Arch Virol (2011) 156:1489–1492
DOI 10.1007/s00705-011-1033-9

ANNOTATED SEQUENCE RECORD

The complete genome sequence of a novel T4-like bacteriophage, IME08

Huanhuan Jiang · Xiaofang Jiang · Sheng Wang · Cun Li · Bin Chen · Xiaoping An · Zhiqiang Mi · Jiankui Chen · Yigang Tong

Received: 4 March 2011 / Accepted: 11 May 2011 / Published online: 26 May 2011

The genus 'T4-like viruses' includes many lytic bacteriophages that infect a wide range of bacteria, including many pathogenic ones. These phages are composed of double-stranded DNA genome held in an icosahedral head attached to a contractile tail with fibers. Sequence analysis of the major head gene (*g23*) has suggested that members of genus 'T4-like viruses' that infect *Escherichia coli* can be divided into four major subgroups, represented by T4, RB69, RB49 and JS98 [1]. To date, there are fewer than 40 T4-like bacteriophages whose whole genome sequences are available [2], and only T4 has been well characterized as model organism in molecular biology [3, 4]. A novel lytic enterobacteria phage, designated IME08, was isolated from hospital sewage with *E. coli* strain 8099 in China in 2008, and its genetic characteristics were described briefly [5]. To further characterize phage IME08 and address the taxonomic status, we sequenced and annotated its complete genome.

The isolation, propagation, purification and titration of bacteriophages were performed as described by Adams [6]. The host range of the phage was initially screened by a spot test and was further confirmed by a plaque assay. The genomic DNA was extracted as described by Sambrook *et al.* [7] and sequenced using the Next-Gen sequencing technique on a Solexa machine at BGI (formerly known as Beijing Genomics Institute). Sequences were assembled using Velvet algorithms [8] with a k-mer of 31. The assembled full-length genome sequence was verified using other assembly software, including ABYSS [9] and Soap [10]. Putative open reading frames within the IME08 full-length sequence were predicted using the software Kodon (Applied Maths, Sint-Martens-Latem, Belgium), with a minimum product size of 50 amino acids, and using the "Bacterial and Plant Plastid Code" as the translation table. The predicted ORFs were then screened against the coding regions of reference sequences in the phage genome database downloaded from EMBL. Coding regions showing the highest homology were used to annotate the IME08 sequence regions. Genomic *tRNA* genes were searched for using tRNAscan-SE (v.1.21) software [11]. The nucleotide distribution was analyzed using the General Codon Usage Analysis program available at http://bioinf.may.ie/gcua/download.html [12].

The complete genome sequence of IME08 consisted of 172,253 nucleotides, with redundant sequences at ends, which is typical for T4-like phage genome [3]. The average G + C content was 39%. Two hundred fifty-three predicted coding sequences (CDSs), along with three *tRNA* genes occupied 92% of the complete genome. About one fifth of the IME08 genes (43 open reading frames [ORFs]) were transcribed in the clockwise direction. There was no difference in the average G + C content between the clockwise- and counterclockwise-transcribed regions. The *tRNA* genes were located in the intergenic regions between ORF129 and ORF130.

Sequences homologous to the essential replication-related genes of phage T4 were identified throughout the genome of IME08 (Fig. 1). These included the following major gene categories: (1) the replisome genes: DNA

H. Jiang and X. Jiang contributed equally to this work.

H. Jiang · X. Jiang · S. Wang · C. Li · B. Chen · X. An · Z. Mi · Y. Tong (✉)
State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China
e-mail: tong.yigang@gmail.com

J. Chen (✉)
Clinical Laboratory, No. 307 Hospital, Beijing, China
e-mail: chenjiankui@hotmail.com

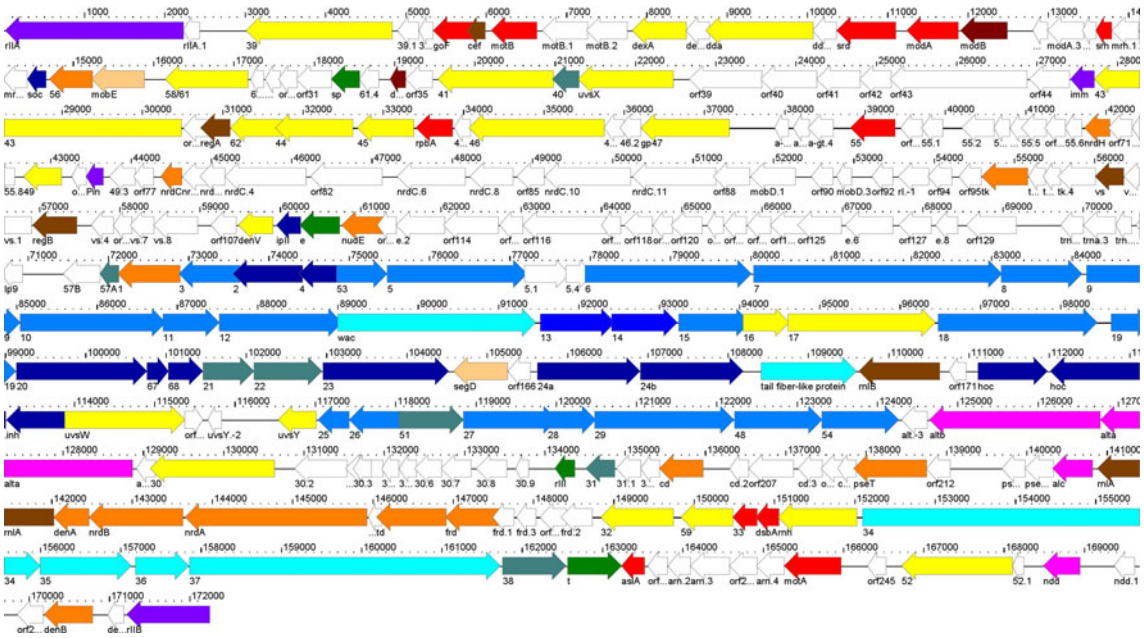

Fig. 1 Bacteriophage IME08 genome. The color scheme (by gene function) is defined as follows: red, transcription; brown, translation; orange, nucleotide metabolism; yellow, DNA replication; dark blue, head; medium blue, neck; light blue, tail; pale blue, tail fiber; pale green, chaperonins, catalysts; green, lysis; purple, host and phage interactions; pink, host alteration/shutoff; peach, homing endonucleases and homologs; white, unknown function

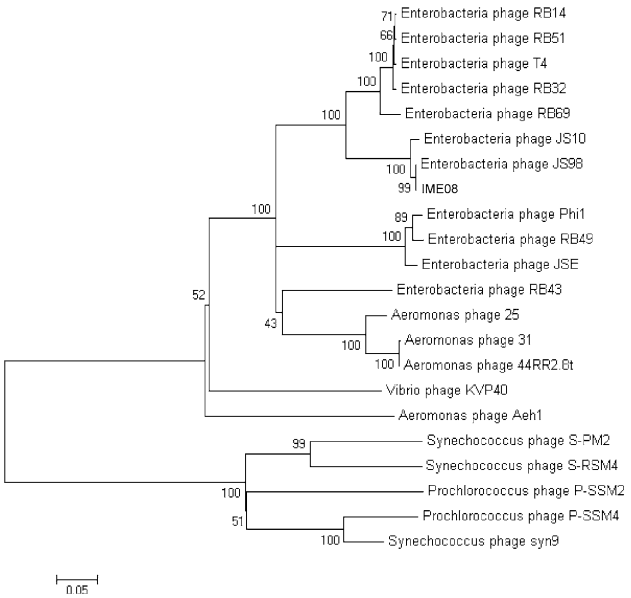

Fig. 2 Neighbor-joining tree analysis based on the alignment of the amino acid sequence of gp23 from T4-type bacteriophages available from the NCBI database. The numbers at the nodes indicate the bootstrap probabilities

polymerase (ORF46), sliding clamp loaders (ORF49 and ORF50), sliding clamp (ORF51), DNA helicase (ORF36), primase for lagging-stand synthesis (ORF27) and ssDNA binding protein (ORF227). (2) The accessory genes for DNA replication: DNA ligase (ORF191), RNase H (ORF231), DNA helicase loader (ORF228) and the topoisomerase II component (gp39 large subunit, ORF3; gp52 medium subunits, ORF246). (3) Join-copy recombination-replication genes: recombination endonuclease (gp46, ORF54; gp47, ORF57), UvsX (ORF38), UvsY (ORF175) and UvsW (ORF175). (4) Join-cut-copy recombination-replication genes: endoVII (ORF73) and terminase (gp16 small subunits, ORF155; gp17 large subunits, ORF156). Phage genomes commonly encode enzymes for nucleotide salvage and modification. In IME08, a major gene cluster for nucleotide metabolism was identified that contained homologues to T4 endonuclease II (ORF217), aerobic NDP reductase (ORF218 and ORF219), thymidylate synthase (ORF220) and dihydrofolate reductase (ORF221). Other genes were scattered throughout the genome, including pyrophosphates (ORF25), glutaredoxin (ORF70), thioredoxin (ORF78), thymidine kinase (ORF96), nudix hydrolase (ORF111), dNMP kinase (ORF136), dCMP deaminase (ORF205), PseT (ORF211) and endonuclease IV (ORF251). These genes were highly similar to those of phage T4, with only a few minor distinctions (i.e., the absence of homologues of *nrdD*, *nrdG* and gene *42* in the IME08 genome). Enzymes for RNA repair and modification were also encoded by the IME08 genome, including two RNA ligases (RnlB, ORF170, and RnlA, ORF216), modifier of suppressor tRNA (Cef, ORF7), valyl-tRNA synthetase modifier (Vs, ORF100) and ADP-ribosylase (ModB, ORF17). Of the genes involved in host-phage

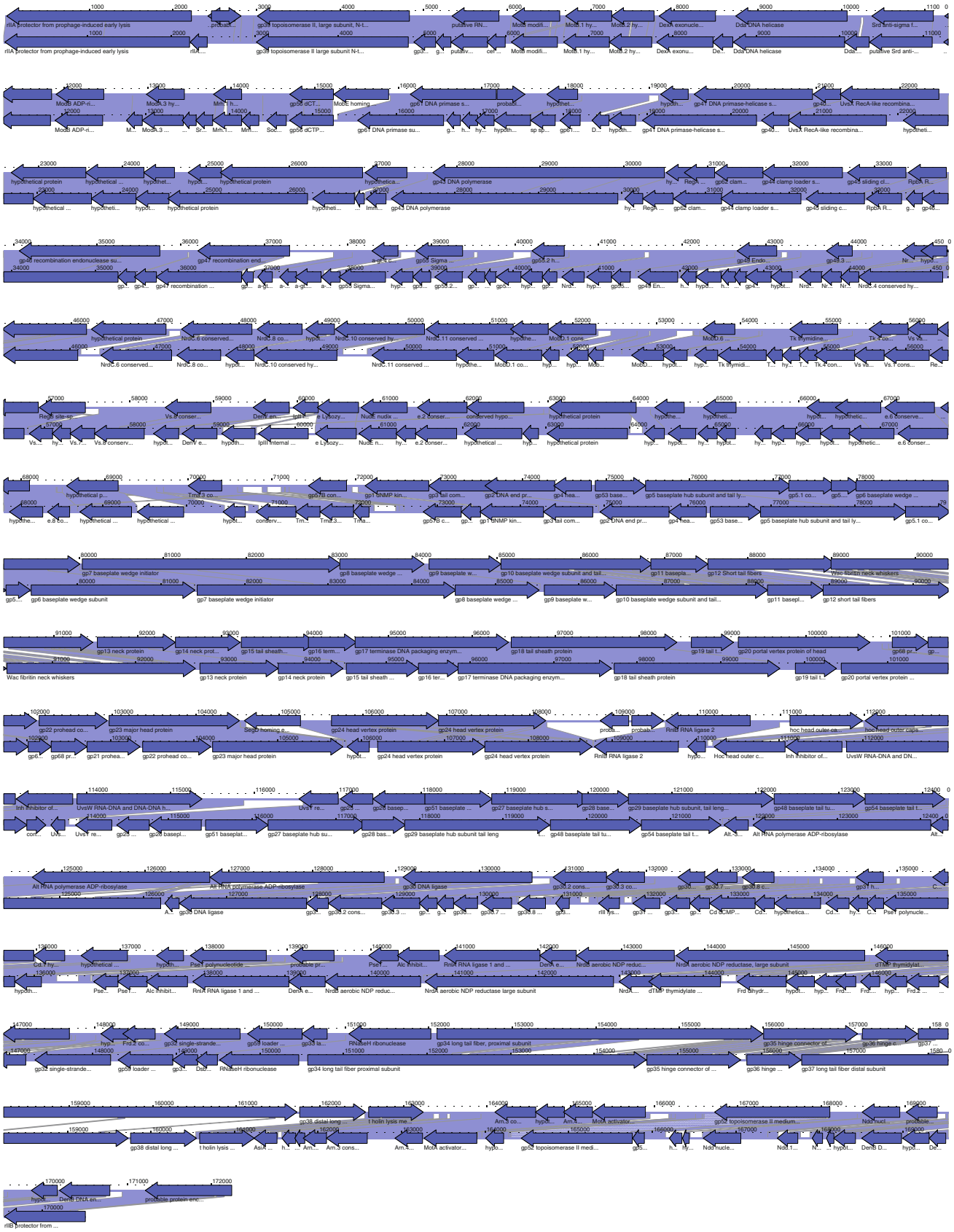

Fig. 3 Comparative illustration of the IME08 and JS98 genomes. The upper genome is the IME08 genome and the lower one is the JS98 genome

interaction, the *rIIA* (*orf*1), *imm* (*orf 45*), *pin* (*orf 75*), *rIIB* (*orf 253*), *alc* (*orf 215*) and *ndd* (*orf 248*) were identified in the IME08 genome, and the *alt* genes (*altb, orf188*; and *alta*, *orf189*) were duplicated. While some other genes, such as *dam* (modifier of DNA adenine in a particular sequence), α- and *β*-glucosyl-transferase genes (modifier of HMC residues) and the *arn* genes (anti-restriction endonucleases) were absent.

The structural genes and the genes involved in the assembly process resided mainly in the largest gene cluster. Although the gene order in this region was almost identical between IME08 and phage T4, there were several differences. In IME08, gene *24* (*g24*), which encodes the head vertex protein, was duplicated (*24a*, *orf167*; *24b*, *orf168)*, and the head outer capsid protein gene was also duplicated. The two opposite ORFs (*orf172*, clockwise; *orf173*, counterclockwise) were found to encode immunoglobulin-like domains. The rest of the structural genes, which are involved in tail fiber morphogenesis, including genes *34* (*orf232*), *35* (*orf233*), *36* (*orf234*), *37* (*orf235*) and *38* (*orf236*), were located near the end of the genome. Gene *37* (encoding the long tail fiber protein) and gene *38* (encoding the assembly catalyst of the distal tail fiber) determine the host range of the phage and are thus of interest for bacteriophage therapy research.

To address the evolutionary status of IME08 among the T4-type phages, we performed phylogenetic analysis of gp23, the typical phylogenetic marker for T4-type phages. The amino acid sequences of the major head protein gp23 from IME08 and T4-type phages (retrieved from the NCBI database) were aligned, and a phylogenetic tree was constructed using MEGA4 software [13]. IME08 phage was shown to be most closely related to phage JS98, which is classified in the JS98 subgroup of T4-like phages of *Escherichia coli* (Figs. 2, 3). However, for gene *37* of IME08, the nucleotide sequence and the deduced amino acid sequence (of 1290 amino acids) showed highest identity to phage T2 (69% and 59%, respectively), whilst sharing only 46% and 25% identity with phage JS98. Like *gp37* of T2 and K3, *gp37* of IME08 did not contain the tripeptide repeats His-Ser-His or His-Thr-His, which are believed to be determinants of host recognition [14]. Gene *38* of IME08 showed 80% nucleotide sequence identity to its counterpart in phage T2. The deduced amino acid sequence of gp38 (262 amino acids) showed the highest identity with *gp38* of phage T2 (88%) whilst sharing only 49% identity with JS98. The gp38 sequence showed a high frequency of alanine (10.62%) and an extremely high frequency of glycine (22.34%), which was coincident with gp38 of phages T2 and K3 [15]. The high degree of identity between IME08 and T2 in these two tail fiber genes indicated that IME08 and T2 may undergo genetic exchange.

The sequence data for phage IME08 were deposited into the GenBank database under accession number HM071924.

References

1. Zuber S, Ngom-Bru C, Barretto C, Bruttin A, Brussow H, Denou E (2007) Genome analysis of phage JS98 defines a fourth major subgroup of T4-like phages in *Escherichia coli*. J Bacteriol 189(22):8206–8214
2. http://phage.ggc.edu/
3. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W (2003) Bacteriophage T4 genome. Microbiol Mol Biol Rev 67(1):86–156
4. Karam J (1994) Molecular biology of bacteriophage T4. ASM Press, Washington, DC
5. Wang S, Jiang H, Chen J, Liu D, Li C, Pan B, An X, Zhang X, Zhou Y, Tong Y (2010) Isolation and rapid genetic characterization of a novel T4-like bacteriophage. J Med Coll PLA 25:331–340
6. Adams MH (1959) Bacteriophages. Interscience Publishers, Inc, New York
7. Sambrook J, Fritsch EF, Maniatis T (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York
8. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821–829
9. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19(6):1117–1123
10. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinformatics 24(5):713–714
11. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955–964
12. McInerney JO (1998) GCUA: general codon usage analysis. Bioinformatics 14(4):372–373
13. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599
14. Riede I, Drexler K, Eschbach ML, Henning U (1986) DNA sequence of the tail fiber genes 37, encoding the receptor recognizing part of the fiber, of bacteriophages T2 and K3. J Mol Biol 191(2):255–266
15. Riede I, Drexler K, Eschbach ML, Henning U (1987) DNA sequence of genes 38 encoding a receptor-recognizing protein of bacteriophages T2, K3 and of K3 host range mutants. J Mol Biol 194(1):31–39