2019;20:185.
42. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*. 2020;48:8883–900.
43. Aroney STN, Newell RJP, Nissen J, Camargo AP, Tyson GW, Woodcroft BJ. CoverM: Read coverage calculator for metagenomics. 2024.
44. Brouwer S, Barnett TC, Ly D, Kasper KJ, De Oliveira DMP, Rivera-Hernandez T, et al. Prophage exotoxins enhance colonization fitness in epidemic scarlet fever-causing *Streptococcus pyogenes*. *Nat Commun*. 2020;11:5018.
45. Brielle R, Pinel-Marie M-L, Felden B. Linking bacterial type I toxins with their actions. *Curr Opin Microbiol*. 2016;30:114–21.
46. Hampton HG, Jackson SA, Fagerlund RD, Vogel AIM, Dy RL, Blower TR, et al. AbiEi binds cooperatively to the type IV *abiE* Toxin–Antitoxin Operator Via a positively-charged surface and causes DNA bending and negative autoregulation. *J Mol Biol*. 2018;430:1141–56.
47. Song Y, Zhang S, Luo G, Shen Y, Li C, Zhu Y, et al. Type II antitoxin HigA is a Key Virulence Regulator in *Pseudomonas aeruginosa*. *ACS Infect Dis*. 2021;7:2930–40.
48. Zhou Z, Martin C, Kosmopoulos JC, Anantharaman K, ViWrap. A modular pipeline to identify, Bin, classify, and predict viral–host relationships for viruses from metagenomes. *iMeta*. 2023;2:e118.
49. Ho SFS, Wheeler NE, Millard AD, van Schaik W. Gauge your phage: benchmarking of bacteriophage identification tools in metagenomic sequencing data. *Microbiome*. 2023;11:84.
50. Zhu Y, Shang J, Peng C, Sun Y. Phage family classification under Caudoviricetes: a review of current tools using the latest ICTV classification framework. *Front Microbiol*. 2022;13.
51. Turner D, Shkoporov AN, Lood C, Millard AD, Dutilh BE, Alfenas-Zerbini P, et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch Virol*. 2023;168:74.
52. Gulyaeva A, Garmaeva S, Kurilshikov A, Vich Vila A, Riksen NP, Netea MG, et al. Diversity and Ecology of Caudoviricetes Phages with Genome Terminal repeats in Fecal metagenomes from four Dutch cohorts. *Viruses*. 2022;14:2305.
53. Wang Y, Zhang Z, Kang J, Chen B, Hong W, Lv B, et al. Phages in different habitats and their ability to carry antibiotic resistance genes. *J Hazard Mater*. 2024;469:133941.
54. Ritz NL, Draper LA, Bastiaanssen TFS, Turkington CJR, Peterson VL, van de Wouw M, et al. The gut virome is associated with stress-induced changes in behaviour and immune responses in mice. *Nat Microbiol*. 2024;9:359–76.
55. Caesar L, Haag KL. Tailed bacteriophages (Caudoviricetes) dominate the microbiome of a diseased stingless bee. *Genet Mol Biol*. 2023;46(3 Suppl 1):e20230120.
56. Benler S, Cobián-Güemes AG, McNair K, Hung S-H, Levi K, Edwards R, et al. A diversity-generating retroelement encoded by a globally ubiquitous *Bacteroides* phage. *Microbiome*. 2018;6:191.
57. Hedžet S, Rupnik M, Accetto T. Novel Siphoviridae bacteriophages Infecting *Bacteroides uniformis* Contain Diversity Generating Retroelement. *Microorganisms*. 2021;9:892.
58. Kondo K, Kawano M, Sugai M. Distribution of Antimicrobial Resistance and Virulence genes within the Prophage-Associated regions in Nosocomial pathogens. *mSphere*. 2021;6:e00452–21.
59. Loh B, Chen J, Manohar P, Yu Y, Hua X, Leptihn S. A Biological Inventory of prophages in *A. baumannii* genomes reveal distinct distributions in classes, length, and genomic positions. *Front Microbiol*. 2020;11:579802.
60. da Silva GC, Gonçalves OS, Rosa JN, França KC, Bossé JT, Santana MF et al. Mobile Genetic Elements Drive Antimicrobial Resistance Gene Spread in *Pasteurellaceae* Species. *Front Microbiol*. 2022;12.
61. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, et al. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun*. 2010;1:147.
62. Hu J, Chen J, Nie Y, Zhou C, Hou Q, Yan X. Characterizing the gut phageome and phage-borne antimicrobial resistance genes in pigs. *Microbiome*. 2024;12:102.
63. Chen J, Novick RP. Phage-mediated intergeneric transfer of toxin genes. *Science*. 2009;323:139–41.
64. Bondy-Denomy J, Davidson AR. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J Microbiol*. 2014;52:235–42.
65. Argov T, Azulay G, Pasechnik A, Stadnyuk O, Ran-Sapir S, Borovok I, et al. Temperate bacteriophages as regulators of host behavior. *Curr Opin Microbiol*. 2017;38:81–7.
66. Wendling CC, Refardt D, Hall AR. Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *Evolution*. 2021;75:515–28.

Publisher's note

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