Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Isolation, genomic and comparative proteomic characterization of three lytic Pseudomonas aeruginosa bacteriophages from wastewater in Bobo-Dioulasso, Burkina Faso

Kobo Gnada
Nazi Boni University
Edmond Wend-Guudi Silga
Nazi Boni University
Christiane You Essoh
Péléforo Gon Coulibaly University
Raymond Karlhis Yao
Nazi Boni University
Arnaud Quetin Sanou
Nazi Boni University
Aïda Ouangraoua
University of Sherbrooke
Michel Kiréopori Gomgnimbou
gomikir@yahoo.fr

Nazi BONI University

Research Article

Keywords: bacteriophage, Pseudomonas aeruginosa, wastewater, comparative genomics, phage therapy, antimicrobial resistance, Burkina Faso

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11433369/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Isolation, genomic and comparative proteomic characterization of three lytic *Pseudomonas aeruginosa* bacteriophages from wastewater in Bobo-Dioulasso, Burkina Faso

Kobo Gnada[1,2,3], Edmond Wend-Guudi Silga[1,2], Christiane You Essoh[5], Raymond Karlhis Yao[1,2,3], Arnaud Quetin Sanou[1,2], Aïda Ouangraoua[6], Michel Kiréopori Gomgnimbou[1,2,3,4*]

[1] Doctoral School of Natural and Agricultural Sciences, Nazi BONI University, Bobo-Dioulasso, Burkina Faso

[2] Molecular Biology and Biotechnology Team, Research Laboratory, MURAZ Center, National Institute of Public Health, Bobo-Dioulasso, Burkina Faso

[3] African Center of Excellence in Biotechnological Innovations for the Elimination of Vector-Borne Diseases (CEA/ITECH-MTV), Nazi BONI University, 01 BP 1091 Bobo-Dioulasso, Burkina Faso

[4] Higher Institute of Health Sciences (IN.S.SA), Nazi BONI University, 01 BP 1091 Bobo-Dioulasso, Burkina Faso

[5] Department of Biochemistry and Genetics, Biological Sciences Division, Péléforo Gon Coulibaly University, Korhogo, Côte d'Ivoire

[6] Department of Computer Science, University of Sherbrooke, Sherbrooke, Québec, Canada

***Corresponding author:** Michel Kiréopori Gomgnimbou, Molecular Biology and Biotechnology Team, Research Laboratory, MURAZ Center, National Institute of Public Health, Bobo-Dioulasso, Burkina Faso. E-mail: gomikir@yahoo.fr. Tel.: +226 70933940. ORCID: 0009-0009-5214-8895

E-mail addresses of all authors: KG: kobognada@gmail.com; EWS: silgaedmond@gmail.com; CYE: christianeessoh13@gmail.com; RKY: karlhisyao@gmail.com; AQS: arnaudsanou07@gmail.com; AO: aida.ouangraoua@usherbrooke.ca; MKG: gomikir@yahoo.fr

Abstract

Background. *Pseudomonas aeruginosa* is a leading cause of opportunistic and healthcare-associated infections and is listed by the World Health Organization among the pathogens for which new therapeutic options are most urgently needed. Lytic bacteriophages are one of the few alternatives that can be deployed rapidly and at low cost, but phage collections and phage genome sequences remain very unevenly distributed geographically, and West Africa is particularly under-represented. Wastewater, which combines high bacterial densities with continuous inputs of faecal, hospital and industrial effluent, is the most productive environment for recovering phages that are locally adapted to circulating clinical strains. We therefore set out to isolate lytic phages active against clinical *P. aeruginosa* isolates from wastewater in Bobo-Dioulasso, Burkina Faso, and to characterize their genomes.

Results. Twenty-eight lytic phages were isolated from five wastewater sites using six *P. aeruginosa* host strains, and screened against a panel of eleven clinical and reference strains. Host ranges were broad: eleven isolates lysed at least ten of the eleven strains and three lysed all eleven, while only two isolates were restricted to four strains. Five phages were selected for whole-genome sequencing and three yielded complete, high-quality assemblies. All three

belong to the class *Caudoviricetes*. vB_PaeM_CM0001 (57,614 bp; 63.5% GC; 88 CDSs) is a member of the family *Mesyanzhinovviridae*, subfamily *Bradleyvirinae*, genus *Abidjanvirus*. vB_PaeP_CM0002 (43,491 bp) and vB_PaeP_CM0003 (43,177 bp), each encoding 65 CDSs with 62.3–62.6% GC, belong to the family *Autographiviridae*, subfamily *Krylovirinae*, genus *Phikmvvirus*. Intergenomic similarities indicate that vB_PaeM_CM0001 is a new isolate of an existing *Abidjanvirus* species, whereas vB_PaeP_CM0002 and vB_PaeP_CM0003 represent two new *Phikmvvirus* species. The two *Phikmvvirus* genomes share 91.9% nucleotide identity and 62 orthogroups and are fully collinear with each other and with their closest relatives; their divergence is confined to nucleotide substitution and to a single 2.0 kb region spanning the tail fibre locus. vB_PaeM_CM0001 shares no detectable orthogroups with them.

Conclusions. These three genomes are the first *P. aeruginosa* phage genomes reported from Burkina Faso. Two of them represent new species. Their broad host ranges, strictly lytic gene repertoires and absence of resistance determinants make them credible starting points for a national phage bank and for future therapeutic evaluation.

Keywords: bacteriophage; *Pseudomonas aeruginosa*; wastewater; comparative genomics; phage therapy; antimicrobial resistance; Burkina Faso

Background

Antimicrobial resistance (AMR) is one of the major public-health problems of the century[1] . Bacterial AMR was associated with an estimated 4.95 million deaths worldwide in 2019, of which 1.27 million were directly attributable to resistance [2] . In its 2024 Bacterial Priority Pathogens List, the World Health Organization (WHO) again identified carbapenem-resistant *Pseudomonas aeruginosa* as a high-priority target for research and development[3].

P. aeruginosa is a non-fermenting Gram-negative bacillus that is ubiquitous in soil, water, plants and built environments, and a formidable opportunistic pathogen responsible for respiratory, urinary, wound, bloodstream and device-associated infections[4] . Its clinical success rests , among other factors, on its ability to colonize surfaces and form biofilms[5], as well as on a combination of intrinsic, acquired and adaptive resistance mechanisms, low outer-membrane permeability, multidrug efflux pumps, antibiotic-inactivating enzymes and horizontal acquisition of resistance genes[6]. In Burkina Faso, *P. aeruginosa* resistance to several antibiotics used in clinical practice, including cefepime, ceftazidime, and ticarcillin/clavulanic acid, has been identified through passive surveillance[7].

Bacteriophage therapy has attracted renewed attention[1] because antibiotic development is slow, costly and increasingly unable to keep pace with resistance. Phages are viruses that infect and lyse bacteria with high host specificity, so that treatment spares the commensal microbiota, and they can be produced locally at modest cost [9] . Advances in sequencing and in phage genome annotation now make it feasible to screen candidate phages for the properties required of a therapeutic agent, a strictly lytic lifestyle, absence of integrases, and absence of virulence or antimicrobial resistance genes, before any in vivo evaluation[10].

The present work has two main motivations. First, the effectiveness of phage therapy depends heavily on the phage bank available. Since the host range is strain-dependent, collections assembled elsewhere may perform poorly against locally circulating isolates, providing a rationale for current efforts to build regional phage banks in Africa [11]. Second, the phage genomes available in public databases come overwhelmingly from a small number of countries, so that sampling under-represented regions is also a direct contribution to viral diversity and to phage taxonomy. Wastewater is the environment of choice for such sampling, as it concentrates both hosts and their phages [12] .

Here, we report the isolation of 28 lytic *P. aeruginosa* phages from five wastewater sites in Bobo-Dioulasso, western Burkina Faso, the characterization of their host ranges against clinical isolates, and the complete genome sequences, comparative genomic and proteomic analyses, taxonomic placement and in silico safety assessment of three of them. To our knowledge these are the first *P. aeruginosa* phage genomes reported from Burkina Faso.

Results

Bacterial hosts and their antibiotic susceptibility profiles

Eleven *P. aeruginosa* strains were used in this study: ten clinical isolates recovered from urine samples at the Clinical Bacteriology Laboratory of Centre MURAZ, and the reference strain ATCC 27853. Disk-diffusion testing showed a homogeneous phenotype across the panel (Table 1). All eleven strains were susceptible to meropenem, and all were categorized as intermediate ("susceptible, increased exposure") to piperacillin/tazobactam, ceftazidime, cefepime, aztreonam and levofloxacin. Three clinical isolates (PA_CM22, PA_CM24 and PA_CM30) were additionally resistant to ticarcillin and piperacillin, a profile consistent with penicillinase production. All strains remained below the usual threshold for multidrug resistance.

Six of these strains, five clinical isolates and ATCC 27853, were used as hosts for phage enrichment and isolation, and the full panel of eleven was used for host-range determination.

Isolation of lytic phages from wastewater

Wastewater was sampled at five sites in Bobo-Dioulasso: the Souro Sanou University Hospital (SSUH), a wastewater discharge station (WDS), the municipal sewage treatment plant (STP), the Houet River backwater (HR) and the Dogona slaughterhouse (SGH) (Additional file 1: Fig. S1). Lytic phages were recovered from all five sites, on every host strain tested, as clear lysis zones on double-layer LB agar (Additional file 1: Fig. S2). Two host strains, PA_CM22 and PA_CM25, were lysed by phages from all five sites, whereas PA_CM26 was lysed by phages from only two (Table 2).

After three rounds of single-plaque purification, 28 distinct lytic phage isolates were obtained. Plaques fell into two visually distinct size classes (Additional file 1: Fig. S3); large plaques predominated (16/28, 57.1%) over small ones (12/28, 42.9%) (Additional file 1: Table S1).

Host range

The 28 purified isolates were spotted on lawns of the eleven *P. aeruginosa* strains (Additional file 1: Fig. S4). Lysis patterns defined 17 distinct host-range profiles (Additional file 1: Table S2). Host ranges were generally broad: 22 isolates (78.6%) lysed at least eight of the eleven strains (≥72.7%), 11 isolates (39.3%) lysed at least ten strains (≥90.9%), and three isolates (PA_BD_S06, PA_BD_S11 and PA_BD_S12, all sharing profile 1) lysed all eleven (100%). At the other extreme, two isolates (PA_BD_S20a and PA_BD_S21b) lysed only four strains (36.4%).

Five isolates were selected for whole-genome sequencing on the basis of their host-range profiles and plaque phenotypes: PA_BD_S01 and PA_BD_S03 (both profile 2), PA_BD_S12 (profile 1), PA_BD_S21 (profile 9) and PA_BD_S29b (profile 7).

Genome sequencing and assembly outcome

The DNA yields from the five selected isolates, measured by fluorometric assay, were 5.92, 85.0, 1.25, 13.18 and 11.4 ng/µL for PA_BD_S01, PA_BD_S03, PA_BD_S12, PA_BD_S21 and PA_BD_S29b, respectively.

Three of the five libraries produced complete, high-quality assemblies. These correspond to the phages designated vB_PaeM_CM0001, vB_PaeP_CM0002 and vB_PaeP_CM0003, corresponding respectively to PA_BD_S03, PA_BD_S12 and PA_BD_S29b (Table 3). These three are the subject of all analyses below. The two remaining datasets, from PA_BD_S01 and PA_BD_S21, were excluded: one yielded a single contig of less than 5 kb, compatible with co-purified host plasmid DNA rather than a phage genome, and the other produced a highly fragmented assembly with a CheckV completeness below 25%.

General features of the three genomes

The three genomes differ substantially in size and gene content and fall into two groups (Table 3, Fig. 1). vB_PaeM_CM0001 (GenBank PZ231876) is 57,614 bp long with 63.54% GC and 88 predicted CDSs over a coding density of 94.5%. vB_PaeP_CM0002 (PZ365779) and vB_PaeP_CM0003 (PZ365800) are 43,491 bp and 43,177 bp long, with 62.63% and 62.25% GC respectively, and each encodes 65 CDSs at coding densities of 95.6% and 96.2%. vB_PaeM_CM0001 is therefore 14,123 bp and 14,437 bp longer than vB_PaeP_CM0002 and vB_PaeP_CM0003 respectively. The three genomes contained no predicted tRNA genes.

vB_PaeP_CM0002 and vB_PaeP_CM0003 are closely related to each other. They differ in length by only 314 bp, encode the same number of CDSs, share 91.9% genome-wide nucleotide identity (Fig. 3) and share 62 orthogroups, corresponding to approximately 95% of their proteomes (Fig. 6). vB_PaeM_CM0001, in contrast, shares no orthogroup with either of them (Fig. 6) and intergenomic similarity with them is negligible (0.1% and 0.2%; Fig. 3).

The lifestyle prediction classified all three phages as strictly virulent, with confidence scores of 97.33% (vB_PaeM_CM0001), 99.97% (vB_PaeP_CM0002) and 99.99% (vB_PaeP_CM0003) (Table 3).

Functional annotation

The coding sequences were annotated by a three-stage consensus approach (see Methods). Across the three genomes, PHROG assigned 13, 18 and 16 products in vB_PaeM_CM0001, vB_PaeP_CM0002 and vB_PaeP_CM0003 respectively, Phold assigned a further 46, 15 and 17, and blastp against viral RefSeq assigned 3, 0 and 0. A putative function could therefore be assigned to 62 of the 88 CDSs of vB_PaeM_CM0001 (70.5%), to 33 of the 65 CDSs of vB_PaeP_CM0002 (50.8%) and to 34 of the 65 CDSs of vB_PaeP_CM0003 (52.3%), leaving 26, 32 and 31 proteins of unknown function respectively (Fig. 2). The annotation source and functional category of every CDS are given in Additional file 2 (Table S7).

Each CDS was assigned to one of eleven functional categories on the basis of its final product name (Table 4), the same scheme used to colour the genome maps in Fig. 1. In vB_PaeM_CM0001 the largest annotated categories are tail assembly and other/auxiliary functions (14 CDSs, 15.9% each), followed by DNA metabolism (11 CDSs, 12.5%) and DNA replication and repair (8 CDSs, 9.1%). The two *Phikmvvirus* genomes share an almost identical profile, dominated by DNA replication and repair (8 and 9 CDSs, 12.3% and 13.8%), with equal numbers of CDSs in head and capsid assembly (5 each, 7.7%) and baseplate and host recognition (5 each, 7.7%), and three CDSs each in DNA injection and in host lysis. The expanded DNA metabolism repertoire in vB_PaeM_CM0001, including three ribonucleotide reductase subunits, thymidylate synthase, dCMP deaminase, and 5-hmdU DNA kinase, distinguishes the two families and is consistent with the larger genome size of vB_PaeM_CM0001.

The annotated products of the three genomes contained no integrase, excisionase, recombinase, or lysogenic transcriptional repressor.

Comparative genomics with the closest available relatives

For each isolate, the five most similar genomes in GenBank were retrieved and ranked by a combined identity × coverage score (Additional file 1: Table S3). All fifteen reference genomes shared more than 96% nucleotide identity over at least 93% of the query length, and differed in length from the corresponding isolate by roughly 1 kb or less.

VIRIDIC intergenomic similarities (Fig. 3) placed vB_PaeM_CM0001 at 90.7–96.3% similarity to its closest relatives, within the same species cluster as *Pseudomonas* phages vB_PaeS_PAO1_Ab20 (LN610585.1) and vB_PaeS_PAO1_Ab18 (NC_026594.1). vB_PaeP_CM0002 and vB_PaeP_CM0003 reached 94.1–94.6% and 90.8–94.6% similarity respectively to their closest relatives, and each fell into its own species cluster.

Whole-genome comparisons with Easyfig (Fig. 4) showed extensive collinearity between each isolate and its reference genomes, interrupted by localized insertions, deletions and gene-order shifts.

Progressive Mauve alignment was performed on genomes normalized for strand and origin (see Methods). Within the Autographiviridae, a single locally collinear block (LCB) spanned the full length of vB_PaeP_CM0002, vB_PaeP_CM0003, vB_PaePA10145phi1_HR2 and MPK7, all in the same orientation (Fig. 5a). Within the Mesyanzhinovviridae, three LCBs were

resolved, occurring in the same order and the same orientation in vB_PaeM_CM0001, vB_PaeS_PAO1_Ab20 and PAKlein3 (Fig. 5b). The alignments did not contain any inverted blocks, and the genomes showed no evidence of structural rearrangements.

The divergence between vB_PaeP_CM0002 and vB_PaeP_CM0003 is limited to nucleotide substitution and to one localized region. The sliding-window sequence identity along the alignment (mean 93.7% in 250 bp windows) fell below 80% over a single interval of approximately 2.0 kb, where it averaged 56.5%. This interval corresponds to the tail fibre locus, comprising CM0002_0054–CM0002_0057 in vB_PaeP_CM0002 and CM0003_0008–CM0003_0011 in vB_PaeP_CM0003. The gene content differs across it: vB_PaeP_CM0002 encodes four annotated tail fiber proteins, whereas vB_PaeP_CM0003 encodes three together with a hypothetical protein of comparable length occupying the equivalent position. The same interval is the most divergent locus between vB_PaeP_CM0002 and MPK7 (56.4% mean identity) and between vB_PaeP_CM0002 and vB_PaePA10145phi1_HR2.

Comparative proteomics

The OrthoFinder clustering of the three isolates against their closest relatives (Table 5) showed that 79 of the 88 proteins of vB_PaeM_CM0001 (89.8%), 63 of the 65 proteins of vB_PaeP_CM0002 (96.9%) and 61 of the 65 proteins of vB_PaeP_CM0003 (93.8%) are shared with at least one relative. Each isolate nevertheless has unique proteins: nine in vB_PaeM_CM0001, two in vB_PaeP_CM0002 and four in vB_PaeP_CM0003. Conversely, one reference protein is absent from vB_PaeM_CM0001, seven from vB_PaeP_CM0002 and two from vB_PaeP_CM0003.

The orthogroup analysis restricted to the three isolates (Fig. 6) confirmed the two-group structure. vB_PaeP_CM0002 and vB_PaeP_CM0003 share 62 orthogroups and have three singletons each, whereas vB_PaeM_CM0001 shares none of its 88 proteins with them. The ViPTree proteomic distance tree (Fig. 7) reproduced this separation, placing vB_PaeM_CM0001 within the *Mesyanzhinovviridae* and vB_PaeP_CM0002/vB_PaeP_CM0003 within the *Autographiviridae*.

Phylogenetic relationships and taxonomic assignment

We carried out a BLASTn analysis, with confirmation by PhaBOX, which assigned all three phages to the class *Caudoviricetes* (Table 3). vB_PaeM_CM0001 belongs to the subfamily *Bradleyvirinae*, genus *Abidjanvirus*, within the family *Mesyanzhinovviridae*. vB_PaeP_CM0002 and vB_PaeP_CM0003 belong to the subfamily *Krylovirinae*, genus *Phikmvvirus*, within the family *Autographiviridae*.

Using the ICTV thresholds of 70% intergenomic similarity for genus and 95% for species, the VIRIDIC results (Fig. 3) show that vB_PaeM_CM0001 shares more than 95% similarity with vB_PaeS_PAO1_Ab18 and vB_PaeS_PAO1_Ab20, indicating that it is a new isolate of an existing *Abidjanvirus* species. In contrast, vB_PaeP_CM0002 and vB_PaeP_CM0003 each show less than 95% similarity to all other genomes and to each other, supporting their classification as two new species within the genus *Phikmvvirus*.

The proteome-based ViPTree analysis (Fig. 7) placed vB_PaeM_CM0001 among the *Mesyanzhinovviridae* reference phages, next to vB_PaeS_PAO1_Ab18, vB_PaeS_PAO1_Ab20 and vB_PaeS_PAO1_Ab05, and vB_PaeP_CM0002 and vB_PaeP_CM0003 adjacent to each other within the *Autographiviridae*, close to vB_PaeP_PPA-ABTNL and MPK7. All three isolates fall within the *Pseudomonadota* host group.

The GBDP phylogenomic tree inferred with VICTOR (Fig. 8) recovered two well-separated clades corresponding to the two families, with 100% bootstrap support for the node separating them. Within the *Autographiviridae* clade, vB_PaeP_CM0002 grouped with vB_PaePA10145phi1_HR2, HX1, BrSP08, Atpa010 and Atpa011, while vB_PaeP_CM0003 grouped with PaFZ4 and Henu_18-20N, close to MPK7 and vB_PaBD_211. The OPTSIL clustering shown alongside the tree assigns vB_PaeP_CM0002 and vB_PaeP_CM0003 to distinct species clusters, in agreement with the VIRIDIC result.

A maximum-likelihood tree built from a conserved structural marker, the major capsid protein (Fig. 9) recovered the expected topology, with vB_PaeM_CM0001 within the *Abidjanvirus* clade and vB_PaeP_CM0002 and vB_PaeP_CM0003 as sister taxa within the *Phikmvvirus* clade.

In silico safety assessment

The three genomes contained no antimicrobial resistance genes, anti-CRISPR proteins, or Acr–Aca operons. Together with the absence of integrase, excisionase and repressor genes, these results are consistent with the strictly virulent lifestyle predicted for all three phages.

A single CDS of vB_PaeM_CM0001, CM0001_0013 (9,749–10,609), returned a structural match against VFDB during Phold annotation and was labelled accordingly in the primary annotation. The match is to VFDB entry VFG015931 at an amino-acid identity of 12.4%, the lowest of any functionally labelled CDS in this genome, against a median of 94.2% for the remaining assignments. Phold's aggregated structural evidence for this protein assigns 74.5% of the total bitscore weight to the "other" category and 16.9% to DNA, RNA and nucleotide metabolism, against 8.5% for the moron, auxiliary metabolic gene and host takeover category to which VFDB matches are attributed. The protein carries two diagnostic FAD-binding motifs, a Rossmann dinucleotide-binding fingerprint at residue 7 and the second conserved flavin-binding region at residue 132, and lies within a contiguous nucleotide-metabolism module that includes ribonucleotide reductase, 5-hmdU DNA kinase, thymidylate synthase and dATP triphosphohydrolase genes. The match therefore reflects the Rossmann fold shared between this protein and flavin-dependent enzymes catalogued in VFDB rather than a virulence function, and the CDS is annotated here as an FAD-dependent oxidoreductase.

Discussion

We isolated 28 lytic phages active against clinical *P. aeruginosa* from five wastewater sites in Bobo-Dioulasso and sequenced three of them to completion. The three genomes fall into two very different groups, one *Abidjanvirus* (*Mesyanzhinovviridae*) and two *Phikmvvirus* (*Autographiviridae*), that share no detectable orthologous protein families.These three

represent new species. The three phages lack resistance determinants, virulence factors, and lysogeny-related genes.

Wastewater as a source of locally adapted *P. aeruginosa* phages

The phages were recovered from every site sampled, which is consistent with their status as the most abundant biological entities on the planet, with global population estimates of 10^{31}–10^{32} particles, roughly ten times the number of bacterial cells[13]. Similar recovery rates from wastewater have been reported for *P. aeruginosa* phages elsewhere[12] . Wastewater combines high bacterial density with continuous inputs of faecal, hospital and industrial material[12], and therefore acts as an interface where phage–host coevolution is sustained[14].

The isolation frequencies were highest at the Dogona slaughterhouse and at the municipal treatment plant, reproducing the pattern we previously observed for *S. aureus* phages sampled at the same sites [15]. These two sites may have the highest and most stable host densities, favouring the persistence of the corresponding phage populations[16] .

The relation of phage distribution relates to host abundance and hydrology can be explained as follows. First, phages from two different subfamilies (*Bradleyvirinae* and *Krylovirinae*) were isolated from the same Houet River sample using two *P. aeruginosa* strains with the same antibiogram. This suggests that phage recovery was more closely related to strain identity than to resistance phenotype. This is expected, as phage adsorption depends on surface features such as receptors, pili, lipopolysaccharides, and outer-membrane proteins, which are not necessarily linked to the resistance mechanisms detected here. Second, two *Krylovirinae* phages were recovered from hydrologically unconnected sites, suggesting that this lineage is widely distributed across the urban water network. Together, these observations are consistent with the idea that phage distribution is influenced by host population composition and is not determined by water connectivity alone.

Host range and therapeutic relevance

Broad host range is one of the main selection criteria for therapeutic phage candidates, since it determines how much of the circulating population a given phage can cover. Twenty-two of our 28 isolates lysed at least eight of eleven strains and three lysed all eleven, a proportion comparable to the 84% and 87% lytic activity against clinical *P. aeruginosa* reported by Garbe et al. and by Majdani and Shams Ghahfarokhi [17, 18]. At the other end of the distribution, two isolates were restricted to four strains. Narrow-spectrum phages of this kind are of limited direct therapeutic use but are valuable as typing tools and as components of cocktails.

Some limitations should be considered when interpreting these findings in a therapeutic context. First, host range was assessed by spot test, which detects lysis from without as well as productive infection. Efficiency-of-plating (EOP) assays on the same panel are needed to confirm that the phages genuinely replicate on each host. Second, the eleven strains all derive from urine samples collected at a single laboratory and share an unusually homogeneous antibiogram, so the panel is not representative of the diversity of *P. aeruginosa* circulating in Burkina Faso. Third, and most importantly, none of the strains in this panel is multidrug-resistant: all were meropenem-susceptible and only three produced a penicillinase.

Genomic diversity within a single wastewater ecosystem

The coexistence of taxonomically distant phages in the same wastewater environment has been documented repeatedly [19], and our data extend this observation to Bobo-Dioulasso. The contrast between vB_PaeM_CM0001 and the two *Phikmvvirus* genomes is striking. They have a 14 kb difference in genome length, a 23-gene difference in coding capacity, and they share no orthogroup [19]. This extreme divergence among co-occurring *P. aeruginosa* phages is frequent. Kwan et al. found that 55% of the open reading frames among 18 *P. aeruginosa* phages were novel and that most of the phages were only distantly related [20] . These patterns are consistent with modular evolution, where functional modules are gained, lost, and exchanged at different rates.

Despite this divergence at the level of gene content, the three genomes share the same overall architecture, a compact, densely coding double-stranded DNA genome organized into contiguous functional modules. Coding densities of 94.5–96.2% are typical of tailed phages, whose short infection cycles and tightly coordinated gene expression favour highly compact genomes[21]. The practical consequence of this compactness for annotation is that overlapping and closely abutting genes are common. Moreover, gene callers that assume canonical intergenic spacing tend to miss them. We used PHANOTATE[22] for this reason.

Functional annotation and hypothetical proteins

A putative function could be assigned to 70.5% of the CDSs of vB_PaeM_CM0001 but to only 50.8% of those of vB_PaeP_CM0002 and vB_PaeP_CM0003. Ratios such as these are common in phage genomics and highlight the limitations of homology-based annotation, given the incomplete representation of viral sequence diversity in reference databases[23, 24]. The PHROG profile of vB_PaeM_CM0001 illustrates the same point: 28.4% of its CDSs fall into the "other" category, i.e. they have a recognizable function that PHROG does not assign to a defined module[25]. Combining remote-homology detection, HMM profiles, and structure-based prediction can help substantially reduce this fraction [22, 26, 27]. The application of these approaches to the 26 and 32 hypothetical proteins identified here would be a useful next step.

In all three genomes, the largest annotated module is DNA, RNA and nucleotide metabolism. This predominance is characteristic of virulent phages, whose infection cycle depends on rapid takeover of host replication and on tight coupling between genome replication and the temporal programme of gene expression. In many phages the activation of late genes is itself contingent on replication progress[28] , and phage-encoded proteins that modulate DNA topology act in concert with host machinery at initiation[29] .

The absence of CDSs assigned to integration/excision and to the moron/AMG/host-takeover categories, in all three genomes and in both PHROG and Phold annotations, is the single most relevant functional observation for therapeutic purposes: it indicates that none of these phages encodes the machinery for stable integration into a host chromosome. This inference is not absolute, a sufficiently divergent integrase could escape homology-based detection[27], but it is consistent with the independent lifestyle predictions and with the absence of a repressor module.

Divergence in the tail fibre module

The two *Phikmvvirus* isolates share 91.9% genome-wide identity, are fully collinear, and share 62 orthogroups, yet a single interval of approximately 2.0 kb drops to 56.5% identity. The diverging interval contains the tail fibre locus, and gene content differs across it. It consists of four annotated tail fibre proteins in vB_PaeP_CM0002 against three plus a hypothetical protein of comparable length in vB_PaeP_CM0003.

Tail fibres are the receptor-binding proteins that determine which host strains a phage can adsorb to, and hypervariability at this locus against a background of high genome-wide conservation is a recurrent feature of the *Autographiviridae*. The same region is also the most divergent part of vB_PaeP_CM0002 compared with its closest relatives, which is consistent with a module shaped by host-driven selection while the rest of the genome remains constrained by replication and virion assembly.

The two phages are collinear and 91.9% identical, with their main difference in the receptor-binding module and host range. This makes the tail fibre locus a likely candidate for explaining the difference in host range. The proteomic data are consistent with this hypothesis, since vB_PaeP_CM0002 and vB_PaeP_CM0003 share 63 (96.9%) and 61 (93.8%) of their proteins respectively with at least one relative, so the gene content across the genus is largely preserved[30].

Taxonomic placement and marker discordance

We used a combination of VIRIDIC intergenomic similarity, proteome-based clustering and marker-gene phylogeny, following follows current ICTV practice for prokaryotic viruses [31]. This is better suited to mosaic genomes than nucleotide comparison alone [32] On this basis, vB_PaeM_CM0001 is a new isolate of an existing *Abidjanvirus* species, while vB_PaeP_CM0002 and vB_PaeP_CM0003 both fall below the 95% species threshold against every available relative and against each other, and therefore constitute two new *Phikmvvirus* species. The VIRIDIC species clustering, VICTOR/OPTSIL clustering, and ViPTree proteome distances all support this conclusion, strengthening the taxonomic assignment.

The relationship between vB_PaeP_CM0002 and vB_PaeP_CM0003 is less uniformly resolved. They are adjacent in the ViPTree proteome tree and sister taxa in the marker-gene tree, but the GBDP whole-genome tree places them in different subclades of the *Phikmvvirus* group. This result is expected for two collinear genomes that are 91.9% identical but have a highly divergent tail fibre locus. GBDP distances are based on high-scoring alignment segments and can therefore be influenced by regions of low similarity, whereas proteome- and marker-based methods rely mainly on the conserved portions of the proteome.

Absence of tRNA genes

We did not detect any tRNA gene in any of the three genomes, which is in line with reports on other *P. aeruginosa* phages .

Safety profile

The three genomes contained no detectable virulence factors, antimicrobial resistance genes, or anti-CRISPR proteins. This is a necessary condition for any therapeutic candidate, since transducing phages can mobilize resistance and virulence determinants, and it is consistent with the strictly lytic repertoire described above [33]. However, this alone is not sufficient. In silico screening is limited by database coverage and the sensitivity of homology-based searches. The clinical application of phage therapy requires a thorough experimental evaluation of several key factors, including the stability of the preparations, control of endotoxin levels, and validation of their efficacy in vitro and in vivo [34].

Limitations

Some limitations of the work should be considered. First, only three of the 28 isolates were sequenced, so the genomic diversity of the collection is unknown. Next, the experimental characterization was limited to host range. We did not assess adsorption rate, latent period, burst size, EOP, pH and thermal stability, and biofilm activity, and transmission electron microscopy was not performed. Therefore, the morphotype was inferred solely from the taxonomy. Also, the host panel was small, showed limited diversity in antibiograms, and was derived from a single sample type and laboratory. Finally, all safety assessments were based on in silico analyses.

Conclusion

We report the isolation of 28 lytic *P. aeruginosa* phages from wastewater in Bobo-Dioulasso and the complete genome sequences of three of them. vB_PaeM_CM0001 is a new isolate of an *Abidjanvirus* species within the *Mesyanzhinovviridae*, while vB_PaeP_CM0002 and vB_PaeP_CM0003 represent two new species of the genus *Phikmvvirus* within the *Autographiviridae*. All three have compact, densely coding genomes, encode no tRNA, no integration or excision functions, no virulence factor and no antimicrobial resistance determinant, and are predicted with high confidence to be strictly virulent. The two *Phikmvvirus* isolates are fully collinear with each other and with their closest relatives, and their divergence is concentrated at the tail fibre locus, the module that determines receptor binding. To our knowledge these are the first *P. aeruginosa* phage genomes described from Burkina Faso, and they extend the known diversity of the genus *Phikmvvirus*.

These genomes provide a documented starting point for a national phage bank. Their translation into therapeutic candidates now requires experimental characterization of their infection kinetics and stability, efficiency-of-plating measurements against a larger and more resistant panel of clinical isolates, in particular carbapenem-resistant strains, and experimental confirmation of the absence of transduction. The most direct and informative next step would be to test whether the tail fibre locus determines the difference in host range. Functional and structural annotation of the 26 and 32 proteins of unknown function reported here would be the second, more exploratory direction.

Methods

Bacterial strains and antibiotic susceptibility testing

Ten *P. aeruginosa* clinical isolates were recovered from urine samples submitted to the Clinical Bacteriology Laboratory of Centre MURAZ for routine diagnostic examination. Isolates were subcultured on Mueller-Hinton agar (Liofilchem, Italy) and incubated for 24 h at 37 °C. Identification was based on Gram staining, oxidase and catalase tests, and the VITEK 2 Compact system (bioMérieux, France) with GN cards. *P. aeruginosa* ATCC 27853 was used as the reference strain and for disk quality control. Antibiotic susceptibility was determined by Kirby-Bauer disk diffusion on Mueller-Hinton agar after 18–24 h of incubation and interpreted according to CA-SFM/EUCAST 2024 breakpoints[35]. The antibiotics tested were ticarcillin, piperacillin, ticarcillin/clavulanic acid, piperacillin/tazobactam, ceftazidime, cefepime, aztreonam, meropenem and levofloxacin.

Wastewater sampling

Five wastewater samples were collected on July 15, 2023, at each of the five sampling sites in Bobo-Dioulasso : the Souro Sanou University Hospital, a wastewater discharge station, the municipal sewage treatment plant, the Houet River backwater and the Dogona slaughterhouse (Additional file 1: Fig. S1). Wastewater was collected in 50 mL Falcon tubes and transported to Centre MURAZ. Samples were left to settle for 1 h; the supernatant was transferred to a fresh tube, centrifuged at approximately 1,650 ×g for 5 min at 4 °C (Centrifuge 5804 R, Eppendorf, Germany), and filtered through 0.45 µm sterile filters.

Phage enrichment, isolation and purification

Clinical strains and ATCC 27853 were streaked on Mueller-Hinton agar and incubated at 37 °C for 18–24 h, then suspended to 0.5 McFarland using a DensiCHEK Plus (bioMérieux, France). For enrichment, 10 mL of wastewater filtrate, 63.75 µL of 2 mM $CaCl_2$ and 500 µL of bacterial suspension were added to 15 mL of tryptic soy broth (Liofilchem, Italy) and incubated at 37 °C for 48 h. Enriched cultures were centrifuged at 3,000 rpm for 20 min at 4 °C (Rotina 380 R, Andreas Hettich, Germany) and the supernatant was filtered through 0.22 µm sterile filters.

Five microlitres of each filtrate were spotted onto double-layer LB agar (0.7% top agar; Liofilchem, Italy) seeded with the corresponding host suspension and incubated at 37 °C for 24 h. Lysis zones were excised, resuspended in 1 mL of phosphate-buffered saline (Oxoid, UK) containing 10 µL of chloroform (Thermo Fisher Scientific, USA), centrifuged at approximately 1,650 ×g for 10 min to remove debris, filtered through a 0.22 µm membrane and stored at 4 °C.

Phages were purified by three successive rounds of single-plaque isolation following Lu et al. [36]. Lysates were serially diluted 1:100 from 10^{-2} to 10^{-8}. 10 µL of the appropriate dilution were mixed with 500 µL of host suspension containing 2 mM $CaCl_2$ (Carlo Erba, France), added to 4 mL of 0.7% soft agar and poured onto LB agar. After 24 h at 37 °C, the plaques were characterized by macroscopic observation based on their size (small or large) and appearance

(clear or turbid), and individual plaques were picked and resuspended in PBS. Purified phages were stored in PBS at 4 °C.

Host range determination

Host range was determined by spot test on double-layer LB agar [37]. Five hundred microlitres of each bacterial suspension were mixed with 4 mL of 0.7% soft agar and poured onto LB agar. Once solidified, 5 µL of each purified lysate from 28 phages-were spotted on the lawn. Plates were incubated at 37 °C for 18–24 h and scored for the presence or absence of a lysis zone. Each phage was tested against all eleven strains.

Phage DNA extraction

Purified isolates were amplified on double-layer LB agar. Those reaching 10^9–10^{11} PFU/mL were retained for sequencing. Plates were overlaid with SM buffer and held at 4 °C for 24 h to allow diffusion. The buffer was collected, filtered through 0.22 µm, and phage particles were precipitated with 10% (w/v) PEG 8000 (Sigma-Aldrich, USA) overnight at 4 °C, then pelleted at 18,000 × g for 15 min at 4 °C. The pellet was resuspended in 500 µL of buffer (100 mM NaCl, 10 mM $MgSO_4$, 10 mM Tris-HCl) and treated with DNase I and proteinase K for 30 min at 37 °C to remove host nucleic acids and protein. DNA was extracted according to Sambrook and Russell [23], and its purity and concentration were assessed on a NanoDrop spectrophotometer and by Qubit dsDNA High Sensitivity assay (Thermo Fisher Scientific).

Whole-genome sequencing

Five purified DNA samples were sequenced by GenoScreen (Lille, France). Libraries were prepared with the Illumina DNA Prep kit according to the manufacturer's instructions. Library quality and size distribution were verified by capillary electrophoresis (Fragment Analyzer, High Sensitivity NGS Fragment Kit, Agilent) and final concentrations were validated fluorometrically with SYBR Green (Thermo Fisher Scientific). Paired-end sequencing was performed on an Illumina platform, yielding 150 bp-length reads. The numbers of raw read pairs were 41,858 for vB_PaeM_CM0001, 39,598 for vB_PaeP_CM0002, and 48,862 for vB_PaeP_CM0003.

Read processing and genome assembly

The raw reads were quality-checked with FastQC and filtered with fastp. Adapters were trimmed, reads with a Phred score below 20 were discarded, and reads shorter than 50 bp were removed. The filtered reads were mapped to the *P. aeruginosa* reference genome (Accession number: NC_002516.2) with Bowtie 2 [38] and non-concordant read pairs, considered of potential viral origin, were extracted with SAMtools [39].

The reads were assembled de novo with SPAdes v4.2.0[40] using multiple k-mer sizes. Assembly graphs were inspected in Bandage [41] to assess circularity, identify ambiguous regions and evaluate alternative paths. The candidate assemblies were validated by read mapping and per-base coverage analysis (`samtools depth`). Contigs were confirmed as viral with VirSorter [42] and their completeness, host contamination and topology were assessed with CheckV [43]. The assembled genomes were compared with closely related reference phages using nucmer from the MUMmer package [44] and visualized as dot plots. In the case

assembly ambiguities remained, the alternative graph paths were evaluated against coverage profiles and reference synteny, and the path with the strongest read support was retained.

All genomes used in comparative analyses, including the reference genomes retrieved from GenBank, were normalized for orientation and origin before alignment. Each genome was oriented so that its coding strand corresponded to the plus strand, determined from the majority strand of its annotated coding sequences and, for records without CDS annotation, by alignment to the annotated genomes of the same family. Each genome was then rotated to begin at a sequence anchor conserved across all members of its family. The two isolates sequenced in this study, vB_PaeM_CM0001 and vB_PaeP_CM0002, served as the reference frames for their respective families and were left in their deposited orientation. The coding sequence coordinates were transformed accordingly, features spanning the new origin were recorded as joins, and all 409 transformed coding sequences were re-translated and verified against the deposited protein sequences.

Genome annotation

The open reading frames were predicted with PHANOTATE [22]. The products were assigned by a three-stage consensus applied in fixed order, each stage annotating only the coding sequences left unassigned by the previous one.

First, the predicted genes were searched against the PHROG database[25] with DIAMOND blastx (e-value ≤ 1e-2, single best hit per query), retaining hits with at least 30% amino-acid identity over at least 50% of the query length. Second, the coding sequences still unassigned were annotated with Phold [45], which predicts a three-dimensional structure for each protein with ProstT5 and compares it against the Phold structural database with Foldseek. Third, any remaining coding sequences were searched with blastp against the NCBI RefSeq protein database restricted to viruses (e-value ≤ 1e-5), retaining hits with at least 60% identity over at least 50% of the query length. Products were curated manually before submission to GenBank.

Each coding sequence was then assigned to one of eleven functional categories on the basis of its final product name: DNA replication and repair, DNA metabolism, transcription and regulation, DNA packaging, head and capsid assembly, tail assembly, baseplate and host recognition, DNA injection, host lysis, other and auxiliary functions, and hypothetical or unknown function. We made this assignment ourselves rather than relying on a single tool, as the annotations were derived from three sources with different classification schemes. The assignment for every coding sequence, together with its annotation source, is given in Additional file 2 (Table S7).

tRNA genes were searched with tRNAScan-SE[46] Annotated CDSs were exported as GFF files and formatted for submission to GenBank. The circular genome maps were drawn with Proksee[47] with CDSs coloured by the functional categories of Table 4 and with GC content and GC skew tracks added.

Selection of reference genomes

The reference genomes were retrieved from NCBI by BLASTn and ranked by a composite score combining percentage identity and query coverage (score = identity × coverage), which penalizes partial and non-specific matches[48]. The five highest-scoring genomes per isolate were retained for all comparative analyses.

Comparative genomics

The genome structures were compared within each genus with Easyfig v2.2.2[49] after aligning all genomes in the same orientation using tBLASTx[49]. The structural rearrangements were detected with progressive Mauve v2.4.1 [50] through the identification of locally collinear blocks, run separately for each family. Both the Mauve alignments and the VIRIDIC analysis were performed on the normalized genomes described above. The intergenomic similarities were computed with VIRIDIC[31] and interpreted against the ICTV demarcation thresholds of 70% for genus and 95% for species. The local nucleotide identity along each alignment was computed in 250 bp windows to delimit divergent regions.

Comparative proteomics

The orthologous groups were identified with OrthoFinder v3.1.4 [51], both among the three isolates and between each isolate and its closest relatives. Results were visualized with OrthoVenn3[52]. A proteome-based distance tree was generated on the ViPTree server [53] , which uses tBLASTx-derived genome-wide similarity.

Phylogenetic analysis and taxonomic classification

The taxonomic assignment was obtained by BLASTn against NCBI and confirmed with PhaBOX v3.0[54]. The whole-genome phylogeny was inferred on the VICTOR server [55] , which computes pairwise GBDP distances[56] and infers balanced minimum-evolution trees with FASTME [57] and 100 bootstrap replicates, using the settings and taxon-delimitation thresholds recommended for prokaryotic viruses. As all genomes analysed were complete, the D0 distance formula recommended for nucleotide sequences was retained. The trees were midpoint-rooted [58] and drawn with ggtree [59]. The species, genus and family clusters were delimited with OPTSIL [60] using an F value of 0.5[61].

A Maximum-likelihood tree was also built from the amino-acid sequences of one conserved structural marker, the major capsid protein. Sequences were aligned with MAFFT[62] and the tree inferred with FastTree[63] under the JTT+CAT model. Trees were visualized and annotated with iTOL[64] .

Lifestyle prediction

Lifestyle (virulent or temperate) was predicted with PhageAI v1.0.2[65] and independently with PhageLeads[66].

In silico safety assessment

Virulence factors were searched online against the VFDB[67] and VICTORS[68] databases using their respective web-based tools and default parameters. The nucleotide and amino acid sequences were queried against the corresponding databases using BLAST-based searches. Antimicrobial resistance genes were detected with PhageLeads[66] and with RGI v6.0.5 against CARD v4.0.1[69] . The analyses related to anti-CRISPR systems were performed using AcRanker, AcrFinder, and CRISPRCasFinder. AcRanker was applied to the predicted phage proteomes using the default settings to rank proteins according to their anti-CRISPR potential. AcrFinder was used with its default parameters, including homology search thresholds of 30% identity, 80% coverage, and an E-value of 0.01, to identify known anti-CRISPR proteins and associated Acr-Aca loci. CRISPR-Cas systems were screened using CRISPRCasFinder with the default parameters[70]. Note that the Phold annotation step described above also compares each predicted structure against VFDB and CARD. The single VFDB match obtained there is reported and interpreted in the Results.

Abbreviations

Acr: anti-CRISPR protein; **AMR:** antimicrobial resistance; **ARG:** antimicrobial resistance gene; **AST:** antibiotic susceptibility testing; **BLAST:** Basic Local Alignment Search Tool; **CARD:** Comprehensive Antibiotic Resistance Database; **CA-SFM:** Comité de l'Antibiogramme de la Société Française de Microbiologie; **CDS:** coding sequence; **CRISPR:** clustered regularly interspaced short palindromic repeats; **dsDNA:** double-stranded DNA; **EOP:** efficiency of plating; **EUCAST:** European Committee on Antimicrobial Susceptibility Testing; **GBDP:** Genome BLAST Distance Phylogeny; **HMM:** hidden Markov model; **HR:** Houet River; **ICTV:** International Committee on Taxonomy of Viruses; **iTOL:** Interactive Tree Of Life; **LB:** Luria-Bertani; **LCB:** locally collinear block; **MAFFT:** Multiple Alignment using Fast Fourier Transform; **NCBI:** National Center for Biotechnology Information; **OPTSIL:** optimization of silhouette; **PBS:** phosphate-buffered saline; **PEG:** polyethylene glycol; **PFU:** plaque-forming unit; **PHROG:** Prokaryotic Virus Remote Homologous Groups; **SGH:** slaughterhouse; **SSUH:** Souro Sanou University Hospital; **STP:** sewage treatment plant; **TerL:** large terminase subunit; **VFDB:** Virulence Factor Database; **VICTOR:** Virus Classification and Tree Building Online Resource; **VIRIDIC:** Virus Intergenomic Distance Calculator; **ViPTree:** Viral Proteomic Tree; **WDS:** wastewater discharge station; **WHO:** World Health Organization

Declarations

Ethics approval and consent to participate

No ethics approval was required for the study. This study did not involve human participants, human subjects research or personally identifiable information. The bacterial isolates used were residual, anonymized clinical isolates obtained during routine diagnostic activity at the Clinical Bacteriology Laboratory of Centre MURAZ. Authorization to collect wastewater within the urban municipality of Bobo-Dioulasso was granted by the Mayor of Bobo-Dioulasso (N°2022 000057 CB/M/SG/DPSHP) and by the general management of the Souro Sanou University Hospital (N°2021- 669 MS/SG/CHUSS/DG).

Consent for publication

Not applicable.

Availability of data and materials

The three complete genome sequences generated in this study have been deposited in GenBank under accession numbers PZ231876 (vB_PaeM_CM0001), PZ365779 (vB_PaeP_CM0002) and PZ365800 (vB_PaeP_CM0003). The sequencing data have been deposited in NCBI BioProject PRJNA1515337, with BioSamples SAMN62578703, SAMN62578704, and SAMN62578705.

Competing interests

The authors declare that they have no competing interests.

Funding

The African Center of Excellence in Biotechnological Innovations for the Elimination of Vector-Borne Diseases (CEA/ITECH-MTV) provided the reagents and consumables, and the molecular biology and biotechnology team at the MURAZ Center funded the whole-genome sequencing. AO was supported by the Canada Research Chair [CRC Tier 2 Grant 950-230577] and the Natural Sciences and Engineering Research Council of Canada [NSERC Discovery Grant RGPIN-2023-05474]. The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript.

Authors' contributions

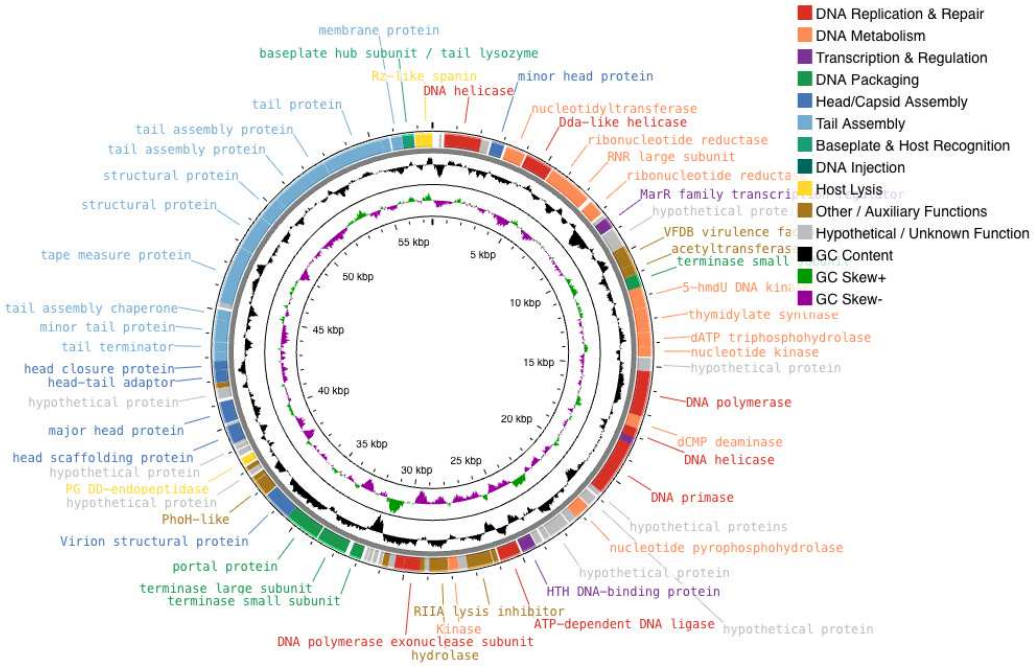
KG performed the wastewater sampling, microbiological analyses and data management, contributed to the methodology, and wrote the first draft. EWS performed the DNA extraction and the bioinformatic analyses, and contributed to the first draft. CYE contributed to the study design and methodology. RKY performed microbiological analyses and contributed to the methodology. AQS performed the DNA extraction. AO contributed to the conceptualization, the bioinformatic analyses, the methodology and the first draft. MKG conceived and supervised the study, contributed to the methodology and to the first draft. All authors reviewed and approved the final manuscript.

Acknowledgements

The authors thank the African Center of Excellence in Biotechnological Innovations for the Elimination of Vector-Borne Diseases, the Molecular Biology and Biotechnology Team of Centre MURAZ, and the University of Sherbrooke for their support.

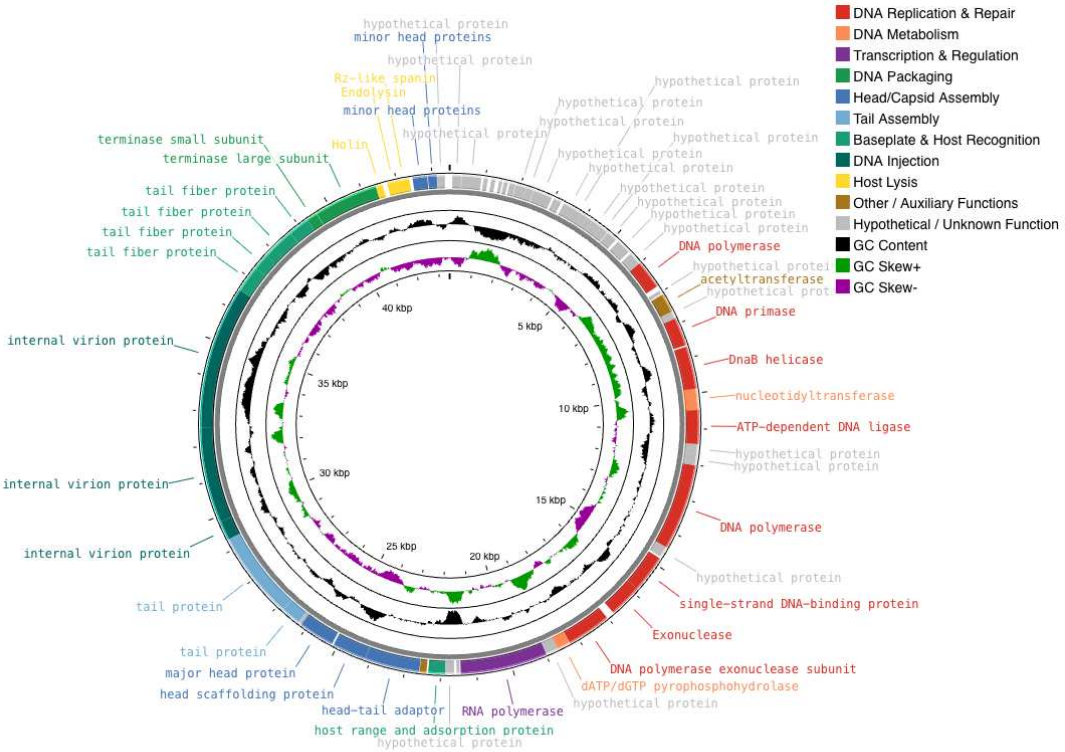
Figures

Fig. 1 Circular maps of the three phage genomes. **a** vB_PaeM_CM0001, **b** vB_PaeP_CM0002, **c** vB_PaeP_CM0003. The outer ring shows predicted CDSs colored by PHROG functional category. The inner tracks show GC content (black) and GC skew (green, positive; purple, negative).



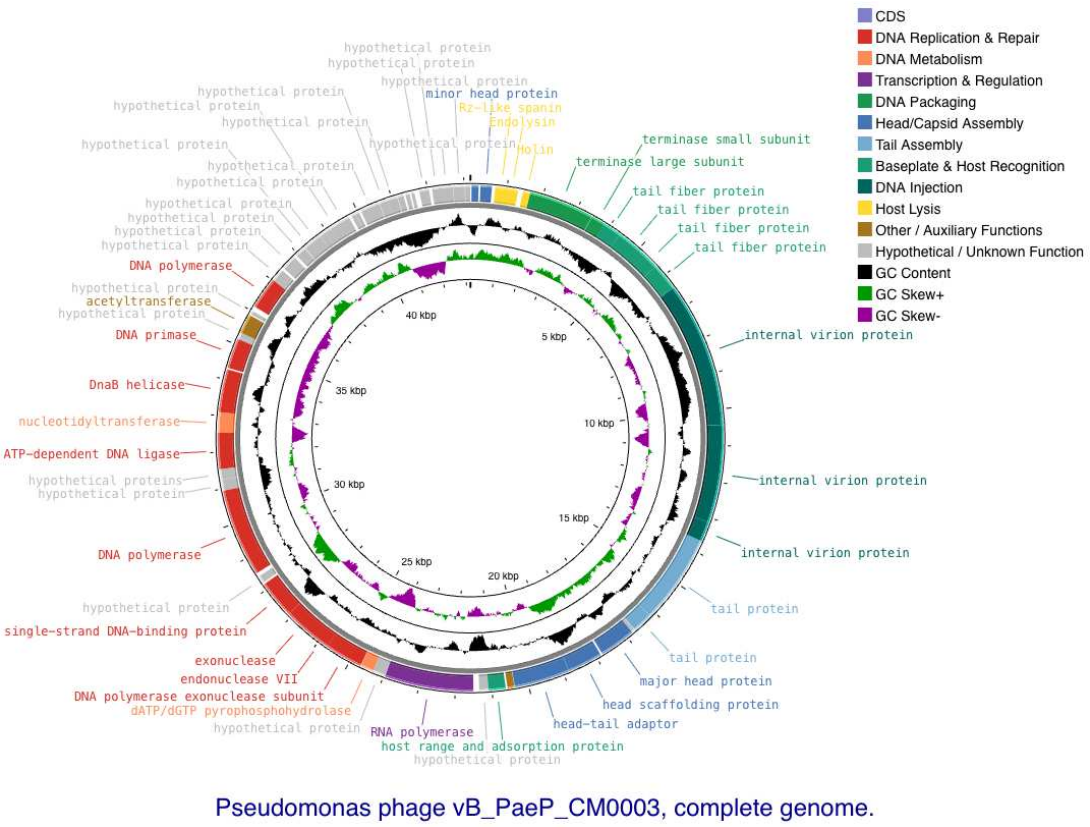
Pseudomonas phage vB_PaeM_CM0001, complete genome.

Fig. 1a: vB_PaeM_CM0001



Pseudomonas phage vB_PaeP_CM0002, complete genome.

Fig. 1b: vB_PaeP_CM0002



Pseudomonas phage vB_PaeP_CM0003, complete genome.

Fig. 1c: vB_PaeP_CM0003

Fig. 2 Functional composition of the three phage genomes. **a** Proportion of predicted coding sequences with an assigned function (blue) and of hypothetical proteins or proteins of unknown function (grey). **b** Distribution of the coding sequences across the eleven functional categories of Table 4, coloured as in the genome maps of Fig. 1.

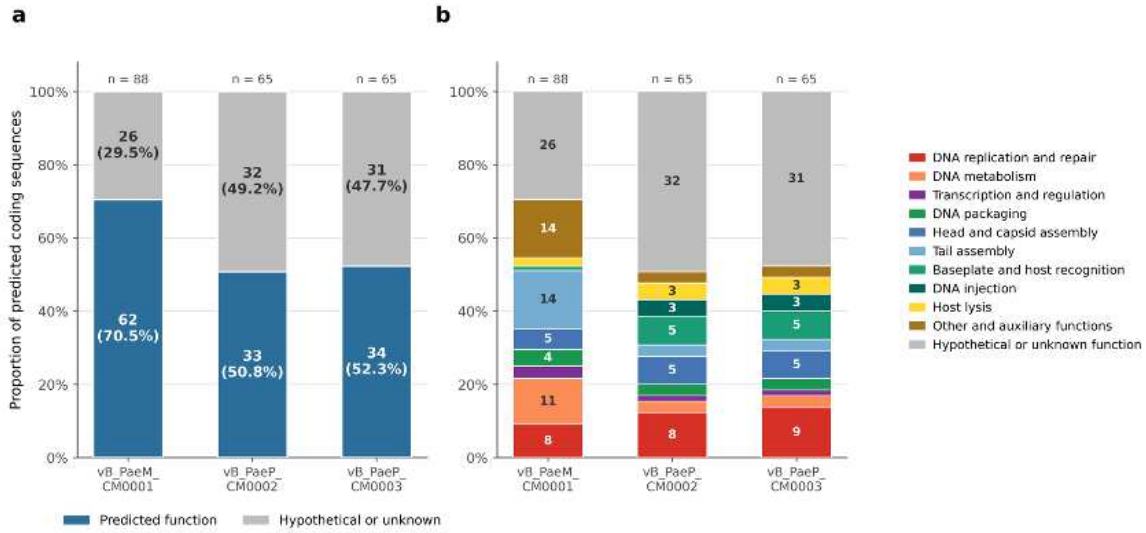


Fig. 3 VIRIDIC intergenomic similarity matrix for the three isolates (underlined) and their closest GenBank relatives. The lower-left triangle shows aligned genome fraction and

genome-length ratio. The upper-right triangle shows pairwise intergenomic similarity (%). Species and genus clusters were assigned using the ICTV thresholds of 95% and 70% respectively.

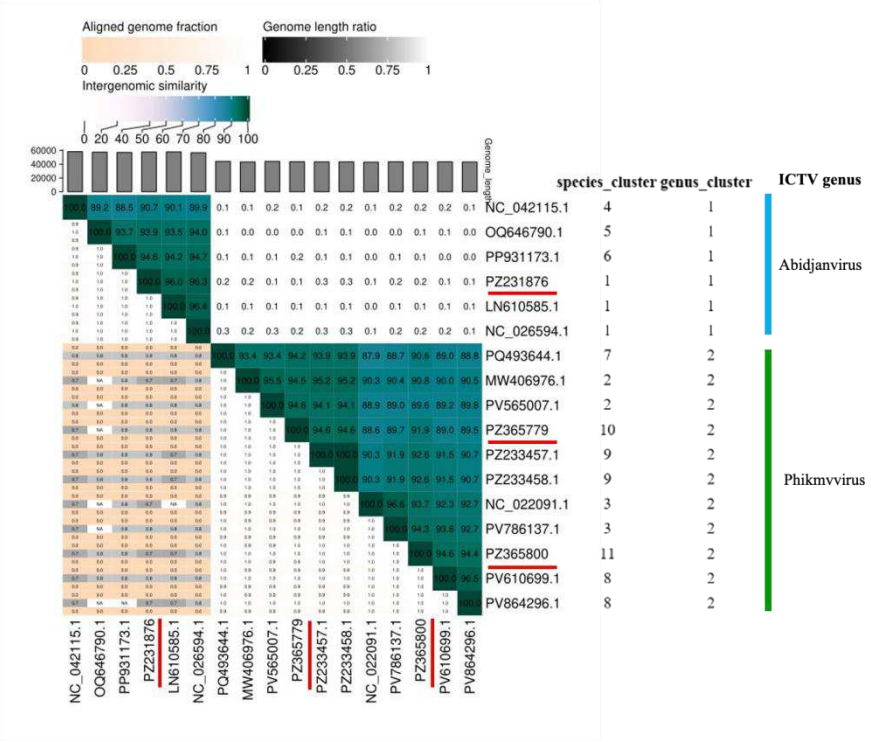


Fig. 4 Whole-genome comparison of the isolates and their closest relatives generated with Easyfig using tBLASTx. **a** *Mesyanzhinovviridae*: vB_PaeS_PAO1_Ab18, vB_PaeM_CM0001 and PAKlein3. **b** *Autographiviridae*: vB_PaePA10145phi1_HR2, vB_PaeP_CM0002, vB_PaeP_CM0003 and MPK7. Arrows represent predicted CDSs. Grey shading indicates regions of tBLASTx similarity, with intensity proportional to identity (scale, 73–100%).

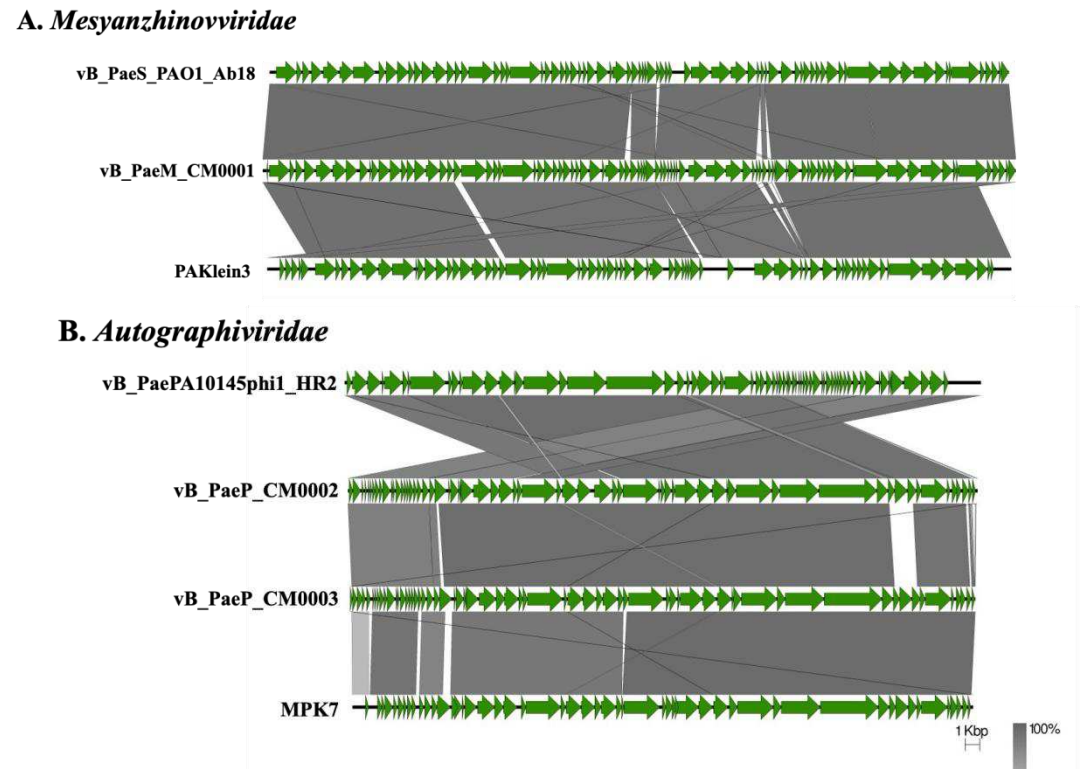


Fig. 5 Progressive Mauve alignments of the three isolates and their closest relatives. **a** *Autographiviridae*: vB_PaePA10145phi1_HR2, vB_PaeP_CM0002, vB_PaeP_CM0003 and MPK7. **b** *Mesyanzhinovviridae*: vB_PaeS_PAO1_Ab20, vB_PaeM_CM0001 and PAKlein3. Isolates sequenced in this study are in bold, with GenBank accessions beneath each name. Coloured boxes are locally collinear blocks (LCBs). Blocks of the same colour are homologous between genomes, and the profile within each block gives the average level of nucleotide identity along it. Blocks drawn above a genome's centre line are in the forward orientation and blocks drawn below it are inverted. Open boxes beneath each genome are predicted coding sequences. The GenBank record for vB_PaeS_PAO1_Ab20 (LN610585) contains no CDS annotation. Before alignment all genomes were oriented to place the coding strand on the plus strand and rotated to a common sequence anchor (see Methods), so coordinates refer to the normalised rather than the deposited sequences.

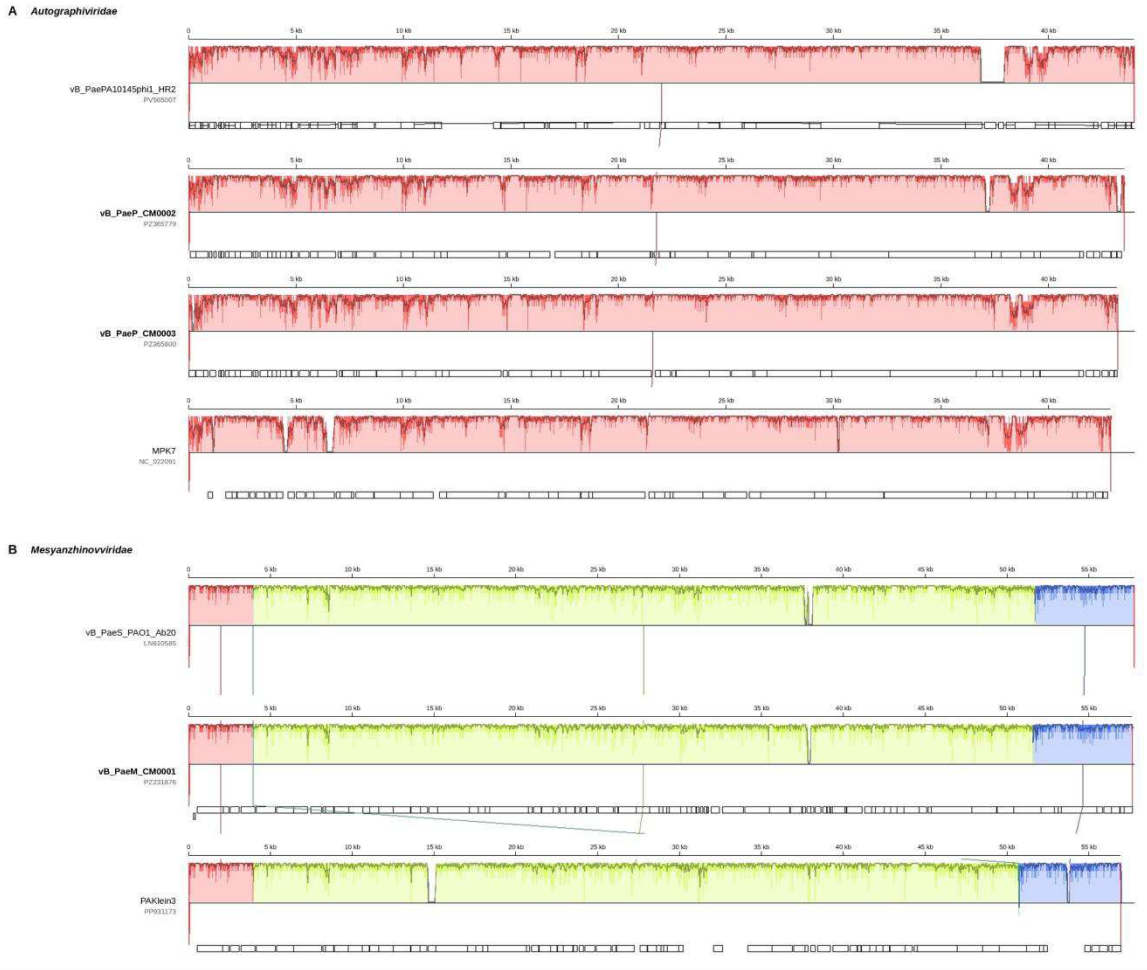
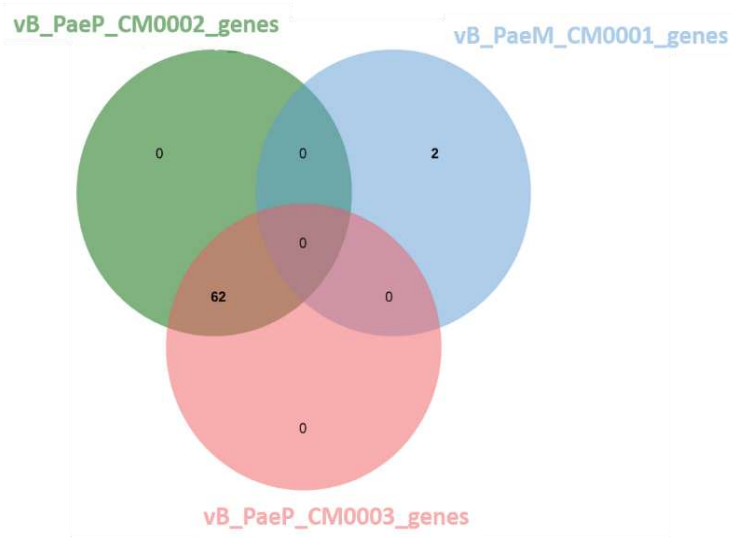


Fig. 6 Orthologous groups shared among the three isolates, identified with OrthoFinder and visualized with OrthoVenn3. The table below the diagram gives, for each genome, the number of proteins, the number of clusters and the number of singletons.



Species	Proteins	Clusters	Singletons
vB_PaeM_CM0001_genes	88	2	84
vB_PaeP_CM0002_genes	65	62	3
vB_PaeP_CM0003_genes	65	62	3

Fig. 7 ViPTree proteomic distance trees. **a** Position of the three isolates (red stars) within the reference tree of dsDNA prokaryotic viruses. The inner ring indicates virus family and the outer ring host group. **b** Detail of the subtree containing the three isolates, with genome lengths in nucleotides. The left colour strip indicates virus family and the right strip host group.

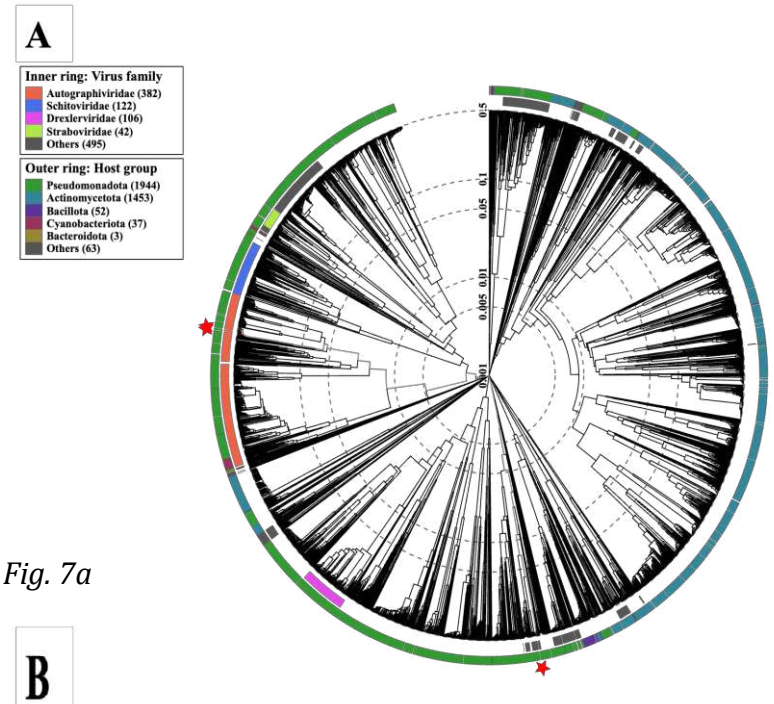


Fig. 7a

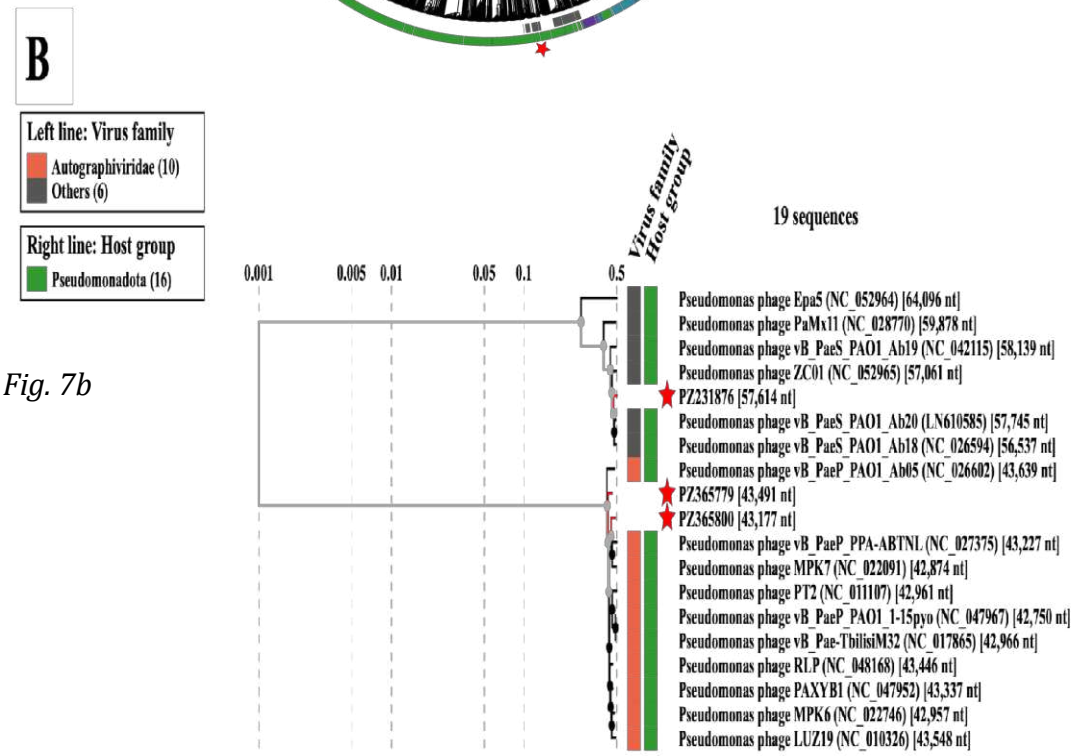


Fig. 7b

Fig. 8 GBDP phylogenomic tree of the three isolates and their closest relatives, inferred with VICTOR from whole-genome distances (formula D6), midpoint-rooted, with support from 100 bootstrap replicates. The colour columns to the right give the OPTSIL family, genus and species clusters, the GC content and the genome length.

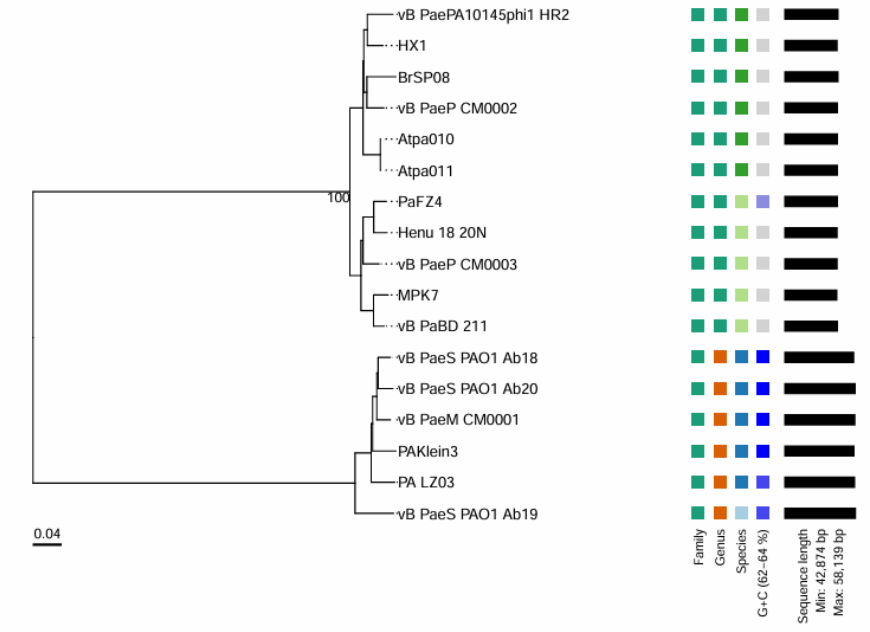
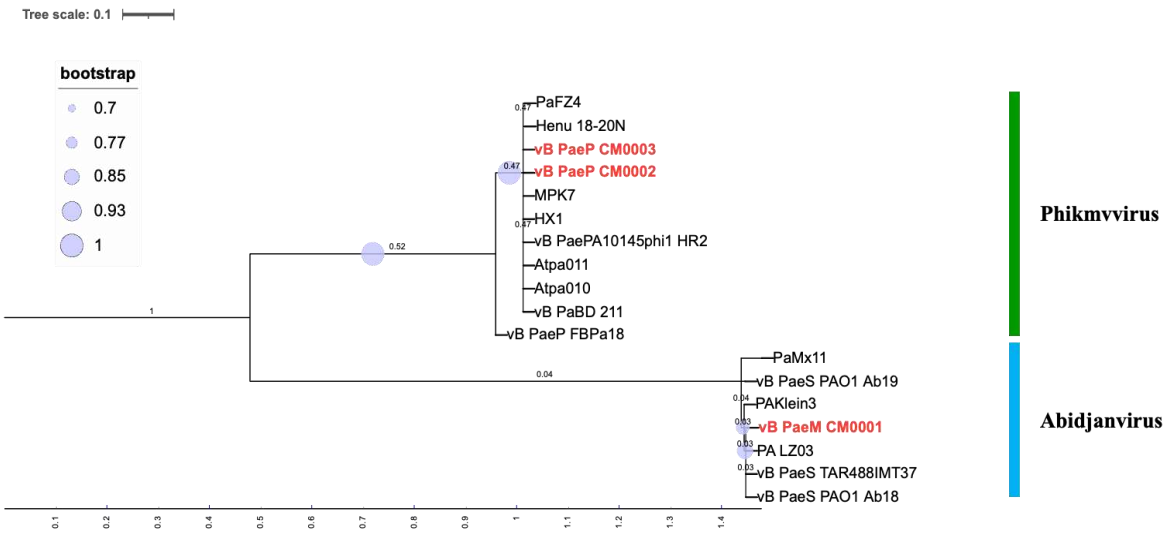


Fig. 9 Maximum-likelihood phylogenies inferred from the major capsid protein, aligned with MAFFT and inferred with FastTree under the JTT+CAT model. Node circles are scaled to bootstrap support. Isolates from this study are shown in red. Colored bars indicate genus assignment.



Tables

Table 1 Antibiotic susceptibility profiles of the eleven *P. aeruginosa* strains used in this study

Strain	Sample type	TC	PRL	TTC	TPZ	CAZ	FEP	ATM	MERO	LEV
27853*	ATCC	NA	NA	A	A	A	A	A	A	A
PA_CM20	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM22*	Urine	R	R	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM23*	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM24*	Urine	R	R	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM25*	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM26	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM27	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM28	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM29	Urine	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT
PA_CM30	Urine	R	R	ZIT	ZIT	ZIT	ZIT	ZIT	S	ZIT

TC: Ticarcillin; PRL: piperacillin; TTC: Ticarcillin/clavulanic acid; TPZ: Piperacillin/tazobactam; CA: Ceftazidime; FEP: Cefepime; ATM: Aztreonam, MERO, Meropenem; LEV: Levofloxacin. S: Susceptible; ZIT:Technical Uncertainty Zone ; R: Resistant. NA : Not Applicable; A : Asset; PA_CM21: *Pseudomonas aeruginosa*_Centre Muraz corresponds to the ATCC 27853 strain used for quality control of antibiotic discs. Asterisks mark strains used as hosts for phage isolation.

Table 2 Recovery of *P. aeruginosa* lytic phages by host strain and sampling site

Host strain	SSUH	WDS	STP	HR	SGH	Isolation frequency
PA_CM21	+	–	+	+	+	4/5
PA_CM22	+	+	+	+	+	5/5
PA_CM23	+	–	–	+	+	3/5
PA_CM24	–	+	+	–	+	3/5
PA_CM25	+	+	+	+	+	5/5
PA_CM26	–	+	+	–	–	2/5

SSUH: Souro Sanou University Hospital; WDS: Wastewater Discharge Station; STP: Sewage Treatment Plant; HR: Houet River; SGH: Slaughterhouse. PA_CM: *P. aeruginosa*_Centre MURAZ collection. +: at least one lytic phage recovered; –: no phage recovered.

Table 3 Genomic features and taxonomic assignment of the three sequenced phages

Feature	vB_PaeM_CM0001	vB_PaeP_CM0002	vB_PaeP_CM0003
Source isolate	PA_BD_S03	PA_BD_S12	PA_BD_S29b
Sampling site	Wastewater_River Houet	Wastewater_River Houet	Wastewater_Hospital Souro Sanou University
Isolation host	*P. aeruginosa*	*P. aeruginosa*	*P. aeruginosa*
GenBank accession	PZ231876	PZ365779	PZ365800
Class	*Caudoviricetes*	*Caudoviricetes*	*Caudoviricetes*
Family	*Mesyanzhinovviridae*	*Autographiviridae*	*Autographiviridae*
Subfamily	*Bradleyvirinae*	*Krylovirinae*	*Krylovirinae*
Genus	*Abidjanvirus*	*Phikmvvirus*	*Phikmvvirus*
Taxonomic status	New isolate of a described species	New species	New species
Nucleic acid	linear dsDNA	linear dsDNA	linear dsDNA
Genome size (bp)	57,614	43,491	43,177
GC content (%)	63.54	62.63	62.25
Coding region (bp, %)	54,431 (94.48)	41,572 (95.59)	41,518 (96.16)
Predicted CDSs	88	65	65
CDSs with assigned function (n, %)	62 (70.45)	33 (50.77)	33 (50.77)
tRNA genes	0	0	0
Predicted lifestyle (confidence)	virulent (97.33%)	virulent (99.97%)	virulent (99.99%)
Sequencing depth (×)	140	132	163
CheckV completeness (%)	94.4	100	100

Légend. vB_PaeP_CM : Bacteria virus_*P. aeruginosa* Podoviridae_Centre Muraz ; vB_PaeM_CM : Bacteria virus _*P. aeruginosa* Myoviridae_Centre Muraz; PA_BD_S03**:** . PA_BD_S: *Pseudomonas aeruginosa* _Bobo Dioulasso_Sewage, PA_CM: *P. aeruginosa*_ Centre MURAZ

Table 4 Distribution of predicted CDSs across functional categories

Functional category	vB_PaeM_CM0001 (n)	%	vB_PaeP_CM0002 (n)	%	vB_PaeP_CM0003 (n)	%
DNA replication and repair	8	9.09	8	12.31	9	13.85
DNA metabolism	11	12.50	2	3.08	2	3.08
Transcription and regulation	3	3.41	1	1.54	1	1.54
DNA packaging	4	4.55	2	3.08	2	3.08
Head and capsid assembly	5	5.68	5	7.69	5	7.69
Tail assembly	14	15.91	2	3.08	2	3.08
Baseplate and host recognition	1	1.14	5	7.69	5	7.69
DNA injection	0	0.00	3	4.62	3	4.62
Host lysis	2	2.27	3	4.62	3	4.62
Other and auxiliary functions	14	15.91	2	3.08	2	3.08
Hypothetical or unknown function	26	29.55	32	49.23	31	47.69
Total	**88**	**100.00**	**65**	**100.00**	**65**	**100.00**

Legend. vB_PaeP_CM : Bacteria virus_*P. aeruginosa* Podoviridae_Centre Muraz ; vB_PaeM_CM : Bacteria virus _*P. aeruginosa* Myoviridae_Centre Muraz. CDSs: Sequences codings;

Each CDS was assigned to a category on the basis of its final product name. The assignment was made by us rather than taken from a single tool's output because products derive from three annotation sources with differing internal classification schemes. The same eleven categories are used to colour the genome maps in Fig. 1. Per-CDS assignments and annotation sources are given in Additional file 2 (Table S7).

Table 5 Proteome comparison of the three isolates with their five closest GenBank relatives

Phage	Total proteins	Proteins shared with ≥1 relative	% shared	Isolate-specific proteins	Reference proteins absent from the isolate
vB_PaeM_CM0001	88	79	89.77	9	1
vB_PaeP_CM0002	65	63	96.92	2	7
vB_PaeP_CM0003	65	61	93.85	4	2

Legend. vB_PaeP_CM : Bacteria virus_*P. aeruginosa* Podoviridae_Centre Muraz ; vB_PaeM_CM : Bacteria virus _*P. aeruginosa* Myoviridae_Centre Muraz;

Additional files

Additional file 1 (DOCX). Supplementary figures and tables: Fig. S1, map of the five wastewater sampling sites in Bobo-Dioulasso; Fig. S2, lysis spots obtained by the spot method; Fig. S3, small and large plaque morphotypes after purification; Fig. S4, host-range spot assay on strain PA_CM21 after 24 h and 48 h of incubation; Table S1, distribution of the 28 isolates by plaque size and host strain; Table S2, complete host-range matrix of the 28 isolates against the eleven strains; Table S3, the five closest GenBank relatives of each sequenced genome.

Additional file 2 (XLSX). Supplementary tables supplied as a spreadsheet: Table S5, protein-level comparison of vB_PaeP_CM0002 and vB_PaeP_CM0003, giving for each protein its closest counterpart in the other genome with percentage identity and length ratio; Table S5b, coding sequences without a counterpart among the available annotated relatives; annotation coverage of every genome used in the comparative analyses; Table S7, complete annotation of the predicted coding sequences of the three isolates (locus tag, coordinates, strand, nucleotide and amino-acid length, product, protein accession, annotation source and assigned functional category), one sheet per genome; together with the product-binning sheet from which the functional categories in Table 4 are derived.

References

1. O'Neill J. Review on antimicrobial Resistance. Antimicrobial Resistance: Talking dug - resistant infections globally. Eur J Hosp Pharm. 2016; 23:245. https://doi.org/10.1136/ejhpharm-2016-001013.

2. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet Lond Engl. 2022; 399:629–55. https://doi.org/10.1016/S0140-6736(21)02724-0.

3. WHO. WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Geneva: World Health Organization.; 2024.

4. Bassetti M, Vena A, Croxatto A, Righi E, Guery B. How to manage Pseudomonas aeruginosa infections. Drugs Context. 2018; 7:212527. https://doi.org/10.7573/dic.212527.

5. Donlan RM. Biofilms: microbial life on surfaces. Emerg Infect Dis. 2002; 8:881–90. https://doi.org/10.3201/eid0809.020063.

6. Breidenstein EBM, de la Fuente-Núñez C, Hancock REW. *Pseudomonas aeruginosa*: all roads lead to resistance. Trends Microbiol. 2011; 19:419–26. https://doi.org/10.1016/j.tim.2011.04.005.

7. Ministry of Health and Public Hygiene. Synthesis report of antimicrobial resistance surveillance in the laboratory 2023. Technical Report. Ouagadougou, Burkina Faso.: Ministry of Health and Public Hygiene; 2024.

8. Wittebole X, De Roock S, Opal SM. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. Virulence. 2014; 5:226–35. https://doi.org/10.4161/viru.25991.

9. Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. Bacteriophage. 2011;1:31–45. https://doi.org/10.4161/bact.1.1.14942.

10. Ly-Chatain MH. The factors affecting effectiveness of treatment in phages therapy. Front Microbiol. 2014;5:51. https://doi.org/10.3389/fmicb.2014.00051.

11. Nagel T, Musila L, Muthoni M, Nikolich M, Nakavuma JL, Clokie MR. Phage banks as potential tools to rapidly and cost-effectively manage antimicrobial resistance in the developing world. Curr Opin Virol. 2022; 53:101208. https://doi.org/10.1016/j.coviro.2022.101208.

12. Adnan M, Ali Shah MR, Jamal M, Jalil F, Andleeb S, Nawaz MA, et al. Isolation and characterization of bacteriophage to control multidrug-resistant Pseudomonas aeruginosa planktonic cells and biofilm. Biol J Int Assoc Biol Stand. 2020;63:89–96. https://doi.org/10.1016/j.biologicals.2019.10.003.

13. Salmond GPC, Fineran PC. A century of the phage: past, present and future. Nat Rev Microbiol. 2015;13:777–86. https://doi.org/10.1038/nrmicro3564.

14. Díaz-Muñoz SL, Koskella B. Bacteria-phage interactions in natural environments. Adv Appl Microbiol. 2014;89:135–83. https://doi.org/10.1016/B978-0-12-800259-9.00004-4.

15. Gnada K, Gomgnimbou K M, Yao RK, Essoh YC. Isolation and Preliminary Characterization of Staphylococcus aureus Lytic Phages from Wastewater Environment in Bobo-Dioulasso, Burkina Faso, West Africa. Adv Microbiol. 2025; 15:42–57. https://doi.org/10.4236/aim.2025.151004.

16. Wang X, Tang J, Dang W, Xie Z, Zhang F, Hao X, et al. Isolation and Characterization of Three Pseudomonas aeruginosa Viruses with Therapeutic Potential. Microbiol Spectr. 2023;11:e0463622. https://doi.org/10.1128/spectrum.04636-22.

17. Garbe J, Wesche A, Bunk B, Kazmierczak M, Selezska K, Rohde C, et al. Characterization of JG024, a pseudomonas aeruginosa PB1-like broad host range phage under simulated infection conditions. BMC Microbiol. 2010; 10:301. https://doi.org/10.1186/1471-2180-10-301.

18. Majdani R, Shams Ghahfarokhi E. Isolation and characterization of lytic bacteriophages against Pseudomonas aeruginosa isolates from human infections in the north-west of Iran. Iran J Microbiol. 2022;14:203–13. https://doi.org/10.18502/ijm.v14i2.9189.

19. Essoh C, Latino L, Midoux C, Blouin Y, Loukou G, Nguetta S-PA, et al. Investigation of a Large Collection of Pseudomonas aeruginosa Bacteriophages Collected from a Single Environmental Source in Abidjan, Côte d'Ivoire. PLOS ONE. 2015;10:e0130548. https://doi.org/10.1371/journal.pone.0130548.

20. Kwan T, Liu J, Dubow M, Gros P, Pelletier J. Comparative genomic analysis of 18 Pseudomonas aeruginosa bacteriophages. J Bacteriol. 2006;188:1184–7. https://doi.org/10.1128/JB.188.3.1184-1187.2006.

21. Jack BR, Boutz DR, Paff ML, Smith BL, Wilke CO. Transcript degradation and codon usage regulate gene expression in a lytic phage. Virus Evol. 2019;5:vez055. https://doi.org/10.1093/ve/vez055.

22. McNair K, Zhou C, Dinsdale EA, Souza B, Edwards RA. PHANOTATE: a novel approach to gene identification in phage genomes. Bioinformatics. 2019;35:4537–42. https://doi.org/10.1093/bioinformatics/btz265.

23. Shen A, Millard A. Phage Genome Annotation: Where to Begin and End. PHAGE. 2021;2:183–93. https://doi.org/10.1089/phage.2021.0015.

24. Ng RN, Grey LJ, Vaitekenas A, McLean SA, Rudrum JD, Laucirica DR, et al. Development and validation of a miniaturized bacteriophage host range screening assay against antibiotic resistant Pseudomonas aeruginosa. J Microbiol Methods. 2021;190:106346. https://doi.org/10.1016/j.mimet.2021.106346.

25. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genomics Bioinforma. 2021;3: lqab067. https://doi.org/10.1093/nargab/lqab067.

26. Çevrim E, Yiğit MG, Ulusoy E, Yılmaz A, Doğan T. A Benchmarking Platform for Assessing Protein Language Models on Function-related Prediction Tasks. 2025;:2025.04.10.648084. https://doi.org/10.1101/2025.04.10.648084.

27. Gonzalez-Tortuero E, Krishnamurthi R, Goodhead I, Allison H, James C. Improving phage genome annotation to understand phage biology: the case of Pseudomonas aeruginosa LES prophages. Access Microbiol. 2022;4:po0076. https://doi.org/10.1099/acmi.ac2021.po0318.

28. Wu R, Geiduschek EP. The role of replication proteins in the regulation of bacteriophage T4 transcription. II. Gene 45 and late transcription uncoupled from replication. J Mol Biol. 1975;96:539–62. https://doi.org/10.1016/0022-2836(75)90138-2.

29. Geiduschek EP, Kassavetis GA. Transcription of the T4 late genes. Virol J. 2010;7:288. https://doi.org/10.1186/1743-422X-7-288.

30. Fendley JM, Molari M, Neher RA, Shraiman BI. Synteny and linkage decay in bacteriophage pangenomes. 2025.08.12.669904. https://doi.org/10.1101/2025.08.12.669904.

31. Moraru C, Varsani A, Kropinski AM. VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. Viruses. 2020;12:1268. https://doi.org/10.3390/v12111268.

32. Low SJ, Džunková M, Chaumeil P-A, Parks DH, Hugenholtz P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. Nat Microbiol. 2019; 4:1306–15. https://doi.org/10.1038/s41564-019-0448-z.

33. Sa S, Gf H, Vk M, Rt S. Phage therapy: From biological mechanisms to future directions. Cell. 2023;186. https://doi.org/10.1016/j.cell.2022.11.017.

34. Strathdee SA, Hatfull GF, Mutalik VK, Schooley RT. Phage therapy: From biological mechanisms to future directions. Cell. 2023 ;186:17–31. https://doi.org/10.1016/j.cell.2022.11.017.

35. Comité de l'antibiogramme de la société Française de Microbiologie/European Committe on Antimicrobial Susceptibilty Test vo 1.0.

https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.sfmmicrobiologie.org/wpcontent/uploads/2019/02/CASFMV2.juillet2015.pdf&ved=2ahUKEwio3sGK1qeMAxUKSkEAHRSlJKAQFnoECBoQAQ&usg=AOvVaw37eeZ5f6ItsYJpu88z_xgL. 2024.

36. Sambrook J, Russell DW. Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press; 2001.

37. Kutter E. Phage host range and efficiency of plating. Methods Mol Biol Clifton NJ. 2009;501:141–9. https://doi.org/10.1007/978-1-60327-164-6_14.

38. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.

39. Hung J-H, Weng Z. Mapping Short Sequence Reads to a Reference Genome. Cold Spring Harb Protoc. 2017;2017. https://doi.org/10.1101/pdb.prot093161.

40. 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 J Comput Mol Cell Biol. 2012; 19:455–77. https://doi.org/10.1089/cmb.2012.0021.

41. Nederbragt, Lex. Spades E coli assembly visualised with bandage. 2015. https://doi.org/10.6084/m9.figshare.1589818.v4.

42. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. PeerJ. 2015;3:e985. https://doi.org/10.7717/peerj.985.

43. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol. 2021; 39:578–85. https://doi.org/10.1038/s41587-020-00774-7.

44. Gao R, Naushad S, Moineau S, Levesque R, Goodridge L, Ogunremi D. Comparative genomic analysis of 142 bacteriophages infecting Salmonella enterica subsp. enterica. BMC Genomics. 2020; 21:374. https://doi.org/10.1186/s12864-020-6765-z.

45. Bouras G, Grigson SR, Mirdita M, Heinzinger M, Papudeshi B, Mallawaarachchi V, et al. Protein structure-informed bacteriophage genome annotation with Phold. Nucleic Acids Res. 2026;54: gkaf1448. https://doi.org/10.1093/nar/gkaf1448.

46. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol Biol Clifton NJ. 2019;1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.

47. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C-Y, et al. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. 2023;51:W484–92. https://doi.org/10.1093/nar/gkad326.

48. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. eLife. 2015;4:e08490. https://doi.org/10.7554/eLife.08490.

49. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27:1009–10. https://doi.org/10.1093/bioinformatics/btr039.

50. Darling AE, Mau B, Perna NT. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS ONE. 2010;5:e11147. https://doi.org/10.1371/journal.pone.0011147.

51. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y.

52. Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res. 2023;51:W397–403. https://doi.org/10.1093/nar/gkad313.

53. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33:2379–80. https://doi.org/10.1093/bioinformatics/btx157.

54. Shang J, Peng C, Liao H, Tang X, Sun Y. PhaBOX: a web server for identifying and characterizing phage contigs in metagenomic data. Bioinforma Adv. 2023;3:vbad101. https://doi.org/10.1093/bioadv/vbad101.

55. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33:3396–404. https://doi.org/10.1093/bioinformatics/btx440.

56. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Highly parallelized inference of large genome-based phylogenies. Concurr Comput Pract Exp. 2014;26:1715–29. https://doi.org/10.1002/cpe.3112.

57. Lefort V, Desper R, Gascuel O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. Mol Biol Evol. 2015;32:2798–800. https://doi.org/10.1093/molbev/msv150.

58. Farris GF. The effect of individual roles on performance in innovative groups*. RD Manag. 1972;3:23–8. https://doi.org/10.1111/j.1467-9310.1972.tb00992.x.

59. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinforma. 2020;69:e96. https://doi.org/10.1002/cpbi.96.

60. Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP. Molecular taxonomy of phytopathogenic fungi: a case study in Peronospora. PloS One. 2009;4:e6319. https://doi.org/10.1371/journal.pone.0006319.

61. Meier-Kolthoff JP, Klenk H-P, Göker M. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol. 2014;64 Pt 2:352–6. https://doi.org/10.1099/ijs.0.056994-0.

62. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.

63. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS One. 2010;5:e9490. https://doi.org/10.1371/journal.pone.0009490.

64. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52:W78–82. https://doi.org/10.1093/nar/gkae268.

65. Tynecki P, Guziński A, Kazimierczak J, Jadczuk M, Dastych J, Onisko A. PhageAI - Bacteriophage Life Cycle Recognition with Machine Learning and Natural Language Processing. 2020;:2020.07.11.198606. https://doi.org/10.1101/2020.07.11.198606.

66. Yukgehnaish K, Rajandas H, Parimannan S, Manickam R, Marimuthu K, Petersen B, et al. PhageLeads: Rapid Assessment of Phage Therapeutic Suitability Using an Ensemble Machine Learning Approach. Viruses. 2022;14:342. https://doi.org/10.3390/v14020342.

67. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 2005;33 Database issue:D325-328. https://doi.org/10.1093/nar/gki008.

68. Sayers S, Li L, Ong E, Deng S, Fu G, Lin Y, et al. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. Nucleic Acids Res. 2019;47:D693–700. https://doi.org/10.1093/nar/gky999.

69. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48:D517–25. https://doi.org/10.1093/nar/gkz935.

70. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51. https://doi.org/10.1093/nar/gky425.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2.xlsx