

RESEARCH

Open Access



Gastrointestinal jumbo phages possess independent synthesis and utilization systems of NAD⁺

Cunyuan Li^{1,3†}, Kaiping Liu^{1†}, Chengxiang Gu^{2,4†}, Ming Li⁵, Ping Zhou³, Linxing Chen^{6,7}, Shize Sun¹, Xiaoyue Li¹, Limin Wang³, Wei Ni^{1*}, Meng Li^{2,4*} and Shengwei Hu^{1*}

Abstract

Background Jumbo phages, phages with genomes > 200 kbp, contain some unique genes for successful reproduction in their bacterial hosts. Due to complex and massive genomes analogous to those of small-celled bacteria, how jumbo phages complete their life cycle remains largely undefined.

Results In this study, we assembled 668 high-quality jumbo phage genomes from over 15 terabytes (TB) of intestinal metagenomic data from 955 samples of 5 animal species (cow, sheep, pig, horse, and deer). Within them, we obtained a complete genome of 716 kbp in length, which is the largest phage genome so far reported in the gut environments. Interestingly, 174 out of the 668 jumbo phages were found to encode all genes required for the synthesis of NAD⁺ by the salvage pathway or Preiss-Handler pathway, referred to as NAD-jumbo phage. Besides synthesis genes of NAD⁺, these NAD-jumbo phages also encode at least 15 types of NAD⁺-consuming enzyme genes involved in DNA replication, DNA repair, and counterdefense, suggesting that these phages not only have the capacity to synthesize NAD⁺ but also redirect NAD⁺ metabolism towards phage propagation need in hosts. Phylogenetic analysis and environmental survey indicated NAD-jumbo phages are widely present in the Earth's ecosystems, including the human gut, lakes, salt ponds, mine tailings, and seawater.

Conclusion In summary, this study expands our understanding of the diversity and survival strategies of phages, and an in-depth study of the NAD-jumbo phages is crucial for understanding their role in ecological regulation.

Keywords Bacteriophage, Animal gut, Metagenome, NAD⁺, Jumbo phage

[†]Cunyuan Li, Kaiping Liu, and Chengxiang Gu contributed equally to this work.

*Correspondence:

Wei Ni
niweiwonderful@sina.com
Meng Li
limeng848@szu.edu.cn
Shengwei Hu
hushengwei@163.com

¹ College of Life Sciences, Shihezi University, Shihezi, Xinjiang 832003, China

² Archaeal Biology Center, Synthetic Biology Research Center, Shenzhen Key Laboratory of Marine Microbiome Engineering, Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Guangdong Higher Education Institutes, Shenzhen University, Shenzhen, China

³ State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural and Reclamation Science, Xinjiang 832003, China

⁴ Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study Shenzhen University, Shenzhen, China

⁵ Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Universitätsstrasse 10, Constance 78457, Germany

⁶ Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA

⁷ Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei 230026, People's Republic of China



© The Author(s) 2024. **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

Bacteriophages (phages), as viruses that infect bacteria, are key components in the natural microbiome because they can shape microbial composition and structure, alter host metabolism, spread antibiotic-resistant genes, and mediate gene transfer [1–3]. Metagenomic investigations have shown the diverse and abundant nature of the gut virome of livestock [4, 5]. Recent studies also demonstrated that phages in livestock microbiome may impact the major functions of the gut such as feed digestion, microbial protein synthesis, and methane emissions [6, 7]. However, we still lack a deep understanding of how microbes interact with phages to affect the gut ecosystem.

Unlike the small genome size commonly understood for phages, some phages have enormous genomes exceeding 200 kbp, referred to as jumbo phage [8], and the largest phage genome reported to date reaches 735 kbp [9]. These jumbo phages are distributed in various environments [9, 10]. Current research indicates that jumbo phages encode auxiliary metabolic genes, such as genes for methane monooxygenases (*pmoC*), to improve the host's survival ability [11]. Jumbo phages may redirect biosynthesis to phage-encoded functions by CRISPR-induced interception of the earliest steps of host translation [9]. Some megaphages with a genome > 500 kbp infect *Prevotella* species enriched in the gut microbiome, which may be common but largely overlooked components of the human and animal gut microbiomes [12, 13]. However, due to the difficulty in assembling large genomes, the number of reported jumbo phages is still relatively scarce in animal guts [14]. We still know very little about the general characteristics of jumbo phages in animal microbiomes.

In this study, we aim to investigate the diversity, genetic repertoires, and survival strategies of jumbo phages in animal microbiomes. We conducted large-scale data mining for identifying jumbo phage genomes by analyzing over 15 TB of metagenomic data from 955 intestinal samples of five animals (cow, sheep, pig, horse, and deer). We identified 174 jumbo phages with all genes required for the synthesis of NAD⁺, referred to as NAD-jumbo phages. These NAD-jumbo phages were further characterized by unique genetic repertoires including NAD⁺-dependent enzymes, DNA replication machinery, DNA repair-related enzymes, and counterdefense systems. We also investigated the distribution of NAD-jumbo phages in natural environments and human microbiomes and detected the appearance of NAD-jumbo phages across Earth's ecosystems. Our study provides new insights into the genomic characteristics and survival strategies of jumbo phages, and the novel path forward to exploring bacteria-phage interactions.

Methods

Metagenomic samples

A total of 424 cattle, 140 sheep, 169 pig, 29 deer, and 193 horse gut metagenomes, comprising 418 cattle, 66 sheep, 29 deer, 163 pig, and 116 horse metagenomes from the NCBI or ENA public databases, and 6 cattle, 6 pig, 74 sheep, and 77 horse metagenomes from this study (Supplementary Table S1). The 6 cattle metagenomic samples were from adult Holstein cattle, including 3 colonic contents and 3 rumen contents. The 6 pig metagenomic samples were from adult Duroc, including 3 colonic contents and 3 stomach contents. The 74 sheep metagenomic samples were from Kazakh sheep, including 33 rumen contents, 5 reticulum contents, 6 omasum contents, 6 abomasum contents, 6 ileal contents, 6 cecum contents, 6 colon contents, and 6 rectal contents. All samples in this study were collected from a slaughterhouse in Shihezi City, China. In addition, the 77 horse metagenomic samples were from the feces of a thoroughbred horse. These samples were collected from the Hazanat International Racecourse in Mulei City, China. All samples were frozen in liquid nitrogen immediately after collection in sterile centrifuge tubes and stored at – 80 °C until genome extraction. All procedures involving animals were approved by the Animal Care Committee of Shihezi University. The study was performed in accordance with the ethical standards established in the 1964 Declaration of Helsinki and subsequent amendments.

DNA extraction and sequencing

To obtain high-quality microbial DNA from the gut contents, the DNA of the samples was extracted by cetyltrimethyl ammonium bromide (CTAB) method [15]. The quality, integrity, and potential contamination of extracted DNA were evaluated by agarose gel electrophoresis and pulsed-field gel electrophoresis. Meanwhile, the Nanodrop Kit (Implen, CA, USA) and Qubit[®]2.0 fluorometer (Life Technologies, CA, USA) were used for purity determination and precise quantification of the extracted DNA. According to the manufacturer's protocol, Illumina sequencing libraries were prepared with a NEBNext[®]Ultra[™] DNA Library preparation kit (New England Biolabs, USA) of 500 mg DNA samples from each sample. These libraries were purified using the AMPure XP system (Beckman Coulter, Brea, CA, USA) and quantitatively analyzed using the Agilent 2100 bioanalyzer and real-time fluorescent quantitative PCR. All libraries were sequenced using the Illumina NovaSeq 6000 platform with a reading length of 150 base pairs (PE150).

Data quality control and phage genome identification

To identify phage sequences in cattle, sheep, horse, pig, and deer gut samples, we performed a rigorous pipeline for phage genome assembly to avoid many false positives in the expected metagenome (Fig. S1) [9, 16]. Obtaining high-quality metagenomic phage contigs is a critical step in subsequent analysis, raw reads were filtered to remove adapter sequences and low-quality sequences using fastp software (Version: 0.20.1, parameters: `-cut_by_quality3 -W 4 -M 20 -n 5 -c -l 50 -w 3`) [17] and aligned them with cattle, sheep horse, pig, and deer reference genomes using bowtie2 (Version: 2.2.4) to remove host sequences, respectively [18]. Next, the cleandata were assembled by the software metaSpades (Version: SPAdes-3.11.1) to obtain genomic contigs [19]. These genomic contigs were screened for metagenomic virus contigs using VirFinder [20], VirSorter 2 [21], VIBRANT [22], and CheckV [23]. When using VirSorter, only contigs classified as 1, 2, 4, or 5 categories were considered. For VirFinder, we retained these contigs with scores > 0.9 and $p < 0.01$ and linear contigs > 5 kbp in length and circular contigs > 1.5 kbp in length. These viral contigs were de-redundant using dRep software (Version: dRep v3.2.2, parameters: `-S_algorithm ANImf, -S_ani 0.95, -cov_thresh 0.8`) [24]. Then only those sequences that VIBRANT identified as viruses with genome size > 200 kbp in length were retained. After filtering, we obtained a total of 1001 jumbo phage genomes from the animal gut. Then, these phages were evaluated for quality using CheckV, and the high-quality and completed genomes (genome completeness > 90%) were retained for subsequent analysis. A total of 668 high-quality jumbo phage genomes were obtained by using these stringent selecting criteria (Additional file 1: Fig. S1).

Protein function annotation

Prodigal software using genetic code 11 (`-m -g 11 -p single`) (Version: V2.6.3) was used to predict ORFs encoded by 668 jumbo phages [25]. Further function annotation of protein sequences encoded by jumbo phage was performed by viral homolog (VOG) [26], KEGG [27], and Pfam [28]. tRNAscan-s.e.v.2.0 with the bacterial model was used to identify tRNA sequences in jumbo phages and their hosts [29]. For detecting the presence of conserved tRNAs in jumbo phages, we used MEGA7.0 software to align the tRNA of jumbo phage with the corresponding tRNAs sequences from bacteria [30].

Jumbo phage classification and phylogenetic analyses

To understand the taxonomic information of 668 jumbo phages, we downloaded the phage genome from the RefSeq prokaryotic database (Prokaryotic Viral RefSeq 211)

and reported jumbo phages. The distance-based hierarchical clustering and virus classification prediction of protein-sharing networks of 668 jumbo phages and Reference phages was performed for population analysis using vConTACT2 (v0.11.3) [31]. The predicted protein sequences were subjected to all-verse-all BLASTp using DIAMOND (more sensitive mode, identity 25%, query cover 50%). Protein clusters were generated using MCL and viral clusters were generated using ClusterONE in vConTACT2 pipeline.

Based on functional annotations on the ORFs using eggnoG-mapper, we select major capsid protein (MCP) and terminase large subunit (TerL) protein of jumbo phages in this study and previously reported [9]. All predicted MCP and TerL protein sequences were examined using PfamScan (v1.6) against Pfam and manual check. DIAMOND BLASTp (`-e 1e-10 -more-sensitive`) was used to search reference sequences against VOG 216 database. The host assignment of reference viruses was based on Virus-Host DB (https://www.genome.jp/virus_hostdb/) [32]. MCP and TerL protein sequences were aligned using linsi mode of the MAFFT (v7.310) software [33]. The alignments were trimmed using trimAL (v 1.4.rev15) to remove positions with gaps over 90% of sequences [34]. The maximum-likelihood phylogenetic trees of MCP and TerL protein sequences were computed by IQ-TREE (v2.0.3) after extended model selection (MCP LG + F + R9; TerL LG + F + R10), and support for nodes was evaluated with 1001 replicates for ultrafast bootstrap [35].

Taxonomy of jumbo phage

The prokaryotic virus from NCBI GenBank, RefSeq, and VOG database were constructed as a reference database using custom scripts. The ORFs encoded by jumbo phage were predicted by Prodigal software (Version: V2.6.3, parameter: `-p meta`) and constructed into a protein database using DIAMOND (Version: v0.9.9) [24]. The protein database was aligned with the reference database using DIAMOND BLASTp (Parameters: `BLASTp -e 1e-5 -query-cover 10 -subject-cover 10`) to obtain the taxonomy of each ORFs encoded by jumbo phage [36]. We assigned taxonomic information for each jumbo phage based on the majority taxonomic assignment [16]. 668 Jumbo phages without the majority taxonomic assignments or with less than two assigned ORFs were considered unassigned phages [16].

Host prediction

We predicted CRISPR sequences in microbial metagenome-assembled genomes (MAGs) from the gut microbiome of animals (Supplementary dataset 1) to obtain CRISPR spacer sequences by using MinCED software

(Version: V.0.4.2), a software based on CRT, to predict the possible host of huge phages [37]. The taxonomic assignment of the MAGs was generated by GTDB-tk (classify_wf workflow with r207 database, v2.1.0) and was converted into the corresponding NCBI phylum taxonomy [38]. In all, we extract 265,818 CRISPR spacers within 12,674 MAGs. CRISPR spacer sequences were queried with jumbo phage using BLASTn with *E*-value of $<10^{-5}$, identity $>95\%$, and coverage $>80\%$ (with ≤ 1 mismatch over the whole spacer) [39]. The host of jumbo phage were assigned to CRISPR spacers taxonomy based on blast hits with the highest bit scores.

Evolutionary analysis and protein structure prediction of NAMPT and NadD

The NAMPT and NadD sequences in NAD-jumbo phages were aligned using MUSCLE [40]. DNA alignments were further polished using trimAl [34]. The maximum-likelihood phylogenetic trees of NAMPT and NadD protein sequences were computed by IQ-TREE (v2.0.3) after extended model selection (MCP LG + F + R9; TerL LG + F + R10), and support for nodes was evaluated with 1001 replicates for ultrafast bootstrap. For RoseTTAFold predictions, run_e2e_ver.sh script was used to predict the three-dimensional structure of NAMPT and NadD proteins [41]. The three-dimensional structure of NAMPT and NadD encoded in NAD-jumbo phages were compared with those encoded in bacteria [42, 43] using Foldseek's easy-to-search command default parameter [44]. AutoDock Vina (V1.5.7, parameter: default) was used to perform molecular docking analysis of the NAMPT protein and its substrate NMN. The docking results were visualized using PyMOL for detailed analysis [45].

Distribution of NAD-jumbo phage in Earth's ecosystem

To study the distribution of NAD jumbo phage genomes in the Earth's ecosystem, we collected human-associated and environmental jumbo phages as an extension of the jumbo phage dataset to analyze the biogeographic information of NAD jumbo phages. We obtained human-associated jumbo phage genomes and metadata from the Metagenomic Gut Virus (MGV) Catalogue [46], Gut Phage Database [16], and an unknown phage genome set in the human gut reported by Benler et al. [47]. We used the jumbo phage genome set reported by Banfield et al. as the source of the environmental jumbo phage genomes [9]. As with the screening criteria for NAD jumbo phages in livestock gut, we used the same strategy to identify NAD jumbo phage genomes from them. We use the information in metadata to represent the distribution of NAD jumbo phage genomes.

Results

Basic features of jumbo phages identified in animal gut microbiomes

To characterize the diversity of jumbo phages in the animal gut, we analyzed 955 metagenomes from a 15.24 TB dataset including 8.71 TB for cattle, 1.96 TB for horses, 1.25 TB for pigs, 2.75 TB for sheep, and 0.56TB for deer (Supplementary Table S1). Based on previous reports [9, 46], we have developed a rigorous phage detection pipeline to identify phages from the metagenomic datasets (Fig. S1, see "Methods" section). A total of 1001 de novo assembled contigs with longer than 200 kbp were classified as jumbo phages (Supplementary Table S2). To improve the quality of subsequent analysis, 668 non-redundant jumbo phages with high-quality (completeness $>90\%$, 378 in total) and complete (290 in total) genomes were selected to form a high-quality gut jumbo phages set of animals.

The 668 jumbo phage genomes were distributed in five animal species (Figs. S2 and S3), representing the largest number of intestinal jumbo phage genomes reported so far to our knowledge. There are 10 phages whose genome size was >500 kbp, which can be classified as megaphages. These megaphages were found in the intestines of cattle, sheep, horses, and pigs, suggesting that the animal intestine may be a megaphage reservoir (Fig. S2). Two bioinformatically validated circular and complete phage genomes, ERR20279001_716809 (Houyi_20) and SRR44355611_713556 (Houyi_34) are 716.8 kbp and 713.6 kbp in length, respectively, and are the largest phages reported to date in the gut (Fig. S4). The previously reported largest intestinal phage genome is 660 kbp in length from horse gut samples [13]. Both phages with a genome length of 716.8 kbp and 713.6 kbp are derived from the rumen metagenome data of cattle, but the average nucleotide homology (ANI) of the two phage genomes is 93.93%, indicating that the two phages are different species in the same genus [48], which suggests that the rumen may be a reservoir for megaphages. The median genome size of the 668 gut jumbo phages was 263.28 kbp, which was much bigger than those of IMG/VR and previous reports [49, 50]. In total, the genome set here expands the catalog of jumbo phages and provides a rich source for jumbo phage biology.

In addition, the jumbo phage genomes averagely encode 23 tRNA genes with predicted functions in translation (Supplementary Table S3), and 82.29% of them with at least 5 tRNA genes. SRR12268561_1_289927 (Houyi_39) encodes up to 68 tRNAs, which covers the full range of amino acids required for protein synthesis. Interestingly, a suppressor tRNA called sup-tRNA is encoded by some jumbo phage genomes, which can induce translational readthrough of nonsense mutations.

It seems that sup-tRNA may help increase the translation efficiency of phage genes under stress [49, 51]. The tRNA sequences encoded by the jumbo phages are different from those of the hosts (Fig. S5), which may help phages redirect the host’s ribosomes to make more of their own proteins [9].

Diversity and hosts

As jumbo phages had expanded genome sizes, the evolutionary relatedness of our newly assembled and published jumbo phages was unclear. To gain insights into the taxonomic information of the discovered 668 jumbo phages, we utilized the most recent Imphared database (September 1, 2024) for DIAMOND BLASTp [52]. After processing the blastp results with vContact2 and Graphanalyzer, we found that only 7 jumbo phage genomes were classified into known families (1 Intestiviridae, 1 Kyanoviridae, and 5 Steigviridae), with the rest only being assignable to Caudoviricetes (Supplementary Table S4). These findings underscore the novelty of our newly discovered

jumbo phages but also highlight the challenges in their classification.

Since the commonly used classification marker genes for phages cannot cover all groups, preliminary grouping can usually be performed according to gene content [31]. A network including the 668 jumbo phages and published phages was generated to group the newly assembled jumbo phages, which showed their novelty (Fig. S6). To further investigate their evolutionary relationship, we constructed phylogenetic trees using major capsid protein (MCP) and terminated large subunit (TerL) (Fig. 1). Out of the 668 jumbo phages, 143 have both MCP and TerL, and 291 have at least one of them (Supplementary Table S5–S6). The proteins of the jumbo phages reported here were well clustered with high guidance support, thus defining the evolutionary clades. Most jumbo phages fall into one of 5 clades in MCP and TerL trees. We named these 5 clades of jumbo phages as Kuafu_jumbo phage, Houyi_jumbo phages, Jingwei_jumbo phages, Pangu_jumbo phages, and Zhurong_jumbo phages (according to

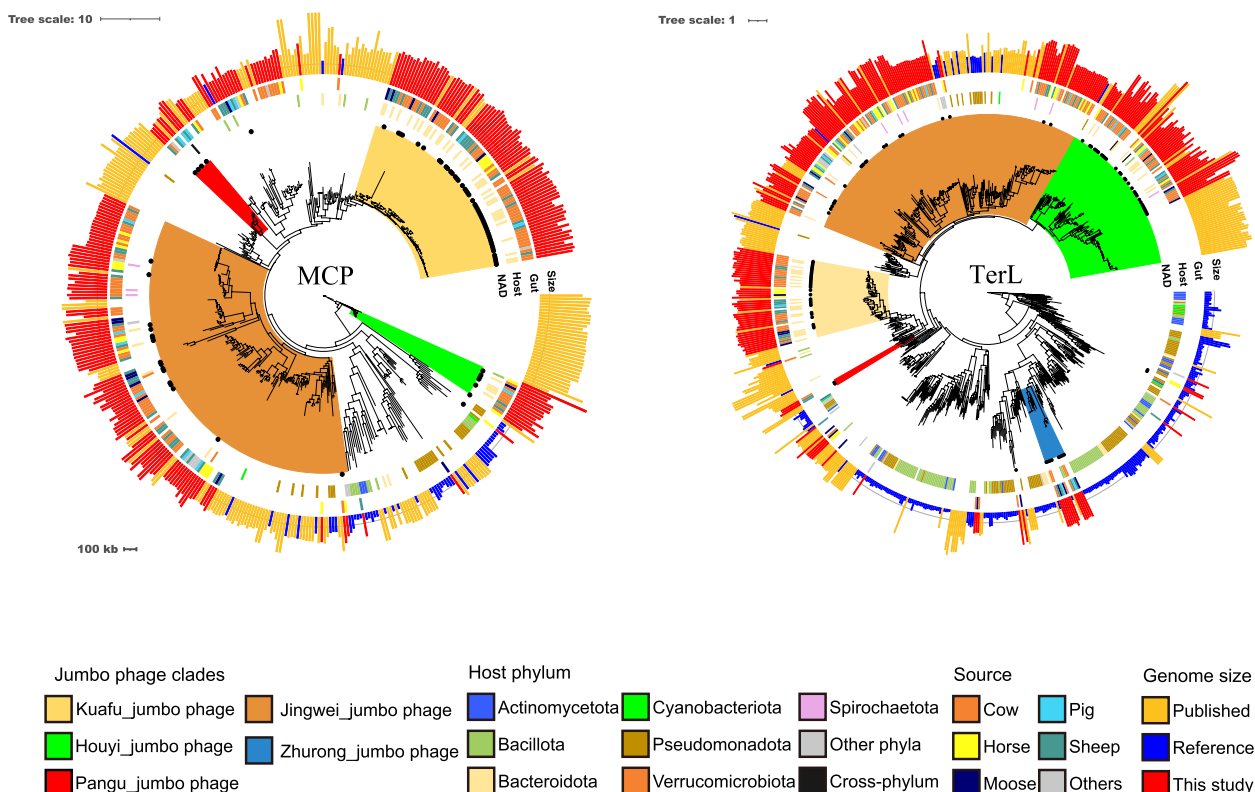


Fig. 1 Phylogenetic tree of jumbo phages. Phylogenetic tree was constructed by major capsid proteins (MCP, left) and large terminase (TerL, right) sequences identified in this study. Homologs references were collected from published jumbo phages [9] and VOG database. From inner to outer ring, the concentric circles indicate the presence of NAD-jumbo phages in this study, host phylum, gut source, and genome size. Jumbo phage clades indicate five different clades of phages identified in this study. Host phylum indicates the predicted host phylum of jumbo phages using CRISPR spacer analysis. The gut source represents which animal guts were sampled to obtain jumbo phages in this study. Genome size is the sequence length of phage genomes used in the tree. Different colors represent different types of clades, host phylum, gut sources, and genome sizes in the legends

the characters in traditional Chinese mythology) (Fig. 1, Supplementary Table S6). Within each clade, phages were sampled from different animal species from different countries in the world, suggesting the diversification of these jumbo phages across gut environments (Fig. 1). The megaphages (genome size >500 kbp) from different animal intestines clustered into the same clade, suggesting that megaphages in animal intestines may share their ancestor (Fig. 1). In addition, no typical virus MCP sequences were found in Zhurong_jumbo phage clade.

Predicting the cellular host of the viruses is important for understanding the virus-host interactions [53]. We used CRISPR spacer targeting to detect phylum-level links between the host and the jumbo phages. We extracted 265,818 CRISPR spacers (Supplementary Table S7) from the metagenome assembly genomes (MAGs) assembled from the animal gut microbiomes (Supplementary dataset 1). We detected 127 links between 123 newly assembled jumbo phages and 10 bacteria phyla. Bacteroidota (63) and Bacillota (34) were the most common host phyla of jumbo phages (Fig. 1, Supplementary Table S7).

Jumbo phages possess independent synthesis pathways of NAD⁺

Annotation of the new jumbo phage genomes revealed they contain a large number of genes involved in DNA replication, transcription, DNA modification, energy synthesis, and metabolism (Supplementary Table S5). We are attracted to the NAD⁺ synthesis pathway-related genes. The KEGG and gene blast analysis revealed that 174 jumbo phages have all the genes necessary for the biosynthesis of NAD⁺ by salvage pathways or Preiss handler pathways (Fig. 2A, Supplementary Table S8). The 174 jumbo phages with at least one complete pathway of NAD⁺ synthesis were named NAD-jumbo phages.

Interestingly, our 10 largest phages (536–716 kbp in length) all belong to NAD-jumbo phages. Moreover, the phages with the genomes of 716 kbp, 713 kbp, 696 kbp, and 648 kbp contained two different types of NAD⁺ synthesis pathways, such as carrying both the NAM salvage pathway and NR salvage pathway (Supplementary Table S8). We further analyzed whether the NAD synthesis pathway is also present in the previously reported largest complete phage genomes with 634 kbp, 636 kbp, 642 kbp, and 735 kbp in length [9]. All but the 735 kbp phages have the complete NR salvage pathway, thus belonging to NAD-jumbo phages as well. However, the 735 kbp phage carries one nicotinate nucleotide adenyltransferase (*nadD*) which plays a vital rate-limiting step in NAD⁺ biosynthesis. Previous experiments documented that the overexpression of bacterial *nadD* improves NAD⁺ synthesis in a variety of bacteria [54, 55].

Notably, NAD⁺ synthesis-related genes are physically clustered together on phage genomes to form biosynthetic gene clusters (BGC) that recur in some NAD-jumbo phages (Fig. 2B, Fig. S7). For example, NAD-jumbo phages *nhp215* carry a BGC clustering *nadR* and *nadM* genes which constitute a complete NR salvage pathway (Fig. 2B); *Kuafu_71* carry a BGC clustering *pncB*, *nadD* and *nadE* genes which are all the genes necessary for the biosynthesis of NAD⁺ by Preiss handler pathway (Fig. 2A, Fig. S7). The BGCs provide a complete pathway of NAD⁺ synthesis, which may allow jumbo phages to synergistically express genes and potentially promote the production of NAD⁺.

Many NAD-jumbo phages encode genes with predicted functions in key regulators of NAD⁺ biosynthesis, especially nicotinamide phosphoribosyltransferase (NAMPT), and nicotinamide riboside transporter (*pnuC*). NAMPT is the rate-limiting enzyme of NAD⁺ salvage pathways. Seventy-one out of 174 NAD-jumbo phages encoded NAMPT genes (Supplementary Table S5), and phage-NAMPT likely accelerated the catalysis of NAM to NMN that was used as a precursor for NAD⁺ biosynthesis in hosts (Fig. 2A). *PnuC*, is a transporter catalyzing cellular uptake of the NAD⁺ precursor nicotinamide riboside (NR) across the bacterial cell membranes. The *pnuC* genes were encoded by 68 out of 174 NAD-jumbo phages (Supplementary Table S5). *PnuC* forms a biosynthetic gene cluster with *nadR* and *nadM* in NAD-jumbo phages *nhp215* (Fig. 2B), and may transport more NR substrate from outside of the bacterial membrane to phage NAD⁺ synthesis.

We found the NAD⁺ synthesis pathway of some jumbo phages was complementary to that of the bacterial hosts (Fig. 2C). For example, *Kuafu_19* has the Preiss handler pathway to use the substrate NA to synthesize NAD⁺, while its host only has the de novo synthesis pathway to use aspartate as a substrate to synthesize NAD⁺ (Fig. 2C). The complementarity of NAD⁺ synthesis between some jumbo phages and their hosts has the opportunity to promote the synthesis efficiency of NAD⁺ and improve the concentration of NAD⁺ available to phages or hosts, which may further facilitate phage propagation.

NAD-jumbo phages were observed to cluster in phylogenetic analysis, which can be roughly divided into 5 categories (Fig. S1). These 174 NAD-jumbo phages were mainly clustered in the clade of *Kuafu_jumbo* phage and *Houyi_jumbo* phage. In particular, *Kuafu_jumbo* phage clade is mainly composed of our newly assembled NAD-jumbo phages. Surprisingly, all predicted hosts for *Kuafu_jumbo* phages were Bacteroidota, suggesting the Phylum-level host specificity of these phages. For *Houyi_jumbo* phage, the NAD-jumbo phages are the most important part of *Houyi* clade. Meanwhile, the largest gut phages ERR20279001_716809

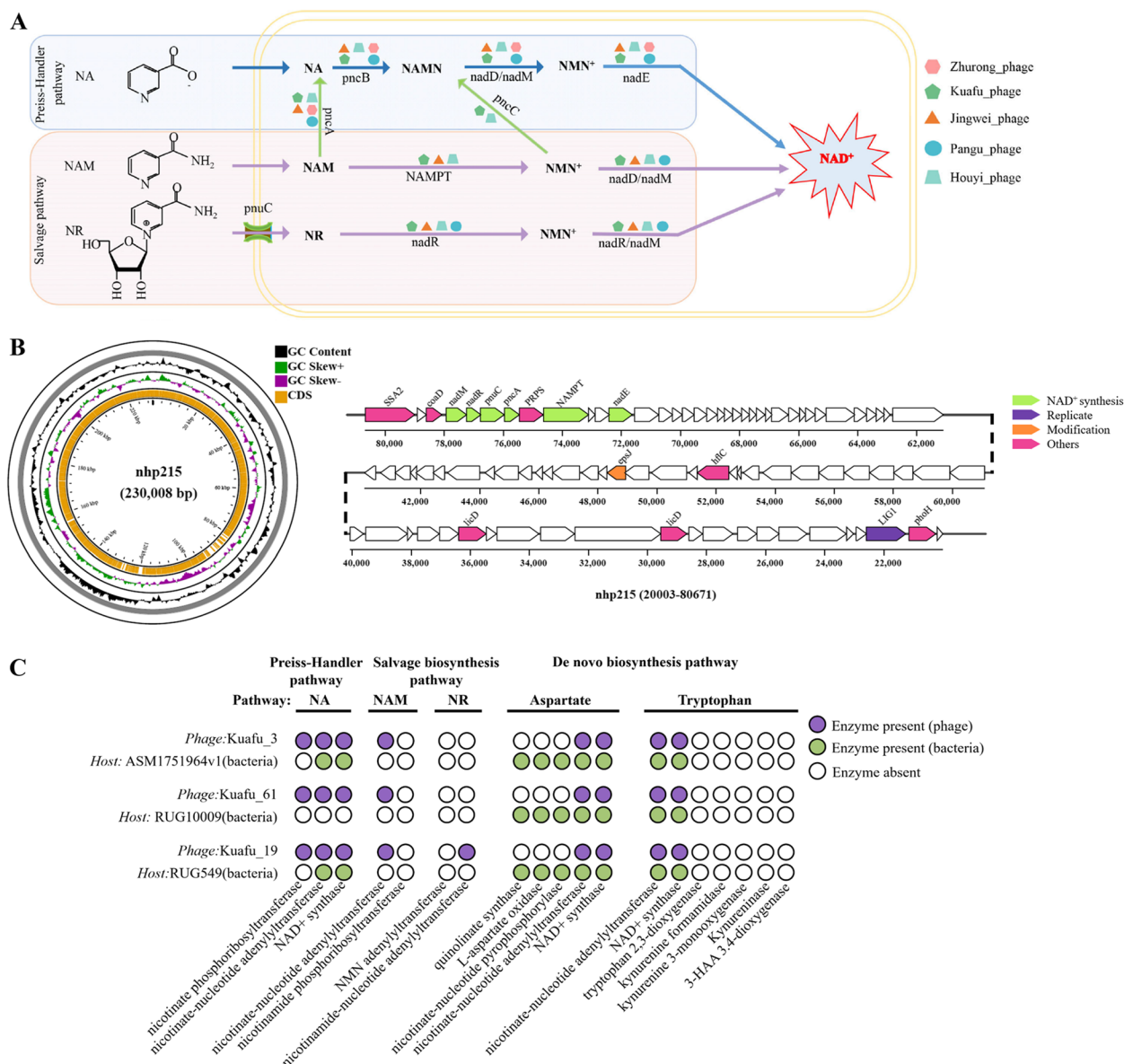


Fig. 2 Jumbo phages possess independent synthetic pathways of NAD⁺. **a** Jumbo phages encode key enzyme genes in each step of NAD⁺ biosynthesis by salvage pathways or Preiss handler pathways. Different colors represent different types of jumbo phages that contain all enzymes necessary for the pathway. The metabolic pathways were obtained from KEGG database. **b** Genetic maps of NAD-jumbo phage nhp215. Genome circos map (left) and gene content (right) of nhp215. The green graph represents the gene for NAD synthesis. The white graph represents the unannotated gene. **c** NAD⁺ biosynthesis pathways are complementary between NAD-jumbo phages and their hosts. NAD-jumbo phages are paired with their host bacteria, and their names are displayed on the left. Enzymes involved in each pathway are listed below the column. Green circles and purple circles represent the presence of enzymes encoded by phages and their hosts, respectively

(716 kbp) and SRR44355611_713556 (713 kbp) belong to Houyi_jumbo phage.

NAD-jumbo phages encode NAD-dependent enzymes for various biological processes

Considering the crucial role of NAD⁺ in various biological processes including DNA replication, DNA repairs,

and redox reactions for all organisms [56], NAD⁺ may be highly significant for the life cycle of NAD-jumbo phages. We performed genome-wide functional annotation of NAD-jumbo phages. Our NAD-jumbo phages encode 15 types of NAD⁺-consuming enzymes (in total of 222 NAD-dependent enzymes or NAD⁺ cofactors) (Supplementary Table S9). These NAD⁺-consuming enzymes

may be involved in DNA replication, DNA repair, and NAD⁺-mediated counterdefense, suggesting that NAD-jumbo phages may mediate various biosynthetic processes by expressing their own NAD-dependent enzymes (Fig. 3).

It is noted that more than 50% (90/174) of the NAD-jumbo phages contain NAD-dependent DNA ligase (Fig. S8, Supplementary Table S8), which is an enzyme necessary for DNA replication and repair. More importantly, these NAD-jumbo phages also encode almost all replication-related genes including DNA topoisomerase, DNA helicase, DNA polymerases family, DNA primase RNase H1, and NAD-dependent DNA ligase (Fig. 3), which are key for DNA replication initiation and elongation in prokaryotes [57, 58], and phage may use these proteins to shift DNA biosynthesis towards phage own

replication. Therefore, besides potentially independent synthesis of NAD⁺, NAD-jumbo phages may express their own NAD-dependent enzymes and replication machinery to direct phage self-replication in bacterial hosts.

Similarly, DNA repair may also be directed towards phage own repair need in bacterial hosts, given that many NAD-jumbo phages carrying genes including DNA damage sensor PARP1 (NAD-dependent type), mismatch recognition factor MutS, as well as excision repair enzymes Vsr and ERCC3 (Fig. 3, Supplementary Table S8 and Table S9). Both MutS and NAD-dependent PARP1 could initiate various forms of DNA repair [59, 60]. The Vsr endonuclease and ERCC3 function in nucleotide excision repair, and these two enzymes were likely used for cutting mismatched DNA in phage

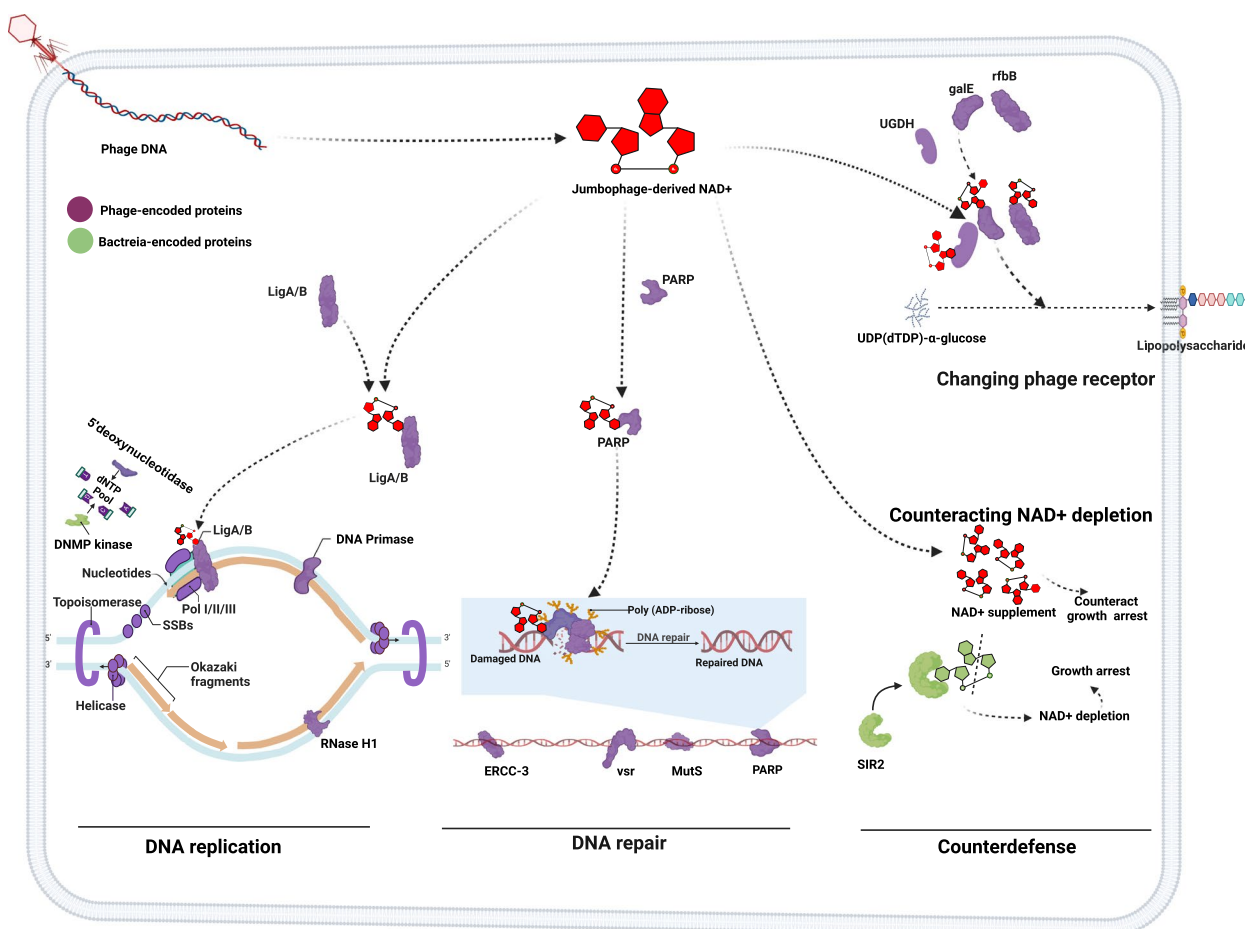


Fig. 3 A model for NAD⁺-mediated jumbo phages direction of DNA replication, DNA repair and counterdefense in bacterial hosts. NAD-jumbo phages independently synthesize NAD⁺ and then utilize NAD-dependent enzymes to potentially direct biosynthesis towards phage propagation needs in host bacteria. Phage-encoded NAD-dependent enzymes were involved in DNA replication, DNA repair, and counterdefense (NAD⁺ in red, phage-encoded proteins in purple, bacteria-encoded proteins in green). In particular, many NAD-jumbo phages encode almost all genes required for DNA replication and some key genes for initiating DNA repair. LigA/B: NAD-dependent DNA ligase. SSBs, single strand DNA-binding protein; PARP, poly [ADP-ribose] polymerase; ERCC-3, DNA excision repair protein ERCC-3; Vsr, DNA mismatch endonuclease; MutS, DNA mismatch repair protein MutS. Figures generated with BioRender (<https://biorender.com/>)

genomes [61, 62]. NAD-dependent DNA ligase of phage seal nicks in genomes in the final stages of DNA-repair pathways. Together, NAD-jumbo phages may use their own NAD⁺ products, NAD-dependent enzymes, and key repair-related factors to direct self-repair in the host.

In addition, it is important to note that many NAD-jumbo phages encode NAD-dependent enzyme genes, such as UDP-glucose 6-dehydrogenase (UGDH), UDP-glucose 4-epimerase (galE), and dTDP-glucose 4,6-dehydratase (rfbB), involved in lipopolysaccharide (LPS) biosynthesis (Fig. 3, Supplementary Table S10). Bacteria LPS is a rich and diverse cell wall polysaccharide, which was adopted by phages as a receptor for infection [63, 64]. Phage-encoded synthesis of LPS may be able to change the modification of cell walls to prevent the infection of competing phages [65, 66]. The survival adaptability of NAD-jumbo phages may be improved by modifying the host cell wall.

Another function of NAD⁺ derived from jumbo phages may be to counteract bacterial abortive infection defense mediated by NAD⁺ depletion. NAD⁺ depletion triggered by the Sir2 domain cause bacteria suicide to limit phage propagation, which has been well documented in anti-phage defense systems [67–69]. NAD-jumbo phages harbor a diverse array of genes involved in the NAD⁺ salvage synthesis pathway, which are absent in its host (Fig. 2C), implying that NAD⁺ may be autonomously synthesized by NAD-jumbo phages. NAD⁺ from NAD-jumbo phages may be used to maintain the normal level of NAD⁺ for propagation of both host and phage, thus counteracting the host NAD⁺ depletion. This suggests that the ability to synthesize NAD⁺ in jumbo phages may be a counter-defense strategy against host NAD⁺ depletion, although this needs further experimental verification.

NAD-jumbo phages are ubiquitous and diverse across Earth's ecosystems

Our NAD-jumbo phages spanned five animal species from different countries in the world, reminding us diverse ecosystem distribution of NAD-jumbo phages.

To end this, we further analyzed whether NAD⁺ synthesis gene pathways were present in human-associated and environment-associated jumbo phages. Of 863 jumbo phages from human and natural environments microbiome, we identified 181 jumbo phages with complete pathways of NAD⁺ synthesis, which belong to NAD-jumbo phage. We identified the distribution of these NAD-jumbo phages across Earth's ecosystems, including cultivated phages, human-associated samples (gut and mouth), and environment-associated samples (freshwater lakes and rivers, marine ecosystems, sediments, hot springs, soil, deep subsurface habitats, and built environment) (Fig. 4A, Supplementary Table S10–S12). The NAD-jumbo phages have been isolated and cultivated on different bacteria hosts, indicating that the identification of such NAD⁺ synthesis pathway in our study is not an artifact of metagenome analyses.

Similar to animal-associated NAD-jumbo phages, the environment- and human-associated NAD-jumbo phages also include NR salvage pathway, NAM salvage pathway, or Preiss handler pathway, showing the diversity of NAD⁺ synthesis pathways across ecosystems (Supplementary Table S10). We further characterized the diversity of NAD⁺ synthesis genes by conducting a phylogenetic analysis of two NAD⁺ synthesis rate-limiting enzymes (NAMPT and nadD) sequences. In most cases, the phylogeny of the NAD⁺ synthesis enzymes from the same ecosystem type displays close clustering, supporting previous findings that phage metabolic enzymes correlate with environment temperature and nutrition for enhancing phage environmental adaptation [70]. In a few cases, the NAD⁺ synthesis enzyme phylogeny of human-associated phages shows close clustering with those of animal-associated phages, and the likely explanation is that horizontal gene transfer occurs in NAD-jumbo phages between animals and humans. Although NAD⁺ synthesis genes were diverse across ecosystems, RoseTTAFold predicted protein structures and conserved amino acid residues of NAMPT and nadD were consistent with those of bacteria determined by cryo-electron microscope (Fig. 4B, C). In addition, Lee et al. reported that *Vibrio* phage KVP40 encoded functional NAMPT protein and

(See figure on next page.)

Fig. 4 The distribution of NAD-jumbo phages across Earth's ecosystems. **a** Global distribution of NAD-jumbo phages identified in 12 different ecosystems types. The different symbols and colored dots represent different ecosystem types in the legend. **b c** Phylogenetic trees (left) and structures (right) of NAD-jumbo phages NAMPT and nadD proteins. Phage-encoded protein sequences were extracted from animal-, human- and environment-associated NAD-jumbo phage genomes. Microbial reference sequences were collected from RefSeq database. Different colors represent the ecosystem origin of protein. The asterisk indicates that the sequence is used for protein structure prediction. Bootstrap values greater than 50 are shown on the branches. For protein structures, blue represents reference protein structures of bacteria *M. tuberculosis* NAMPT and *Xanthomonas campestris* pv. *campestris* (Xcc) nadD from cryo-electron microscope results; Yellow represents RoseTTAFold-predicted viral protein structures of NAMPT and nadD from different ecosystem types and cultivated phages. The red dot on protein structures indicates bioactive amino acid residues

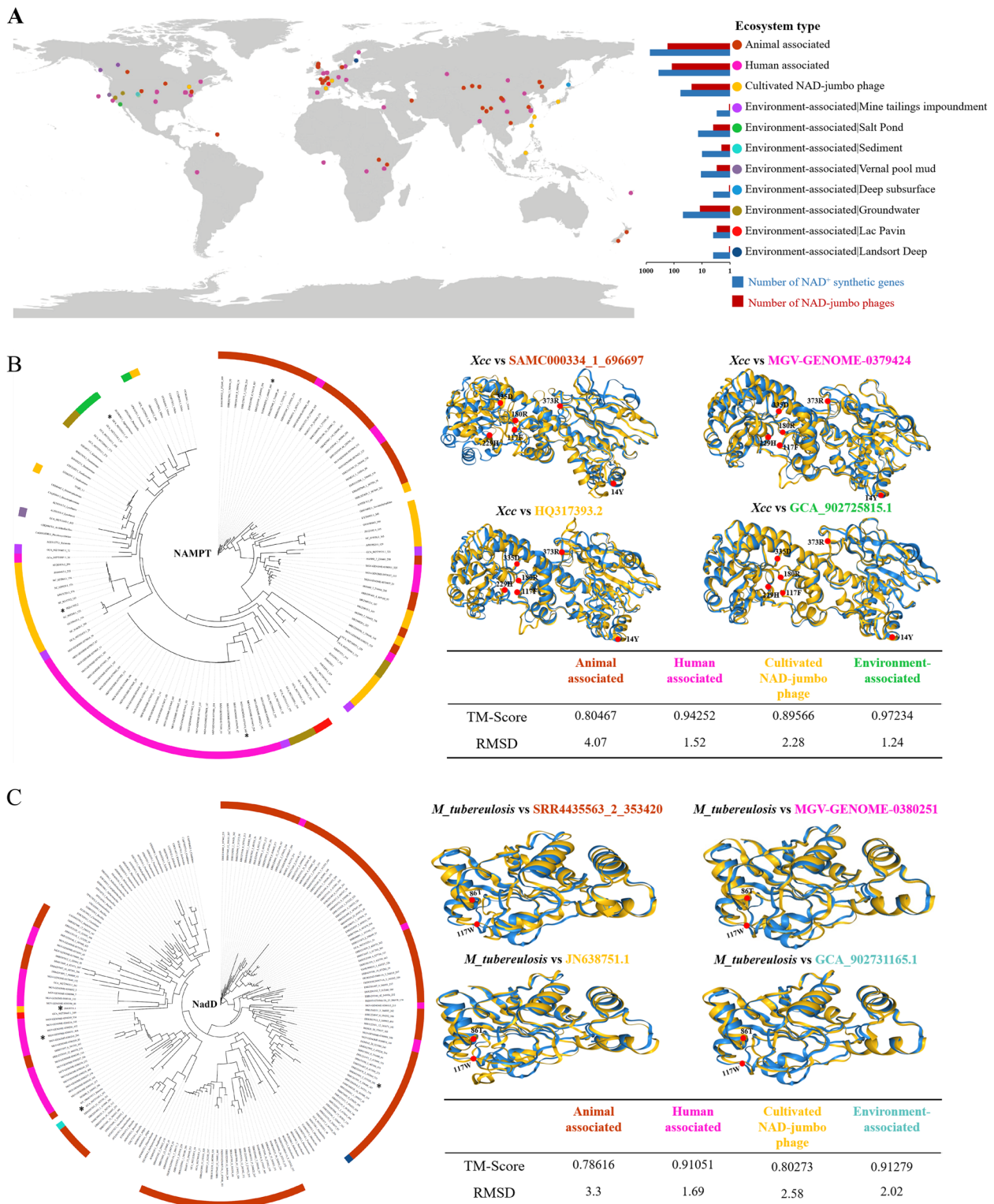


Fig. 4 (See legend on previous page.)

nadD protein for NAD⁺ synthesis in vitro. AutoDock Vina results showed that both the NAMPT proteins from NAD-jumbo phages and KVP40 phage had a high binding affinity for their substrate NAM; both the nadD proteins from NAD-jumbo phages and KVP40 phage had a high binding affinity for their substrate NMN⁺ (Figs. S9 and S10) [71]. Besides the synthesis genes of NAD⁺, the environment- and human-associated NAD-jumbo phages also carried NAD-consuming enzyme genes involved in DNA replication, DNA repair, and counterdefense (Supplementary Table S12), which was consistent with those found in the animal gut. For NAD-jumbo phages in the human gut, the hosts of these phages are mainly concentrated in Bacteroidota (Supplementary Table S13). NAD-jumbo phages in human microbiomes have a complement NAD⁺ synthesis pathway with their hosts (Fig. S11). Collectively, phage-encoded NAD⁺ synthesis and utilization systems are widely distributed across Earth's ecosystems, and potentially assist jumbo phage functions in various environments.

Discussion

Our study indicates that NAD-jumbo phages are widely distributed across Earth's ecosystems. We have successfully assembled a total of 1001 jumbo phage genomes. Among them, 668 non-redundant jumbo phage genomes were high-quality or complete, which is the highest quality and largest quantity of jumbo phage genome resources to date. By integrating 863 jumbo phage genomes from the public database and previous reports [9], we form the global Jumbo phage database (JumbophageDB), which contains 1864 viral draft genomes across ecosystems (Supplementary dataset 2). According to the taxonomic standard of ICTV, the analysis results of VIRIDIC software showed that JumbophageDB contains a total of 1317 genus-level clusters and 1680 species-level clusters of jumbo phage genomes [48, 72]. This resource further expands the unexplored viral diversity not found in other databases, and it will improve the detection of viral reads in the microbiome and may help to investigate their roles in gut health and the natural environment.

It has been reported that bacteriophage *Vibrio* KVP40-encoded NAMPT has shown the highest activity, resulting in a threefold increase in the production of NAD⁺ precursor (NMN), when compared with eight NAMPTs from different sources of eukaryotes and prokaryotes [73]. NMN, as an effective anti-aging product to increase the NAD⁺ levels in humans, ameliorates the detrimental effects of NAD⁺ reduction with age and significantly improves longevity as well as ameliorates age-related complications [74]. KVP40-encoded NAMPT has been used for large-scale production of NMN in engineering bacteria. NAD-jumbo phages include 71 NAMPTs and

other hundreds of NMN synthesis genes (*nadR*, *nadM*, *nadD*) (Table S4), which provides a valuable gene repository for producing active NMN by genetically engineered bacteria or yeast.

NAD-jumbo phages possess expanded genetic repertoires including all enzyme genes required for the synthesis of NAD⁺ through the salvage pathway and the Preiss-Handler pathway rather than the de novo pathway. It seems to be related to the high efficiency of NAD⁺ synthesis by the salvage pathway and the Preiss-Handler pathway [75]. Two studies have reported that phage-encoded enzymes of NAD⁺ synthesis show high biological activity in vitro and in bacteria [71, 73], implying that NAD-jumbo phages may synthesize true NAD⁺ in hosts. In addition, NAD-jumbo phages also encode at least 15 types of NAD⁺-consuming enzyme genes involved in DNA replication, DNA repair, and counterdefense, suggesting that phages have the capacity to redirect NAD⁺ metabolism towards phage propagation needs in bacterial hosts. The remodeling of NAD⁺ metabolism is aligned with the findings for viruses infecting archaea and cyanobacteria which redirect the chemical energy (ATP) synthesis to increase viral production [76]. The remodeling metabolism of NAD-jumbo phages may represent a survival strategy or counterdefense system against bacterial hosts and might be used to improve human and animal health. For example, phages equipped with NAD⁺ synthetic pathways may be used for regulating gut microbiome or killing drug-resistant bacteria.

The host of many NAD-jumbo phages were Bacteroidota which are dominant symbiotic bacteria in the gut of animals and humans, suggesting that NAD-jumbo phage may be an important component of the intestinal microbiome. Moreover, NAD-jumbo phages have complemented the NAD⁺ synthesis pathway with their bacterial hosts in animal and human gut, implying that NAD-jumbo phages may affect NAD⁺ synthesis of gut bacteria and even mammalian cells. Recent studies report trans-kingdom cooperation of NAD⁺ synthesis between the gut microbiome and mammalian cells wherein gut microbiota boost NAD⁺ biosynthesis of the mammalian host, which suggests the important role of this interaction in NAD-mediated antiaging or antitumor [77–79]. Therefore, the metabolic interaction between NAD-jumbo phages and gut core bacteria (such as Bacteroidota) discovered in the present study has some important implications for NAD-mediated human health. Firstly, to improve the level of NAD⁺ in tissues for antiaging in humans, future studies need to consider the effects of gut phage NAD⁺ metabolism on mammalian NAD⁺ biosynthesis. Secondly, NAMPT or *pncA* expression of gut bacteria promotes mammalian NAD synthesis [80], which may lead to tumor cell resistance and tumor cell survival

or migration. Therefore, phage-encoded NAMPT or *pncA* genes likely affect microbe-mediated drug resistance in tumor microbiome [81].

It is important to note that although large-scale metagenomic data analysis has been performed and a considerable number of jumbo phage genomes have been identified, the total number of available jumbo phage genomes remains limited. This restricts our ability to explore and gain insights into other conserved features and diverse functions of jumbo phages. In particular, the results and conclusions presented in this study are based on analyses performed using bioinformatics tools. The NAD⁺ synthesis and utilization in jumbo phages require further experimental validation.

Overall, there is still a poor understanding of jumbo phages, and our expanded high-quality jumbo phages provide important resources for jumbo phage biology. Al-Shayeb et al. suggested that jumbo phages are completely different from small phages and more similar to symbiotic bacteria, blurring the distinctions between living and non-living [9]. Our findings reinforce the point given that one of the characteristics of life is the ability to produce metabolic energy such as NAD⁺ to sustain themselves. In summary, our study of NAD-jumbo phages provides a greater understanding of life evolution and phage-host interactions.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01984-w>.

Supplementary Material 1. Figure S1. Pipeline for high-quality jumbo phage construction. Figure S2. Genome size distribution of jumbo phage in the gut of animal. Figure S3. Distribution of genome completeness and classification of jumbo phage genomes into quality tiers. The abscissa represents the genomes length, and the ordinate represents the genomes completeness. Figure S4. Genome circos map of 716.8kbp and 713.6kbp jumbo phages. Figure S5. Sequence alignment of tRNAs encoded by jumbo phages and hosts. Figure S6. Network images of the newly assembled jumbo phages based on shared protein ortholog families. vConTACT2 v0.11.3 was used for clustering each phage. Our jumbo phages were highlighted in different colors. Figure S7. Genetic maps of NAD-jumbo phages. The white graph represents the unannotated protein. Figure S8. The phylogeny of *ligA/B* genes encoded by jumbo phages. The *ligA/B* genes encoded by the jumbo phages in this study are shown in green. Figure S9. Docking of NAM to NAMPT protein verified by Lee et al. (A) and encoded by NAD jumbo phages (B, C). Figure S10. Docking of NMN to *nadD* protein verified by Lee et al. (A) and encoded by NAD jumbo phages (B, C). Figure S11. NAD⁺ biosynthesis pathways are complementary between NAD-jumbo phages and their bacterial hosts. NAD-jumbo phages from human gut are paired with their hosts. The purple circles represent the presence of the corresponding enzyme in the NAD-jumbo phages. The green circles represent the presence of the corresponding enzyme in the hosts. The space circles represent the absence of the corresponding enzyme in the genomes.

Supplementary Material 2. Supplementary Table 1: All detailed information on samples and metagenomic data. Supplementary Table 2: Basic characteristics of the assembled jumbo phages. Supplementary Table 3: Putative transfer RNA information encoded by the jumbo phages. Supplementary Table 4: ICTV taxonomic annotation of gut jumbo phage

genomes. Supplementary Table 4: Functional annotation of our assembled jumbo phage genomes. Supplementary Table 5: Clade distribution of our assembled jumbo phages in phylogenetic trees. Supplementary Table 6: Proto-spacer and host predicted in the jumbo phage genomes. Supplementary Table 7: Basic information and gene function annotation of NAD-jumbo phages. Supplementary Table 8: The NAD-jumbo phages encode a large number of NAD-dependent enzymes and NAD⁺ cofactors. Supplementary Table 9: Jumbo phages and NAD-jumbo phages assembled from the human body samples. Supplementary Table 10: Basic information of jumbo phages and NAD-jumbo phages assembled from metagenomic data of different environments. Supplementary Table 11: Genomic annotation of NAD-jumbo phages from different environmental metagenomic samples. Supplementary Table 12: Host prediction of NAD-jumbo phages assembled from human metagenomic samples. Supplementary dataset 1-2. Supplementary dataset 1: animal gut MAG dataset contains 29,599 genomes of bacteria and archaea (DOI:10.5281/zenodo.11127795). Supplementary dataset 2: JumbophageDB includes 1864 draft genomes of jumbo phages (DOI:10.5281/zenodo.11127795).

Acknowledgements

Not applicable.

Authors' contribution

C.L., K.L., W.N., M.L. and S.H. designed the research. C.L., K.L. and C.G. performed analysis. C.L., K.L., M.L., P.Z., Z.S., X.L. and L.W. performed the collection of samples. W.N., M.L. and S.H. supervised the study. C.L., K.L., W.N., M.L. and S.H. wrote the manuscript. C.L., M.L. L.C and S.H. improved the manuscript. All authors approved the final version of the manuscript.

Funding

This work was supported by the Science and Technology Program of Xinjiang Production and Construction Corps (2024AB050 to S.W.H), the Tianshan Talent Project [to S.W.H and to W. N], the Tianchi Talent Project [to C.Y.L and to X.Y.L], the National Natural Science Foundation of China [32360016 to C.Y.L, 32225003 and 32393971 to M.L], the National Key Research and Development Program of China (2022YFA0912200 to M.L), Guangdong Major Project of Basic and Applied Basic Research (2023B0303000017 to M.L) and Shenzhen University 2035 Program for Excellent Research (2022B002 to M.L).

Data availability

All of the original sequences obtained in this work have been deposited in the National Center for Biotechnology Information (NCBI) under project number PRJNA1069655, PRJNA1089803 and PRJNA1116133. All the original sequences published in this work can be found in the NCBI (under project number PRJNA291523, PRJNA322715, PRJNA631951, PRJNA656673, PRJNA526070, PRJNA448333, PRJNA214227, PRJNA647157, PRJNA702231, PRJNA202380 and PRJNA657455), the European Nucleotide Archive (under project number PRJEB10338, PRJEB21624, PRJEB31266, PRJEB39057, PRJEB11755, PRJEB12797 and PRJEB34458), the Genome Sequence Archive (PRJCA000114) and the Beijing Institute of Genomics under project PRJCA002018. Supplementary dataset 1 and 2 have been deposited in the Zenodo (DOI:10.5281/zenodo.11127795).

Declarations

Ethics approval and consent to participate

All of the procedures involving animals were approved by the Animal Care Committee of Shihezi University. The study was performed in accordance with the ethical standards established in the 1964 Declaration of Helsinki and subsequent amendments.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 17 May 2024 Accepted: 20 November 2024
Published online: 20 December 2024

References

- Chevallereau A, Pons BJ, van Houte S, Westra ER. Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol.* 2022;20(1):49–62.
- Kieft K, Breister AM, Huss P, Linz AM, Zanetakos E, Zhou Z, et al. Virus-associated organosulfur metabolism in human and environmental systems. *Cell Rep.* 2021;36: 109471.
- Chee MSJ, Serrano E, Chiang YN, Harling-Lee J, Man R, Bacigalupe R, et al. Dual pathogenicity island transfer by piggybacking lateral transduction. *Cell.* 2023;186(16):3414–26.
- Yan M, Pratama AA, Somasundaram S, Li Z, Jiang Y, Sullivan MB, et al. Interrogating the viral dark matter of the rumen ecosystem with a global virome database. *Nat Commun.* 2023;14(1):5254.
- Li C, Li X, Guo R, Ni W, Liu K, Liu Z, et al. Expanded catalogue of metagenome-assembled genomes reveals resistome characteristics and athletic performance-associated microbes in horse. *Microbiome.* 2023;11(1):7.
- Gilbert RA, Townsend EM, Crew KS, Hitch TC, Friedersdorff JC, Creevey CJ, et al. Rumen virus populations: technological advances enhancing current understanding. *Front Microbiol.* 2020;11: 514980.
- Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Inter-species cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat Microbiol.* 2018;3(11):1274–84.
- Yuan Y, Gao M. Jumbo bacteriophages: an overview. *Front Microbiol.* 2017;8: 226872.
- Al-Shayeb B, Sachdeva R, Chen LX, Ward F, Munk P, Devoto A, et al. Clades of huge phages from across Earth's ecosystems. *Nature.* 2020;578(7795):425–31.
- Weinheimer AR, Aylward FO. Infection strategy and biogeography distinguish cosmopolitan groups of marine jumbo bacteriophages. *ISME J.* 2022;16(6):1657–67.
- Chen LX, Méheust R, Crits-Christoph A, McMahon KD, Nelson TC, Slater GF, et al. Large freshwater phages with the potential to augment aerobic methane oxidation. *Nat Microbiol.* 2020;5(12):1504–15.
- Devoto AE, Santini JM, Olm MR, Anantharaman K, Munk P, Tung J, et al. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nat Microbiol.* 2019;4(4):693–700.
- Crisci MA, Chen LX, Devoto AE, Borges AL, Bordin N, Sachdeva R, et al. Closely related Lak megaphages replicate in the microbiomes of diverse animals. *Iscience.* 2021;24(8):102875.
- Shkoporov AN, Hill C. Bacteriophages of the human gut: the “known unknown” of the microbiome. *Cell Host Microbe.* 2019;25(2):195–209.
- Griffith GW, Ozkose E, Theodorou MK, Davies DR. Diversity of anaerobic fungal populations in cattle revealed by selective enrichment culture using different carbon sources. *Fungal Ecol.* 2009;2(2):87–97.
- Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. *Cell.* 2021;184(4):1098–109 e9.
- Chen C, Zhou Y, Fu H, Xiong X, Fang S, Jiang H, et al. Expanded catalog of microbial genes and metagenome-assembled genomes from the pig gut microbiome. *Nat Commun.* 2021;12(1):1106.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–90.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017;27(5):824–34.
- Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome.* 2017;5:1–20.
- 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;9:1–13.
- 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;8:1–23.
- Nayfach S, Camargo AP, Schulz F, Eloie-Fadrosch E, Roux S, Kyrpidis NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol.* 2021;39(5):578–85.
- Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 2017;11(12):2864–8.
- Hyatt D, Chen GL, LoCasio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010;11:1–11.
- Grazziotin AL, Koonin EV, Kristensen DM. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic acids Res.* 2016;45: gkw975.
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30.
- 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):D427–32.
- Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research.* 2021;49(16):9077–96.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution.* 2016;33(7):1870–4.
- Bin Jang H, 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.
- Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, et al. Linking virus genomes with host taxonomy. *Viruses.* 2016;8(3): 66.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25(15):1972–3.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59–60.
- MinCED-Mining CRISPRs in Environmental Datasets 2022. Available online: <https://github.com/ctSkenneron/minced>. Accessed 15 Jan 2023.
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics.* 2022;38(23):5315–6.
- Guillemet M, Chabas H, Nicot A, Gatchich F, Ortega-Abboud E, Buus C, et al. Competition and coevolution drive the evolution and the diversification of CRISPR immunity. *Nature Ecology & Evolution.* 2022;6(10):1480–8.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
- Baek M, DiMaio F, Anishchenko I, Duparaj J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;373(6557):871–6.
- Rodionova IA, Zuccola HJ, Sorci L, Aleshin AE, Kazanov MD, Ma CT, et al. Mycobacterial nicotinate mononucleotide adenylyltransferase: structure, mechanism, and implications for drug discovery. *J Biol Chem.* 2015;290(12):7693–706.
- Xu G, Ma J, Fang Q, Peng Q, Jiao X, Hu W, et al. Structural insights into *Xanthomonas campestris* pv. *campestris* NAD⁺ biosynthesis via the NAM salvage pathway. *Communications Biology.* 2024;7(1):255.
- Barrio-Hernandez I, Yeo J, Jänes J, Mirdita M, Gilchrist CLM, Wein T, et al. Clustering predicted structures at the scale of the known protein universe. *Nature.* 2023;622(7983):637–45.
- Li W, Guan Y, Shi L, Chen Y, Huang H, Zhen H, et al. Identification of angiotensin-converting enzyme inhibitory peptides from peanut meal (*Arachis hypogaea* Linn) fermented by *Lactobacillus pentosus* using MALDI-TOF-MS and LC-MS/MS. *Food Frontiers.* 2024;5(2):820–32.
- Nayfach S, Páez-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(7):960–70.
- Benler S, Yutin N, Antipov D, Rayko M, Shmakov S, Gussow AB, et al. Thousands of previously unknown phages discovered in whole-community human gut metagenomes. *Microbiome.* 2021;9:1–17.
- Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. *Viruses.* 2021;13(3):506.
- Camargo AP, Nayfach S, Chen IMA, Palaniappan K, Ratner A, Chu K, et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* 2023;51(D1):D733–43.

50. Paez-Espino D, Eloee-Fadrosch EA, Pavlopoulos GA, Thomas AD, Hunte-mann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature*. 2016;536(7617):425–30.
51. Wang J, Zhang Y, Mendonca CA, Yukselen O, Muneeruddin K, Ren L, et al. AAV-delivered suppressor tRNA overcomes a nonsense mutation in mice. *Nature*. 2022;604(7905):343–8.
52. Pandolfo M, Telatin A, Lazzari G, Adriaenssens EM, Vitulo N. Meta-Phage: an automated pipeline for analyzing, annotating, and classifying bacteriophages in metagenomics sequencing data. *Msystems*. 2022;7(5):e00741–822.
53. De Smet J, Hendrix H, Blasdel BG, Danis-Wlodarczyk K, Lavigne R. Pseudomonas predators: understanding and exploiting phage–host interactions. *Nat Rev Microbiol*. 2017;15(9):517–30.
54. Ajunwa OM, Odeniyi OA, Garuba EO, Nair M, Marsili E, Onilude AA. Evaluation of extracellular electron transfer in *Pseudomonas aeruginosa* by co-expression of intermediate genes in NAD synthetase production pathway. *World J Microbiol Biotechnol*. 2022;38(5):90.
55. Li F, Li YX, Cao YX, Wang L, Liu CG, Shi L, et al. Modular engineering to increase intracellular NAD (H/+) promotes rate of extracellular electron transfer of *Shewanella oneidensis*. *Nat Commun*. 2018;9(1):3637.
56. Chen Y, Ying Y, Lalsiamthara J, Zhao Y, Imani S, Li X, et al. From bacteria to biomedicine: Developing therapies exploiting NAD⁺ metabolism. *Bioorg Chem*. 2024;142:106974.
57. Sanyal G, Doig P. Bacterial DNA replication enzymes as targets for anti-bacterial drug discovery. *Expert Opin Drug Discov*. 2012;7(4):327–39.
58. Yi L, Lü X. New strategy on antimicrobial-resistance: Inhibitors of DNA replication enzymes. *Curr Med Chem*. 2019;26(10):1761–87.
59. Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK. The crystal structure of DNA mismatch repair protein MutS binding to a G-T mismatch. *Nature*. 2000;407(6805):711–7.
60. Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol*. 2017;18(10):610–21.
61. Macintyre G, Doiron KM, Cupples CG. The Vsr endonuclease of *Escherichia coli*: an efficient DNA repair enzyme and a potent mutagen. *J Bacteriol*. 1997;179(19):6048–52.
62. Balasingham SV, Zegeye ED, Homberset H, Rossi ML, Laerdahl JK, Bohr VA, et al. Enzymatic activities and DNA substrate specificity of *Mycobacterium tuberculosis* DNA helicase XPB. *PLoS ONE*. 2012;7(5): e36960.
63. Warring SL, Malone LM, Jayaraman J, Easingwood RA, Rigano LA, Frampton RA, et al. A lipopolysaccharide-dependent phage infects a pseudomonad phytopathogen and can evolve to evade phage resistance. *Environ Microbiol*. 2022;24(10):4834–52.
64. Ongena V, Briegel A, Claessen D. Cell wall deficiency as an escape mechanism from phage infection. *Open Biol*. 2021;11(9): 210199.
65. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nat Rev Microbiol*. 2010;8(5):317–27.
66. Stokar-Avihail A, Fedorenko T, Hör J, Garb J, Leavitt A, Millman A, et al. Discovery of phage determinants that confer sensitivity to bacterial immune systems. *Cell*. 2023;186(9):1863–76 e16.
67. Garb J, Lopatina A, Bernheim A, Zaremba M, Siksnys V, Melamed S, et al. Multiple phage resistance systems inhibit infection via SIR2-dependent NAD⁺ depletion. *Nat Microbiol*. 2022;7(11):1849–56.
68. Tamulaitiene G, Sabonis D, Sasnauskas G, Ruksenaite A, Silanskas A, Avraham C, Ofir G, Sorek R, Zaremba M, Siksnys V. Activation of Thoreris antiviral system via SIR2 effector filament assembly. *Nature*. 2024;627(8003):431–6. <https://doi.org/10.1038/s41586-024-07092-x>.
69. Osterman I, Samra H, Rousset F, Loseva E, Itkin M, Malitsky S, Yirmiya E, Millman A, Sorek R. Phages reconstitute NAD⁺ to counter bacterial immunity. *Nature*. 2024;634(8036):1160–7. <https://doi.org/10.1038/s41586-024-07986-w>.
70. Huang X, Jiao N, Zhang R. The genomic content and context of auxiliary metabolic genes in roseophages. *Environ Microbiol*. 2021;23(7):3743–57.
71. Lee JY, Li Z, Miller ES. *Vibrio* phage KVP40 encodes a functional NAD⁺ salvage pathway. *Journal of bacteriology*. 2017;199(9):10.1128.
72. Moraru C, Varsani A, Kropinski AM. VIRIDIC—A novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. *Viruses*. 2020;12(11): 1268.
73. Huang Z, Li N, Yu S, Zhang W, Zhang T, Zhou J. Systematic engineering of *Escherichia coli* for efficient production of nicotinamide mononucleotide from nicotinamide. *ACS Synth Biol*. 2022;11(9):2979–88.
74. Covarrubias AJ, Perrone R, Grozio A, Verdin E. NAD⁺ metabolism and its roles in cellular processes during ageing. *Nat Rev Mol Cell Biol*. 2021;22(2):119–41.
75. Yang L, Mu X, Nie Y, Xu Y. Improving the production of NAD⁺ via multi-strategy metabolic engineering in *Escherichia coli*. *Metab Eng*. 2021;64:122–33.
76. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology journal*. 2019;16:1–13.
77. Rook G, Bäckhed F, Levin BR, McFall-Ngai MJ, McLean AR. Evolution, human-microbe interactions, and life history plasticity. *The Lancet*. 2017;390(10093):521–30.
78. Shkoporov AN, Turkington CJ, Hill C. Mutualistic interplay between bacteriophages and bacteria in the human gut. *Nat Rev Microbiol*. 2022;20(12):737–49.
79. Piel D, Bruto M, Labreuche Y, Blanquart F, Goudenège D, Barcia-Cruz R, et al. Phage–host coevolution in natural populations. *Nat Microbiol*. 2022;7(7):1075–86.
80. Khaidizar FD, Bessho Y, Nakahata Y. Nicotinamide phosphoribosyltransferase as a key molecule of the aging/senescence process. *Int J Mol Sci*. 2021;22(7): 3709.
81. Pylaeva E, Harati MD, Spyra I, Bordbari S, Strachan S, Thakur BK, et al. NAMPT signaling is critical for the proangiogenic activity of tumor-associated neutrophils. *Int J Cancer*. 2019;144(1):136–49.

Publisher's Note

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