Characterization and genome sequence of the genetically unique 
Escherichia bacteriophage vB_EcoM_IME392

Yunjia Hu1 · Shanwei Tong3 · Ping Li4 · Xiaoping An1 · Lihua Song1,2 · Huahao Fan1,2 · Yigang Tong1,2

Received: 25 August 2020 / Accepted: 17 May 2021 / Published online: 8 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract
In this study, a novel Escherichia coli-specific bacteriophage, vB_EcoM_IME392, was isolated from chicken farm sewage in Qingdao, China. The genome of IME392 was found by next-generation sequencing to be 116,460 base pairs in length with a G+C content of 45.4% (GenBank accession number MH719082). BLASTn results revealed that only 2% of the genome sequence of IME392 shows sequence similarity to known phage sequences in the GenBank database, which indicates that IME392 is a novel bacteriophage. Transmission electron microscopy showed that IME392 belongs to the family Myoviridae. The host range, the multiplicity of infection, and a one-step growth curve were also determined.

Introduction
Since the German paediatrician Theodor Escherich isolated Escherichia coli from healthy human faeces in 1885, this bacterium has been extensively and thoroughly studied [7, 8, 21, 22, 24]. E. coli, as a model organism, currently plays a vital role in life science research and in biotechnology industries such as pharmaceuticals and industrial chemicals [5, 13, 18]. E. coli is an important microorganism that is ubiquitous in the natural environment and the mammalian gastrointestinal tract, and it is part of the normal intestinal flora. E. coli and other facultative anaerobes constitute approximately 0.1% of the gut microbiota [9]. Most E. coli strains are harmless, but certain serotypes can cause severe food poisoning, septic shock, meningitis, and/or urinary tract infections [16, 29], which seriously threaten human life and property safety. The discovery and use of antibiotics alleviated these dangers, but at the same time, antibiotic abuse has brought a new challenge to the clinical treatment of antibiotic resistance. There have been reports of E. coli strains resistant to all major antibiotic types, including extended-spectrum beta-lactams, carbapenems, fluoroquinolones, aminoglycosides, and trimethoprim-sulfamethoxazole [28]. Recently, even plasmid-mediated colistin resistance has emerged [23]. Bacteriophages that can lyse bacteria [14] seem to be a promising solution to the prevalence of multidrug-resistant bacteria. After their discovery by Twort and D’Hérelle, phages were soon used to treat bacterial infections. Phage therapy has certain advantages over antibiotic therapy, including low cost, easy availability, specificity, and few side effects.

Bacteriophages are the most widespread biological entities in nature, and their number is 10 times greater than that of bacteria. In co-evolution with their hosts, bacteriophages have developed tremendous diversity. The genome of E. coli phage MS2, with 3,569 nucleotides of positive-sense single-stranded RNA, was the first genome to be completely sequenced [10]. The following year, Sanger et al. completed the sequencing of bacteriophage Φ-X174, which was the first DNA genome to be sequenced [30]. M13 is a filamentous
bacteriophage composed of circular single-stranded DNA (ssDNA) that is 6,407 nucleotides long, encapsulated in approximately 2,700 copies of the major coat protein P8, and capped with five copies of two different minor coat proteins (P9, P6, P3) on the ends [26]. In 1951, Esther Lederberg serendipitously discovered that, after ultraviolet irradiation, the laboratory E. coli K12 strain released a bacteriophage, which was later named lambda. Subsequently, the entire life cycle, including lytic and lysogenic phases, has been deeply studied. In addition, bacteriophages with large genomes, such as T2, T4, T5, and T6 have also been studied in many aspects. Of these, T4-like phages are the most representative and widely studied phages with large genomes. The genomes of these phages are more than 100 kb in length and encode 100-300 – or even more – proteins and a variety of tRNAs. Most of these large-genome phages share well-organized and highly conserved core genes, especially those encoding DNA replication and virion structural proteins [6, 27]. Research on these known phages will help us understand and discover the mysteries of life and provide guidance for future research.

Thanks to the rapid development of high-throughput sequencing technology, the number of phage sequences in the GenBank database has grown geometrically. However, most of these phage sequences have a high degree of similarity to previously known sequences from bacteria or phages. In this study, we isolated and identified a genetically unique E. coli phage whose genome sequence is only 2% identical to those of the most similar sequences in GenBank, indicating that it may have completely different characteristics and functions from existing phages.

Materials and methods

Sampling, isolation, and purification of Escherichia phage IME392

Phage IME392 and its host strain, E. coli E2, were isolated from sewage samples from a chicken farm in Qingdao, China. For the purification of phage particles, the sewage samples were first centrifuged at 12,000 × g for 5 min and then filtered through a 0.22-μm filter to remove host cells. The filtrate was serially diluted tenfold in sterile PBS, and 100 μL of each dilution was mixed with 200 μL of the log-phase host bacterial culture, followed by incubation at room temperature for 5 minutes. The mixture was added to 5 mL of preheated 0.75% LB soft agar and poured onto the surface of 1.5% hard agar plates. After solidification, the plates were incubated overnight at 37 °C. Single plaques were isolated from the plates and again incubated overnight with a liquid culture of E2 with shaking at 37 °C. Cultures were re-centrifuged and sterile filtered, and the filtrates were subjected to another round of plaque assays. This process was repeated three times to obtain pure phage stocks.

Multilocus sequence typing (MLST)

Based on previous reports [15], primer pairs for eight house-keeping genes, dinB, icdA, pabB, polB, putP, trpA, trpB, and uidA, were designed for PCR amplification. All PCR products were purified by gel extraction and then sequenced by Beijing Ruiboxingke Biotechnology Co., Ltd., using the universal sequencing primers OF and/or OR. Further details about the MLST procedure can be found at http://www.pasteur.fr/mlst.

DNA extraction, gene sequencing, and bioinformatic analysis

Phage DNA was extracted using a modified standard phenol-chloroform extraction protocol [34]. First, DNase I and RNase A (Thermo Scientific, USA) were added to the purified phage IME392 preparation to a final concentration of 1 μg/mL and incubated overnight at 37 °C. After incubation at 80 °C for 15 minutes to inactivate DNase I and cooling to room temperature, lysis buffer with a final concentration of 0.5% SDS, 50 μg of protease K per ml, and 20 mM EDTA was added. The solution was incubated for 1 hour at 56 °C, and an equal volume of Tris-saturated phenol was added. The mixture was vortexed to form a uniform emulsion. After centrifugation at 10,000 × g at 4 °C for 5 minutes, the upper aqueous phase was collected and transferred to a new tube, and an equal volume of extraction agent (phenol:chloroform:isoamyl alcohol, 25:24:1) was added. The mixture was centrifuged again (10,000 × g, 5 min), and the aqueous phase was collected and added to an equal volume of isopropanol. The mixture was centrifuged again (10,000 × g, 4 °C, 5 min), and the aqueous phase was collected and added to an equal volume of isopropanol. The mixture was incubated at -20 °C for more than 1 hour, followed by centrifugation at 10,000 × g at 4 °C for 20 minutes, which precipitated the phage DNA. The pellet was washed twice with 1 mL of 75% cold ethanol, resuspended in 30 μL of deionized water, and stored at -20 °C.

A 2×300-nt paired-end DNA library was prepared using an NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® according to the manufacturer’s instructions. A Bioruptor UCD-200TS ultrasonic system was used to fragment 50 μL of DNA (approximately 100 ng) into 300- to 600-bp fragments. The resulting fragmented DNA was end-repaired and ligated to the NEBNext adaptor. Cleanup
of adaptor-ligated DNA was performed using AMPure XP beads. Finally, the cleaned DNA was amplified by PCR for 4 to 5 cycles, and the PCR product was purified again using AMPure XP Beads. An Agilent 2100 Bioanalyzer system was used to measure the size distribution of the constructed library fragments, and the library was quantified using a KAPA Library Quantification Kit. Whole-genome sequencing was performed on an Illumina MiSeq sequencing platform (San Diego, CA, United States) with a 600-cycle MiSeq v3 Reagent kit to generate 2×300-bp paired-end reads. In total, 555,564 raw reads were generated.

The raw sequencing data quality was analysed using the quality control software FastQC v0.11.5 and filtered for low-quality reads and adaptor regions using Trimmomatic 0.36 with default parameters [4]. The high-quality reads were assembled using SPAdes v3.13.0 with default parameters, and approximately 4,872 contigs were generated [3]. For the assembled contigs, Bandage v0.8.1 [31], which is a tool for visualizing assembly graphs with connections, was used to display the connections between those contigs. Only three of the contigs were circular, with lengths of 116,460, 39,440 and 4,888 bp and coverage of 352×, 7×, and 18×, respectively. BLASTn analysis confirmed that the two shorter contigs were lysogenic phages and plasmids. Mapping was carried out using CLC Genomics Workbench 12.0.2 (length fraction = 0.95; similarity fraction = 0.95), which was also used to adjust the sequences and for result checking. A consensus genome sequence was generated that spanned 100% of the reference genome, and the 425,959 mapped reads had an 884.1 mean read coverage. A nucleic acid sequence similarity search was performed using BLASTn (https://blast.ncbi.nlm.nih.gov/blast.cgi). Gene annotation was first run on RAST [2] (http://rast.nmpdr.org/) and then refined by amino acid sequence comparisons in BLASTp. A genome function map was generated using the laboratory’s self-built script and optimized using Inkscape 0.92.1.

The amino acid sequences of the major capsid protein and the terminase large subunit of the bacteriophage IME392 were used to construct a neighbor-joining phylogenetic tree via MEGA 7.0 with 1000 bootstrap replicates, which was optimized using the online website tool EvolView (https://www.evolgenius.info/evolview/).

Transmission electron microscopy

After centrifugation of the coculture of the phage and its host at 12,000 × g and passage through a 0.22-μm filter, the phage particles were purified by sucrose density gradient centrifugation [33]. Approximately 20 μL of purified, enriched phage samples were deposited on carbon-coated copper grids, allowed to absorb for 15 minutes, and then dried using filter paper. The phage particles were negatively stained with 2% (w/v) phosphotungstic acid (pH 7.0) for 2 min and examined using a JEM-1200EX transmission electron microscope (Jeol Ltd., Tokyo, Japan) at an acceleration voltage of 80 kV.

Host range determination

The host range of phage IME392 was determined by spot assay and confirmed by plaque assay. Suspected hosts were cultured at 37 °C to reach an optical density of 1.0. Three hundred milliliters of bacterial culture was added to 5 mL of preheated 0.7% LB agar and poured onto 1.5% agar plates.

After solidification, each plate was tested by pipetting 5 μL of phage suspension onto the bacterial lawn. Possible hosts were identified by plaque formation after overnight incubation at 37 °C. The plaque assay procedure is as described in the phage purification section.

Determination of the optimal multiplicity of infection

The multiplicity of infection (MOI), or the ratio of phage particles to bacterial cells prior to culture, affects the final level of bacteriophage produced. At the optimal MOI, a cultured product contains the most phage particles after reaching stationary phase. To determine the optimal MOI, first, the number of colony-forming units of the log-phase (OD600 = 0.6) host bacterial E2 culture and the number of plaque-forming units of the bacteriophage IME392 stock solution were calculated separately. A bacteria-phrase mixture was added to 5 mL of LB medium at different ratios to achieve MOIs of 10, 1, 0.1, 0.01, 0.001, and 0.0001 and was subsequently incubated at 37 °C with shaking at 220 rpm for 4 hours. After centrifugation at 12,000 × g for 1 min and passing the culture through a 0.22-μm filter, the phage titer was calculated using the double-layer plate method after serial dilution. Three replicates were performed, and the MOI that produced the highest phage titer was considered the best MOI for this phage.

One-step growth curve

A one-step growth curve was generated by the following method. A mixture of phage and bacteria at the optimal MOI (0.1) was incubated at 37 °C for 10 minutes for absorption. After centrifugation at 12,000 × g for 1 min, the supernatant containing unabsorbed phage particles was discarded, and the pellet was then washed twice with LB medium and resuspended in 20 mL of LB medium. The moment when the precipitation was resuspended in medium was defined as time zero. The suspension was then cultured at 37 °C with shaking at 220 rpm for 140 min. Samples (200 μL) were collected every 10 minutes (every 5 minutes in the
first 30 minutes) and then centrifuged and plated on double agar plates to determine the phage titer. Each sample was plated on three separate plates. Finally, the resulting one-step growth curve was plotted by GraphPad Prism 8.0.

**Determination of pH and temperature tolerance**

To determine the tolerance of phage particles to different pH values, LB medium was adjusted to a variety of pH values ranging from 2 to 13 with 5 M HCl or NaOH solution and then passed through a 0.22-μm filter. Then, 100 μL of purified phage suspension was added to 900 μL of LB medium with different pH values and incubated at 37 °C for 1 hour. To investigate the thermal stability of the phage, 100 μL of purified phage suspension was added to 900 μL of LB medium and incubated at 30 °C, 37 °C, 40 °C, 50 °C, 60 °C, 70 °C, or 80 °C for 1 hour. The titer of the remaining viable phage particles was determined using the double-layer agar method. All assays were performed in triplicate.

**Proteomic analysis**

The phage particles were concentrated using polyethylene glycol (PEG) and then purified by sucrose density gradient centrifugation [33]. Purified phages were mixed with 6x protein loading buffer (TransGen Biotech Co., LTD) and then boiled for 10 minutes, followed by concentration at 12,000 x g and 4 °C for 3 min. The proteins were separated by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and the bands were visualized by staining the gels with Coomassie brilliant blue. Gel slices were then excised and trypsinized. Briefly, 2.5 μg of trypsin was added to 100 μg of protein solution at a protein:trypsin ratio of 40:1 and incubated at 37 °C for 4 hours. Trypsin was added one more time at the above ratio, and enzymatic digestion was continued at 37 °C for 8 hours. The enzymatically digested peptides were desalted using a Strata X column and vacuum dried. The dried peptide sample was analyzed by liquid chromatography mass spectrometry (LC-MS). The full spectrum identification of proteins was mainly based on experimental tandem mass spectrometry data matched with theoretical mass spectrometry data obtained by database simulation. First, the original mass spectrum data were converted to a mass spectrum peak file. Then, the sequence in the database was searched and matched using Mascot v2.3.02 (parameters: enzyme, trypsin; fragment mass tolerance, 0.05 Da; fixed modifications, carboxamidomethyl (C), variable modifications oxidation (M), Gln->pyro-Glu (N-term Q), and deamidated (NQ); max missed cleavages, 1; instrument type, ESI-FTICR; database, bacteriophage_392_nr.fasta), and filtering and quality control (Mascot evaluation ≤ 0.05) were performed on the search results to give reliable protein identification results.

**Results**

**Morphological features of phage IME392**

After 10 hours of culture on double agar plates, the bacteriophage IME392 formed visible but small plaques reaching approximately 0.3-0.5 millimeters in diameter. Transmission electron microscopy results suggested that bacteriophage IME392 should be classified morphologically as a member of the family Myoviridae, possessing an icosahedral head approximately 83.93 ± 0.55 nm in diameter (n = 10) and a contractile tail 122.23 ± 3.55 nm in length (n = 10) (Fig. 1).

![Fig. 1](image)

**Fig. 1** Morphological features of bacteriophage vB_EcoM_IME392. IME392 possesses an icosahedral head approximately 83.93 ± 0.55 nm in diameter and a contractile tail 122.23 ± 3.55 nm in length (n = 10).
Host range

In this study, the ability of bacteriophage IME392 to lyse strains was determined by spot assay and plaque assay. A host range test was conducted on 33 clinically isolated pathogenic strains or environmentally isolated strains, including 28 *E. coli* strains and other bacteria (*Salmonella*). As shown in Table 1, among 28 strains of *E. coli*, only five strains produced bright and clear plaques, and the plaques formed on the lawns of eight strains was slightly turbid. The phage could not infect the other 15 *E. coli* strains or strains of other species. This indicates that the bacteriophage IME392 has a relatively narrow and specific host range.

**Physiological features of phage IME392**

The optimal MOI was determined to be 0.1, and this was used to generate the one-step growth curve for IME392 shown in Figure 2. The latent period and burst period were both approximately 15 minutes. In comparison to other known phages [11, 12, 19, 35], IME392 has a lower titer when reaching the stationary phase, at approximately $10^8$ plaque-forming units per milliliter (PFU/mL), which is consistent with the small plaque size of IME392.

| Bacterium  | Strain                  | Relevant characteristic(s) or source | ST | Spot assay |
|------------|-------------------------|--------------------------------------|----|------------|
| *Escherichia coli* |                         |                                      |    |            |
| 94         | Laboratory strain collection | 1 | +++        |
| 108        | Clinical isolates        | 43 | +          |
| 109        | Clinical isolates        | 477 | -         |
| 156        | Laboratory strain collection | 169 | +         |
| 161        | ATCC25922                | 52 | -          |
| 196        | Isolated from a dairy farm | Non-typeable | -   |
| 239        | DH5α                    | 262 | +          |
| 611        | Laboratory strain collection | 19 | +++        |
| 1186       | Clinical isolates        | 8 | +          |
| 1196       | Clinical isolates        | 44 | -          |
| 1644       | Isolated from chicken manure | 357 | +          |
| 1645       | Isolated from chicken manure | Non-typeable | +   |
| 1646       | Isolated from milk       | 357 | +++        |
| 1647       | Isolated from milk       | 466 | -          |
| 1648       | Isolated from milk       | 533 | -          |
| 2042       | BL21                    | 83 | -          |
| 2621       | MG1655                  | Non-typeable | +++ |
| 2724       | BL21(DE3)               | 83 | -          |
| 2726       | Laboratory strain collection | 945 | +        |
| 2727       | Nissle1917              | 4 | -          |
| 2734       | Laboratory strain collection | Non-typeable | -   |
| 2735       | Laboratory strain collection | Non-typeable | -   |
| 2736       | Laboratory strain collection | 88 | +          |
| 2738       | Laboratory strain collection | 132 | -        |
| 2739       | Laboratory strain collection | 303 | -        |
| 2740       | Laboratory strain collection | 466 | -        |
| 2741       | Laboratory strain collection | 24 | -        |
| 2743       | Ocean University of China | Non-typeable | +++ |
| *Salmonella* |                         |                                      |    |            |
| 2693       | CMCC50001               | / | -          |
| 2694       | ATCC13311               | / | -          |
| 2695       | ATCC14028               | / | -          |
| 2696       | CMCC50115               | / | -          |
| 2697       | CMCC50094               | / | -          |

+++ , clear plaques; +, plaques with slight turbidity; -, no plaques formed
The temperature and pH stability data for IME392 phage particles are shown in Figure 3A and B, respectively. IME392 is extremely sensitive to heat treatment. Incubation at 60 °C for 1 hour reduced the phage titer by 99.94%, while no phage activity remained if the incubation temperature reached 70 °C or higher. IME392 was also sensitive to both low- and high-pH environments. No live phage was detected when the phage particles were incubated at pH 2.0 or 13.0 at 37 °C for 1 hour.

**Genome sequencing and analysis**

The complete genome sequence of phage IME392 was 116,460 base pairs in length, with a GC content of 45.4%. A total of 160 potential open reading frames (ORFs) that could encode proteins were predicted by RAST. The genome sequence was submitted to GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and is available under the sequence ID MH719082.1.

The genome sequence of IME392 was analyzed using BLASTn. The results revealed that only 2% of the genome sequence (nt 91,907-94,497) was significantly similar to known nucleotide sequences in public databases. Homologous sequences included both phage and bacterial sequences. Eighty-nine potential CDSs were further analyzed using BLASTp, among which 34 were identified as functional proteins, including 15 morphogenesis-related proteins, 16 replication-related proteins, and five lysis-related proteins. Two tailspikes were classified as both morphogenesis- and lysis-related proteins. Twenty-three ORFs encoding proteins were highly homologous to phage proteins of unknown function, and two others were homologous to only hypothetical bacterial proteins. The products of the remaining 30 ORFs had no significant sequence similarity to proteins from the public database. The annotated genes are shown on the genome function map (Fig. 4) and in Table 2.

The replication-related modules of IME392 are mainly distributed between nt 7,051 and 54,779 of the genome, with a total length of approximately 47 kb, which is approximately one-third of the length of the entire genome. This is a relatively rare observation. A total of 60 open reading frames were predicted in this region, of which 16 ORF-encoded proteins have known functions, nine are hypothetical proteins, and the remaining 35 did not show homology to currently known proteins. The particularly large size of the replication-related region may be related to a large number of genes...
Escherichia bacteriophage vB_EcoM_IME392

for non-homologous proteins interspersed between genes encoding replication-related enzymes such as ligase, DNA polymerase, RNA polymerase, helicase, and topoisomerase. The next two genes encode two tRNAs, specifically for Met (CAT) and Arg (TCT). Packaging-related virion structure and lysis proteins are closely linked in the genome, covering the range of 61-104 kb, and they are all located in the positive strand of the IME392 genome.

The major capsid protein and terminase large subunit were chosen for phylogenetic tree construction (Fig. 5). In all cases, vB_EcoM_IME392 clustered with Vibrio phage Va2 and Vibrio phage 1.031.O._10N.261.46.F8 but belonged to different branches, indicating that they might have a common ancestor. Vibrio phage Va2 was isolated from aquaculture waters in Qingdao, Shandong, China, and IME392 was also isolated from Qingdao. This geographical similarity reveals that they may have an evolutionary connection. The Va2 genome is 128,360 bp in length and contains 134 open reading frames, similar to IME392. The 152,942-bp Vibrio phage 1.031.O._10N.261.46.F8 partial genome is larger than the others. Like IME392, 1.031.O._10N.261.46.F8 also has two tRNA genes [17]. However, the phylogenetic relationship to other phages was clearly distant, indicating that IME392 is a novel phage. Moreover, bacteriophage IME392
| ORF | Start | End  | Strand | Length | Top BLASTp hit                                      | Family                      | Accession no. | E-value   | MW (kDa) | pI       |
|-----|-------|------|--------|--------|---------------------------------------------------|----------------------------|---------------|-----------|----------|----------|
| 1   | 240   | 61   | -      | 180    | Hypothetical protein [Pectobacterium brasiliense]  | WP_172644783.1              | 6.00E-16      | 6.70     | 4.17     |          |
| 2   | 467   | 249  | -      | 219    | No hit                                            |                            |               | 7.53     |          |          |
| 3   | 962   | 630  | -      | 333    | No hit                                            |                            |               | 12.63    |          |          |
| 4   | 1270  | 1025 | -      | 246    | Hypothetical protein                              | WP_063439228.1              | 1.00E-18      | 9.22     | 4.87     |          |
| 5   | 1460  | 1272 | -      | 189    | DUF1653 domain-containing protein [Enterobacter asburiae] | NQW57987.1                  | 1.00E-07      | 7.43     | 6.27     |          |
| 6   | 1645  | 1460 | -      | 186    | No hit                                            |                            |               | 6.83     |          | 5.08     |
| 7   | 1899  | 1642 | -      | 258    | No hit                                            |                            |               | 9.61     |          | 6.27     |
| 8   | 2062  | 1886 | -      | 177    | No hit                                            |                            |               | 7.02     |          | 10.69    |
| 9   | 2746  | 2117 | -      | 630    | Hypothetical protein [Raoultella planticola]      | WP_032697006.1              | 5.00E-46      | 24.17    |          | 5.49     |
| 10  | 3642  | 2737 | -      | 906    | Glycosyltransferase family 4 protein [Escherichia coli] | WP_137536985.1              | 1.00E-98      | 34.86    |          | 5.66     |
| 11  | 4306  | 3671 | -      | 636    | Morphogenic protein [Enterobacter cancerogenus]    | WP_137271981.1              | 9.00E-63      | 23.57    |          | 5.67     |
| 12  | 4791  | 4270 | -      | 522    | Hypothetical protein [Enterobacter rogenkampii]    | WP_095429045.1              | 2.00E-33      | 19.64    |          | 4.87     |
| 13  | 5510  | 4803 | -      | 708    | Putative membrane protein [Escherichia phage adrianh] | WP_032697006.1              | 4.00E-27      | 24.89    |          | 4.31     |
| 14  | 5740  | 5510 | -      | 231    | Hypothetical protein [Salmonella phage barely]     | YP_009885781.1              | 1.00E-17      | 8.50     |          | 7.62     |
| 15  | 6106  | 5789 | -      | 318    | Hypothetical protein [Erwinia phage vB_EamM-Bue1]  | YP_009837251.1              | 3.00E-46      | 12.35    |          | 7.63     |
| 16  | 6522  | 6103 | -      | 420    | Inhibitor of host Lon protease [Escherichia phage EcS1] | BBC78124.1                  | 2.00E-08      | 15.65    |          | 4.72     |
| 17  | 6742  | 6509 | -      | 234    | Hypothetical protein [Enterobacter phage Arya]     | YP_095429045.1              | 2.00E-33      | 19.64    |          | 4.87     |
| 18  | 7051  | 6797 | -      | 255    | Acyl carrier protein [Candidatus Acetothermum autotrophicum] | BAL59693.1                  | 6.00E-06      | 9.31     |          | 3.60     |
| 19  | 8994  | 7051 | -      | 1944   | NAD-dependent DNA ligase LigA [Ralstonia phage RP13] | BCG50031.1                  | 7.00E-75      | 72.06    |          | 5.16     |
| 20  | 9817  | 9059 | -      | 759    | DNA polymerase/3'-5' exonuclease PolX [Thermobispora bispora] | WP_013132370.1              | 6.00E-26      | 28.56    |          | 5.86     |
| 21  | 9981  | 9814 | -      | 168    | No hit                                            |                            |               | 6.45     |          | 11.07    |
| 22  | 10354 | 9983 | -      | 372    | No hit                                            |                            |               | 13.82    |          | 5.32     |
| 23  | 10473 | 10366| -      | 108    | No hit                                            |                            |               | 4.08     |          | 9.98     |
| 24  | 10831 | 10589| -      | 243    | No hit                                            |                            |               | 9.29     |          | 8.85     |
| 25  | 11208 | 10831| -      | 378    | No hit                                            |                            |               | 14.32    |          | 8.52     |
| 26  | 11359 | 11198| -      | 162    | No hit                                            |                            |               | 6.21     |          | 6.53     |
| 27  | 11565 | 11359| -      | 207    | No hit                                            |                            |               | 7.81     |          | 9.66     |
| 28  | 11869 | 11573| -      | 297    | Hypothetical protein [Vibrio phage 1.031.O._10N.261.46.F8] | AUR83081.1                  | 4.00E-04      | 11.10    |          | 8.06     |
| 29  | 12252 | 11872| -      | 381    | No hit                                            |                            |               | 13.92    |          | 4.82     |
| 30  | 12573 | 12262| -      | 312    | No hit                                            |                            |               | 11.59    |          | 9.43     |
| 31  | 13018 | 12605| -      | 414    | No hit                                            |                            |               | 14.49    |          | 5.11     |
| 32  | 13249 | 13031| -      | 219    | No hit                                            |                            |               | 8.49     |          | 9.15     |
| 33  | 13634 | 13260| -      | 375    | No hit                                            |                            |               | 14.01    |          | 5.51     |
| 34  | 13978 | 13694| -      | 285    | No hit                                            |                            |               | 10.11    |          | 3.95     |
| 35  | 14442 | 14092| -      | 351    | No hit                                            |                            |               | 13.85    |          | 5.76     |
| 36  | 14728 | 14423| -      | 306    | No hit                                            |                            |               | 11.20    |          | 9.65     |
| 37  | 15148 | 14804| -      | 345    | No hit                                            |                            |               | 12.43    |          | 5.62     |
Table 2 (continued)

| ORF | Start | End  | Strand | Length | Top BLASTp hit | Family            | Accession no. | E-value | MW (kDa) | pI     |
|-----|-------|------|--------|--------|----------------|-------------------|---------------|---------|----------|--------|
| 38  | 15659 | 15141| -      | 519    | No hit         |                   |               |         | 18.97    | 4.33   |
| 39  | 15952 | 15653| -      | 300    | No hit         |                   |               |         | 11.09    | 4.32   |
| 40  | 16457 | 15927| -      | 531    | Hypothetical protein [Halomonas sp. S2151] |                   |               |         | 24.96    | 6.44   |
| 41  | 18025 | 16457| -      | 1569   | DNA-directed RNA polymerase beta subunit [Ralstonia phage RP13] | Myoviridae       | BCG50162.1   | 7.00E-85 | 58.73    | 8.79   |
| 42  | 18697 | 18041| -      | 657    | No hit         |                   |               |         | 26.59    | 8.72   |
| 43  | 19095 | 18757| -      | 339    | No hit         |                   |               |         | 12.77    | 4.26   |
| 44  | 19889 | 19203| -      | 687    | No hit         |                   |               |         | 26.59    | 8.72   |
| 45  | 21418 | 19904| -      | 1515   | DNA-directed RNA polymerase beta' subunit [Leptospira santarosai] |                   |               |         | 12.30    | 9.47   |
| 46  | 21752 | 21438| -      | 315    | No hit         |                   |               |         | 12.61    | 4.42   |
| 47  | 22083 | 21739| -      | 345    | No hit         |                   |               |         | 17.66    | 6.51   |
| 48  | 22547 | 22080| -      | 468    | No hit         |                   |               |         | 12.61    | 9.47   |
| 49  | 23118 | 22679| -      | 240    | No hit         |                   |               |         | 8.63     | 4.28   |
| 50  | 24533 | 23073| -      | 1461   | No hit         |                   |               |         | 54.16    | 6.89   |
| 51  | 24941 | 24672| -      | 270    | No hit         |                   |               |         | 9.66     | 10.91  |
| 52  | 28209 | 24922| -      | 3288   | DNA polymerase I [Ralstonia phage RP13] | Myoviridae       | BCG50106.1   | 4.00E-13 | 125.53   | 5.03   |
| 53  | 29916 | 28279| -      | 1638   | DEAD/DEAH box helicase family protein [Ralstonia phage RP13] | Myoviridae       | BCG50107.1   | 4.00E-12 | 62.13    | 5.23   |
| 54  | 30872 | 29880| -      | 993    | Hypothetical protein [Ralstonia phage RP13] | Myoviridae       | BCG50109.1   | 3.00E-35 | 38.40    | 6.28   |
| 55  | 31185 | 30862| -      | 324    | No hit         |                   |               |         | 11.60    | 6.82   |
| 56  | 31774 | 31175| -      | 600    | HD domain-containing protein [Ralstonia phage RP13] | Myoviridae       | BCG50111.1   | 1.00E-24 | 22.53    | 4.98   |
| 57  | 32931 | 31858| -      | 1074   | Thymidylate synthase [Salmonella phage SeSz-1] | Ackermannviridae | YP_009881884.1 | 2.00E-51 | 40.37    | 5.38   |
| 58  | 33650 | 32928| -      | 723    | Hypothetical protein [Ralstonia phage RP13] | Myoviridae       | BCG50110.1   | 6.00E-25 | 27.11    | 5.66   |
| 59  | 34711 | 33740| -      | 972    | Hypothetical protein [Ralstonia phage RP13] | Myoviridae       | BCG50138.1   | 2.00E-47 | 38.60    | 6.03   |
| 60  | 35790 | 34711| -      | 1080   | No hit         |                   |               |         | 41.14    | 5.42   |
| 61  | 36872 | 35838| -      | 1035   | No hit         |                   |               |         | 40.12    | 5.04   |
| 62  | 37490 | 36879| -      | 612    | PIG-L family deacetylase [Kocuria rhizophila] |                   | WP_144801709.1 | 5.00E-23 | 23.21    | 5.25   |
| 63  | 38959 | 37577| -      | 1383   | BCCT family transporter [Halomonas sp. Y2R2] |                   | WP_149284844.1 | 1.00E-107 | 50.70    | 9.21   |
| 64  | 39871 | 39863| -      | 909    | No hit         |                   |               |         | 33.51    | 4.83   |
| 65  | 40713 | 39973| -      | 741    | Hypothetical protein [Candidatus Woesearchaeota archaeon] |                   |             |         |         |        |
| 66  | 42065 | 40818| -      | 1248   | DNA topoisomerase, type IIA subunit A [Vibrio phage] | Myoviridae       | AUR82996.1   | 9.00E-51 | 46.87    | 8.17   |
| 67  | 44053 | 42068| -      | 1986   | DNA topoisomerase (ATP-hydrolyzing) subunit B [Thermosiphon atlanticus] |                   | WP_073073159.1 | 8.00E-54 | 75.27    | 5.14   |
| 68  | 44527 | 44090| -      | 438    | No hit         |                   |               |         | 16.56    | 8.66   |
| 69  | 44990 | 44736| -      | 255    | No hit         |                   |               |         | 9.71     | 4.81   |
| 70  | 45226 | 44987| -      | 240    | No hit         |                   |               |         | 8.83     | 5.79   |
| 71  | 48139 | 45374| -      | 2766   | SMC family ATPase [Ralstonia phage RP13] | Myoviridae       | BCG50154.1   | 0.0      | 104.00   | 5.17   |
| 72  | 48993 | 48136| -      | 858    | DNA polymerase [Bacillus phage SP-10] | Herelleviridae   | YP_007003453.1 | 3.00E-12 | 32.56    | 5.24   |
| 73  | 50056 | 49046| -      | 1011   | Hypothetical protein [Ralstonia phage RP13] | Myoviridae       | BCG50116.1   | 4.00E-39 | 37.52    | 4.77   |
| ORF | Start | End   | Strand | Length | Top BLASTp hit                                                                 | Family      | Accession no. | E-value | MW (kDa) | pI  |
|-----|-------|-------|--------|--------|--------------------------------------------------------------------------------|-------------|---------------|---------|----------|-----|
| 74  | 51193 | 50153 | -      | 1041   | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50123.1    | 4.00E-07 | 39.79    | 5.90 |
| 75  | 51417 | 51190 | -      | 228    | No hit                                                                        |             |               | 8.18    | 4.55     |     |
| 76  | 51870 | 51421 | -      | 450    | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50121.1    | 6.00E-29 | 17.21    | 5.05 |
| 77  | 53129 | 51870 | -      | 1260   | Replicative helicase [Ralstonia phage RP13]                                   | Myoviridae  | BCG50120.1    | 2.00E-110| 46.60    | 5.30 |
| 78  | 54779 | 53139 | -      | 1260   | DEAD/DEAH box helicase family protein [Ralstonia phage RP13]                  | Myoviridae  | BCG50119.1    | 6.00E-87 | 60.65    | 6.14 |
| 79  | 55300 | 54782 | -      | 519    | No hit                                                                        |             |               | 19.63   | 5.32     |     |
| 80  | 55657 | 55346 | -      | 312    | No hit                                                                        |             |               | 11.82   | 6.13     |     |
| 81  | 55947 | 55702 | -      | 246    | No hit                                                                        |             |               | 9.16    | 9.45     |     |
| 82  | 56254 | 55937 | -      | 318    | No hit                                                                        |             |               | 12.11   | 6.72     |     |
| 83  | 57140 | 56349 | -      | 792    | No hit                                                                        |             |               | 31.38   | 6.15     |     |
| 84  | 57285 | 57163 | -      | 123    | No hit                                                                        |             |               | 4.62    | 9.78     |     |
| 85  | 58180 | 58108 | -      | 73     | tRNA-Met-CAT                                                                  |             |               |         |          |     |
| 86  | 58725 | 58653 | -      | 73     | tRNA-Arg-TCT                                                                  |             |               |         |          |     |
| 87  | 60064 | 59522 | -      | 543    | Hypothetical protein [Vibrio phage PWH3a-P1]                                  | Myoviridae  | YP_007675922.1| 8.00E-41 | 20.84    | 7.07 |
| 88  | 60343 | 60182 | -      | 162    | No hit                                                                        |             |               | 5.83    | 3.96     |     |
| 89  | 60812 | 60330 | -      | 483    | Putative ssDNA binding protein [Pectobacterium phage PP99]                    | Autographiviridae | YP_009788767.1| 7.00E-41 | 18.26    | 7.63 |
| 90  | 61153 | 60812 | -      | 342    | No hit                                                                        |             |               | 12.31   | 9.26     |     |
| 91  | 61350 | 61823 | +      | 474    | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50099.1    | 3.00E-20 | 16.87    | 5.34 |
| 92  | 61827 | 63653 | +      | 1827   | Terminase large subunit [Vibrio phage 1.031.O_10N.261.46.F8]                  | Myoviridae  | AUR83174.1    | 7.00E-86 | 68.52    | 5.40 |
| 93  | 63654 | 65231 | +      | 1578   | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50094.1    | 4.00E-112| 58.86    | 4.58 |
| 94  | 65224 | 65478 | +      | 255    | No hit                                                                        |             |               | 9.16    | 4.37     |     |
| 95  | 65491 | 66927 | +      | 1437   | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50091.1    | 1.00E-73 | 52.33    | 4.70 |
| 96  | 66983 | 68410 | +      | 1428   | Major capsid protein [Vibrio phage 1.031.O_10N.261.46.F8]                    | Myoviridae  | AUR83170.1    | 4.00E-71 | 50.73    | 4.74 |
| 97  | 68560 | 69276 | +      | 717    | No hit                                                                        |             |               | 25.01   | 4.40     |     |
| 98  | 69310 | 69846 | +      | 537    | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50086.1    | 3.00E-39 | 20.16    | 9.87 |
| 99  | 69846 | 70133 | +      | 288    | No hit                                                                        |             |               | 10.89   | 11.65    |     |
| 100 | 70130 | 70759 | +      | 630    | Hypothetical protein [Ralstonia phage RP13]                                   | Myoviridae  | BCG50082.1    | 6.00E-29 | 23.90    | 4.78 |
| 101 | 70769 | 71332 | +      | 564    | No hit                                                                        |             |               | 20.53   | 4.84     |     |
| 102 | 71334 | 71999 | +      | 666    | Hypothetical protein [Chloroflexi bacterium]                                 | RLC62271.1 | 6.00E-04      | 24.65   | 4.71     |     |
| 103 | 72001 | 72192 | +      | 192    | No hit                                                                        |             |               | 7.02    | 6.18     |     |
| 104 | 72203 | 74002 | +      | 1800   | Phage tail sheath domain-containing protein [Pseudoalteromonas citrea]       | WP_138594708.1| 9.00E-122     | 64.39   | 4.80     |     |
| 105 | 74013 | 74537 | +      | 525    | Hypothetical protein [Pseudoalteromonas citrea]                              | WP_138594710.1| 2.00E-37      | 19.20   | 4.98     |     |
| 106 | 74552 | 75028 | +      | 477    | Hypothetical protein [Pseudoalteromonas]                                      | WP_125251889.1| 2.00E-20      | 18.43   | 4.90     |     |
| 107 | 75038 | 75865 | +      | 828    | Hypothetical protein [Pseudoalteromonas sp. McH1-7]                          | WP_176033179.1| 2.00E-28      | 31.11   | 5.11     |     |
| 108 | 75944 | 78802 | +      | 2859   | Glycoside hydrolase [Acinetobacter radioresistens]                           | WP_138009711.1| 1.00E-39      | 102.37  | 5.75     |     |
| 109 | 78802 | 79437 | +      | 636    | Hypothetical protein [Vibrio phage 1.031.O_10N.261.46.F8]                    | Myoviridae  | AUR83149.1    | 3.00E-09 | 23.26    | 8.86 |
| ORF Start | ORF End | Strand | Length | Top BLASTp hit | Family | Accession no. | E-value | MW (kDa) | pI |
|-----------|---------|--------|--------|----------------|--------|---------------|---------|----------|----|
| 108 7945  | 7982    | +      | 37     | Hypothetical protein [Pseudomonas aeruginosa PSL121770] | Myoviridae | WP_007860038.1 | 1.00E-34 | 26.40    | 7.53 |
| 109 7981  | 8046    | +      | 66     | Hypothetical protein [Pseudomonas aeruginosa PSL121770] | Myoviridae | WP_007860039.1 | 1.00E-34 | 26.40    | 7.53 |
| 110 8046  | 8101    | +      | 66     | Hypothetical protein [Pseudomonas aeruginosa PSL121770] | Myoviridae | WP_007860040.1 | 1.00E-34 | 26.40    | 7.53 |
| 111 8101  | 8156    | +      | 56     | No hit | | | | | |
| 112 8156  | 8211    | +      | 56     | No hit | | | | | |
| 113 8211  | 8266    | +      | 56     | No hit | | | | | |
| 114 8266  | 8321    | +      | 56     | No hit | | | | | |
| 115 8321  | 8376    | +      | 56     | No hit | | | | | |
| 116 8376  | 8431    | +      | 56     | No hit | | | | | |
| 117 8431  | 8486    | +      | 56     | No hit | | | | | |
| 118 8486  | 8541    | +      | 56     | No hit | | | | | |
| 119 8541  | 8596    | +      | 56     | No hit | | | | | |
| 120 8596  | 8651    | +      | 56     | No hit | | | | | |
| 121 8651  | 8706    | +      | 56     | No hit | | | | | |
| 122 8706  | 8761    | +      | 56     | No hit | | | | | |
| 123 8761  | 8816    | +      | 56     | No hit | | | | | |
| 124 8816  | 8871    | +      | 56     | No hit | | | | | |
| 125 8871  | 8926    | +      | 56     | No hit | | | | | |
| 126 8926  | 8981    | +      | 56     | No hit | | | | | |
| 127 8981  | 9036    | +      | 56     | No hit | | | | | |
| 128 9036  | 9091    | +      | 56     | No hit | | | | | |
| 129 9091  | 9146    | +      | 56     | No hit | | | | | |
| 130 9146  | 9201    | +      | 56     | No hit | | | | | |
| 131 9201  | 9256    | +      | 56     | No hit | | | | | |
| 132 9256  | 9311    | +      | 56     | No hit | | | | | |
| 133 9311  | 9366    | +      | 56     | No hit | | | | | |
| 134 9366  | 9421    | +      | 56     | No hit | | | | | |
| 135 9421  | 9476    | +      | 56     | No hit | | | | | |
| 136 9476  | 9531    | +      | 56     | No hit | | | | | |
| 137 9531  | 9586    | +      | 56     | No hit | | | | | |
| 138 9586  | 9641    | +      | 56     | No hit | | | | | |
| 139 9641  | 9696    | +      | 56     | No hit | | | | | |
| 140 9696  | 9751    | +      | 56     | No hit | | | | | |
| 141 9751  | 9806    | +      | 56     | No hit | | | | | |
| 142 9806  | 9861    | +      | 56     | No hit | | | | | |
| 143 9861  | 9916    | +      | 56     | No hit | | | | | |
| 144 9916  | 9971    | +      | 56     | No hit | | | | | |
| 145 9971  | 10026   | +      | 56     | No hit | | | | | |
| 146 10026 | 10081   | +      | 56     | No hit | | | | | |
| 147 10081 | 10136   | +      | 56     | No hit | | | | | |
| 148 10136 | 10191   | +      | 56     | No hit | | | | | |
| 149 10191 | 10246   | +      | 56     | No hit | | | | | |
| 150 10246 | 10301   | +      | 56     | No hit | | | | | |
| 151 10301 | 10356   | +      | 56     | No hit | | | | | |
| 152 10356 | 10411   | +      | 56     | No hit | | | | | |
| 153 10411 | 10466   | +      | 56     | No hit | | | | | |
| 154 10466 | 10521   | +      | 56     | No hit | | | | | |
| 155 10521 | 10576   | +      | 56     | No hit | | | | | |
| 156 10576 | 10631   | +      | 56     | No hit | | | | | |
| 157 10631 | 10686   | +      | 56     | No hit | | | | | |
| 158 10686 | 10741   | +      | 56     | No hit | | | | | |
| 159 10741 | 10796   | +      | 56     | No hit | | | | | |
| 160 10796 | 10851   | +      | 56     | No hit | | | | | |
| 161 10851 | 10906   | +      | 56     | No hit | | | | | |
| 162 10906 | 10961   | +      | 56     | No hit | | | | | |
| 163 10961 | 11016   | +      | 56     | No hit | | | | | |
| 164 11016 | 11071   | +      | 56     | No hit | | | | | |
| 165 11071 | 11126   | +      | 56     | No hit | | | | | |
| 166 11126 | 11181   | +      | 56     | No hit | | | | | |
| 167 11181 | 11236   | +      | 56     | No hit | | | | | |
| 168 11236 | 11291   | +      | 56     | No hit | | | | | |
| ORF | Start | End   | Strand | Length | Top BLASTp hit                                      | Family                | Accession no. | E-value  | MW (kDa) | pI   |
|-----|-------|-------|--------|--------|---------------------------------------------------|-----------------------|---------------|----------|----------|------|
| 144 | 109838| 109599| 240    | -      | Hypothetical protein                               | WP_157760109.1        | 6.00E-08      | 8.90     | 10.49    |      |
| 145 | 110066| 109893| 174    | -      | Hypothetical protein                               | WP_001563024.1        | 1.00E-17      | 6.60     | 4.34     |      |
| 146 | 110549| 110085| 465    | -      | NADAR family protein [Delftia acidovorans]         | NIT77307.1            | 8.00E-31      | 17.84    | 8.78     |      |
| 147 | 111118| 110549| 570    | -      | Hypothetical protein [Klebsiella phage ZCKP1]      | Myoviridae            | YP_009803533.1| 2.00E-21 | 22.48    | 5.48 |
| 148 | 111362| 111108| 255    | -      | No hit                                            |                       |               | 9.35     | 9.21     |      |
| 149 | 111536| 111372| 165    | -      | No hit                                            |                       |               | 6.24     | 9.39     |      |
| 150 | 111729| 111577| 153    | -      | No hit                                            |                       |               | 5.66     | 3.85     |      |
| 151 | 111910| 111716| 195    | -      | Hypothetical protein [Vibrio phage vB_VchM_Kuja]   | Ackermannviridae       | YP_009854103.1| 6.00E-08 | 7.66     | 4.44 |
| 152 | 112204| 111971| 234    | -      | No hit                                            |                       |               | 8.61     | 7.96     |      |
| 153 | 112904| 112734| 171    | -      | Hypothetical protein [Citrobacter phage Merlin]    | Myoviridae            | YP_009203833.1| 2.00E-04 | 6.31     | 6.53 |
| 154 | 113256| 112897| 360    | -      | No hit                                            |                       |               | 13.74    | 5.67     |      |
| 155 | 113600| 113367| 234    | -      | Hypothetical protein [Klebsiella phage vB_KpnM_KpS110] | Ackermannviridae       | YP_009798919.1| 4.00E-08 | 8.38     | 4.57 |
| 156 | 114141| 113611| 531    | -      | LysM peptidoglycan-binding domain-containing protein |                       | WP_094245731.1| 6.00E-04 | 20.10    | 6.82 |
| 157 | 114497| 114159| 339    | -      | No hit                                            |                       |               | 13.19    | 4.56     |      |
| 158 | 114724| 114497| 228    | -      | No hit                                            |                       |               | 8.39     | 6.11     |      |
| 159 | 115275| 114721| 555    | -      | Hypothetical protein [Klebsiella oxytoca 10-5244]  | KMV90563.1            | 5.00E-26      | 20.62    | 5.86     |      |
| 160 | 115696| 115286| 411    | -      | Hypothetical protein [Cronobacter phage CR9]       | Myoviridae            | YP_009015051.1| 3.00E-13 | 15.52    | 6.04 |
| 161 | 115874| 115686| 189    | -      | No hit                                            |                       |               | 6.89     | 4.04     |      |
| 162 | 116130| 115933| 198    | -      | No hit                                            |                       |               | 7.21     | 4.44     |      |
forms an independent branch in the phylogenetic tree, has low similarity to known phages in the family Myoviridae, and can be classified as a member of a new genus.

Proteomic analysis

To identify the predicted proteins by our genomic analysis, the phage was concentrated and analysed using mass spectrometry. Sixty proteins were identified, and 28 out of these were identified as having homologues of known function (Table 3). Nine of the identified proteins can be categorized as structural proteins or proteins involved in the morphogenesis of the phage (CDS1, CDS3, CDS48, CDS53, CDS65, CDS66, CDS68, CDS69, CDS72). In addition to the terminase large subunit, all replication-related proteins were identified. However, only two lysis-related proteins were identified. Mass spectrometry-based proteomics identified 28 of the 34 proteins predicted in our genomic analysis (Table 2) with a known function. In total, 60 out of 89 (67.4%) predicted proteins were identified by mass spectrometry proteomics, most of which were encoded by identified structural genes.

Discussion

The invention of antibiotics has greatly improved the living conditions of humans and saved tens of thousands of lives. However, at the same time, the problems of drug resistance and even multidrug resistance caused by antibiotic abuse have posed new challenges. It has been estimated that 700,000 people die from drug-resistant infections each year, and this number may rise to 10 million by 2050. The speed of development of new antibiotics cannot match the speed of antibiotic resistance development [20, 25]. Phage therapy has gradually become a research hotspot due to its high efficiency, excellent specificity, and easy availability. A variety of phage preparations have been successfully developed and used in clinical treatments. In this study, we isolated and identified a new bacteriophage, IME392, that can infect E. coli. The phage infected only some of the E. coli strains tested, and none of the strains from other species. In addition, no toxin genes, antibiotic resistance genes, phage lysogenic factors, or other pathogen-related genes were found among the genes with known functions in the phage genome. However, the functions of many genes are still unknown, so it is important to identify the functions of these genes. The use of this phage to treat infections caused by resistant E. coli still has a long way to go.

Genomic analysis shows that bacteriophage IME392 has low similarity to existing biological entities (only 2%) and might have numerous novel features. The genome annotation of IME392 revealed only 13 predicted phage structural proteins, while the genes encoding other phage structural proteins are still unknown. In addition, the genome of phage IME392 also encodes a variety of replication-related proteins, including DNA polymerase, DNA ligase, DNA helicase, topoisomerase, 5’-3’ DNA exonuclease, and other replication-related proteins. Therefore, it can rely on its own encoded enzymes to replicate, and we speculate that it may have its own replication mechanism. It is worth noting that we also found two genes encoding a DNA-directed RNA polymerase in the genome of the bacteriophage IME392, which is not common in bacteriophages.

Glycosyltransferase is an enzyme that can catalyze the transfer of the glycosyl moiety from an activated nucleotide sugar to a nucleophosphilic glycosyl acceptor molecule [32]. Some bacteriophages modify the glycome by influencing the expression of host glycosyltransferases, while other phages are unique in that they can express their own glycosyltransferases. These bacteriophages glucosylate their DNA...
Furthermore, glycosyltransferase may function in the puncturing or lysis of the cell wall peptidoglycan. We were surprised to find that the genome of bacteriophage IME392 also contains a gene encoding glycosyltransferase, which may prevent its removal by the host, but its exact function in the life cycle of the bacteriophage still needs experimental study. The IME392 genome also encodes two distinct tailspike proteins. The tailspike protein of enterobacteria phage P22 mediates the recognition and adhesion between the bacteriophage and the surface of *Salmonella enterica* cells [1].

It is speculated that the tailspike protein of bacteriophage IME392 has a similar function. We believe that the presence of the glycosyltransferase and tailspike protein helps phage IME392 infect and adsorb to the host more easily.

| CDS no. | Molecular mass (kDa) | No. of unique peptides | No. of unique spectra | Coverage (%) | Predicted function | Abundance | iBAQ |
|---------|----------------------|------------------------|-----------------------|--------------|--------------------|-----------|------|
| 48      | 50.699               | 11                     | 180                   | 33.05%       | Putative major capsid protein | 54692.29223 | 2878.541696 |
| 3       | 23.553               | 4                      | 8                     | 28.44%       | Putative morphogenic protein | 6750.338238 | 843.7922797 |
| 22      | 40.343               | 6                      | 18                    | 25.49%       | Putative thymidylate synthase | 51649.30422 | 2347.695646 |
| 53      | 64.349               | 10                     | 88                    | 24.21%       | Putative tail sheath protein | 25463.46055 | 943.0911316 |
| 76      | 24.319               | 5                      | 8                     | 22.79%       | Putative lysis protein       | 21046.06895 | 1315.379309 |
| 65      | 33.880               | 4                      | 19                    | 22.62%       | Putative tail fiber protein  | 23590.13879 | 1387.655223 |
| 31      | 46.839               | 7                      | 19                    | 22.17%       | Putative DNA topoisomerase subunit A | 32046.70718 | 1144.525257 |
| 81      | 16.686               | 2                      | 5                     | 20.00%       | Putative dUTP diphosphatase  | 3151.507709 | 315.1507709 |
| 32      | 75.226               | 7                      | 19                    | 19.06%       | Putative DNA topoisomerase subunit B | 18602.95892 | 489.5515504 |
| 8       | 28.546               | 3                      | 7                     | 18.25%       | Putative DNA-directed DNA polymerase family X | 5244.725872 | 374.6232766 |
| 68      | 52.295               | 5                      | 14                    | 17.60%       | Putative collagen fiber protein | 18607.30015 | 744.292006 |
| 19      | 62.092               | 6                      | 15                    | 16.70%       | Putative DNA helicase         | 19464.30079 | 648.8100263 |
| 1       | 24.150               | 2                      | 5                     | 16.27%       | Putative morphogenic protein   | 6115.22847 | 436.8020336 |
| 66      | 47.984               | 4                      | 17                    | 12.95%       | Putative tail fiber protein   | 6728.588573 | 373.8104763 |
| 71      | 111.264              | 10                     | 35                    | 12.76%       | Putative endosialidase tailspike | 41860.88542 | 872.1017796 |
| 39      | 46.568               | 4                      | 7                     | 12.41%       | Putative DNA helicase         | 32529.96823 | 1355.415343 |
| 69      | 22.999               | 1                      | 16                    | 11.39%       | Putative tail fiber assembly protein | 2197.969707 | 199.8154279 |
| 18      | 125.445              | 10                     | 25                    | 10.59%       | Putative DNA polymerase I     | 56612.4992 | 870.9615262 |
| 80      | 23.642               | 2                      | 9                     | 10.14%       | Putative dCTP deaminase       | 1812.964954 | 129.4974967 |
| 70      | 23.087               | 2                      | 3                     | 9.66%        | Putative tail fiber protein   | 6499.410564 | 464.2436117 |
| 72      | 94.901               | 7                      | 12                    | 8.71%        | Putative colonidase tailspike | 19163.57878 | 491.373815 |
| 7       | 72.014               | 4                      | 8                     | 6.65%        | Putative DNA ligase           | 17073.1183 | 474.2532861 |
| 12      | 58.693               | 3                      | 7                     | 5.56%        | Putative DNA-directed RNA polymerase beta’ subunit | 10412.92608 | 359.0664165 |
| 15      | 54.876               | 2                      | 3                     | 5.56%        | Putative DNA-directed RNA polymerase beta’ subunit | 4614.473378 | 128.1798161 |
| 57      | 102.312              | 4                      | 4                     | 5.36%        | Putative glycoside hydrolase  | 21515.99538 | 352.7212358 |
| 2       | 32.959               | 1                      | 2                     | 4.56%        | Putative glycosyltransferase  | 5288.362023 | 278.3348433 |
| 35      | 32.535               | 1                      | 1                     | 4.56%        | Putative 5’-3’ DNA exonuclease | 2315.063498 | 136.1802058 |
| 40      | 60.611               | 1                      | 1                     | 1.65%        | Putative DNA helicase         | 8802.824665 | 352.1129866 |

Table 3 Characteristics of the vB_EcoM_IME392 virion proteome identified by LC-MS/MS

In conclusion, we present here the biology and the genomic and proteomic characteristics of *E. coli* phage IME392, which was isolated from sewage samples from a chicken farm in Qingdao, China. The newly isolated phage IME392 was identified as a member of the family *Myoviridae*. The findings of this study not only provide new phage resources for the development of phage therapy against *E. coli* but also show that there are still many completely novel phages waiting to be discovered.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00705-021-05160-5.

Acknowledgements This research was supported by the Key Project of Beijing University of Chemical Technology (XK1803-06), the National Key Research and Development Program of China (2018YFA0903000, 2020YFC2005405, 2020YFA0712100, 2020YFC0840805), Funds for
First-Class Discipline Construction (XK1805), the Inner Mongolia Key Research and Development Program (2019ZD006), the National Natural Science Foundation of China (81672001, 81621005), the NSFC-MFST Project (China-Mongolia) (31961143024), and Fundamental Research Funds for Central Universities (BUCTRC201917, BUCTZY2022).

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Andres D, Baxa U, Hanke C, Seckler R, Barbiriz S (2010) Carbohydrate binding of Salmonella phage P22 tailspike protein and its role during host cell infection. Biochem Soc Trans 38:1386–1389
2. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formuska T, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genom 9:75
3. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pyshkin AV, Sierotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477
4. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
5. Chen X, Zhou L, Tian K, Kumar A, Singh S, Prior BA, Wang Z (2013) Metabolic engineering of Escherichia coli: a sustainable industrial platform for bio-based chemical production. Biotechnol Adv 31:1200–1223
6. Chibani-Chennoufi S, Canchaya C, Bruttin A, Brüssow H (2004) Comparative genomics of the T4-Like Escherichia coli phage JS98: implications for the evolution of T4 phages. J Bacteriol 186:8276–8286
7. Cohen SN, Chang AC, Boyer HW, Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci USA 70:3240–3244
8. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ (1961) General nature of the genetic code for proteins. Nature 192:1227–1228
9. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) Diversity of the human intestinal microbial flora. Science 308:1635–1638
10. Fiers W, Contreras R, Duerinck F, Haegeman G, Isertant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berge A, Volckaert G, Ysebaert M (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 260:500–507
11. Fu P, Zhao Q, Shi L, Xiong Q, Ren Z, Xu H, Chai S, Xu Q, Sun X, Sang M (2021) Identification and characterization of two bacteriophages with lytic activity against multidrug-resistant Escherichia coli. Virus Res 291:198696
12. Guo Z, Huang J, Yan G, Lei L, Wang S, Yu L, Zhou L, Gao A, Feng X, Han W, Gu J, Yang J (2017) Identification and characterization of Dpo42, a novel depolymerase derived from the Escherichia coli phage vB_EcoM_ECOO78. Front Microbiol 8:1460
13. Huang CJ, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in Escherichia coli and its recent advances. J Ind Microbiol Biotechnol 39:383–399
14. Jassim SA, Limoges RG (2014) Natural solution to antibiotic resistance: bacteriophages “The Living Drugs.” World J Microbiol Biotechnol 30:2153–2170
15. Jauregui F, Landraud L, Passet V, Diancourt L, Frapy E, Guignon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S (2008) Phylogenetic and genomic diversity of human bacteriophage Escherichia coli strains. BMC Genom 9:560
16. Kaper JB, Nataro JP, Mobley HL (2004) Pathogenic Escherichia coli. Nat Rev Microbiol 2:123–140
17. Kauffman KM, Brown JM, Sharma RS, Vaninsberghe D, Elsherbini J, Polz M, Kelly L (2018) Viruses of the Nahant Collection, characterization of 251 marine Vibrioaceae viruses. Sci Data 5:180114
18. Kim B, Park H, Na D, Lee SY (2014) Metabolic engineering of Escherichia coli for the production of phenol from glucose. BioTechnol J 9:621–629
19. Kutter EM, Skutt–Kakaria K, Blasdel B, El-Shibiny A, Castano A, Bryan D, Kropinski AM, Villegas A, Ackermann HW, Toribio AL, Pickard D, Anany H, Callaway T, Brabban AD (2011) Characterization of a VI-like phage specific to Escherichia coli O157:H7. Virol J 8:430
20. Laxminarayanan R, Amáible-Cuevas CF, Cao O, Evans T, Heymann DL, Hoffman S, Holmes A, Mendelson M, Sridhar D, Woolhouse M, Rottingen JA (2016) UN High-Level Meeting on antimicrobials—what do we need? Lancet 388:218–220
21. Linn S, Arber W (1968) Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proc Natl Acad Sci USA 59:1300–1306
22. Liu T, Khosla C (2010) Genetic engineering of Escherichia coli for biofuel production. Annu Rev Genet 44:53–69
23. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH, Shen J (2016) Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis 16:161–168
24. Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. Genetics 28:491–511
25. O’Neill J (2016) Tackling drug-resistant infections globally: final report and recommendations. Government of the United Kingdom
26. Opella SJ, Stewart PL, Valentine KG (1987) Protein structure by solid-state NMR spectroscopy. Q Rev Biophys 19:455–477
27. Petrov VM, Nolan JM, Bertrand C, Levy D, Desplats C, Krish HM, Karam JD (2006) Plasticity of the gene functions for DNA replication, characterization of 251 marine Vibrionaceae viruses. Sci Commun 5:180114
28. Pechman D, Bertrand C, Levy D, Desplats C, Krish HM, Karam JD (2006) Plasticity of the gene functions for DNA replication, characterization of 251 marine Vibrionaceae viruses. Sci Commun 5:180114
29. Russo TA, Johnson JR (2003) Medical and economic impact of drug-resistant infections. J Ind Microbiol Biotechnol 30:5–16
30. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265:687–695
31. Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352

Escherichia bacteriophage vB_EcoM_IME392

2519
32. Williams GJ, Thorson JS (2009) Natural product glycosyltransferases: properties and applications. Adv Enzymol Relat Areas Mol Biol 76:55–119
33. Yuan W, Zhang Y, Wang G, Bai J, Wang X, Li Y, Jiang P (2016) Genomic and proteomic characterization of SE-I, a temperate bacteriophage infecting Erysipelothrix rhusiopathiae. Arch Virol 161:3137–3150
34. Zhang Q, Xing S, Sun Q, Pei G, Cheng S, Liu Y, An X, Zhang X, Qu Y, Tong Y (2017) Characterization and complete genome sequence analysis of a novel virulent Siphoviridae phage against Staphylococcus aureus isolated from bovine mastitis in Xinjiang, China. Virus Genes 53:464–476
35. Zhou Y, Bao H, Zhang H, Wang R (2015) Isolation and Characterization of Lytic Phage vB_EcoM_JS09 against Clinically Isolated Antibiotic-Resistant Avian Pathogenic Escherichia coli and Enterotoxigenic Escherichia coli. Intervirology 58:218–231

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