Characterization and Genome Sequence of a Genetically Unique Escherichia Bacteriophage vB_EcoM_IME392

Yunjia Hu

Beijing University of Chemical Technology

Shanwei Tong

The University of British Columbia

Ping Li

Yangzhou University

Xiaoping An

Beijing University of Chemical Technology

Lihua Song

Beijing University of Chemical Technology

Huahao Fan (✉ fanhuahao@mail.buct.edu.cn)

Beijing University of Chemical Technology <https://orcid.org/0000-0001-5007-2158>

Yigang Tong

Beijing University of Chemical Technology

Research Article

Keywords: Escherichia coli-specific bacteriophage vB_EcoM_IME392, electron microscopy, German pediatrician Theodor Escherich isolated

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-306526/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Archives of Virology on July 8th, 2021. See the published version at <https://doi.org/10.1007/s00705-021-05160-5>.

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 sequenced by Next Generation Sequencing is found to contain 116,460 base pairs in length with a G+C content of 45.4% (deposited in GenBank under the accession MH719082). BLASTN results reveal that only 2% of the genome sequence of IME392 shows homology to known phage sequences in GenBank, which indicates IME392 is a very novel bacteriophage. Transmission electron microscopy showed that IME392 belongs to the family *Myoviridae*. The host range, multiplicity of infection and one-step growth curve was also determined.

Introduction

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

Bacteriophages are the most widespread living entities in nature, and their number is ten times more than that of bacteria. In co-evolution with the host, bacteriophages have produced tremendous diversity. The genome of *Escherichia coli* phage MS2 with 3,569 nucleotides of positive-sense single-stranded RNA is the first genome to be completely sequenced [10]. The following year, Sanger completed the sequencing of the Φ -X174 bacteriophage, which is the first DNA genome to be sequenced [30]. M13 is a filamentous bacteriophage composed of circular single-stranded DNA (ssDNA) which is 6,407 nucleotides long encapsulated in approximately 2,700 copies of the major coat protein P8, and capped with 5 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 *Escherichia coli* K-12 strain released a bacteriophage that was later named lambda. Subsequently, the entire life cycle of its lysis and lysogen has been deeply studied. In addition, bacteriophages with large genome such as T2, T4, T5 and T6 have also been studied in many aspects. Among them, the T4-like phages are the most representative and the widely studied phages with large genome. The genomes of these phages are more than 100 kb and encode 100-300 or even more kinds of proteins and a variety of tRNAs. Most of these large genome phages share the well-organized and highly conserved core genes, especially those encode DNA replication and virion structural proteins [6, 27]. The 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 phage sequence in GenBank has grown geometrically. However, most of the phage sequences in GenBank have a high degree of homology with known bacteria or phage, while phage with lower homology are rare. In this study, we isolated and identified a genetically unique *Escherichia coli* phage, whose genome is only 2% homologous to the phage in GenBank, indicating that it may have completely different characteristics and functions from the existing phage.

Materials And Methods

Sampling, Isolation and Purification of *Escherichia coli* Phage IME392

Phage IME392 and its host cell, *Escherichia coli* E2, were isolated from sewage samples taken from a chicken farm in Qingdao, China. For the purification of phage particles, the sewage samples were firstly centrifuged at 12,000×g for 5 min and then filtered through a 0.22- μ m membrane. After that, an equal amount of the filtrate was added to the 3×LB medium containing the log-phase host bacteria E2 ($OD_{600} = 0.4$), and cultured overnight at 37 °C, 180 rpm. The culture was centrifuged at 12,000×g for 5 minutes and precipitation was discarded while the supernatant was filtrated by passing through a 0.22- μ m filter to remove host cells. The filtrate was serially diluted with sterile PBS in a 10-fold dilution method, and then 100 μ L of the dilution was mixed with 200 μ L of the log-phase host bacteria 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 on the surface of 1.5% hard agar plates. After solidification, the plates were incubated at 37 °C overnight. Single plaques were isolated from the plates and again incubated over night with a liquid culture of E2 with shaking at 37 °C. Cultures were re-centrifuged, sterile filtrated and the filtrates were subjected to another round of plaque assays. This process is repeated three times to obtain pure phage stocks.

Multilocus sequence typing

According to the previous reports [15], primer pairs for eight housekeeping genes including *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 universal sequencing primer OF and/or OR. Further details on this MLST scheme can be found at <http://www.pasteur.fr/mlst>.

DNA Extraction, Gene Sequencing and Bioinformatic Analysis

According to a modified standard phenol-chloroform extraction protocol [34], DNase I and RNase A (Thermo Scientific, USA) with a final concentration of 1 µg/mL were added to the purified phage IME392 stock solution and incubated overnight at 37 °C. After incubating at 80 °C for 15 minutes to inactivate DNase I and cooling down to room temperature, the lysis buffer with a final concentration of 0.5% SDS, 50 µg/ml protease K and 20 mM EDTA was added. The solution was incubated for 1 hour at 56 °C, and added an equal volume of Tris-saturated phenol. The mixture was vortexed to form a uniform emulsion. After 10,000×g centrifugation at 4 °C for 5 minutes, the upper aqueous phase was collected and transferred to a new tube, while also adding an equal volume of extraction agent (phenol: chloroform: isoamyl alcohol= 25:24:1). 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 10,000×g centrifugation at 4 °C for 20 minutes, precipitating phage DNA. The precipitation was collected and washed twice with 1 mL of 75% cold ethanol, then resuspended in 30 µL of deionized water and stored at -20 °C.

According to the manufacturer's instructions, a 2×300 nt paired-end DNA library was prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina®. The Bioruptor UCD-200TS ultrasound system is used to fragment a total volume of 50 µL DNA (about 100 ng) into 300-600 bp fragments. The resulting fragmented DNA was end repaired and ligated to NEBNext adaptor. Cleanup of Adaptor-ligated DNA was used 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. Agilent 2100 Bioanalyzer system was used to measure the size distribution of the constructed library fragments, and the library was quantified using the KAPA Library Quantification Kits. Whole genome sequencing was performed on 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. Overall, 555,564 raw reads were generated.

The raw sequencing data quality was analyzed using the quality control software FastQC v0.11.5 and filtered for low quality reads and adapter regions using Trimmomatic 0.36 with default parameters [4]. The generated high-quality reads were assembled by 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. Among them, only 3 contigs are circularity with lengths of 116,460, 39,440 and 4,888 bp respectively and coverage of 352×, 7× and 18× respectively. By BLASTn analysis, it is confirmed that the two shorter contigs are lysogenic phage and plasmid respectively. The mapping was carried out with 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 was generated that spanned 100% of the reference genome and the 425,959 mapped reads had an 884.1 mean read coverage. Nucleic acid sequence similarity search was performed by BLASTn (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Gene annotation was first run on RAST [2] (<http://rast.nmpdr.org/>), then refined by amino acid sequence comparisons on BLASTp. The genome function map was generated by the laboratory's self-built script and optimized using Inkscape 0.92.1.

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

Transmission Electron Microscopy

After centrifuging the co-culture of the phage and its host at 12,000×g and filtering with a 0.22-µm filter, the phage particles were purified by sucrose density gradient centrifugation [33]. Approximately 20 µL purified, enriched phage sample were deposited on carbon-coated copper grids for a 15 minutes absorption and then dried using filter paper. The phage particles were negatively stained by 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 the acceleration voltage of 80 kV.

Host Range Determination

The host range of phage IME392 was determined by spot assay and was confirmed with the plaque assays. Suspected hosts were cultured at 37 °C to reach the optical density of 1.0. 300 mL of bacterial cultures was added to 5 mL of preheated 0.7% LB agar, and poured on 1.5% agar plates respectively. After solidification, each plate was examined by pipetting 5 µL phage suspension onto the bacterial lawn. Eligible hosts are detected by plaque formation after incubating overnight at 37 °C. The plaque assay procedure is as described in the phage purification section.

Determination of Optimal Multiplicity of Infection

Multiplicity of infection (MOI), or the ratio of phage particles to bacteria cells prior to culture, affects the final production of bacteriophages. At optimal MOI, the cultured product contains the largest amount of phage particles after reaching stationary phase. In order to determine the optimal MOI, firstly, the colony forming unit of the log-phase ($OD_{600} = 0.6$) host bacteria E2 culture and the plaque forming unit of the bacteriophage IME392 stock solution were calculated separately. The bacteria-phage mixture was added to 5mL LB medium according to different ratios to achieve a MOI of 10, 1, 0.1, 0.01, 0.001, or 0.0001, which were subsequently incubated at 37 °C, 220 rpm for 4 hours. After centrifuging at 12,000×g for 1min and filtering 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 conducted for determination and the MOI produced the highest phage titer was considered the best MOI of the phage.

One-step Growth Curve

The one-step growth curve was examined by the following method. A mixture of phage and bacteria at optimal MOI (0.1) was allowed to incubate in 37 °C for 10 minutes for absorption. After centrifugation at 12,000×g for 1 min, the supernatant which contained the unabsorbed phage particles were discarded, while the precipitation was then washed twice with LB medium and re-suspended in 20 mL of LB medium. The moment when the precipitation

was re-suspended in medium was defined as time zero. The obtained culture was subsequently transferred to a shaker and cultured at 37 °C, 220 rpm for 140 min. The 200 µL samples 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 were plated on 3 separate plates respectively. Finally, the 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, LB medium were adjusted to a variety of pH (2.0, 3.0, 4.0 ... 13.0) with 5 M HCl or NaOH solution and then filtered through a 0.22-µm filter. 100 µl purified phage suspension were added into the 900 µl LB medium of different pHs and incubated at 37 °C for 1 hour. To investigate the thermal stability of the phage, 100 µl purified phage suspension were added to 900 µl LB medium and incubated at a variety of temperatures (30 °C, 37 °C, 40 °C, 50 °C, 60 °C, 70 °C, 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 by PEG and then purified by sucrose density gradient centrifugation [33]. Purified phages were mixed with 6× protein loading buffer (TransGen Biotech Co., LTD) and then boiled for 10 minutes, followed by concentrating at 12,000×g, 4 °C for 3 min. The proteins were separated by 12% SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) and the bands were visualized by staining the gels with Coomassie brilliant blue. Gel slices were then excised and trypsinized. Briefly, 2.5µg Trypsin enzyme was added to 100 µg protein solution according the ratio of protein:enzyme = 40:1, and enzymatically digested at 37 °C for 4 hours. Add Trypsin one more time according to the above ratio, and continue enzymatic digested at 37 °C for 8 hours. The enzymatically digested peptides are desalted by Strata X column and vacuumed to dry. The dried peptide sample was analyzed by using Liquid chromatography–mass spectrometry (LC-MS). The full spectrum identification of protein is mainly based on experimental tandem mass spectrometry data, matching with theoretical mass spectrometry data obtained by database simulation, so as to obtain protein identification results. First the original mass spectrum data was converted into a mass spectrum peak file, and then searched and matched the sequence in the database using Mascot v2.3.02 (parameter: Enzyme Trypsin, Fragment Mass Tolerance 0.05 Da, Fixed modifications Carbamidomethyl (C), Variable modifications Oxidation (M); Gln->pyro-Glu (N-term Q); Deamidated (NQ), Max Missed Cleavages 1, Instrument type ESI-FTICR, Database Bacteriophage_392_nr.fasta) and performed some filtering and quality control (Mascot evalue <= 0.05) 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 forms visible but small plaques, reaching around 0.3-0.5 millimeters in diameter. Transmission electron microscopy results suggest that bacteriophage IME392 is morphologically classified as *Myoviridae*, possessing an icosahedral head of approximately 83.93 ± 0.55 nm in diameter (n=10) and a contractile tail of 122.23 ± 3.55 nm in length (n=10) (Fig. 1).

Host range

In this study, the ability of bacteriophage IME392 to lyse strains was determined by the spot assay as well as plaque assay. The host range test was conducted on 33 clinically isolated pathogenic strains or environmentally isolated strains, including 28 *E. coli* strains, as well as other bacteria consisted of *Salmonella*. As the results are shown in **Table 1**, among 28 strains of *E. coli*, only 5 strains produced bright and clear plaques and the plaque formed on the lawns of 8 strains was slightly turbid. The phage couldn't infect other 15 *E. coli* strains and strains from other species. This indicates that the bacteriophage IME392 has a relatively narrow and more specific host range.

Physiological Features of Phage IME392

The optimal MOI was determined as 0.1, and it was the original phage to host cell ratio of the one-step growth curve experiment. The one-step growth curve of IME392 is shown in figure 2. It can be observed that the latent period and burst period are both approximately 15 minutes. In comparison with other known phages [11, 12, 19, 35], IME392 has lower phage titer when reaching stationary phase, at about 10⁸ plaque forming units per milliliter (PFU/mL), and this result is consistent with the small plaque size of IME392.

The temperature and pH stability of IME392 phage particles are shown in figures 3A and 3B, respectively. IME392 is extremely sensitive to heat treatment. Incubating up to 60 °C for 1 hour is able to reduce phage titer by 99.94%, while no phage titer was observed if the incubation temperature reached 70 °C or higher. IME392 is also sensitive to both low and high pH environment. No live phage was detected when the phage particles were incubated in pH = 2.0 or 13.0 environment at 37 °C for 1 hour.

Genome Sequencing and Analysis

The complete genome sequence of phage IME392 consists of 116,460 base pairs in length, with a GC content of 45.4%. 160 potential open reading frames (ORFs) were predicted by RAST. The genome sequence is submitted to Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>), available to public access under the sequence ID MH719082.1.

The genome sequence of IME392 was analyzed by BLASTN. Results have revealed that only 2% of the genome sequence (91,907-94,497) are homologous to known nucleotide sequences in public databases. Homologous sequences contain both phage and bacterial sequences. The 89 potential CDS were further analyzed by BLASTP, among which 34 were identified as functional protein, including 15 morphogenesis-related protein, 16 replication-related protein, and 5 lysis-related protein. Two tailspikes were classified as both morphogenesis and lysis related protein. 23 ORFs encoding proteins are highly homologous to phage protein of unknown function, and 2 others are only homologous to bacterial hypothetical protein. The products of the remaining 30 ORFs have no significant homology to proteins from the public database. The annotated genes were shown on the genome function map (Fig.4) and Table 2.

The replication-related modules of IME392 are mainly distributed in the 7,051-54,779 region of the genome, with a total length of about 47 kb, which is about one-third of the length of the entire genome, and the length of the span is relatively rare. A total of 60 open reading frames are predicted in this region, of which 16 ORF-encoded proteins have known functions, 9 are hypothetical proteins, and the remaining 35 are not hit and have no homology with currently known proteins. The situation where the length span is particularly large may be related to a large number of non-homology 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, 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. This unique organization of the IME392 genome may be one of the reasons for its low homology with existing bacteriophages.

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 belong to different branches, indicating that they may 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 carries a total of 134 open reading frames, which is basically the same as IME392. The 152,942 bp Vibrio phage 1.031.O_10N.261.46.F8 partial genome is larger than the two of them. Like IME392, 1.031.O_10N.261.46.F8 also has 2 tRNA genes [17]. However, the phylogenetic relationship with other phages is clearly distant, indicating that IME392 is a novel phage. Moreover, the bacteriophage IME392 presents an independent branch on the phylogenetic tree, and has low homology with the existing phages in the *Myoviridae* family, and can be classified as a new genus.

Proteomic Analysis

To identify the predicted proteins by our genomic analysis, the phage was concentrated and analysed using mass spectrometry. 60 proteins were identified and 28 out of these were identified as having homologs of known function (Table 3). Nine of the identified proteins can be categorized as structural proteins, 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. The 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 them encoded by identified structural genes.

Discussion

The invention of antibiotics has greatly improved the living conditions of mankind and saved tens of thousands of lives. But at the same time, the problem of drug resistance and even multi-drug resistance caused by the antibiotics abuse has brought new challenges to mankind. And it was 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 [20, 25]. Phage therapy has gradually become a research hotspot due to its high efficiency, strong specificity, and easy availability. At present, a variety of phage preparations have been successfully developed and used in clinical treatment. In this study, we isolated and identified a new type of bacteriophage IME392 that can infect *Escherichia coli*. The phage can only infect part of the tested *E. coli* strains, but not other *E. coli* strains or strains from other species. In addition, no toxin genes, antibiotic resistance genes, phage lysogenic factors and other pathogen-related genes have been found in 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 the bacteriophage IME392 has low similarity with existing biological entities (only 2%) and may have numerous uncharted features. The genome annotation of IME392 only revealed 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 encoding enzymes to replicate, we speculate that it may have its own replication mechanism. It is worth noting that we also found two genes encoding the 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 the activated nucleotide sugar to the ribophilic 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 to protect it from host restriction endonuclease systems. Furthermore, glycosyltransferase may function in the puncturing or lysis of the cell wall peptidoglycan. We were surprised to find that the genome of the bacteriophage IME392 also contains a gene encoding glycosyltransferase, which may prevent it from the removal by the host, but its exact function in the life cycle of bacteriophage IME392 still needs further experimental verification. The *Escherichia coli* bacteriophage 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 glycosyltransferase and tailspike protein helps phage IME392 infect and adsorb the host more easily.

In conclusion, we present here the biology and genomic and proteomic characteristics of *Escherichia coli* phage IME392 isolated from sewage samples taken from a chicken farm in Qingdao, China. The newly isolated phage IME392 was assigned as a member of the *Myoviridae* family. 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 for our discovery and research in nature.

Declarations

Acknowledgments:

This research was supported by Key Project of Beijing University of Chemical Technology(No. XK1803-06), National Key Research and Development Program of China (2018YFA0903000, 2020YFC2005405, 2020YFA0712100, 2020YFC0840805), Funds for First-class Discipline Construction (No. XK1805), Inner Mongolia Key Research and Development Program (NO. 2019ZD006), National Natural Science Foundation of China (81672001, 81621005), NSFC-MFST project (China–Mongolia) (No. 31961143024), Fundamental Research Funds for Central Universities (No. BUCTRC201917, BUCTZY2022).

Conflict of interest

The authors declare that they have no conflict 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, Barbirz 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, Formsma K, 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 Genomics* 9:75
3. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin 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 U S A* 70:3240-3244
8. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ (1961) General nature of the genetic code for proteins. *Nature* 192:1227-1232
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, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe 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* 291:198196
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_ECO078. *Front Microbiol* 8:1460
13. Huang CJ, Lin H, Yang X (2012) Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *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. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbone E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S (2008) Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 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 *Vibrionaceae* 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 *Vil*-like phage specific to *Escherichia coli* O157:H7. *Virology* 430:430
20. Laxminarayan R, Amabile-Cuevas CF, Cars O, Evans T, Heymann DL, Hoffman S, Holmes A, Mendelson M, Sridhar D, Woolhouse M, Røttingen 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 ϕ d replicative form. *Proc Natl Acad Sci U S A* 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* 19:7-49
27. Petrov VM, Nolan JM, Bertrand C, Levy D, Desplats C, Krisch HM, Karam JD (2006) Plasticity of the gene functions for DNA replication in the T4-like phages. *J Mol Biol* 361:46-68
28. Pitout JD (2012) Extraintestinal pathogenic *Escherichia coli*: an update on antimicrobial resistance, laboratory diagnosis and treatment. *Expert Rev Anti Infect Ther* 10:1165-1176
29. Russo TA, Johnson JR (2003) Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes Infect* 5:449-456
30. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage ϕ 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
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-1, 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 ν B_EcoM_JS09 against Clinically Isolated Antibiotic-Resistant Avian Pathogenic *Escherichia coli* and Enterotoxigenic *Escherichia coli*. *Intervirology* 58:218-231

Tables

Table 1 Host range and lytic capability of phage IME392

Bacteria	Strains	Relevant characteristic(s) or source	ST	Spot assay
<i>Escherichia coli</i>				
	94	Laboratory strain collection	1	+++
	108	Clinical isolates	43	+
	109	Clinical isolates	477	-
	156	Laboratory strain collection	169	+
	161	ATCC25922	52	-
	196	Isolated from 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	+++
<i>Salmonella</i>				
	2693	CMCC50001	/	-
	2694	ATCC13311	/	-
	2695	ATCC14028	/	-
	2696	CMCC50115	/	-
	2697	CMCC50094	/	-

+++; Clear plaques; +, plaques with slight turbidity; -, no plaque formed.

Table 2 Functional protein of bacteriophage vB_EcoM_IME392 and homology to proteins databases.

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	Not hit				7.53	6.03
3	962	630	-	333	Not hit				12.63	6.50
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	Not hit				6.83	5.08
7	1899	1642	-	258	Not hit				9.61	6.27
8	2062	1886	-	177	Not 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	norphogenetic protein [Enterobacter cancerogenus]		WP_137271981.1	9.00E-63	23.57	5.67
12	4791	4270	-	522	hypothetical protein [Enterobacter roggenkampii]		WP_095429045.1	2.00E-33	19.64	4.87
13	5510	4803	-	708	putative membrane protein [Escherichia phage adrianh]	Myoviridae	QHR71453.1	4.00E-27	24.89	4.31
14	5740	5510	-	231	hypothetical protein [Salmonella phage barely]	Ackermannviridae	YP_009888578.1	1.00E-17	8.50	7.62
15	6106	5789	-	318	hypothetical protein [Erwinia phage vB_EamM-Bue1]	Ackermannviridae	YP_009837725.1	3.00E-46	12.35	7.63
16	6522	6103	-	420	Inhibitor of host Lon protease [Escherichia phage EcS1]	Myoviridae	BBC78124.1	2.00E-08	15.65	4.72
17	6742	6509	-	234	hypothetical protein [Enterobacter phage Arya]	Myoviridae	YP_009284295.1	4.00E-10	8.82	6.27
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]	Myoviridae	BCG50031.1	7.00E-75	72.06	5.16
20	9817	9059	-	759	DNA polymerase/3'-5' exonuclease PolX [Thermobispora bisporea]		WP_013132370.1	6.00E-26	28.56	5.86
21	9981	9814	-	168	Not hit				6.45	11.07
22	10354	9983	-	372	Not hit				13.82	5.32
23	10473	10366	-	108	Not hit				4.08	9.98
24	10831	10589	-	243	Not hit				9.29	8.85
25	11208	10831	-	378	Not hit				14.32	8.52
26	11359	11198	-	162	Not hit				6.21	6.53
27	11565	11359	-	207	Not hit				7.81	9.66

28	11869	11573	-	297	hypothetical protein [Vibrio phage 1.031.O_10N.261.46.F8]	Myoviridae	AUR83081.1	4.00E-04	11.10	8.06
29	12252	11872	-	381	Not hit				13.92	4.82
30	12573	12262	-	312	Not hit				11.59	9.43
31	13018	12605	-	414	Not hit				14.49	5.11
32	13249	13031	-	219	Not hit				8.49	9.15
33	13634	13260	-	375	Not hit				14.01	5.51
34	13978	13694	-	285	Not hit				10.11	3.95
35	14442	14092	-	351	Not hit				13.85	5.76
36	14728	14423	-	306	Not hit				11.20	9.65
37	15148	14804	-	345	Not hit				12.43	5.62
38	15659	15141	-	519	Not hit				18.97	4.33
39	15952	15653	-	300	Not hit				11.09	4.32
40	16457	15927	-	531	hypothetical protein [Halomonas sp. S2151]		WP_045991529.1	9.00E-19	19.91	9.23
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	Not hit				24.96	6.44
43	19095	18757	-	339	Not hit				12.77	4.26
44	19889	19203	-	687	Not hit				26.59	8.72
45	21418	19904	-	1515	DNA-directed RNA polymerase beta' subunit [Leptospira santarosai]		WP_075917787.1	9.00E-84	54.91	8.61
46	21752	21438	-	315	Not hit				12.30	9.47
47	22083	21739	-	345	Not hit				12.61	4.42
48	22547	22080	-	468	Not hit				17.66	6.51
49	22918	22679	-	240	Not hit				8.63	4.28
50	24533	23073	-	1461	Not hit				54.16	6.89
51	24941	24672	-	270	Not hit				9.66	10.91
52	28209	24922	-	3288	DNA polymerase I [Ralstonia phage RP13]	Myoviridae	BCG50106.1	4.00E-139	125.53	5.03
53	29916	28279	-	1638	DEAD/DEAH box helicase family protein [Ralstonia phage RP13]	Myoviridae	BCG50107.1	4.00E-125	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	Not 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]	Ackermannvirida	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	38.63	6.03
60	35790	34711	-	1080	Not hit				41.14	5.42

61	36872	35838	-	1035	Not 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	38963	-	909	Not hit				33.51	4.83
65	40713	39973	-	741	hypothetical protein [Candidatus Woesearchaeota archaeon]		OIO62389.1	1.00E-11	27.70	5.03
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	Not hit				16.56	8.66
69	44990	44736	-	255	Not hit				9.71	4.81
70	45226	44987	-	240	Not 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
74	51193	50153	-	1041	hypothetical protein [Ralstonia phage RP13]	Myoviridae	BCG50123.1	4.00E-07	39.79	5.90
75	51417	51190	-	228	Not 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	-	1641	DEAD/DEAH box helicase family protein [Ralstonia phage RP13]	Myoviridae	BCG50119.1	6.00E-87	60.65	6.14
79	55300	54782	-	519	Not hit				19.63	5.32
80	55657	55346	-	312	Not hit				11.82	6.13
81	55947	55702	-	246	Not hit				9.16	9.45
82	56254	55937	-	318	Not hit				12.11	6.72
83	57140	56349	-	792	Not hit				31.38	6.15
84	57285	57163	-	123	Not hit				4.62	9.78
	58180	58108	-	73	tRNA-Met-CAT					
	58725	58653	-	73	tRNA-Arg-TCT					
85	60064	59522	-	543	hypothetical protein [Vibrio phage PWH3a-P1]	Myoviridae	YP_007675922.1	8.00E-41	20.84	7.07
86	60343	60182	-	162	Not hit				5.83	3.96
87	60812	60330	-	483	putative ssDNA binding protein [Pectobacterium phage PP99]	Autographiviridae	YP_009788767.1	7.00E-41	18.26	7.63
88	61153	60812	-	342	Not hit				12.31	9.26
89	61350	61823	+	474	hypothetical protein [Ralstonia phage RP13]	Myoviridae	BCG50099.1	3.00E-20	16.87	5.34

90	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
91	63654	65231	+	1578	hypothetical protein [Ralstonia phage RP13]	Myoviridae	BCG50094.1	4.00E- 112	58.86	4.58
92	65224	65478	+	255	Not hit				9.16	4.37
93	65491	66927	+	1437	hypothetical protein [Ralstonia phage RP13]	Myoviridae	BCG50091.1	1.00E- 73	52.33	4.70
94	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
95	68560	69276	+	717	Not hit				25.01	4.40
96	69310	69846	+	537	hypothetical protein [Ralstonia phage RP13]	Myoviridae	BCG50086.1	3.00E- 39	20.16	9.87
97	69846	70133	+	288	Not hit				10.89	11.65
98	70130	70759	+	630	hypothetical protein [Ralstonia phage RP13]	Myoviridae	BCG50082.1	6.00E- 29	23.90	4.78
99	70769	71332	+	564	Not hit				20.53	4.84
100	71334	71999	+	666	hypothetical protein [Chloroflexi bacterium]		RLC62271.1	6.00E- 04	24.65	4.71
101	72001	72192	+	192	Not hit				7.02	6.18
102	72203	74002	+	1800	phage tail sheath domain-containing protein [Pseudoalteromonas citrea]		WP_138594708.1	9.00E- 122	64.39	4.80
103	74013	74537	+	525	hypothetical protein [Pseudoalteromonas citrea]		WP_138594710.1	2.00E- 37	19.20	4.98
104	74552	75028	+	477	hypothetical protein [Pseudoalteromonas]		WP_125251889.1	2.00E- 20	18.43	4.90
105	75038	75865	+	828	hypothetical protein [Pseudoalteromonas sp. McH1-7]		WP_176033179.1	2.00E- 28	31.11	5.11
106	75944	78802	+	2859	glycoside hydrolase [Acinetobacter radioresistens]		WP_138000971.1	1.00E- 39	102.37	5.75
107	78802	79437	+	636	hypothetical protein [Vibrio phage 1.031.O_10N.261.46.F8]	Myoviridae	AUR83149.1	3.00E- 09	23.26	8.86
108	79445	79822	+	378	hypothetical protein [Pseudoalteromonas sp.]		NRA76876.1	1.00E- 13	14.49	4.53
109	79819	80943	+	1125	hypothetical protein [Pseudoalteromonas peptidolytica]		WP_147390429.1	1.00E- 34	41.96	8.64
110	80946	81605	+	660	putative tail protein [Thermus phage phiLo]	Myoviridae	AYJ73937.1	7.00E- 19	24.15	5.48
111	81613	81981	+	369	phage baseplate assembly protein [Thiovulum sp. ES]		EJF07099.1	2.00E- 24	14.09	4.96
112	81978	83351	+	1374	hypothetical protein [Vibrio phage 1.031.O_10N.261.46.F8]	Myoviridae	AUR83140.1	7.00E- 63	50.82	4.77
113	83344	84822	+	1479	Not hit				56.57	4.65
114	84831	85598	+	768	hypothetical protein [Planctomycetes bacterium]		RIK79130.1	5.00E- 10	27.17	4.72
115	85611	86528	+	918	tail spike protein [Salmonella phage moki]	Ackermannviridae	YP_009888731.1	8.00E- 26	33.90	4.45

116	86538	86879	+	342	Not hit			11.70	5.62
117	86881	88227	+	1347	putative tail fibers protein [Salmonella enterica]	EAM1616335.1	1.00E-32	48.01	5.52
118	88227	88655	+	429	tail fiber assembly protein [Escherichia sp. ESNIH1]	WP_103819369.1	2.00E-20	16.11	4.29
119	88667	90118	+	1452	putative phage collar protein [Escherichia coli]	STL39829.1	3.00E-64	52.33	5.59
120	90118	90726	+	609	tail fiber assembly protein [Escherichia coli]	WP_181193511.1	2.00E-50	23.01	4.45
121	90802	91425	+	624	tail fiber protein [Escherichia coli]	HAI7147756.1	1.00E-38	23.10	8.75
122	91505	94492	+	2988	tail fibers protein [Salmonella phage IME207]	Siphoviridae YP_009322784.1	0.0	111.33	5.41
123	94616	97237	+	2622	colanic acid-degrading protein [Cronobacter phage vB_CsaM_GAP32]	Myoviridae YP_006987233.1	1.00E-125	94.96	5.17
124	97262	98713	+	1452	hypothetical protein [Gimesia algae]	QDT94270.1	1.00E-07	52.14	6.16
125	98730	99551	+	822	hypothetical protein [Klebsiella variicola]	WP_135714845.1	6.00E-20	29.48	5.10
126	99563	102607	+	3045	DUF4815 domain-containing protein [Syntrophobacteraceae bacterium]	NSW86778.1	9.00E-145	109.48	4.71
127	102617	102793	+	177	Not hit			6.68	8.19
128	102818	103465	+	648	putative endolysin [Burkholderia phage BcepSaruman]	Myoviridae YP_009904215.1	1.00E-66	24.33	9.20
129	103480	103896	+	417	hypothetical protein [Aeromonas enteropelogenes]	WP_042027513.1	3.00E-11	15.06	5.39
130	103900	104130	+	231	Not hit			8.50	5.34
131	104199	104531	+	333	Not hit			12.70	5.06
132	104826	104584	-	243	Not hit			9.33	4.90
133	105380	104826	-	555	Not hit			20.84	5.02
134	105739	105377	-	363	cell wall hydrolase [Pseudomonas panacis]	WP_154841641.1	2.00E-20	13.86	10.12
135	106148	105771	-	378	Not hit			13.62	4.93
136	106784	106161	-	624	putative dCTP deaminase [Dehalococcoidia bacterium]	MXV43540.1	7.00E-15	23.66	6.51
137	106966	106784	-	183	Not hit			6.85	4.90
138	107522	107052	-	471	dUTP diphosphatase [Shewanella donghaensis]	WP_144212174.1	4.00E-61	16.83	5.09
139	107899	107519	-	381	hypothetical protein [Acidobacteria bacterium]	MBF844480.1	2.00E-35	14.60	7.08
140	108123	107899	-	225	Not hit			8.01	12.02
141	108464	108162	-	303	hypothetical protein [Klebsiella phage Muenster]	Myoviridae QOE32388.1	2.00E-16	11.14	9.12
142	109188	108475	-	714	putative 3'-5' exoribonuclease [Pseudomonas peli]	WP_090248742.1	6.00E-07	27.41	5.08

143	109606	109253	-	354	Not hit			13.31	5.44
144	109838	109599	-	240	hypothetical protein		WP_157760109.1	6.00E-08	10.49
145	110066	109893	-	174	hypothetical protein		WP_001563024.1	1.00E-17	4.34
146	110549	110085	-	465	NADAR family protein [Delftia acidovorans]		NIT77307.1	8.00E-31	8.78
147	111118	110549	-	570	hypothetical protein [Klebsiella phage ZCKP1]	Myoviridae	YP_009803533.1	2.00E-21	5.48
148	111362	111108	-	255	Not hit			9.35	9.21
149	111536	111372	-	165	Not hit			6.24	9.39
150	111729	111577	-	153	Not hit			5.66	3.85
151	111910	111716	-	195	hypothetical protein [Vibrio phage vB_VchM_Kuja]	Ackermannviridae	YP_009854103.1	6.00E-08	4.44
152	112204	111971	-	234	Not hit			8.61	7.96
153	112904	112734	-	171	hypothetical protein [Citrobacter phage Merlin]	Myoviridae	YP_009203833.1	2.00E-04	6.53
154	113256	112897	-	360	Not hit			13.74	5.67
155	113600	113367	-	234	hypothetical protein [Klebsiella phage vB_KpnM_KpS110]	Ackermannviridae	YP_009798919.1	4.00E-08	4.57
156	114141	113611	-	531	LysM peptidoglycan-binding domain-containing protein		WP_094245731.1	6.00E-04	6.82
157	114497	114159	-	339	Not hit			13.19	4.56
158	114724	114497	-	228	Not hit			8.39	6.11
159	115275	114721	-	555	hypothetical protein [Klebsiella oxytoca 10-5244]		KMV90563.1	5.00E-26	5.86
160	115696	115286	-	411	hypothetical protein [Cronobacter phage CR9]	Myoviridae	YP_009015051.1	3.00E-13	6.04
161	115874	115686	-	189	Not hit			6.89	4.04
162	116130	115933	-	198	Not hit			7.21	4.44

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

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 fibers 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 collar 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 fibers 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 fibers protein	6499.410564	464.2436117
72	94.901	7	12	8.71%	putative colanidase 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

Figures

Figure 1

Morphological Features of Bacteriophage vB_EcoM_IME392. IME392 possess an icosahedral head of approximately 83.93 ± 0.55 nm in diameter and a contractile tail of 122.23 ± 3.55 nm in length (n=10).

Figure 2

One-step Growth Curve of Bacteriophage vB_EcoM_IME392. The latent period and burst period of IME392 are both approximately 15 minutes.

Figure 3

The temperature and pH stability of Bacteriophage vB_EcoM_IME392. IME392 is extremely sensitive to heat treatment. Incubating up to 60 °C for 1 hour is able to reduce phage titer by 99.94%, while no phage titer was observed if the incubation temperature reached 70 °C or higher (3A). It is also sensitive to both low and high pH environment. No phage titer was detected when the phage particles were incubated in pH=2.0 or 13.0 environment at 37 °C for 1 hour (3B).

Figure 4

Page 17/18

The genome function map of Bacteriophage vB_EcoM_IME392. The outermost circle represents the gene coding region. The arrows indicate the direction of transcription of each gene. Different colors refer to different functional categories: hypothetical protein (gray), lysis-related protein (yellow), phage morphogenesis protein (blue), replication and regulation-related protein (green). The two innermost circles represent GC-skew $[(G-C) / (G+C)]$ and G+C content. The gray circle represents the G+C content. The outward direction indicates that the G+C content of this region is larger than the average G+C content of the whole genome, and the inward direction indicates that the G+C content of that region is lower than the average G+C content. The black circle represents the GC-skew $[(G-C) / (G+C)]$. The outward direction indicates that the GC-skew $[(G-C) / (G+C)]$ is larger than zero, and the inward direction indicates that the GC-skew $[(G-C) / (G+C)]$ is lower than zero. The scale units are base pairs.

Figure 5

Phylogenetic analysis of phages in different genus of Myoviridae family based on their major capsid proteins (A) and terminase large subunit (B). Phylogenetic trees were constructed using the neighbor-joining method with 1000 bootstrap replications.

Supplementary Files

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

- [SupplementaryTable120210405.doc](#)