



OPEN ACCESS

EDITED BY

Tasha M. Santiago-Rodriguez,
Baylor College of Medicine, United States

REVIEWED BY

Swapnil Ganesh Sanmukh,
Université Clermont Auvergne, France
Marie-Agnès Petit,
Institut Nationale de la recherche
Agronomique, France

*CORRESPONDENCE

Fernando Fernández-Ramírez
✉ ffernandez@ciencias.unam.mx
Eva Martínez-Peñañiel
✉ penafieleva@gmail.com

†These authors have contributed equally to
this work

†These authors share first authorship

RECEIVED 13 August 2024

ACCEPTED 27 January 2025

PUBLISHED 24 February 2025

CITATION

Negrete-Méndez H, Valencia-Toxqui G,
Sepúlveda-Robles OA, Ríos-Castro E,
Hurtado-Cortés JC, Flores V, Cázares A,
Kameyama L, Martínez-Peñañiel E and
Fernández-Ramírez F (2025) Genomic and
proteomic analyses of Nus-dependent
non-lambdaoid phages reveal a novel
coliphage group prevalent in gut: mEp_{immI}.
Front. Microbiol. 16:1480411.
doi: 10.3389/fmicb.2025.1480411

COPYRIGHT

© 2025 Negrete-Méndez, Valencia-Toxqui,
Sepúlveda-Robles, Ríos-Castro,
Hurtado-Cortés, Flores, Cázares, Kameyama,
Martínez-Peñañiel and Fernández-Ramírez.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Genomic and proteomic analyses of Nus-dependent non-lambdaoid phages reveal a novel coliphage group prevalent in gut: mEp_{immI}

Honorio Negrete-Méndez^{1††}, Guadalupe Valencia-Toxqui^{2††},
Omar A. Sepúlveda-Robles^{3†}, Emmanuel Ríos-Castro⁴,
Jairo C. Hurtado-Cortés¹, Víctor Flores⁵, Adrián Cázares⁶,
Luis Kameyama^{1†}, Eva Martínez-Peñañiel^{1*†} and
Fernando Fernández-Ramírez^{7*†}

¹Departamento de Genética y Biología Molecular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav), México City, México, ²Department of Biology, Center for Phage Technology, Texas A&M University, College Station, TX, United States, ³Unidad de Investigación Médica en Genética Humana, UMAE Hospital de Pediatría, Centro Médico Nacional "Siglo XXI", Instituto Mexicano del Seguro Social (IMSS), México City, México, ⁴Genomics, Proteomics and Metabolomic Core Facility (UGPM), LaNSE, Cinvestav-IPN, México City, México, ⁵Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom, ⁶Parasites and Microbes Programme, Wellcome Sanger Institute, Cambridge, United Kingdom, ⁷Unidad de Genética, Hospital General de México "Dr. Eduardo Liceaga", México City, México

Introduction: Nus-dependent Mexican *Escherichia coli* phages (mEp) were previously isolated from clinical samples of human feces. Approximately 50% corresponded to non-lambdaoid temperate phages integrating a single immunity group, namely immunity I (mEp_{immI}), and these were as prevalent as the lambdaoid phages identified in such collection.

Methods: In this work, we present the structural and functional characterization of six representative mEp_{immI} phages (mEp010, mEp013, mEp021, mEp044, mEp515, and mEp554). In addition, we searched for homologous phages and prophages in the GenBank sequence database, and performed extensive phylogenetic analyses on the compiled genomes.

Results: A biological feature-based characterization of these phages was carried out, focusing on proteins relevant to phage biological activities. This included mass spectrometry analysis of mEp021 virion structural proteins, and a series of infection assays to characterize the function of the main repressor protein and the lipoproteins associated with superinfection-exclusion; to identify the main host receptor proteins recognized by these phages and the prophage insertion sites within the host genome, which were associated with specific integrase sequence-types present in the viral genomes. Further, we compiled 42 complete homologous genomes corresponding to 38 prophages from *E. coli* strains and 4 phages from metagenomes, displaying a wide geographical distribution. Intergenomic distance analyses revealed that these phages differ from previously established phage clades, and whole-proteome similarity analyses yielded a cohesive and monophyletic branch, when compared to >5,600 phages with dsDNA genomes.

Discussion: According to current taxonomic criteria, our results are consistent with a novel family demarcation, and the studied genomes correspond to 9 genera and 45 distinct species. Further, we identified 50 core genes displaying high synteny among the mEp_{immI} genomes, and these genes were

found arranged in functional clusters. Furthermore, a biological feature-based characterization of these phages was carried out, with experiments focusing on proteins relevant to phage biological activities, revealing common traits as well as diversity within the group. With the integration of all these experimental and bioinformatics findings, our results indicate that the mEp_{immI} phages constitute a novel branch of *Caudoviricetes* distinct to other known siphovirus, contributing to the current knowledge on the diversity of phages infecting *Escherichia coli*.

KEYWORDS

mEp_{immI} phages, phage phylogenomics, Nus-dependent phages, phage repressor, phage integration, antitermination, outer membrane receptors, phage superinfection-exclusion

1 Introduction

Bacteriophages or phages constitute the most abundant biological entity in our planet, with 10^{31} estimated viral particles in the biosphere, displaying a wide genetic diversity due to high recombination rates (Mushegian, 2020). Phage morphological features were one of the earliest criteria used for taxonomic classification, allowing the recognition of the order *Caudovirales* (i.e., tailed virus), which includes families *Myoviridae*, *Podoviridae* and *Siphoviridae* (Ackermann and Prangishvili, 2012; Mavrich and Hatfull, 2017). With the advent of second-generation sequencing methodologies, the number of phage and prophage genomes deposited in databases exponentially increased. Consequently, the morphology-based classification was revised in 2022, and genomic and proteomic features were incorporated as main taxonomic criteria, which in turn allow for a consistent and automated classification (Turner et al., 2021, 2024). Currently, 48 phage families have been integrated into a new class named *Caudoviricetes*, which includes all bacterial and archaeal tailed virus with icosahedral capsids and double-stranded DNA genomes (Turner et al., 2021, 2023; Zhu et al., 2022). Phages displaying substantial genome mosaicism may represent important challenges for taxonomic classification, as their genomes are the result of multiple and diverse horizontal transfer events that have occurred throughout their evolution, which have led to similarities in genome organization and regulation among phages from distinct origins (Hatfull, 2008; Cázares et al., 2014). Therefore, various unrelated phages were eventually clustered together in larger groups based on their biological features (Comeau et al., 2007; Evseev et al., 2021; Kupczok et al., 2022). Former terms such as Mu-like, T4-like, T7-like or lambdoid were traditionally used to designate groups of phages with functional resemblance, and refer to phages sharing similarities in genome organization, regulatory systems and protein functions, rather than a strict taxonomic identity based on nucleotide and protein sequence analyses (Comeau et al., 2007). For instance, phage P22 that infects *Salmonella* spp. is a podovirus (genus *Lederbergvirus*) morphologically different to phage λ (genus *Lambdavirus*), however, it is usually classified as a lambdoid phage, a large biological group mainly constituted by siphovirus infecting *Escherichia coli*. The similarities in genome organization among lambdoid phages have led to the ability to produce fertile hybrids (Botstein and Herskowitz, 1974; Campbell, 1994; Schicklmaier

et al., 1999). Thus, the obtention of recombinant virions capable of producing viable progeny constitutes one basic genetic approach commonly used to discriminate between lambdoid and non-lambdoid coliphages; for this, co-infection assays using a reference lambdoid phage and a tested phage are performed. Other methods for phage grouping include the assessment of similarity in the N and Q antitermination systems, and in the repressor-based immunity control (Fattah et al., 2000; Hendrix, 2002; Degnan et al., 2007).

Lambdoid phages have been extensively characterized, being λ the prototype of this group. Their genomes are sequentially organized in functional clusters of conserved genes (regulation, replication, lysis, structural proteins, etc.), with several interspersed accessory genes (Casjens and Hendrix, 2015). The gene regulation of these phages is organized in four main transcriptional units: the left operon which includes the antitermination N gene, among others; the right operon with early regulators *cro* and *cII*, and replication genes such as *O* and *P*; the late operon involved in transcription of structural and lysis genes; and the immunity operon containing the *cI* repressor gene, and others (Dydecka et al., 2020; Kupczok et al., 2022).

One particular trait of lambdoid phages is the requirement of host Nus factors and the viral N and Q proteins for their antitermination function. Kameyama et al. (1999) reported a collection of 96 Nus-dependent coliphages named mexican *Escherichia coli* phages (mEp), which were originally isolated from clinical samples of human feces. From these, 46 corresponded to lambdoid phages that were further classified in 19 immunity groups (II-XX). The remaining 50 phages integrated the most numerous immunity group (I) which displayed no similarity to lambdoid phages, other than the use of host Nus factors. Conversely, the DNA of these phages did not hybridize with λ DNA probes; their structural proteins (i.e., those related to the capsid and tail viral structures) were not recognized by anti- λ antibodies, most were not induced by UV light treatment, and no viable progeny was obtained in co-infection assays using λ -BLK20, a phage λ derivative harboring a *N-IacZ* gene fusion at the left side of *att* site that produces faint-blue lysis plaques on standard X-gal plates, allowing the screening of recombination events between phages (Kameyama et al., 1991). All these results indicated that the mEp immunity group I phages (mEp_{immI}) were distinct from the lambdoid phages. Noteworthy, these phages

were as prevalent in biological samples as the lambdoid groups (Kameyama et al., 1999, 2001).

In this work, we extended the characterization of the mEp_{imm1} group by experimental approaches and whole-genome sequencing of 6 representative coliphages (12%), including mEp010, mEp013, mEp021, mEp044, mEp515 and mEp554. We also identified 42 homologous complete phage and prophage genomes from sequences deposited in the GenBank database. Following current taxonomic criteria, we present extensive comparative genomics and proteomics analyses, revealing that the mEp_{imm1} group constitutes a novel branch of *Caudoviricetes* distinct from previously described coliphages. In order to explore their similarities and diversity, and reinforce their status as an independent group, we performed a traditional biological-feature characterization of these phages, focusing on proteins whose function is related to the virion structure as well as to the most relevant biological activities of phages, namely receptor usage and recognition, genome integration, lysogeny regulation, antitermination, and superinfection-exclusion. Taken together, our results provide evidence of the uniqueness and diversity within this novel phage group.

2 Materials and methods

2.1 Bacteria, plasmids, and media

The bacterial strains and plasmids used in this study are listed in [Supplementary Table S1](#). Derivative *E. coli* K-12 wild type strain W3110 was used for phage propagation and infection assays. Keio strains JW0940-6 (*ompA*⁻), JW3996-1 (*lamB*⁻) and JW2203-1 (*ompC*⁻) were used in infection assays in order to assess the different membrane receptor proteins used by the studied phages (Datsenko and Wanner, 2000; Baba et al., 2006). The DH5 α strain was used for plasmid manipulation. Lysogenic broth (LB; 10 g/L tryptone, 5 g/L yeast extract 10 g/L NaCl) and TMG phage-dilution media (10 mM Tris-HCl, pH 7.2, 10 mM MgSO₄, and 0.1% gelatin) were prepared as previously described (Silhavy et al., 1984). LB media was supplemented with ampicillin (100 μ g/mL), kanamycin (30 μ g/mL) or isopropyl β -D-1-thiogalactopyranoside (IPTG 0.1 mM) (Sigma-Aldrich, St. Louis, MA, USA) when required. All media were purchased from BD Difco™ (Franklin Lakes, NJ, USA). T4 DNA ligase and restriction endonucleases *Nde*I, *Eco*RI, and *Hind*III were purchased from New England Biolabs (Ipswich, MA, USA).

2.2 Phages and prophages

Phages mEp021, mEp010, mEp013, mEp044, mEp515 and mEp554 were selected as representative phages from the immunity group I (Kameyama et al., 1999). For the experimental approaches described in this study, the phages are listed in [Supplementary Table S1](#). The λ phage was used as negative control in all the infection assays. The phages and prophages used for comparative genome analyses are listed in [Table 1](#). These genomes were retrieved from the GenBank database (NCBI) through BLASTn searches (Altschul et al., 1997) using as query a phage genome belonging to immunity group

I (mEp021). We selected hits with query coverage threshold = 60%, *E*-value = 0.0, and percent identity threshold = 70%. For phage genomes ($n = 4$), the keywords “Viruses (taxid:10239)” or “Caudoviricetes (taxid:2731619)” were used in the query, and for prophage genomes ($n = 38$) the keyword was “*Escherichia coli* (taxid:562).” In all cases, the retrieved genomes were manually inspected to assure their completeness. Genomes corresponding to 11 siphovirus (λ , HK022, HK97, N15, T1, T5, psiM2, SPBc2, c2, phi-C31, and L5) and 1 podovirus (P22) were used as external references in the main comparative analyses ([Supplementary Table S2](#)), and an additional larger dataset, including these reference phages and 50 additional neighboring phages (as observed in the ViPTree analysis, [Supplementary Figure S4A](#)), was used to increase the coverage of our analyses ([Supplementary Table S3](#)).

2.3 Phage propagation, phage DNA extraction and RFLP analysis

For phage propagation, 1 mL of *Escherichia coli* W3110 (Bachmann, 1972; Asakura and Kobayashi, 2009) was cultured overnight (O/N) and mixed with ~ 10 plaque-forming units (PFU) of phage, and 1 mL of 0.5 M CaCl₂:MgCl₂ (1:1). After allowing the phages to adsorb for 15 min, 50 mL of LB broth was added to obtain a final concentration of 10 mM CaCl₂:MgCl₂ (1:1). The culture was incubated at 37°C with shaking at 200 rpm until the cells were lysed, 5 mL of chloroform was added, the sample was centrifuged at 4,200 $\times g$ for 10 min, and the supernatant was recovered. For phage DNA extraction, 50 mL of supernatant was treated with PEG-8000 (Sigma-Aldrich, St. Louis, MA, USA) and NaCl at final concentration of 10% and 1 M, respectively. The solution was incubated for 8 h at 4°C and centrifuged at 10,590 $\times g$ for 10 min. The obtained pellet was resuspended in 1 mL of TMG, treated with chloroform (v:v) (J.T. Baker, Phillipsburg, NJ, USA) and centrifuged at 9,300 $\times g$ in order to eliminate PEG-8000 (Sigma-Aldrich, St. Louis, MA, USA). The recovered phages were purified by cesium chloride gradient as indicated in previous protocols (Shevchenko et al., 2006). The collected band was dialyzed in Tris 50 mM, NaCl 10 mM and MgCl₂ 10 mM solution for 12–18 h at 4°C. The phage DNA was extracted as described previously (Sambrook et al., 1989) and the pellet was resuspended in 50 μ L of nuclease-free water for restriction analysis and DNA sequencing. For restriction fragment length polymorphism (RFLP) analysis, 1 μ g of genomic DNA was restricted with *Nde*I (10 U), then the products were resolved and visualized in 1% agarose gel electrophoresis.

2.4 Transmission electron microscopy (TEM)

10 μ L of dialyzed CsCl-purified bacteriophages (1×10^{11} PFU/mL) were deposited on a Formvar carbon-coated grid and incubated for 1 min. The excess liquid was removed with filter paper and then 10 μ L of 1% uranyl acetate (Electron Microscopy Science, Hatfield, PA, USA) was applied for 30 and 60 s and removed. The preparations were analyzed using the electron transmission microscopes JEM-1400 (JEOL, Akishima, Japan) at 100 kV.

TABLE 1 Studied phages and prophages of mEp_{imm1} group.

mEp (our laboratory)	Host	Bacterial host	Isolation source	Country	Accession number
mEp021	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	México	MH706966.1
mEp010	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	México	PP180001
mEp013	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	México	PP180002
mEp044	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	México	PP180004
mEp515	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	México	PP180003
mEp554	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	México	PP180005

Bacteriophages (metagenomes)	Host	Bacterial host	Isolation source	Country	Accession number
0621_18038	<i>Homo sapiens</i>	ND	Feces	Japan	OP073365.1
ct11k1	<i>Homo sapiens</i>	ND	Human gut	Denmark	BK025398.1
ctPXR1	<i>Homo sapiens</i>	ND	Human gut	USA	BK024291.1
ctx8n3	<i>Homo sapiens</i>	ND	Human gut	Finland	BK020724.1

Prophages (bacterial genomes)	Host	Bacterial host	Isolation source	Country	Accession number
FHI99	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	Norway	LM997270.1
UMNK88	<i>Sus scrofa domesticus</i>	<i>Escherichia coli</i>	Porcine Neonatal Diarrhea	USA	NC_017641.1
504005_aEPEC	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	India	NZ_CYBZ01000006.1
09-00049	ND	<i>Escherichia coli</i>	Lettuce	USA	NZ_CP015228.1
Ecol_244	<i>Homo sapiens</i>	<i>Escherichia coli</i>	ND	Argentina	CP019020.1
1-110-08_S4_C1	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	Tanzania	NZ_JHDK01000106.1
1-392-07_S4_C3	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	Tanzania	NZ_JOSH01000199.1
HVH 69 (4-2837072)	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (UTI induced bacteremia)	Denmark	AVUV01000007.1
HVH 103 (4-5904188)	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (UTI induced bacteremia)	Denmark	AVVS01000012.1
KOEGE 62 (175a)	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (UTI induced bacteremia)	Denmark	AWAL01000002.1
KTE83	<i>Homo sapiens</i>	<i>Escherichia coli</i>	ND	Denmark	ANUT01000018.1
MS21-1	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Gastrointestinal_tract	USA	NZ_GG772625.1
FBPI	<i>Sus scrofa domesticus</i>	<i>Escherichia coli</i>	Feces	The Netherlands	AYKC01000020.1
O5:K4(L):H4 str. ATCC235002	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Urine	NA	NZ_CAPL01000001.1
G180	<i>Bos taurus</i>	<i>Escherichia coli</i>	Mastitis	France	NZ_LONT01000016.1
HVH 110 (4-6978754)	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (UTI induced bacteremia)	Denmark	AVVY01000006.1
UMEA 3323-1	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Urine	Sweden	NZ_KI530679.1
MIN12	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Human postoperative wound swab	Poland	CP069657.1
LSU61	<i>Cervidae</i>	<i>Escherichia coli</i>	ND	USA	CP038336.1
2-331	ND	<i>Escherichia coli</i>	Sewage	Norway	CP110117.1
127	<i>Canis lupus familiaris</i>	<i>Escherichia coli</i>	Whole organism	United Kingdom	CP023377.1
190	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Urine	USA	CP020520.1

(Continued)

TABLE 1 (Continued)

Prophages (bacterial genomes)	Host	Bacterial host	Isolation source	Country	Accession number
C16EC0292	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (bloodstream infection)	South Korea	CP088668.1
E21845	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Diarrhea	Thailand	NZ_CP076272.1
FAH	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces from healthy subject	China	CP116035.1
GN02175	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (bloodstream infection)	USA	CP041550.1
GN04592	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (bloodstream infection)	USA	CP095538.1
GN05505	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Blood (bloodstream infection)	USA	CP095526.1
KKa019	ND	<i>Escherichia coli</i>	River water	Japan	CP091664.1
M1/5	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	Germany	CP053296.1
MLI107	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Feces	Mali	CP116987.1
PAR	<i>Cacatuidae</i>	<i>Escherichia coli</i>	Feces	China	CP012379.1
RHBSTW-00392	ND	<i>Escherichia coli</i>	Wastewater effluent sample	United Kingdom	CP056564.1
RM9245	ND	<i>Escherichia coli</i>	Feces	USA	CP044314.1
STEC306	<i>Ovis aries</i>	<i>Escherichia coli</i>	Mutton	China	NZ_CP091016.1
STEC307	<i>Ovis aries</i>	<i>Escherichia coli</i>	Mutton	China	NZ_CP091018.1
STEC308	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Patient	China	NZ_CP041435.1
WCHEC025970	<i>Homo sapiens</i>	<i>Escherichia coli</i>	Pure culture	China	CP036177.1

Six mEp_{imm1} phages are from the collection previously reported (Kameyama et al., 1999), and were experimentally characterized in this work. Four homologous bacteriophage and 38 prophage complete genomes were identified from metagenomes and *E. coli* genomes from GenBank, respectively, and were integrated in subsequent phylogenomics analyses.

2.5 Virion structural protein analysis and detection by mass spectrometry (MS)

2.5.1 Sample processing of the structural proteins of the virus

30 μ L of CsCl-purified bacteriophage particles were resuspended in Laemmli buffer and were boiled in a water bath for 10 min followed by 5 min on ice. The sample (20 μ L) was resolved by 16% SDS-PAGE at 90 V for 2.5 h. Protein bands were visualized by staining with Silver Stain Plus (Bio-Rad, Hercules, CA, USA), following the protocol provided by the supplier. Proteins from a mEp021 virion sample were resolved by 10% SDS-PAGE. Fifteen gel slides corresponding to discrete protein bands were enzymatically digested with trypsin, according to the modified protocol of Shevchenko et al. (2006) and Barrera-Rojas et al. (2023). Briefly, bands were excised from the gel and transferred into centrifuge microtubes to be destained using a solution containing 2.5% formic acid (FA), 50% methanol (MeOH), and were subsequently dehydrated with acetonitrile (ACN), and the remaining solvent was eliminated in a Savant DNA120 SpeedVac Concentrator (Thermo Fisher Scientific, Waltham, MA, USA) for 10 min. Then, proteins were reduced with 10 mM DTT (Sigma-Aldrich, St. Louis, MO, USA) in 100 mM ammonium bicarbonate (ABC) (Sigma-Aldrich, St. Louis, MO, USA) and alkylated with 50 mM iodoacetamide (IAA) (Sigma-Aldrich, St. Louis, MO, USA) in 100 mM ABC. Afterward, bands were washed with 100 mM ABC and dehydrated with ACN; subsequently, bands were hydrated and washed again with 100 mM ABC and dehydrated with

ACN. Then, the excess solvent was removed using the SpeedVac for 10 min. Proteins were enzymatically digested overnight using 20 ng/ μ L of trypsin (Sigma-Aldrich, St. Louis, MA, USA) in 50 mM ABC at 37°C in a Precision water bath (Thermo Fisher Scientific, Waltham, MA, USA). Once the time had passed, the reaction was stopped with 40 μ L of a solution of 5% FA for 10 min at room temperature; subsequently, peptides were eluted from the gel for two cycles using 40 μ L of a solution of 5% FA and 50% ACN. Peptides were concentrated in the SpeedVac and desalted using ZipTips C18 (Millipore, Burlington, MA, USA).

2.5.2 Mass spectrometry analysis

The resulting tryptic peptides were concentrated to an approximate volume of 10 μ L; 4 μ L were loaded into a Symmetry C18 Trap V/M precolumn (Waters, Milford, MA, USA); 180 μ m X 20 mm, 100 Å pore size, 5 μ m particle size and desalted using as a mobile phase A, 0.1% formic acid (FA) in H₂O and mobile phase B, 0.1% FA in acetonitrile (ACN) under the following isocratic gradient: 95% mobile phase A and 5% of mobile phase B at a flow of 15 μ L/min during 5 min. Then, peptides were loaded and separated on a HSS T3 C18 Column (Waters, Milford, MA, USA); 75 μ m X 150 mm, 100 Å pore size, 1.8 μ m particle size; using an UPLC ACQUITY M-Class (Waters, Milford, MA, USA) with the same mobile phases mentioned above under the following gradient: 0–3 min 10% B (90% A), 15 min 20% B (80% A), 60 min 60% B (40% A), 61–64 min 90% B (10% A), 65 to 70 min 10% B (90% A) at a flow of 250 nL/min and 30°C. The spectra data were acquired in a mass spectrometer with electrospray ionization (ESI) and ion mobility

separation (IMS) Synapt G2- Si (Waters, Milford, MA, USA) using data-independent acquisition (DIA) through HDMS^E mode (Full-Scan DIA) (Lou and Shui, 2024). The tune page for the ionization source was set with the following parameters: 2.60 kV in the sampling capillary, 30 V in the sampling cone, 30 V in the source offset, 70°C for the source temperature, 0.6 Bar for the nano flow gas and 120 L/h for the purge gas flow. Two chromatograms were acquired (low and high energy) in positive mode in a *m/z* range of 50–2000 with a scan time of 500 ms. No collision energy was applied to obtain the low energy chromatogram, while for the high energy chromatograms the precursor ions were fragmented in the transfer using a collision energy ramp of 30–45 V. Synapt G2-Si was calibrated with [Glu¹]-Fibrinopeptide fragments, through the precursor ion $[M + 2H]^{2+} = 785.84261$ fragmentation of 32 eV with a result of ≤ 1 ppm across all MS/MS measurements.

2.5.3 Database search

The generated mEp021.raw files containing MS and MS/MS spectra were deconvoluted and compared using ProteinLynx Global SERVER (PLGS) (Li et al., 2009) v3.0.3 software (Waters, Milford, MA, USA) using a target decoy strategy against an in-house fasta database of 82 predicted proteins from the mEp021 genome (Elias and Gygi, 2009; Käll et al., 2007). Workflow parameters were Trypsin as a cut enzyme and one missed cleavage allowed: carbamidomethyl (C) as a fixed modification and acetyl (K), acetyl (N-term), amidation (N-term), deamidation (N, Q), oxidation (M), Phosphoryl (S, T, Y) as variable modifications. Automatic peptide and fragment tolerance, minimum fragment ion matches per peptide: 2, minimum fragment ion matches per protein: 5, minimum peptide matches per protein: 1, and false discovery rate of 4%.

2.6 Genome sequencing and *de novo* assembly of mEp021

The genome of prototype phage mEp021 was sequenced at the National Laboratory of Genomics for Biodiversity (Cinvestav, Irapuato, México) using the AB SOLiD technology (Applied Biosystems, Waltham, MA, USA). The sequencing reads were preprocessed for base calling, quality score assignment and filtering using the Applied Biosystems (Waltham, MA, USA) *de novo* assembly accessory software with default values. The phage genome was assembled *de novo* using Velvet v1.1 (Zerbino and Birney, 2008) with default settings, resulting in eight contigs. The assembly was examined using Tablet (Milne et al., 2015) and manually edited to remove errors. The complete mEp021 genome was then assembled into a single scaffold by Sanger sequencing, using primers designed to amplify outwards of each contig in order to fill the sequence gaps (see Supplementary Table S4 and Supplementary Figure S1). Sequencing products were read in an ABI PRISM 310 automatic sequencer (Applied Biosystems, Waltham, MA, USA) of the Genetics and Molecular Biology Department facility (Cinvestav-IPN, México City, México).

Coding sequences in the genome of mEp021 were predicted with heuristic Hidden Markov Models using GeneMark v4.3 (Borodovsky and Lomsadze, 2011), and the corresponding ribosome-binding sites were detected with RBS_finder (Suzek et al., 2001). Gene calling was carried out via comparison of the predicted mEp021 ORF products against the non-redundant protein database from NCBI using

BLASTp, with a 50% coverage, 1E-4 e-value and > 0% identity threshold (Altschul et al., 1997). InterProScan (Paysan-Lafosse et al., 2022) was used to identify conserved domains and protein families among the identified ORFs. The Artemis annotation tool (Carver et al., 2011) was used to integrate the results from BLASTp and InterProScan, and to visualize the genome. Final annotation was achieved by comparing these annotations with the one obtained through the PHAge Search Tool (PHAST) server (Zhou et al., 2011).

2.7 Genome sequencing, assembly, and annotation of mEp010, mEp013, mEp044, mEp515, mEp554

Phage genomes were sequenced using Illumina technology, with services provided by the Microbial Genome Sequencing Center (Pittsburgh, PA, USA) using a MiSeq instrument (for: mEp010, mEp013, mEp044, mEp554) and in the SeqCenter (Pittsburgh, PA, USA) using a NovaSeq 6000 sequencer (for: mEp515). In all cases, the sequencing was based on the protocol by Baym et al. (2015). The genome assembly was performed using the Galaxy interface,¹ a graphic environment that provides well established bioinformatics tools for genome assembly and analyses. The quality of the raw reads was evaluated with FastQC version 0.12.1.² Low quality bases were removed using Trimmomatic version 0.38.1 using the default parameters (Q20), and the reads were assembled into contigs using the Shovill genome assembly pipeline, which is dedicated to microorganisms with small genome sizes. This pipeline uses the SPAdes assembler version 1.1.0 at its core, which is currently considered as the standard *de novo* genome assembler for Illumina platforms; K-mer was set as default. Once assembled, the complete genomes were annotated using the Genome Annotation Service available in the Bacterial and Viral Bioinformatics Resource Center (BV-BRC),³ using the following parameters: annotation recipe: “bacteriophages”; taxonomy name: “Enterobacteria phage mEp021” (the mEp021 genome was set as reference), and Taxonomy ID: “1150757.” The annotation recipe for bacteriophages is based on the PHANOTATE method, which is specifically designed for phage gene calling (McNair et al., 2019). The accession numbers of mEp010, mEp013, mEp021, mEp044, mEp515 and mEp554 are, respectively, indicated in Table 1.

2.8 *In silico* comparative phylogenomics

Two complementary approaches were used to analyze the phylogenetic relationships of the studied mEp_{imm1} phages: one at the nucleotide sequence-level, including whole-genome alignments and distance-based clustering using different tools, and another at the whole proteome-level; in both cases we included sets of bacteriophage genomes as external references. Whole-genome alignments were performed using the NGPhylogeny server (Lemoine et al., 2019), which includes the Multiple Alignment Program for amino acid or

1 <https://usegalaxy.org>

2 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

3 <https://www.bv-brc.org/>

nucleotide sequences (MAFFT), the Block Mapping and Gathering with Entropy program (BMGE) for aligned sequence cleaning, and Maximum likelihood-based Inference of phylogenetic trees with Smart Model Selection (PhyML+SMS), to construct intergenomic distance-based phylogenetic trees. The VIRIDIC tool (Moraru et al., 2020) was used to calculate intergenomic distance matrices to allow the identification of genera and species within the analyzed phages.

The Viral Proteomic Tree server (ViPtree) v4.0 (Nishimura et al., 2017), which is a tBLASTx-based tool recommended by International Committee on Taxonomy of Viruses (ICTV) for family-level classification of viruses (Turner et al., 2024), was used to perform comparative whole-proteome analysis of the 48 mEp_{imm1} phages against the proteomes of >5,600 dsDNA reference phages deposited in its database.⁴ VirClust, a tool for higher taxon-level analyses (Moraru, 2023), was used to calculate intergenomic distances based on conserved protein clusters (PC) and to identify viral genome clusters (VGC), using 100 bootstrap resampling and default parameters for the 60 genomes dataset, and a 0.7 distance threshold for the 110 genomes dataset.

Analyses of the repressor, integrase, Gp17-like, J and Lpp putative proteins of the mEp_{imm1} phages were carried out using MultAlin⁵ for multiple protein sequence alignments. Analysis of the sequences surrounding the *attR* and *attL* of the prophages were carried out through multiple nucleotide-sequence alignments, which were performed with Clustal Omega (v1.0)⁶ (Sievers et al., 2011). All of these alignments were visualized using ESPript (v3.0) (Robert and Gouet, 2014). For the analysis of *nut* sites, the sequences of *nutR1*, *nutR2* and *nutL* from mEp021 were searched in the genomes of the other mEp_{imm1} phages, and the consensus sequence was visualized using WebLogo v2.8.2 (Crooks et al., 2004).

2.9 Identification of the core genome of the mEp_{imm1} phages

We manually determined the percentages of core genes, flexible genes and unique genes in the set of 48 phages and coliphages genomes. In order to homogenize gene annotations, the 42 complete genome sequences of the homologous phages and prophages from GenBank and mEp021 were reannotated, using PHANOTATE via BV-BRC (for details see the previous subsection). According to the relative position of the genes identified in the phage genomes, all of the predicted proteins were manually inspected. Those sharing similar amino acid sequences in either terminal segment, were selected and pairwise-aligned against the corresponding mEp021 proteins; this allowed the verification of their identities, which in all cases was above 58%. In this manner, we identified gene products that were common to all the studied genomes (i.e., core genes); gene products that were identified in more than one genome but not all genomes were considered as flexible, and those only present in a single genome were labeled as unique. We performed pairwise-alignments of the core proteins, taking their corresponding mEp021 homologs as reference, using Geneious Prime® 2024.0.5. The resulting identity percentage

scores were manually compiled and used to generate a correlation heatmap using ChiPlot.⁷ Automated analysis of the core genome of the 48 mEp_{imm1} phages was performed using VirClust with default parameters (Moraru, 2023); the gene calling pipeline in this tool is based on the MetaGeneAnnotator for bacterial and phage genes (Noguchi et al., 2008), and allows the identification of conserved protein clusters among the input phage genomes.

2.10 Construction of plasmids pRep₀₂₁, pLpp₀₂₁ and pLpp₀₁₀

The biological function of the predicted repressor (Rep) and lipoprotein (Lpp) proteins of the mEp_{imm1} phages was tested by means of phage infection assays. For this, we used our sequenced phages as templates to amplify and clone the corresponding coding genes in expression vectors. Considering the strong sequence similarity that the repressor proteins displayed among all the mEp_{imm1} phages, we selected the repressor gene of the archetype phage mEp021, which is the best characterized phage of our collection. Hence, the *gp15* repressor gene of mEp021 was amplified using specific primers (Supplementary Table S4). For lipoprotein genes, we based our selection on the following criteria: first, two out of the four identified lipoprotein sequence-types were represented in the phages of our collection (mEp021 and mEp515 for type-A lipoproteins, and mEp010, mEp013, mEp044 and mEp554 for type-C lipoproteins); notably, the amino acid sequences of the type-A lipoproteins were identical, while the ancestral sequence for type-C lipoproteins corresponded to that of mEp010. Hence, the *gp81* gene of mEp021 and the *gp116* gene of mEp010 were amplified via PCR using the primers described in Supplementary Table S4. In the case of *gp81*, the reverse primer was designed with a TWA codon (W = A or T) positioned upstream to the sequence coding for the 6xHisTag, and a TAA codon after it (Supplementary Figure S2). For the experiments described in this work, we only selected the construction with a stop codon before the 6xHisTag, which yielded a wild-type Gp81 lipoprotein. Amplicons were inserted into the transit plasmid pJET1.2blunt (Thermo Fisher Scientific, Waltham, MA, USA). Subsequently, the inserts corresponding to *gp15*, *gp81* and *gp116* genes were recovered from the transit plasmid by digestion with *EcoRI* and *HindIII* and were ligated into the *EcoRI*–*HindIII* sites of the low copy-number vector pKQV4 (Strauch et al., 1989) for IPTG-inducible expression from the *Ptac* promoter, generating the plasmids pRep₀₂₁, pLpp₀₂₁, and pLpp₀₁₀ (Supplementary Figure S2).

2.11 Phage infection assays

Several aspects related to the mEp_{imm1} most relevant biological activities were tested by means of phage infection assays, including the repressor function, outer membrane receptor (OMR) recognition, and superinfection-exclusion. Infection assays were performed as described previously (Arguijo-Hernández et al., 2018). For each case, Keio collection strains and W3110 were transformed with pRep₀₂₁,

4 <https://www.genome.jp/viptree/>

5 <http://multalin.toulouse.inra.fr/>

6 <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>

7 <https://www.chiplot.online>

pLpp₀₂₁, pLpp₀₁₀ or pKQV4, respectively. Cultures of each strain were grown O/N in LB broth at 37°C in a rotary shaker at ~200 rpm, and 300 µL of each one were mixed with 3 mL of soft top agar and poured on LB-agar plates, forming a bacterial lawn. Serial dilutions of phages were prepared in TMG solution and spotted onto the bacterial lawn, and the plates were incubated overnight at 37°C.

2.12 *In silico* three-dimensional structure analyses of the repressor proteins

Three-dimensional models of the repressor protein of mEp021 (namely Rep₀₂₁) were generated using the ColabFold software (Mirdita et al., 2023) available in UCSF ChimeraX v1.7.1 (Meng et al., 2023), the DI-TASSER (Deep learning-based Iterative Threading ASSEMBly Refinement) server (Zheng et al., 2023) and the Phyre² Protein Homology/analogY Recognition Engine V 2.0 server (Kelley et al., 2015). The crystallographic structure of the λ repressor (3BDN) was used for further superposition analysis with the different Rep₀₂₁ models, using Phyre² and PyMOL (Schrödinger and DeLano, 2020).

2.13 Phage attachment site validation by PCR and sequencing

During the lysogenic cycle of phages, the phage DNA may integrate into the host genome, using an integrase protein that recognizes specific insertion motifs within the host genome and promotes the integration of the phage DNA. In order to explore the integration process of the mEp_{imm1} phages, we predicted the attachment sites and integrase types through sequence analyses, and validated these findings in the W3110 (mEp021) lysogen by means of DNA sequencing. Briefly, bacterial colonies from W3110 (mEp021) lysogens and the W3110 isogenic strain were selected and resuspended in 100 µL of distilled H₂O. Genomic DNA of these cells was obtained by mechanical disruption using glass beads (SIGMA, G-1634) and vortexing for 5 min. Then, 5 µL of each lysate were mixed with 10 nM dNTP mix, and 10 µM primers T1, B1, T2 and B2 in different paired combinations (Supplementary Table S4), and Dream Taq polymerase (Thermo Fisher Scientific, Waltham, MA, USA), according to the PCR protocol provided for the enzyme. Reactions were performed using a T100 Thermal Cycler (Bio-Rad, Hercules, CA, USA) under the following conditions: a denaturing step at 94°C/5 min, followed by 35 amplification cycles of 94°C/30 s, 58°C/60 s and 68°C/30 s for each cycle, and a final extension step at 72°C/10 min. The PCR products were then sequenced using the *attR* and *attL* primers (described in Supplementary Table S4), in an automatic sequencer Perkin Elmer™ ABI PRISM™ 310 (Applied Biosystems, Waltham, MA, USA), at the sequencing facility of the Genetics and Molecular Biology Department (Cinvestav-IPN, México City, México).

2.14 Verification of Keio mutant strains

The genotypes of the JW0940-6, JW3996-1 and JW2203-1 Keio strains were verified by PCR (Supplementary Figure S9), using specific primers listed in Supplementary Table S4. Briefly, the forward primer was directed to the middle region of the kanamycin cassette and the

reverse primer was directed to the 3' region of the *ompA*, *lamB* or *ompC* genes, respectively. The bacterial DNA was obtained by mechanical disruption with glass beads (Sigma-Aldrich, St. Louis, MA, USA) and vortexing for 5 min. The PCR reaction mix was prepared following the protocol supplied for DreamTaq DNA polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Reactions were performed using a T100 Thermal Cycler (Bio-Rad, Hercules, CA, USA) under the following conditions: a denaturation step for 5 min at 95°C, followed by 34 cycles of denaturation for 45 seg at 95°C, annealing for 30 seg at 56.5°C and elongation for 60 seg at 68°C, and a final step of 7 min at 68°C. The amplification products were resolved by electrophoresis in 1% agarose gels stained with ethidium bromide, and visualized using a UV transilluminator.

3 Results

3.1 Structural characterization of representative mEp_{imm1} phages

We experimentally characterized six mEp_{imm1} phages that were selected from the collection previously reported by Kameyama et al. (1999): mEp010, mEp013, mEp021, mEp044, mEp515, and mEp554. These phages present a siphovirus morphology, with ~60 nm capsids and ~130 nm tails, according to TEM (Figure 1A). Unique RFLP and structural protein electrophoretic patterns were observed for each of these phages (Figures 1B,C). In the case of mEp021, we identified 16 structural proteins by mass spectrometry (MS) analysis, as indicated in Figure 1D. Whole genome sequencing of the six mEp_{imm1} phages was performed, and tBLASTx analysis revealed a conserved genome organization among these phages and that their proteins share high similarity rates (>80%, Figure 2A). Annotation of the sequenced genomes and mapping of the structural protein-coding genes (Figure 1D, top) indicated the presence of clusters of genes with related functions (Figure 2A).

3.2 Comparative phylogenomics analyses of mEp_{imm1} phages

The genome of the archetype mEp021 phage (accession number MH706966.1), which was deposited in GenBank in 2018, was used to retrieve homologous phages and prophages in the GenBank (NCBI) database using BLASTn, with the parameters described in the Methods section. This allowed the identification of 42 additional viral complete genomes (Table 1), which displayed homogeneous GC content (average 46 ± 0.003%), a conserved genome architecture and high similarity rates in protein sequence (Supplementary Figure S3). Hence, these viral genomes were included in subsequent analyses as members of the mEp_{imm1} group. Whole-genome phylogenetic analysis indicated that all mEp_{imm1} phages and prophages (*n* = 48) constitute part of the same evolutionary branch, that bifurcates into two main subgroups (Figure 2B); several lambdoid and non-lambdoid phages were included as an external *Caudoviricetes* reference group. The segregation pattern of the six mEp_{imm1} phages sequenced in this work indicated evolutionary closeness among them (Figure 2B). Intergenomic distance calculations using VIRIDIC, confirmed this observation, and indicated the presence of 9 genera and 45 species

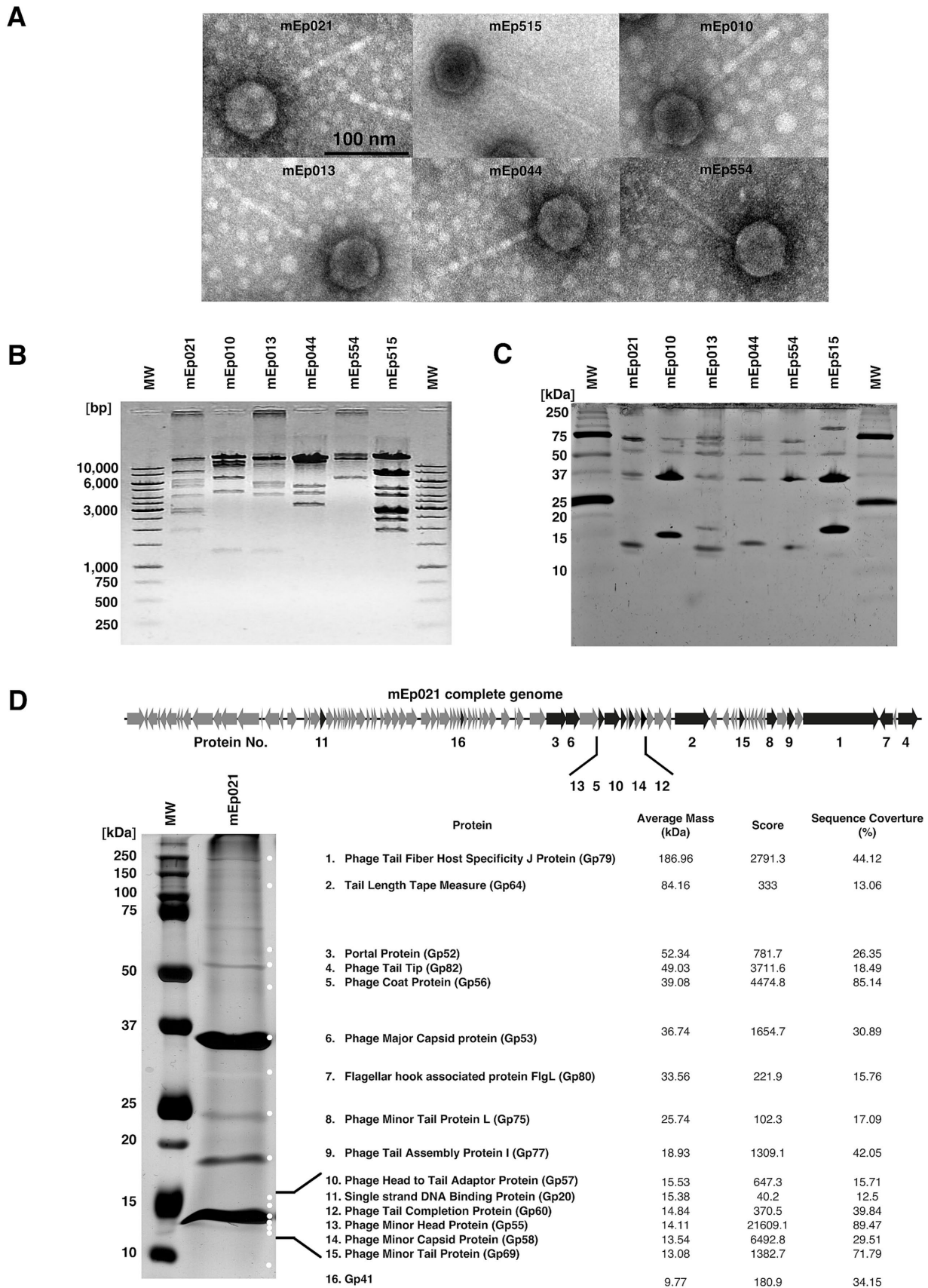
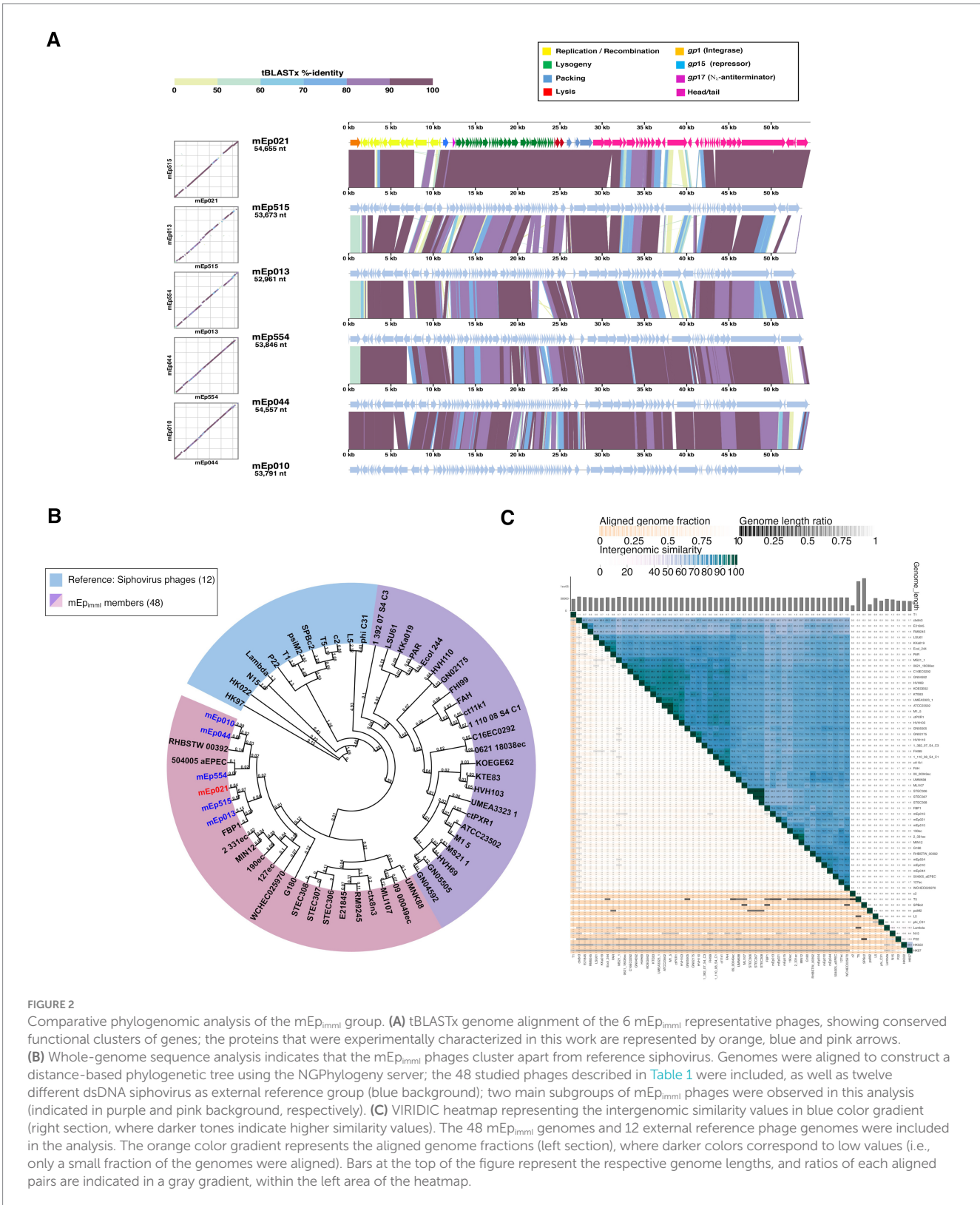


FIGURE 1 Structural characterization of six representative phages of the mEp_{mm1} group. (A) Transmission electron microscopy (150x magnification) revealed ~60 nm phage particle sizes. (B) RFLP analysis using *NdeI* indicated genomic differences among the studied phages. (C) Structural proteins of variable

(Continued)

FIGURE 1 (Continued)
size were observed among the 6 phages, in total virion proteins 16% SDS-PAGE analysis. (D) Sixteen discrete protein bands were identified from the 10% SDS-PAGE analysis of the mEp021 structural proteins by HPLC/MS/MS. These proteins were subsequently mapped in the viral genome, which is illustrated at the top of the figure.



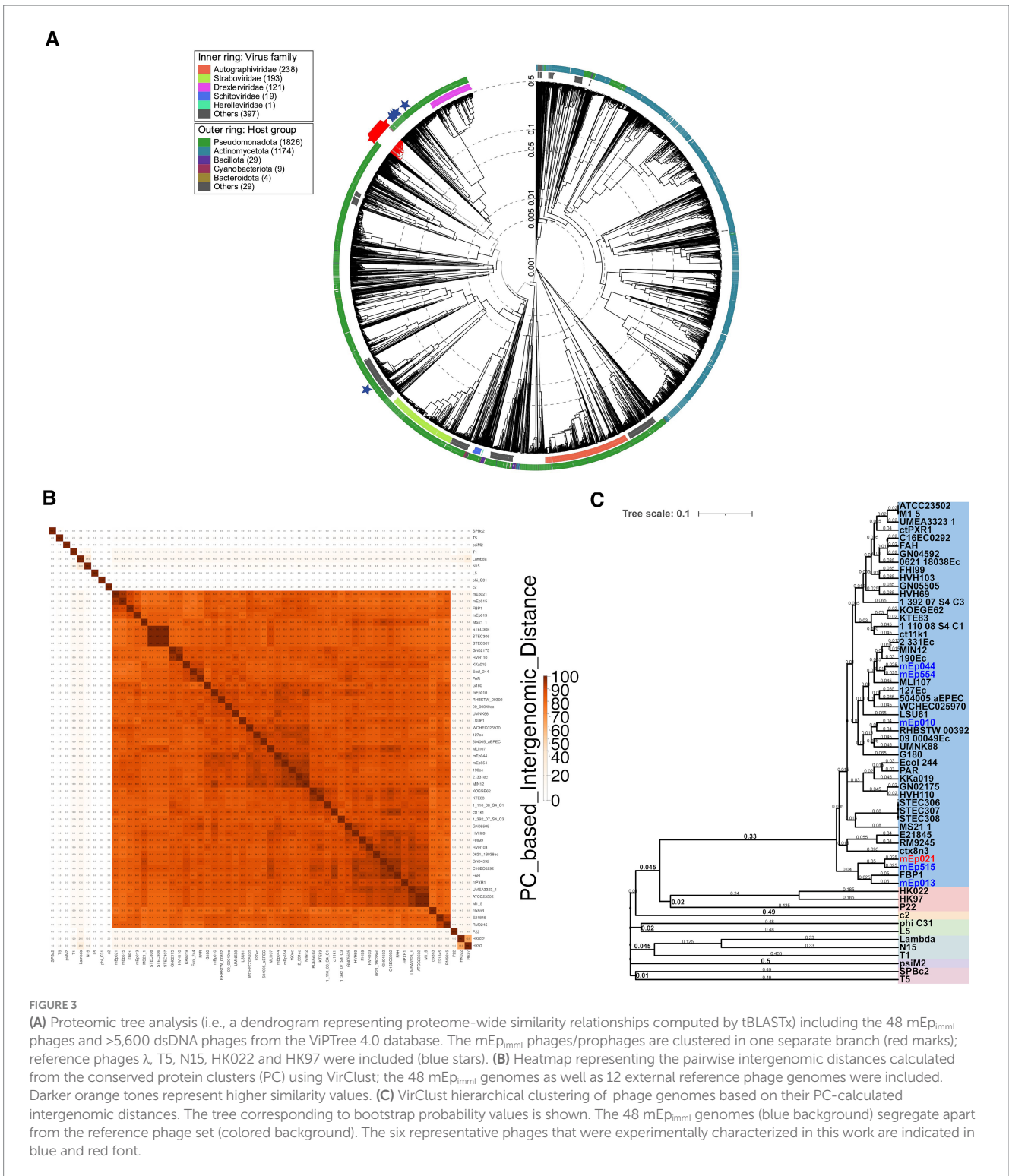


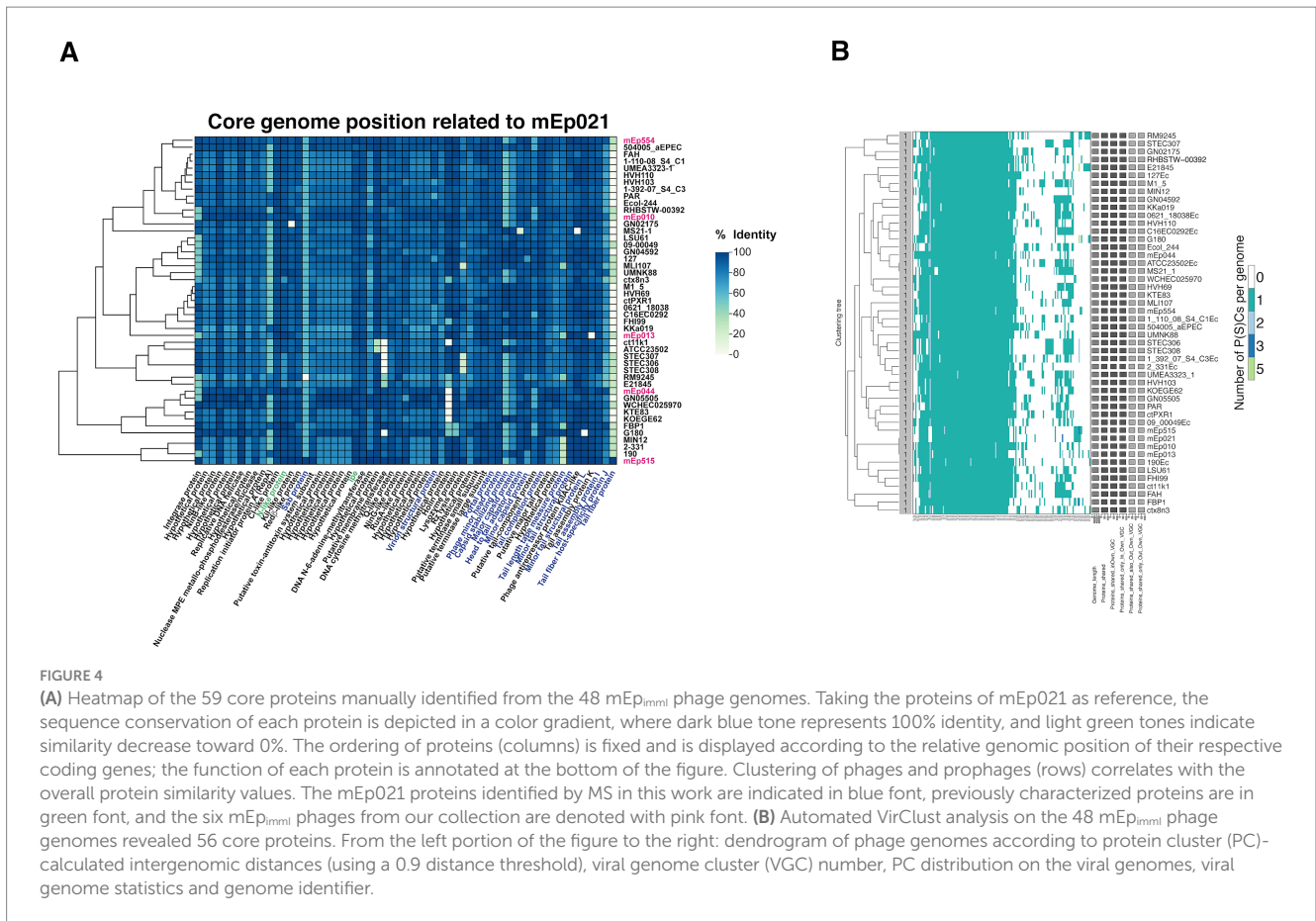
FIGURE 3

(A) Proteomic tree analysis (i.e., a dendrogram representing proteome-wide similarity relationships computed by tBLASTx) including the 48 mEp_{imm1} phages and >5,600 dsDNA phages from the ViPTree 4.0 database. The mEp_{imm1} phages/prophages are clustered in one separate branch (red marks); reference phages λ, T5, N15, HK022 and HK97 were included (blue stars). (B) Heatmap representing the pairwise intergenomic distances calculated from the conserved protein clusters (PC) using VirClust; the 48 mEp_{imm1} genomes as well as 12 external reference phage genomes were included. Darker orange tones represent higher similarity values. (C) VirClust hierarchical clustering of phage genomes based on their PC-calculated intergenomic distances. The tree corresponding to bootstrap probability values is shown. The 48 mEp_{imm1} genomes (blue background) segregate apart from the reference phage set (colored background). The six representative phages that were experimentally characterized in this work are indicated in blue and red font.

among the 48 genomes analyzed (Figure 2C), according to current criteria for phage taxonomy (Turner et al., 2024).

Further, tBLASTx analysis integrating the available proteomes from >5,600 dsDNA reference phages deposited in the ViPTree server, revealed that the mEp_{imm1} group constitutes a compact and independent branch, and further separation from the reference group was observed when the 42 homologous phages and prophages were incorporated into the analyses (Figure 3A and

Supplementary Figures S4A,B). These results demonstrate that the mEp_{imm1} phages constitute a novel group, distinct to previously described bacteriophage clades. Intergenomic distance calculations based on conserved protein clusters using VirClust (Supplementary Figures S4C,D), confirmed that the 48 mEp_{imm1} genomes belong to a separate branch, when compared against 62 phage genomes corresponding to the 12 external reference phages and 50 neighboring phage genomes observed in the ViPTree analysis



(Supplementary Tables S2, S3 and Supplementary Figure S4A); this analysis indicated that the mEp_{immi} phages correspond to a unique viral genome cluster (VGC 1, Supplementary Table S5), while reference phages were coherently clustered, such as HK022 and HK97 (VGC 7, subfamily Hendrixvirinae), P22 (VGC 12, genus Lederbergvirus). At this distance threshold, phage λ (genus Lambdavirus) is clustered together with phage N15 (genus Ravivirus), however, at 0.65 distance threshold these phages are properly segregated to distinct VGC, while the VGCs corresponding to the aforementioned reference phages and the mEp_{immi} group maintain their composition (data not shown).

The 48 mEp_{immi} genomes showed high synteny, containing from 106 to 128 genes according to our gene calling procedure (see Methods). Taking into account the relative position of each gene in the 48 phage genomes, the identity of all the predicted proteins was manually assessed in order to determine the core, flexible and unique genes in this group. We used the mEp021 genome as a reference, as the virion structural proteins were consistently mapped by MS analysis. This analysis revealed that 47.16% of the total genes (2813/5965) corresponded to core genes, representing 59 distinct gene products that were shared among all the analyzed genomes (Figure 4A). The gene products that were coded in several genomes but not all, were identified as flexible (49.14%; 2,931/5,965 genes), and were mainly located in the variable regions of these phages. Finally, 3.7% (221/5,965) corresponded to unique gene products. Automated analysis using the VirClust tool yielded from 87 to 103 called genes per genome, and indicated that the core genome included 57.4% of the

total genes (2649/4619), corresponding to 56 core proteins (Figure 4B); 42.1% (1946/4619) were accessory genes and 0.5% (24/4619) were unique gene products. It should be noted that the core gene sets from the manual and automated analyses shared 50 common gene products (84.7 and 89.2% coincident genes, respectively), which were further considered as the confirmed core genome. The genes that were not common to both analyses mostly corresponded to hypothetical proteins.

3.3 The mEp_{immi} phages use a C_I_λ-like repressor protein

Lysogen immunity is regulated by the phage repressor protein, and constitutes one of the primary characteristics that allowed the identification of mEp_{immi} phage group (Kameyama et al., 1999; Stayrook et al., 2008; Thomason et al., 2019; Sedhom and Solomon, 2023). Alignment of the predicted repressor proteins of these phages indicated a high sequence similarity among them (95.95%), displaying conserved Cro/CI-type HTH and S24 LexA-like domains, according to InterProScan analysis (Figure 5A); these domains are required for DNA binding and protein dimerization, respectively. In order to demonstrate its functionality, the coding sequence of a representative repressor protein, Rep₀₂₁ of phage mEp021, was inserted in a low copy-number expression vector; as expected, expression of Rep₀₂₁ in a transformed W3110 strain blocked the infection of other mEp_{immi} phages (Figure 5B).

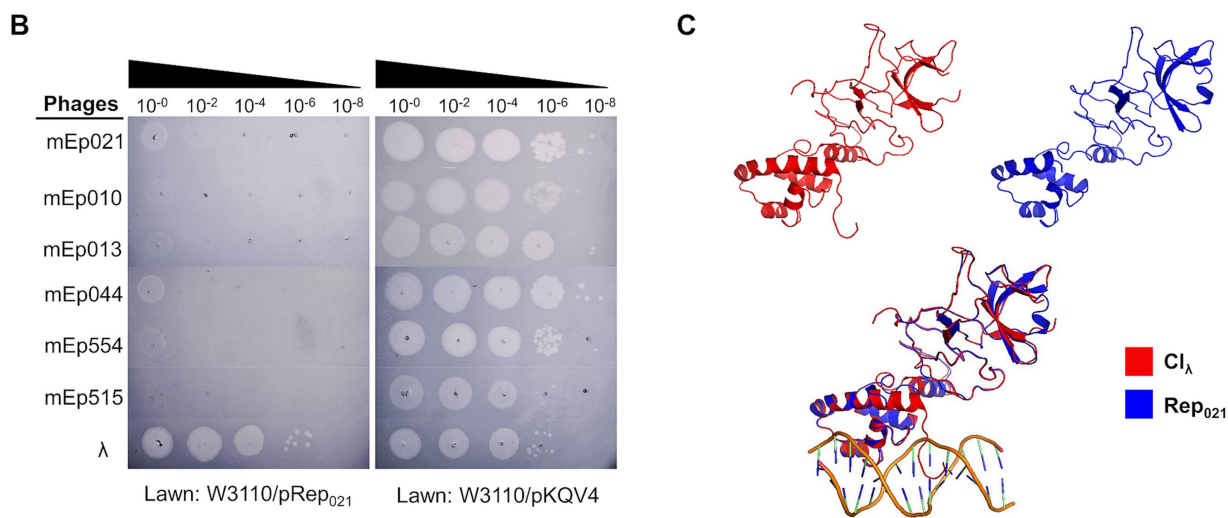


FIGURE 5
The mEp021 repressor protein is unique and is highly conserved in the mEp_{imm1} group. **(A)** Alignment of the 48 mEp phages/prophages repressor proteins, showing high conservation among all the group. **(B)** Phage exclusion assays were performed in the transformed W3110 host strain expressing the repressor Rep₀₂₁ from phage mEp021. Infection of phages mEp021, mEp010, mEp013, mEp044, mEp554, and mEp515 was inhibited; conversely, λ phage proliferated as expected. **(C)** The Rep₀₂₁ repressor protein (blue) was modeled using the Phyre² server, and the resulting structure was superposed with the CI_λ (red) crystallographic model (accession number 3BDN); the functional domains retain high structural similarity.

Moreover, superposition of the Rep₀₂₁ computational protein model and the CI_λ repressor crystallographic model (Figure 5C and Supplementary Figure S6B) revealed a remarkable structural similarity, especially in the regions corresponding to both functional domains of these proteins, despite that the amino acid sequences of the Cro/CI-type HTH domain at the amino-terminal region are less conserved.

3.4 mEp_{imm1} phages and prophages have four types of integrases

Once the phage has ejected its DNA into the host cell, it may undergo lysogeny. During this process, the phage DNA may integrate into the host genome, using an integrase protein which is generally phage-encoded and is located nearby to the insertion

sequence motifs. These phage insertion sequence motifs are similar to the host insertion sites and may correspond to distinct loci in the bacterial genome; these motifs are specifically recognized by the integrase proteins. Hence, we mapped the insertion site in the host genome and identified the insertion sequence motifs of each phage, as well as the integrase gene they, respectively, harbored. For this, we analyzed the host sequences located upstream and downstream to the 38 homologous mEp_{imm1} prophage genomes that were found integrated in *E. coli* genomes from GenBank, and the W3110 (mEp021) lysogen was included (Supplementary Figure S7A). Different att sites and motifs were identified. For most phages, the integration site was located within the 5'-terminal region of the ydaM gene, while for other phages the integration site was found at the 3'-terminal region of the abgT gene, or at the initiation region of the mppA gene (Figure 6A and Table 2).

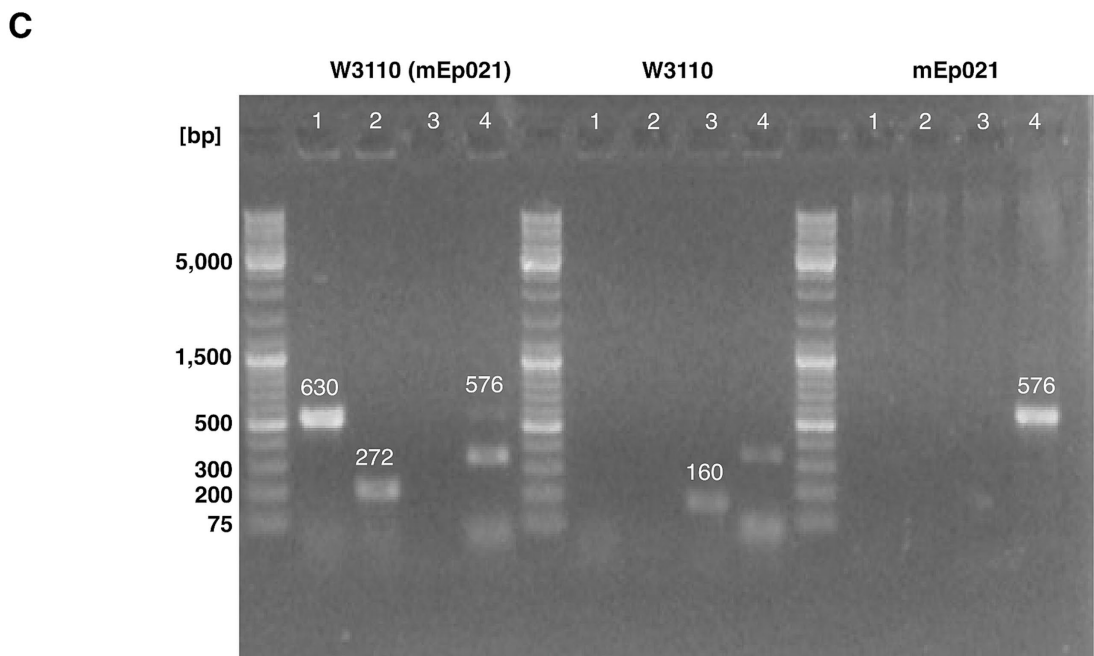
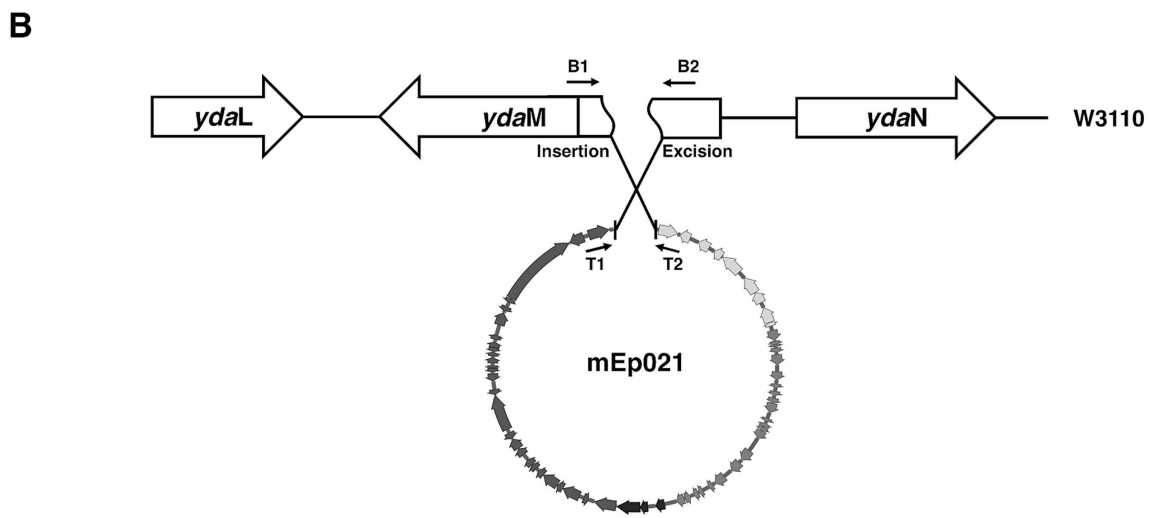
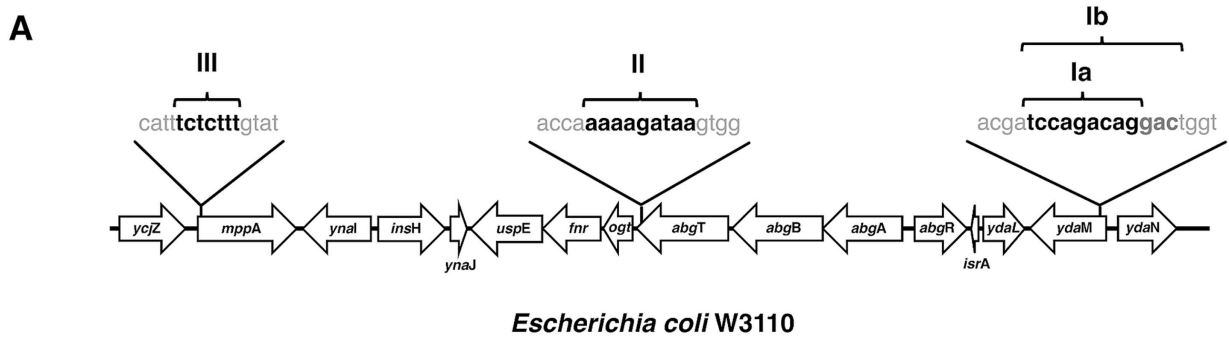


FIGURE 6 Insertion sites of mEp021 and mEpmmi prophages in the host genome. **(A)** *E. coli* genes are indicated in white arrows outlined in black, and the genomic positions of the prophage insertion sites are indicated with black lines; integration sequence motifs are shown. The associated integrase protein type is indicated above each integration site, respectively. **(B)** Schematic representation of the experimental strategy to validate the insertion site of the mEp021 bacteriophage in the W3110 strain. PCR primers (black arrows) were designed to amplify from the att site at the left and right ends of the mEp021 DNA (T1 and T2), and from the predicted insertion sites within the *ydaM* host gene (B1 and B2). W3110 genes are represented by black arrows; the integration site disrupts the *ydaM* gene (segmented arrow). The mEp021 genome is shown in circular form at the bottom. Diverse primer

(Continued)

FIGURE 6 (Continued)

combinations were tested, and the predicted product sizes are indicated. (C) PCR validations of mEp021 integration site in its W3110 lysogen. The W3110 (mEp021) lysogen, W3110 and mEp021 DNA templates were evaluated, as indicated in the top of the gel images. The numbers 1, 2, 3 and 4 indicated on the top of the gels correspond to the primer combinations denoted in (B). For W3110 (mEp021) template: lane 1, amplification of the right attachment site yielded the expected 630 bp amplicon (primers T1/B2); lane 2, amplification of the left attachment site (272 bp amplicon, primers T2/B1); lanes 3, no amplification; lane 4, a faint 576 bp band corresponding to episomal mEp021 was detected, as well as one minor unspecific amplicon. For W3110 template: only the 160 pb product corresponding to the *ydaM* region amplification was observed (lane 3, primers B1/B2). For mEp021 template: only the 576 bp amplicon corresponding to the integration region of the mEp021 genome was observed (lane 4, primers T1/T2); the mEp021 DNA was obtained from CsCl-gradient virion purification to avoid bacterial DNA contamination.

TABLE 2 Distribution of phages and prophages according to the identified integrase groups.

mEp _{imm1} integrases group				
Ia	Ib	II	III	IV
*ct11K1	*0621-18038	*mEp010	LSU61	ctx8n3
*ctPXR1	*mEp515	*mEp044		
1-110-08_S4_C1	*mEp554	*mEp013		
ATCC23502	mEp021	09-0049		
Ecol_244	1-392-07_S4_C3	UMNK88		
FHI99	504005_aEPEC	RM9245		
HVH110	MIN12	RHBSTW-00392		
HVH69	2-331	GNO5505		
HVH103	FBP1	GNO4592		
KOEGE62	G180	E21845		
KTE83	STEC306	190		
MS21-1	STEC307	127		
C16EC0292	STEC308	▷ MLI107		
FAH	WCHEC025970			
GN02175				
KKa019				
M1/5				
PAR				
UMEA3323-1				

*The host insertion site was inferred from the viral integrase homology analysis.

▷ Insertion site with ~ 7000 bp deletion.

Phages are in bold and prophages in black font.

The mEp021 archetype phage for the mEp_{imm1} group is indicated in red font.

We identified genes coding for integrase proteins in all the mEp_{imm1} phages. All of these corresponded to the tyrosine-recombinase family (Groth and Calos, 2004) and amino acid sequence alignments allowed the recognition of four main types or clusters of these enzymes: types I (a and b), II, III and IV (Supplementary Figure S7B). We hypothesized that the use of different attachment sites would be related to the different integrase sequence-types. The results were consistent with such hypothesis, since the phages displaying the same attachment sites harbored the same integrase sequence-types: phages integrating at the *ydaM* locus code for integrases types Ia and Ib; those integrating at *abgT* code for type II integrases, and the ones integrating at *mppA* bear type III integrases (Table 2). Further, a unique integrase was found in the genome of phage ctx8n3

(integrase IV), a phage identified from metagenome sequences (Table 1); however, its host integration site could not be determined because the sequence upstream to the phage integrase (i.e., phage genome 5' terminus) did not match the W3110 reference strain genome, and no sequence downstream to the phage genome was available.

As revealed by this analysis, phages mEp010, mEp013, and mEp044 harbor type II integrases, while phages mEp021, mEp515 and mEp554 code for type Ia integrases. Hence, the host integration site of archetype mEp021 phage was experimentally verified by colony PCR assays in a fresh W3110 lysogen, using primers flanking the *attR* (T1 and B2) and *attL* (T2 y B1) sites (Figure 6B), and the results confirmed the predicted attachment sites (Figure 6C).

3.5 The mEp_{imm1} phages contain an N_λ-like early antitermination system

It has been shown that the N_λ-like protein Gp17 is responsible for the early antitermination process observed in phage mEp021 (Valencia-Toxqui et al., 2023). Its coding gene *gp17* was present in all the studied mEp_{imm1} genomes, and the arginine residues at positions 17, 19, 20 and 24 were fully conserved, indicating that this domain is fundamental to protein function (Supplementary Figure S8); arginine in position 21 was conserved to a lesser extent (45 out of 48 phages). Gp17 acts upon three *Nut* sites (*nutR1*, *nutR2*, and *nutL*), which are located in a similar genomic position among all the mEp_{imm1} phages. The *nutR2* and *nutL* sites show less sequence variability than the *nutR1* (Figure 7), however, the adenine-rich region of the boxB recognition loop was well conserved in all three *Nut* sites, suggesting that the early antitermination process is similar among all the mEp_{imm1} phage group.

3.6 Bacterial outer membrane proteins involved in the infection of mEp_{imm1} phages

One of the fundamental steps for phage multiplication is the infection of their hosts. For this, different OMR proteins of the bacterial cell may be recognized by the phage particles, in order to adsorb and eject their genome into the host. Hence, we determined the host OMR proteins used by our six mEp_{imm1} phages and analyzed the phage structural protein that is involved in such recognition. For this, we performed a series of infection assays using several outer membrane protein null-mutant strains of the Keio collection (Datsenko and Wanner, 2000; Baba et al.,

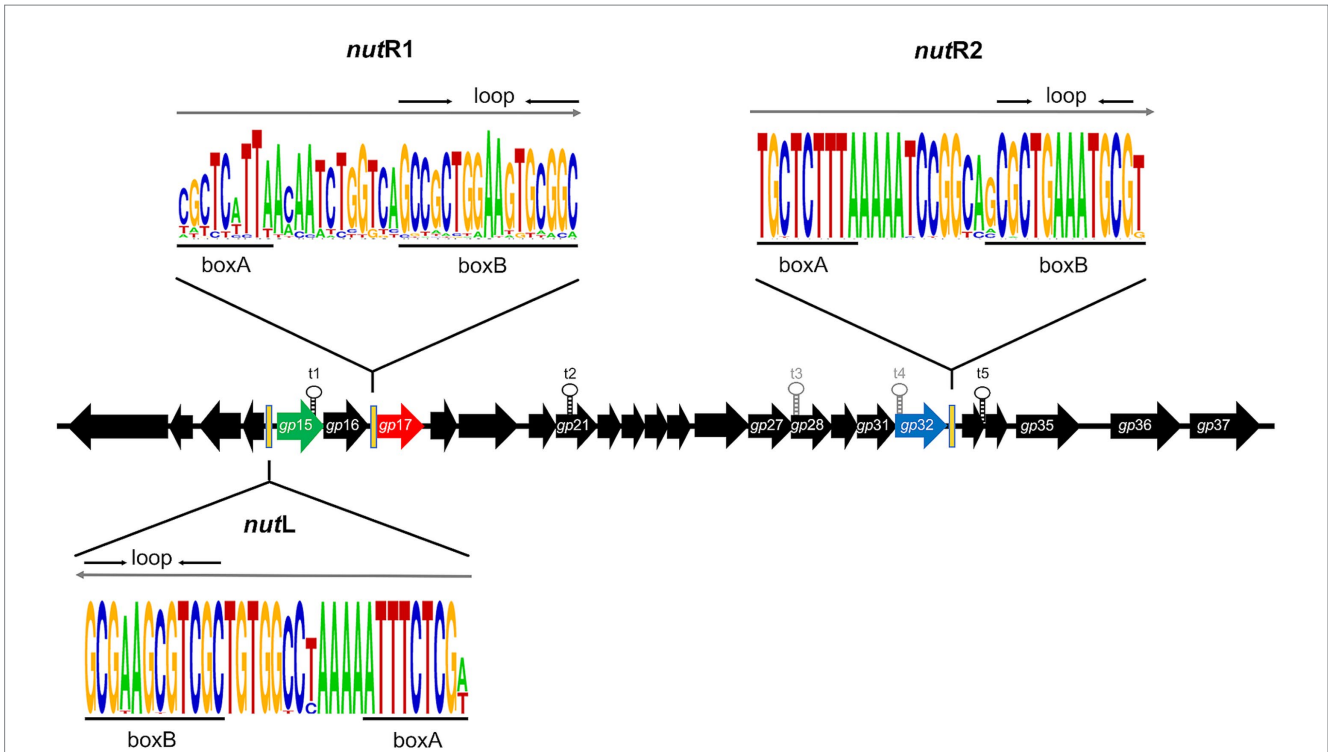


FIGURE 7
 The antitermination system is highly conserved among the mEp phage group. Consensus sequences of the three *nut* sites (containing the *boxA* and *boxB* motifs, respectively) among the mEp_{mmi} group is represented in the WebLogo format (<https://weblogo.berkeley.edu/logo.cgi>) (Crooks et al., 2004). The direction of transcription is indicated with gray arrows. The positions of these *nut* sites in the mEp021 genome are indicated by black bars; the *nutL* site is located upstream of the *gp15* gene (coding the Rep₀₂₁ repressor, represented as green arrow); the *nutR1* site is upstream of the *gp17* gene (*N_s*-type antiterminator, represented by a red arrow), and the *nutR2* site is downstream of the *gp32* gene (a putative N-cytosine-methyltransferase, represented by a blue arrow). The predicted (*t3* and *t4*) and experimentally demonstrated transcriptional terminators (*t1*, *t2*, and *t5*) are indicated in gray and black font, respectively.

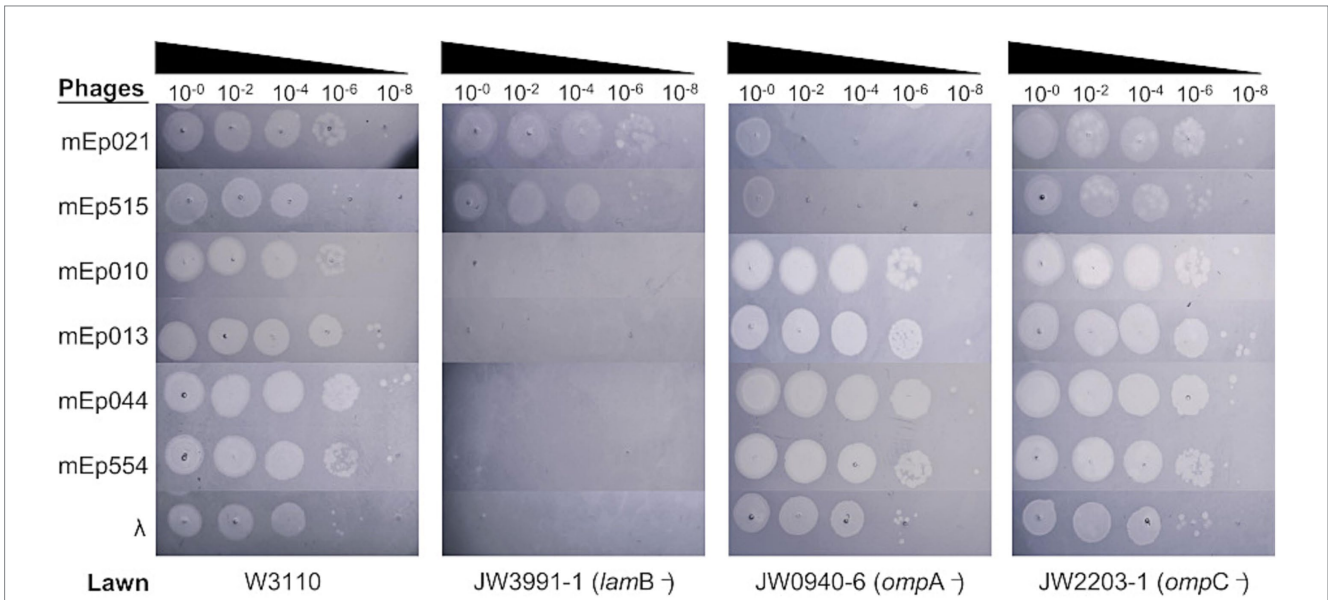


FIGURE 8
 Identification of the main host (OMR) used by the 6 representative mEp_{mmi} phages. Infection assays were performed on wild-type strain W3110 and three different Keio mutants (*ompA*⁻, *lamB*⁻, and *ompC*⁻), as labeled. Serial dilutions of bacteriophages were spotted onto bacterial lawns, as indicated. According to the observed lysis spots, the OMR *OmpA* is required for infection of phages mEp021 and mEp515, while the OMR *LamB* is required for mEp010, mEp013, mEp044, and mEp554.

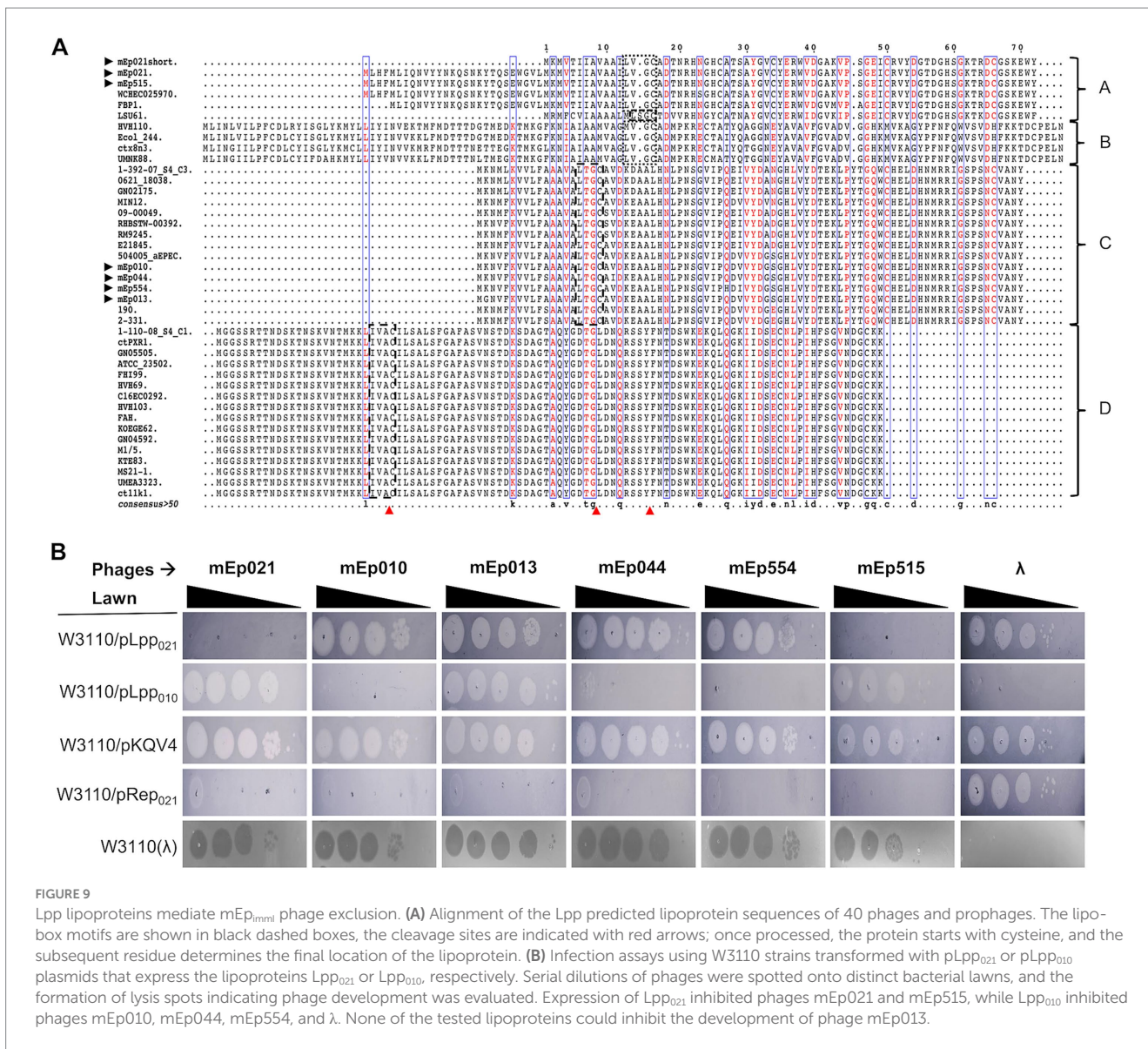


FIGURE 9
Lpp lipoproteins mediate mEp_{inmm1} phage exclusion. **(A)** Alignment of the Lpp predicted lipoprotein sequences of 40 phages and prophages. The lipobox motifs are shown in black dashed boxes, the cleavage sites are indicated with red arrows; once processed, the protein starts with cysteine, and the subsequent residue determines the final location of the lipoprotein. **(B)** Infection assays using W3110 strains transformed with pLpp₀₂₁ or pLpp₀₁₀ plasmids that express the lipoproteins Lpp₀₂₁ or Lpp₀₁₀, respectively. Serial dilutions of phages were spotted onto distinct bacterial lawns, and the formation of lysis spots indicating phage development was evaluated. Expression of Lpp₀₂₁ inhibited phages mEp021 and mEp515, while Lpp₀₁₀ inhibited phages mEp010, mEp044, mEp554, and λ. None of the tested lipoproteins could inhibit the development of phage mEp013.

2006), which are knockouts for different non-essential genes of *E. coli*. As observed in Figure 8, in the *lamB* knockout strain the infection of phages mEp010, mEp013, mEp044 and mEp554 was completely inhibited, while the *ompA* knockout strain displayed a strong inhibition of mEp021 and mEp515 infections (barely yielding ~10² PFU/mL).

The recognition and binding to the bacterial OMR proteins is mediated by phage J proteins. A multiple alignment of the predicted J proteins of the mEp_{inmm1} phages revealed that, despite all these proteins display a different size, a notable sequence conservation is observed, mainly in the amino-terminal region (Supplementary Figure S10); this region also shares structural similarity with the J protein from phage λ (data not shown). The carboxy-terminal region shows more sequence variation, and contains repeats of different length (throughout positions 950–1,475), a conserved stretch (position 1,476–1,690), and a less conserved terminus. The latter is involved in physical contact with the bacterial receptor, as shown in phage λ studies (Ge and Wang, 2024; Wang et al., 2024).

3.7 The Lpp exclusion proteins of the mEp_{inmm1} phages

The superinfection-exclusion mechanism has been described in several lysogens and it can be mediated by lipoproteins, among other types of proteins. We searched the mEp_{inmm1} genomes for lipoprotein genes through the identification of the lipobox signature motif ([LVI][ASTVI][GAS][C]), which is recognized by the lipoprotein signal peptidase (Lsp) that cleaves the signal peptide and promotes further allocation of the lipoprotein on the bacterial membrane (LoVullo et al., 2015). Our results indicated that 40 out of the 48 mEp_{inmm1} phages contain at least one gene that putatively codes for a lipoprotein. A multiple alignment of the predicted lipoproteins allowed the identification of four sequence types (Figure 9A). Two of these types were represented in the phages sequenced in this work: type-A lipoprotein was found in mEp021 and mEp515, and the respective amino acid sequences were identical; we selected the coding gene from mEp021, as this is the mEp_{inmm1} archetype phage that has been best described so far. Type-C lipoproteins were found in mEp010,

mEp013, mEp044 and mEp554, and the respective amino acid sequences were not identical: the mEp013 lipoprotein harbored a K > G substitution at position 2, the mEp044 an A > S substitution at position 11, the mEp554 lipoprotein displayed a Q > H substitution at position 38 and a conservative substitution V > I at position 40 (see Figure 9A). The consensus sequence of this alignment corresponded to that of the mEp010 lipoprotein, which contained the most conserved amino acids in each position in relation to the other lipoprotein sequences of this type. In addition, the mEp010 phage was the most distant to the archetype phage mEp021 in the ViPTree analysis (Supplementary Figure S4A), therefore we selected its lipoprotein coding gene as representative for the type-C lipoproteins. We cloned the aforementioned genes in an expression vector in order to assess their function through phage infection assays (Arguijo-Hernández et al., 2018), and the results showed that the mEp021 type-A lipoprotein (Lpp₀₂₁) excluded phages mEp021 and mEp515, while the mEp010 type-C lipoprotein (Lpp₀₁₀) was able to exclude phages mEp010, mEp044, mEp554 and λ (Figure 9B).

4 Discussion

Phages of the order *Caudoviricetes* (tailed virus) have been reported as the most frequent communities in both healthy individuals and inflammatory bowel syndrome patients; consistently, the morphotypes myovirus, podovirus and siphovirus were recognized as the most predominant entities in gut (Ansari et al., 2020; Coughlan et al., 2021). As such, the mEp_{immi} phages were found to be as prevalent as the lambdoid phages in clinical samples of human feces (Kameyama et al., 1999), and we also identified complete and incomplete homologous phage genomes in two metagenome projects (NCBI Bioprojects 285453 and 862966, registered in 2015 and 2022, respectively) (Tisza and Buck, 2021). We started the characterization of these phages using mEp021 as a representative of the group, due to its ability to induce a hemolytic phenotype in its *E. coli* MC4100 lysogen, which is mediated by the phage-encoded Ipe protein (Kameyama et al., 2001; Martínez-Peñañiel et al., 2012). In addition, mEp021 has an antitermination mechanism similar to that of λ , but with the use of three different *nut* sites instead of the typical two (Valencia-Toxqui et al., 2023).

Grouping of phages based on similarities in their genome organization and gene regulation can provide insights into their biological behavior (Valencia-Toxqui and Ramsey, 2024). In this work, we extended the characterization of the mEp_{immi} phages, including mEp021, mEp010, mEp013, mEp044, mEp515, and mEp554, to provide structural and functional data supporting the establishment of a new group of coliphages, following current taxonomical criteria and traditional biological characterization approaches. TEM analysis (Figure 1A) confirmed a common siphoviral morphotype in the analyzed coliphages, however, the differential host range reported (Kameyama et al., 1999) and the electrophoretic patterns observed both in the DNA restriction and structural protein SDS-PAGE analyses presented herein, indicated that each of these phages is a particular entity (Figures 1B,C). Whole genome sequencing confirmed this, and despite the fact that these phages display diverse characteristics (Kameyama et al., 1999, 2001), their overall sequence similarity is high and they share a common genome architecture (Figure 2A).

MS analysis of the mEp021 structural proteins allowed the mapping of the structural protein gene cluster in the phage genomes (Figure 1D, black arrows), including those coding for the phage tail fiber protein J, tail length tape measure protein, portal protein, tail tip protein, coat protein, among others. Interestingly, the coding genes of two structural proteins were distant to the main cluster, both being located upstream to the lysis region. One of such genes codes for a putative single-stranded DNA binding protein (Ssb); to our knowledge, this type of proteins does not form part of the viral structure but are rather involved in the stimulation of viral DNA replication, as it was previously shown for a Ssb protein from phage ϕ 29 (Soengas et al., 1994). We therefore suggest that this protein may participate in DNA protection or organization. The other mislocated gene codes for a hypothetical virion structural protein of unknown function, which is conserved in all the analyzed genomes with a > 60% similarity. Distant homology searches using PLMSearch (Liu et al., 2024) indicated that this protein is similar to a translation repressor protein from *E. coli* phage RB69, which is an RNA-binding protein, and to the uncharacterized protein YorC from the SPbeta prophage (similarity scores = 0.9963, in both cases); a helix-turn-helix DNA binding motif was recognized in the first hit (Uniprot accession Q01751), suggesting that this phage protein may physically interact with DNA. As this protein was detected in the MS analysis as part of the viral structural proteins (which were obtained from phages purified through a CsCl-gradient), we suggest it is possibly involved in the conformation of viral particles.

A BLASTn search in the GenBank database using mEp021 as reference, allowed the identification of 42 additional homologous mEp_{immi} prophages and phages (Table 1), from *E. coli* genomes ($n = 38$) and metagenome sequences ($n = 4$). This set of viral genomes substantially differed from reference phage genomes of the order *Caudoviricetes*, as demonstrated by subsequent phylogenetic analysis, which indicated that these phages warrant a formal taxonomic classification. Several analysis tools such as VIRIDIC, ViPTree, and VirClust have shown to be consistent with the criteria set by the Bacterial Virus Subcommittee of the International Committee on Taxonomy of Viruses (ICTV) to assign taxonomic ranks at different levels (Turner et al., 2021, 2024). Despite their widely diverse geographical origins, an elementary alignment analysis using whole genome sequences, showed that the mEp_{immi} phages constitute a well delineated group that clusters apart from reference siphovirus (Figure 2B); this was confirmed by intergenomic distance analyses using VIRIDIC (Figure 2C). These results allowed the identification of 8 genera and 45 species among the 48 genomes examined, according to the demarcation criteria of 95% sequence similarity for species and 70% for genus in whole-genome analyses (Turner et al., 2024). Further, whole-proteome comparative analysis showed that the 48 mEp_{immi} phages form a cohesive and monophyletic group, sharing a significant proportion of orthologous genes (Figure 3A and Supplementary Figure S4A,B), when compared to >5,600 dsDNA phage proteomes; this is consistent with a family-level classification (Turner et al., 2021, 2024). These results were further supported by the analysis using VirClust, which calculates intergenomic distances based on the identified protein clusters, confirming the segregation of the mEp_{immi} group as an independent branch, in both analyses using the dataset of 12 reference phages (Figures 3B,C) and a larger dataset including the neighboring viral genomes observed in the ViPTree analysis ($n = 110$) (Supplementary Figures S4B,C). VirClust analysis

confirmed that the mEp_{imm1} group constituted a single viral genome cluster (VGC) (Supplementary Table S5). Additional analyses will be required to determine whether viral sub-families are present among this phage group.

All the mEp_{imm1} phages displayed high sequence similarity and genome synteny, as revealed by the tBLASTx analysis (Supplementary Figure S3). This indicates that these phages may share numerous biological similarities and, considering their common ecological niche, it is possible that these phages course an early phase of evolutionary diversification, as previously suggested (Kameyama et al., 1999). Their genes are clustered according to related functions, and this includes the genes involved in replication/recombination, the lysis-lysogeny regulation, packaging, lysis and viral morphology gene clusters (Figure 2A). However, sequence and gene content variation was also observed, involving genes such as integrases, those involved in lysis/lysogeny regulation, the J protein and the tail tip protein genes, among others. Considering this, we analyzed the whole gene content and its conservation among all these genomes. A manual pangenome analysis indicated that 47.16% of the total number of identified genes are present in all the 48 genomes analyzed, and these correspond to 59 core gene products that are related to all the functional clusters (Figure 4A). Variable or accessory genes constituted 49.14% of the total genes and were mainly located in the regulation cluster; we suggest that this region is the main contributor to group diversification. In addition, an automated analysis using VirClust indicated the presence of 56 core genes (Figure 4B). The majority of the genes that were not coincident between these analyses (9 in the manual analysis and 6 in VirClust, respectively) corresponded to hypothetical proteins. From these, 4 of the proteins identified in the manual analysis are probably related to phage functions, including a Kil-like protein, DNA cytosine methyltransferase, phage antirepressor protein KilAC-like, and a phage tail assembly protein K. Conversely, two proteins were recognized among the non-coincident proteins from the VirClust analysis, including an AlpA-family transcriptional regulator from bacteria and phages, and a centrosome-associated protein, which is a heterologous finding but may refer to a nucleic acid binding protein as well. Taking these results into account, we confirmed the presence of at least 50 core genes in the studied mEp_{imm1} genomes, which were coincident between the manual and automated analyses.

In order to validate relevant genes from the functional clusters, we performed a series of experiments using the six sequenced mEp_{imm1} phages from our collection. A prime method in phage characterization is to evaluate the auto-exclusion mechanism of a phage from its own lysogenic strain, as well as the exclusion of phages from the same group or family (Dulbecco, 1952). The 50 original phages from the immunity I group were partly characterized using this approach (Kameyama et al., 1999), which implies that there is similarity in their main repressor proteins. Consistently, we found high similarity among the sequences of the repressor proteins of the 48 mEp_{imm1} phages in this study, showing the two highly conserved functional domains: Cro/CI-type HTH and S24 LexA-like (Figure 5A and Supplementary Figure S6), which are typical features of CI-type repressors. We demonstrated Rep₀₂₁ exclusion activity using transformed strains expressing this protein, which inhibited mEp021, mEp010, mEp013, mEp044, mEp515 and mEp554 infections (Figure 5B). Noteworthy, the use of highly concentrated phage preparations produced a growth inhibition effect, but no isolated lysis

plaques were observed. This could be result of a non-specific lysis due to lysis from without (Abedon, 2011). Notable amino acid sequence differences were observed between the CI_λ repressor and the 48 Rep predicted proteins (Supplementary Figure S5A), especially regarding the Cro/CI-type HTH domain. However, the three-dimensional structure of both functional domains is highly conserved (Figures 5C and Supplementary Figure S6B), according to the structural models of Rep₀₂₁ obtained from three different prediction tools (see Materials and Methods). Despite the high structural similarity at the domain-level, the global conformation of Rep₀₂₁ and CI_λ differs, probably due to the positioning of both domains, which is mediated by a hinge region in between. All these structural differences may account for specific biological properties. For instance, sequence variation in the HTH domains of CI_λ and Rep021 may lead to the recognition of different operator sequences; the six initial amino acid residues of CI_λ play a critical role in the recognition of the operator site, where Lys3, Lys4, and Lys5 were shown to be involved in direct physical interaction with the DNA bases (Clarke et al., 1991); a shorter N-terminal arm with two highly conserved lysine residues is observed in Rep021 (Supplementary Figures S6A,B), thus it is probably involved in a similar function. Moreover, it is possible that the self-cleavage activity of Rep₀₂₁ is different to that of CI_λ, since the C-terminal region contains the catalytic site for self-cleavage but lacks the specific amino acids of the cleavage site at the hinge region (A111-G112) (Sauer et al., 1982; Supplementary Figure S6A). This may explain previous observations in which UV radiation fails to induce mEp_{imm1} lysogens to the lytic cycle (Kameyama et al., 1999).

The mEp_{imm1} group harbors a post-transcriptional regulation system similar to the λ antitermination system, which requires host Nus factors as well as the phage N_λ-like and Q_λ-like proteins, to allow the development of these phages. In previous work, we experimentally characterized the N_λ-like protein of mEp021 phage, namely Gp17, which is responsible for the early antitermination process in this phage. Unlike N_λ, Gp17 acts upon 3 *nut* sites (*nutR1*, *nutR2* and *nutL*) (Valencia-Toxqui et al., 2023). In this work, we show that Gp17 is highly similar among the mEp_{imm1} phages, and conserves the ARM involved in the interaction with the boxB region of the *nut* site, which is determinant for its binding to RNA (Supplementary Figure S8). Consistently, the *nut* sites are also conserved in these phages (Figure 7), especially regarding the adenine-rich region in the boxB motifs. *nutR2* and *nutL* are the most similar, while *nutR1* displays higher sequence variability although it retains the boxA and boxB binding motifs. These results indicate that the mEp_{imm1} phages harbor a similar antitermination system, that differs to that of λ due to the requirement of an additional *nut* site.

Despite the observed similarities in sequence, genome organization, repressor-related immunity control and antitermination, the mEp_{imm1} phages displayed several biological differences among them. One of such differences involved the host genome integration sites, which were analyzed and compared between the 38 prophage genomes (corresponding to *E. coli* lysogen and metagenome sequences). We identified attachment sites within four distinct bacterial loci: *ydaM* (motifs Ia and Ib), *abgT* (motif II), *mppA* (motif III) and one which could not be mapped (motif IV), due to insufficient metagenome sequence data; each of these was separated by approximately 7 kb to the others. Most of the analyzed prophages were found inserted into the *ydaM* locus, which codes for

a putative diguanylate cyclase enzyme related to the expression of the biofilm-associated curli fimbriae (Weber et al., 2006; Figure 6A and Table 2). We assume that this represents the ancestral insertion site, and that subsequent variation has allowed the recognition of distinct insertion sites, which were less prevalent (Supplementary Figure S7A). For instance, the difference between sites Ia and Ib, is the presence of 3 additional base-pairs in the *att* motif of the latter (Figure 6A). The archetype phage mEp021 insertion site, corresponding to the Ib motif, was experimentally confirmed by PCR in W3110 (mEp021) lysogens (Figures 6B,C), but further experimental approaches will be required to determine whether the 3 additional base-pairs are essential for phage integration. Noteworthy, these integration sites differed from those reported for phage λ , involving a 16 kb stretch between the *gal* and *bio* loci (Landy and Ross, 1977). We refined the λ insertion sites through a similar lysogen analysis ($n = 15$; data not shown), revealing that this phage integrates in between the *ybhC* and *ybhA* loci, unlike the mEp_{imm1} phages. Moreover, we identified four types of integrases in the mEp_{imm1} genomes, according to the sequence alignments of predicted proteins (Supplementary Figure S7B). Interestingly, each of the four sequence-types of integrases specifically corresponded to an integration site observed in the lysogens, therefore the observed amino acid sequence variation could explain their specificities. Amino acid sequence analysis of the integrases revealed that the N-terminal region containing the DNA-binding domain was conserved, while the central and C-terminal regions harboring the catalytic domain displayed the most variability, which may account for the differential recognition and insertion to the diverse target regions. It is noteworthy that the insertion region recognized by prophage MLI107 displays a > 7 kb deletion, despite its integrase displays no relevant amino acid sequence differences when compared to group II integrases; therefore, this may be the result of a bacterial recombination event in the target region.

This novel group of phages also require different OMR proteins for infection (Figure 8). Similar to λ , the mEp010, mEp013, mEp044, and mEp554 phages use the bacterial LamB maltoporin, despite being non-lambdaoid phages. Interestingly, the mEp021 and mEp515 phages require the bacterial OmpA OMR protein; few phages are known to use this receptor protein, mainly belonging to the T-even group (Schwarz et al., 1983). Noteworthy, the absence of OmpA does not completely inhibit the infection of these phages, suggesting that this protein acts as the main OMR, but alternative receptor proteins may also be used. These findings prove the diversity within the mEp_{imm1} phages and reveal discrete biological similarities with heterologous phages. Sequence analyses of the predicted phage J proteins showed that the C-terminal region, which harbors the Receptor Binding Domain (RBD) (Wang et al., 2024), is variable (Supplementary Figure S10). Phylogenetic clusters based on the J protein alignments were consistent with the OMR groups observed and the presence of additional clusters suggests the possibility that even other receptor proteins, different to the ones characterized here, may be used by the remaining phages of the mEp_{imm1} group (Supplementary Figure S11). This is analogous to what has been observed in the lambdaoid group of the mEp collection, in which the use of FhuA rather than the LamB receptor is predominant, and OmpC is secondarily used (Hernández-Sánchez et al., 2008).

Finally, another differential characteristic among the mEp_{imm1} phages is the superinfection-exclusion system, which can be mediated by lipoproteins that interfere with DNA ejection into the cytoplasm or block the host receptor protein. This system was present in 83% of the studied genomes and at least three distinct lipoprotein types were identified, according to amino acid sequence alignments (Figure 9A). These lipoprotein genes were located within the structural protein gene cluster of these phages and prophages. We experimentally tested the Lpp₀₂₁ lipoprotein, corresponding to phages mEp021 and mEp515, belonging to the type-A lipoprotein present in phages which also use the OmpA OMR protein for infection; the amino acid sequences of these lipoproteins were identical. Consistently, expression of Lpp₀₂₁ excluded phages mEp021 and mEp515. Likewise, expression of the type-C ancestral lipoprotein Lpp₀₁₀ excluded almost all of the corresponding group phages which use the LamB receptor: mEp010, mEp044, mEp554 and, unexpectedly, phage λ . Despite the latter uses the same OMR, to our current knowledge this is the first instance in which expression of a non-lambdaoid lipoprotein promotes phage λ exclusion, besides that there are no lipoproteins from lambdaoid phages reported to exclude λ (Figure 9B). Remarkably, Lpp₀₁₀ expression did not exclude phage mEp013 (also LamB-dependent), despite the amino acid sequence between the Lpp from mEp010 and the one from mEp013 differs in only one residue at position 2 (Figure 9A); this region belongs to the hydrophobic peptide leader that is processed by Lsp and is not expected to be involved in the exclusion mechanism. Thus, it is possible that the observed evasion of lipoprotein exclusion is mediated by the variation in the mEp013 protein that recognizes the LamB receptor, probably the J protein (see Supplementary Figure S10). However, with our current approaches, it is difficult to determine which domain of the J protein is involved in the physical recognition of the OMR.

In this work, we have characterized several fundamental aspects of the mEp_{imm1} phages according to current taxonomic criteria, which include comparative genomics and proteomics analyses, and compiled evidence related to the most relevant biological traits of these phages, such as morphotype, phage infection, integration, immunity and lysogeny regulation, and super-infection exclusion (i.e., “lifestyle” and host factors) (Turner et al., 2024). According to whole genome alignments, intergenomic distance calculations and hierarchical clustering, these coliphages constitute a novel branch of *Caudoviricetes*, distinct from other known coliphages, representing 9 genera and 45 species. Whole proteome comparison against >5,600 dsDNA phages indicated that these phages form a separate branch with internal cohesion, consistent with a family-level classification. Traditional phage characterization often allows for the association of biological behaviors that may not be evident in taxonomic classifications, constituting a complementary approach that helps biologists understand the relationships and characteristics of specific phages or groups of phages. Therefore, we additionally presented a biological feature-based characterization of these phages, and the results revealed particular traits in which these phages differ from reference phages like λ , such as distinct genome integration sites or the use of an additional *nut* site for antitermination, but also revealed diversity within the group, as observed in the use of different OMR proteins or lipoproteins for infection and superinfection exclusion, respectively. Noteworthy, these phages were as prevalent as lambdaoid phages in clinical

samples of human feces and homologous phage genomes were identified from sources with diverse geographical locations, including metagenome samples. It is possible that this group has undergone minor diversification events, as indicated by the high similarity rates of their repressor proteins (95.95%), which account for the fact that these phages share the same immunity group. Contrastingly, lambdoid phages have developed widely diverse immunity groups (Kameyama et al., 1999, 2001). This conservation could also indicate that the emergence of the mEp_{imm1} phages as *E. coli* parasites is recent, and that the original host could be another bacterial species. According to the ViPTree analysis, the branches closest to the mEp_{imm1} cluster are mainly constituted by *Salmonella* and *Klebsiella* phages (Supplementary Figure S4A). In addition, we found two complete homologous prophage genomes in *Shigella flexneri* and *Shigella sonnei* strains (accession numbers: ABMBTU010000008 and AAZUZA010000004, respectively), displaying intergenomic similarities of 69.7 and 73.7%, respectively, according to the VIRIDIC analysis. Further exploration and genomic characterization of mEp_{imm1} phages from different ecological niches will help to clarify the origins and evolution of this novel group of coliphages.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

HN-M: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. GV-T: Conceptualization, Formal analysis, Investigation, Methodology, Software, Data curation, Visualization, Writing – original draft, Writing – review & editing. OS-R: Formal analysis, Funding acquisition, Resources, Writing – original draft, Writing – review & editing. ER-C: Formal analysis, Methodology, Software, Validation, Writing – review & editing. JH-C: Methodology, Software, Visualization, Writing – review & editing. VF: Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing. AC: Investigation, Methodology, Software, Visualization, Writing – review & editing. LK: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. EM-P: Supervision, Visualization, Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Software. FF-R: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study described in this manuscript was supported by the Instituto Mexicano del Seguro Social through the registration of the protocol R-2023-3603-026, sponsored by Fundación IMSS. HN-M conducted this study to fulfill the requirements of the doctoral program in Genetics and Molecular Biology (Genetics and Molecular Biology department at Cinvestav), and received a doctoral scholarship from Conahcyt (773632). JH-C conducted part of this study to fulfill the requirements of master's program in the same Department and received a scholarship from Conahcyt (921940). This study was supported by Hospital General de México “Dr. Eduardo Liceaga” through the registered protocol DI/21/501/03/59.

Acknowledgments

We especially thank the National Institute of Genetics (Japan) and Dimitris Georgellis for providing us the derivative knockout mutant strains (Keio Collection). We acknowledge to Guadalupe Aguilar for technical assistance in DNA sequencing and to Fernando Riveros Mckay Aguilera and Leonardo Collado-Torres of Winter Genomics for preassembling assistance. To Ma. de Lourdes Rojas-Morales of LaNSE (Cinvestav-Zacatenco) for her technical assistance in electron microscopy experiments. To Carlos Eduardo Martínez-Soto for the construction of pLpp₀₂₁ and to Gerson Isaac Caraballo-Hernández for the construction of pRep₀₂₁.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1480411/full#supplementary-material>

References

- Abedon, S. T. (2011). Lysis from without. *Bacteriophage* 1, 46–49. doi: 10.4161/bact.1.1.13980
- Ackermann, H.-W., and Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. *Arch. Virol.* 157, 1843–1849. doi: 10.1007/s00705-012-1383-y
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Ansari, M. H., Ebrahimi, M., Fattahi, M. R., Gardner, M. G., Safarpour, A. R., Faghihi, M. A., et al. (2020). Viral metagenomic analysis of fecal samples reveals an enteric virome signature in irritable bowel syndrome. *BMC Microbiol.* 20:123. doi: 10.1186/s12866-020-01817-4
- Arguijo-Hernández, E. S., Hernandez-Sanchez, J., Briones-Peña, S. J., Oviedo, N., Mendoza-Hernández, G., Guarneros, G., et al. (2018). Cor interacts with outer membrane proteins to exclude FhuA-dependent phages. *Arch. Virol.* 163, 2959–2969. doi: 10.1007/s00705-018-3954-z
- Asakura, Y., and Kobayashi, I. (2009). From damaged genome to cell surface: transcriptome changes during bacterial cell death triggered by loss of a restriction-modification gene complex. *Nucleic Acids Res.* 37, 3021–3031. doi: 10.1093/nar/gkp148
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2:2006.0008. doi: 10.1038/msb4100050
- Bachmann, B. J. (1972). Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol. Rev.* 36, 525–557. doi: 10.1128/br.36.4.525-557.1972
- Barrera-Rojas, J., Gurubel-Tun, K. J., Ríos-Castro, E., López-Méndez, M. C., and Sulbarán-Rangel, B. (2023). An initial proteomic analysis of biogas-related metabolism of Euryarchaeota consortia in sediments from the Santiago river, México. *Microorganisms* 11:1640. doi: 10.3390/microorganisms11071640
- Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., and Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 10:e0128036. doi: 10.1371/journal.pone.0128036
- Borodovsky, M., and Lomsadze, A. (2011). Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. *Bioinform.* 35, 4.5.1–4.5.17. doi: 10.1002/0471250953.bi0405s35
- Botstein, D., and Herskowitz, I. (1974). Properties of hybrids between *Salmonella* phage P22 and coliphage λ . *Nature* 251, 584–589. doi: 10.1038/251584a0
- Campbell, A. (1994). Comparative molecular biology of lambdoid phages. *Ann. Rev. Microbiol.* 48, 193–222. doi: 10.1146/annurev.mi.48.100194.001205
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2011). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469. doi: 10.1093/bioinformatics/btr703
- Casjens, S. R., and Hendrix, R. W. (2015). Bacteriophage lambda: early pioneer and still relevant. *Virology* 479–480, 310–330. doi: 10.1016/j.virol.2015.02.010
- Cázares, A., Mendoza-Hernández, G., and Guarneros, G. (2014). Core and accessory genome architecture in a group of *Pseudomonas aeruginosa* Mu-like phages. *BMC Genomics* 15:1146. doi: 10.1186/1471-2164-15-1146
- Clarke, N., Beamer, L., Goldberg, H., Berkower, C., and Pabo, C. (1991). The DNA binding arm of lambda repressor: critical contacts from a flexible region. *Science* 254, 267–270. doi: 10.1126/science.254.5029.267
- Comeau, A. M., Bertrand, C., Letarov, A., Tétart, F., and Krusch, H. M. (2007). Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* 362, 384–396. doi: 10.1016/j.virol.2006.12.031
- Coughlan, S., Das, A., O’Herlihy, E., Shanahan, F., O’Toole, P. W., and Jeffery, I. B. (2021). The gut virome in irritable bowel syndrome differs from that of controls. *Gut Microbes* 13, 1–15. doi: 10.1080/19490976.2021.1887719
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6640–6645. doi: 10.1073/pnas.120163297
- Degnan, P. H., Michalowski, C. B., Babić, A. C., Cordes, M. H. J., and Little, J. W. (2007). Conservation and diversity in the immunity regions of wild phages with the immunity specificity of phage λ . *Mol. Microbiol.* 64, 232–244. doi: 10.1111/j.1365-2958.2007.05650.x
- Dulbecco, R. (1952). Mutual exclusion between related phages. *J. Bacteriol.* 63, 209–217. doi: 10.1128/jb.63.2.209-217.1952
- Dydecka, A., Bloch, S., Necel, A., Topka, G., Węgrzyn, A., Tong, J., et al. (2020). The ea22 gene of lambdoid phages: preserved polysogenic function despite of high sequence diversity. *Virus Genes* 56, 266–277. doi: 10.1007/s11262-020-01734-8
- Elias, J. E., and Gygi, S. P. (2009). Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* 604, 55–71. doi: 10.1007/978-1-60761-444-9_5
- Evseev, P., Lukianova, A., Sykilinda, N., Gorshkova, A., Bondar, A., Shneider, M., et al. (2021). *Pseudomonas* phage MD8: genetic mosaicism and challenges of taxonomic classification of lambdoid bacteriophages. *Int. J. Mol. Sci.* 22:10350. doi: 10.3390/ijms221910350
- Fattah, K. R., Mizutani, S., Fattah, F. J., Matsushiro, A., and Sugino, Y. (2000). A comparative study of the immunity region of lambdoid phages including Shiga-toxin-converting phages: molecular basis for cross immunity. *Genes Genet. Syst.* 75, 223–232. doi: 10.1266/ggs.75.223
- Ge, X., and Wang, J. (2024). Structural mechanism of bacteriophage lambda tail’s interaction with the bacterial receptor. *Nat. Commun.* 15:4185. doi: 10.1038/s41467-024-48686-3
- Groth, A. C., and Calos, M. P. (2004). Phage integrases: biology and applications. *J. Mol. Biol.* 335, 667–678. doi: 10.1016/j.jmb.2003.09.082
- Hatfull, G. F. (2008). Bacteriophage genomics. *Curr. Opin. Microbiol.* 11, 447–453. doi: 10.1016/j.mib.2008.09.004
- Hendrix, R. W. (2002). Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* 61, 471–480. doi: 10.1006/tpbi.2002.1590
- Hernández-Sánchez, J., Bautista-Santos, A., Fernández, L., Bermúdez-Cruz, R. M., Uc-Mass, A., Martínez-Peñañel, E., et al. (2008). Analysis of some phenotypic traits of feces-borne temperate lambdoid bacteriophages from different immunity groups: a high incidence of cor+, FhuA-dependent phages. *Arch. Virol.* 153, 1271–1280. doi: 10.1007/s00705-008-0111-0
- Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4, 923–925. doi: 10.1038/nmeth1113
- Kameyama, L., Bermúdez, R. M., García-Mena, J., Ishida, C., and Guarneros, G. (2001). Properties of a new coliphage group from human intestinal flora. *Recent Res. Dev. Virol.* 3, 297–303.
- Kameyama, L., Fernandez, L., Calderon, J., Ortiz-Rojas, A., and Patterson, T. A. (1999). Characterization of wild lambdoid bacteriophages: detection of a wide distribution of phage immunity groups and identification of a nus-dependent, nonlambdoid phage group. *Virology* 263, 100–111. doi: 10.1006/viro.1999.9888
- Kameyama, L., Fernandez, L., Court, D. L., and Guarneros, G. (1991). RNaseIII activation of bacteriophage lambda N synthesis. *Mol. Microbiol.* 5, 2953–2963. doi: 10.1111/j.1365-2958.1991.tb01855.x
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Kupczok, A., Bailey, Z. M., Refardt, D., and Wendling, C. C. (2022). Co-transfer of functionally interdependent genes contributes to genome mosaicism in lambdoid phages. *Microb. Genom.* 8:mgen000915. doi: 10.1099/mgen.0.000915
- Landy, A., and Ross, W. (1977). Viral integration and excision: structure of the lambda att sites. *Science* 197, 1147–1160. doi: 10.1126/science.331474
- Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., et al. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* 47, W260–W265. doi: 10.1093/nar/gkz303
- Li, G., Vissers, J. P. C., Silva, J. C., Golick, D., Gorenstein, M. V., and Geromanos, S. J. (2009). Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* 9, 1696–1719. doi: 10.1002/pmic.200800564
- Liu, W., Wang, Z., You, R., Xie, C., Wei, H., Xiong, Y., et al. (2024). PLMSearch: protein language model powers accurate and fast sequence search for remote homology. *Nat. Commun.* 15:2775. doi: 10.1038/s41467-024-46808-5
- Lou, R., and Shui, W. (2024). Acquisition and analysis of DIA-based proteomic data: a comprehensive survey in 2023. *Mol. Cell. Proteomics* 23:100712. doi: 10.1016/j.mcp.2024.100712
- LoVullo, E. D., Wright, L. F., Isabella, V., Huntley, J. F., and Pavelka, M. S. (2015). Revisiting the gram-negative lipoprotein paradigm. *J. Bacteriol.* 197, 1705–1715. doi: 10.1128/jb.02414-14
- Martínez-Peñañel, E., Fernández-Ramírez, F., Ishida, C., Reyes-Cortés, R., Sepúlveda-Robles, O., Guarneros-Peña, G., et al. (2012). Overexpression of Ipe protein from the coliphage mEp021 induces pleiotropic effects involving haemolysis by HlyE-containing vesicles and cell death. *Biochimie* 94, 1262–1273. doi: 10.1016/j.biochi.2012.02.004
- Mavrich, T. N., and Hatfull, G. F. (2017). Bacteriophage evolution differs by host, lifestyle and genome. *Nat. Microbiol.* 2, –17112. doi: 10.1038/nmicrobiol.2017.112
- McNair, K., Zhou, C., Dinsdale, E. A., Souza, B., and Edwards, R. A. (2019). PHANOTATE: a novel approach to gene identification in phage genomes. *Bioinformatics* 35, 4537–4542. doi: 10.1093/bioinformatics/btz265

- Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., et al. (2023). UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* 32:e4792. doi: 10.1002/pro.4792
- Milne, I., Bayer, M., Stephen, G., Cardle, L., and Marshall, D. (2015). Tablet: visualizing next-generation sequence assemblies and mappings. *Methods Mol. Biol.* 1374, 253–268. doi: 10.1007/978-1-4939-3167-5_14
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2023). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1
- Moraru, C. (2023). VirClust—A tool for hierarchical clustering, Core protein detection and annotation of (prokaryotic) viruses. *Viruses* 15:1007. doi: 10.3390/v15041007
- Moraru, C., Varsani, A., and Kropinski, A. M. (2020). VIRIDIC—A novel tool to calculate the Intergenomic similarities of prokaryote-infecting viruses. *Viruses* 12:1268. doi: 10.3390/v12111268
- Mushegian, A. R. (2020). Are there 10³¹ virus particles on earth, or more, or fewer? *J. Bacteriol.* 202:e00052-20. doi: 10.1128/jb.00052-20
- Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., and Goto, S. (2017). ViPTree: the viral proteomic tree server. *Bioinformatics* 33, 2379–2380. doi: 10.1093/bioinformatics/btx157
- Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15, 387–396. doi: 10.1093/dnares/dsn027
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., et al. (2022). InterPro in 2022. *Nucleic Acids Res.* 51, D418–D427. doi: 10.1093/nar/gkac993
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316
- Sambrook, J., Fritsch, E. R., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. 2nd Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sauer, R. T., Ross, M. J., and Ptashne, M. (1982). Cleavage of the lambda and P22 repressors by recA protein. *J. Biol. Chem.* 257, 4458–4462. doi: 10.1016/s0021-9258(18)34744-6
- Schicklmaier, P., Wieland, T., and Schmieger, H. (1999). Molecular characterization and module composition of P22-related Salmonella phage genomes. *J. Biotechnol.* 73, 185–194. doi: 10.1016/s0168-1656(99)00120-0
- Schrödinger, L., and DeLano, W. (2020). "PyMOL." Available at: <http://www.pymol.org/pymol>.
- Schwarz, H., Riede, I., Sonntag, I., and Henning, U. (1983). Degrees of relatedness of T-even type *E. coli* phages using different or the same receptors and topology of serologically cross-reacting sites. *EMBO J.* 2, 375–380. doi: 10.1002/j.1460-2075.1983.tb01433.x
- Sedhom, J., and Solomon, L. A. (2023). Lambda CI binding to related phage operator sequences validates alignment algorithm and highlights the importance of overlooked bonds. *Genes* 14:2221. doi: 10.3390/genes1412221
- Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860. doi: 10.1038/nprot.2006.468
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol. Syst. Biol.* 7:MSB201175. doi: 10.1038/msb.2011.75
- Silhavy, T. J., Berman, M. L., and Enquist, L. W. (1984). Experiments with gene fusions. New York: Cold Spring Harbor Laboratory Press.
- Soengas, M. S., Esteban, J. A., Salas, M., and Gutiérrez, C. (1994). Complex formation between phage phi 29 single-stranded DNA binding protein and DNA. *J. Mol. Biol.* 239, 213–226. doi: 10.1006/jmbi.1994.1364
- Stayrook, S., Jaru-Ampornpan, P., Ni, J., Hochschild, A., and Lewis, M. (2008). Crystal structure of the lambda repressor and a model for pairwise cooperative operator binding. *Nature* 452, 1022–1025. doi: 10.1038/nature06831
- Strauch, M. A., Spiegelman, G. B., Perego, M., Johnson, W. C., Burbulys, D., and Hoch, J. A. (1989). The transition state transcription regulator *abrB* of *Bacillus subtilis* is a DNA binding protein. *EMBO J.* 8, 1615–1621. doi: 10.1002/j.1460-2075.1989.tb03546.x
- Suzek, B. E., Ermolaeva, M. D., Schreiber, M., and Salzberg, S. L. (2001). A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* 17, 1123–1130. doi: 10.1093/bioinformatics/17.12.1123
- Thomason, L. C., Morrill, K., Murray, G., Court, C., Shafer, B., Schneider, T. D., et al. (2019). Elements in the lambda immunity region regulate phage development: beyond the 'Genetic Switch'. *Mol. Microbiol.* 112, 1798–1813. doi: 10.1111/mmi.14394
- Tisza, M. J., and Buck, C. B. (2021). A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc. Natl. Acad. Sci.* 118:e2023202118. doi: 10.1073/pnas.2023202118
- Turner, D., Adriaenssens, E. M., Lehman, S. M., Moraru, C., and Kropinski, A. M. (2024). Bacteriophage taxonomy: a continually evolving discipline. *Methods Mol. Biol.* 2734, 27–45. doi: 10.1007/978-1-0716-3523-0_3
- Turner, D., Kropinski, A. M., and Adriaenssens, E. M. (2021). A roadmap for genome-based phage taxonomy. *Viruses* 13:506. doi: 10.3390/v13030506
- Turner, D., Shkoporov, A. N., Lood, C., Millard, A. D., Dutilh, B. E., Alfenas-Zerbini, P., et al. (2023). Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Arch. Virol.* 168:74. doi: 10.1007/s00705-022-05694-2
- Valencia-Toxqui, G., Ballinas-Turrén, E. P., Bermúdez-Cruz, R. M., Martínez-Peñafiel, E., Guarneros, G., and Kameyama, L. (2023). Early antitermination in the atypical coliphage mEp021 mediated by the Gp17 protein. *Arch. Virol.* 168:92. doi: 10.1007/s00705-023-05721-w
- Valencia-Toxqui, G., and Ramsey, J. (2024). How to introduce a new bacteriophage on the block: a short guide to phage classification. *J. Virol.* 98:e0182123. doi: 10.1128/jvi.01821-23
- Wang, C., Duan, J., Gu, Z., Ge, X., Zeng, J., and Wang, J. (2024). Architecture of the bacteriophage lambda tail. *Structure* 32, 35–46.e3. doi: 10.1016/j.str.2023.10.006
- Weber, H., Pesavento, C., Possling, A., Tischendorf, G., and Hengge, R. (2006). Cyclic-di-GMP-mediated signalling within the sigma network of *Escherichia coli*. *Mol. Microbiol.* 62, 1014–1034. doi: 10.1111/j.1365-2958.2006.05440.x
- Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107
- Zheng, W., Wuyun, Q., Freddolino, P. L., and Zhang, Y. (2023). Integrating deep learning, threading alignments, and a multi-MSA strategy for high-quality protein monomer and complex structure prediction in CASP15. *Proteins Struct. Funct. Bioinform.* 91, 1684–1703. doi: 10.1002/prot.26585
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39, W347–W352. doi: 10.1093/nar/gkr485
- Zhu, Y., Shang, J., Peng, C., and Sun, Y. (2022). Phage family classification under Caudoviricetes: a review of current tools using the latest ICTV classification framework. *Front. Microbiol.* 13:1032186. doi: 10.3389/fmicb.2022.1032186