GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Bioinformatic analysis of phage AB3, a phiKMV-like virus infecting *Acinetobacter baumannii*

J. Zhang[1]*, X. Liu[1]* and X.-J. Li[2]

[1]Department of Geriatrics Medicine,
The Third People's Hospital of Chongqing, Chongqing, China
[2]Department of General Internal Medicine, The First Brand,
The First Affiliated Hospital of Chongqing Medical University,
Chongqing, China

*These authors contributed equally to this study.
Corresponding author: X.-J. Li
E-mail: lixiujuancq@gmail.com

Genet. Mol. Res. 14 (1): 190-198 (2015)
Received March 10, 2014
Accepted August 4, 2014
Published January 16, 2015
DOI http://dx.doi.org/10.4238/2015.January.16.2

ABSTRACT. The phages of *Acinetobacter baumannii* has drawn increasing attention because of the multi-drug resistance of *A. baumanni*. The aim of this study was to sequence *Acinetobacter baumannii* phage AB3 and conduct bioinformatic analysis to lay a foundation for genome remodeling and phage therapy. We isolated and sequenced *A. baumannii* phage AB3 and attempted to annotate and analyze its genome. The results showed that the genome is a double-stranded DNA with a total length of 31,185 base pairs (bp) and 97 open reading frames greater than 100 bp. The genome includes 28 predicted genes, of which 24 are homologous to phage AB1. The entire coding sequence is located on the negative strand, representing 90.8% of the total length. The G+C mol% was 39.18%, without areas of high G+C content over 200 bp in

length. No GC island, tRNA gene, or repeated sequence was identified. Gene lengths were 120-3099 bp, with an average of 1011 bp. Six genes were found to be greater than 2000 bp in length. Genomic alignment and phylogenetic analysis of the RNA polymerase gene showed that similar to phage AB1, phage AB3 is a phiKMV-like virus in the T7 phage family.

Key words: *Acinetobacter baumannii*; Phage; Bioinformatic analysis

INTRODUCTION

Acinetobacter baumannii has been increasingly detected worldwide and has become an important opportunistic pathogen. The ever-growing resistance of *A. baumannii* to antimicrobial agents and the occasional outbreak of multi-drug-resistant strains pose a challenge to clinical treatment. A phage is a virus that infects and kills bacteria. Extensive efforts have been made to use phages to treat infectious diseases as clinical drug-resistant bacteria are becoming an increasing problem (Gupta and Prasad, 2011). However, additional studies on phages are necessary because of the extremely large population and diversity of phages. In this study, we isolated and sequenced *A. baumannii* phage AB3 and carried out preliminary annotation and analysis of its genome.

MATERIAL AND METHODS

Isolation and sequencing of the phage AB3 genome

Forty strains of *A. baumannii* suspension in the early logarithmic growth phase and 50 mL LB culture medium were added to 1000 mL wastewater collected from the wastewater treatment center in our hospital and cultured at 37°C for 24 h. Bacteria were removed by filtration after centrifugation. Polyethylene glycol 8000 was added to the supernatant and the resulting mixture was placed in an ice bath overnight. Subsequently, the solution was centrifuged at 10,000 *g* for 10 min at 4°C; pellets were resuspended in 1 mL phage buffer. Next, 10 µL processed sewage water was mixed with 200 µL host bacteria suspension for 20 min before being uniformly seeded onto a culture dish containing 2 mL top agar. Homogeneous phages were obtained from single plaques after 5 rounds of purification. Phage AB3 showing the broadest spectrum of bacteria lysis was chosen for further studies.

Next, 20 µL phage AB3 suspension was added dropwise onto a copper mesh grid; after 20 min of natural precipitation, 1 drop of 2% phosphotungstic acid was added to the copper mesh grid to stain the phage for 8 min. After drying, phages were observed using a transmission electron microscope (Hitachi H-7500, Tokyo, Japan) (Figure 1). The phage AB3 genome was sequenced by BGI (Shenzhen, China) using the shotgun technique (Dong et al., 2001) and the sequencing results were uploaded to GenBank (accession No. KC311669).

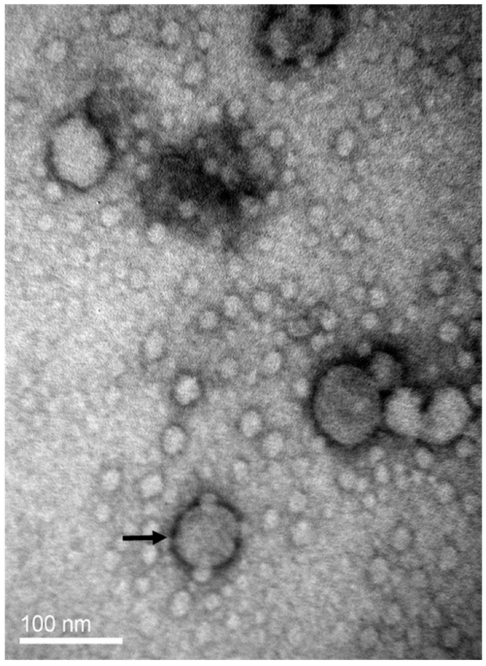

Figure 1. Phage AB3 visualized by transmission electron microscopy (magnification 200,000X). Phage AB3 was negatively stained with 2% phosphotungstic acid. The white arrow indicates phage AB3.

Analysis of phage AB3 genetic composition

Genetic composition analysis was conducted using EditSeq contained in the Lasergene package. tRNA genes in the phage genome were searched using the tRNAscan-SE program developed by the University of Washington (Lowe and Eddy, 1997). Tandem repeats were predicted using Tandem Repeat Finder. Transcription terminators in the phage AB3 genome were predicted using the FindTerm program available from Softberry and promoters in the phage AB3 genome were predicted using the BPROM program on the Softberry website (linux1.softberry.com/berry.phtml). The GC profile in the phage AB3 genome was analyzed using the CpGPlot/CpGReport/Isochore (Rice et al., 2000) functions in the EMBOSS software suite on the website of the European Bioinformatics Institute (www.ebi.ac.uk/emboss/cpgplot/).

Prediction of phage AB3 coding genes

The open reading frames (ORFs) of the phage AB3 genome were analyzed using the ORF Finder program available on the NCBI website (ncbi.nlm.nih.gov/gorf/gorf.html) of the U.S. National Institutes of Health (Rombel et al., 2002). The full-length genomic sequence measuring 31,185 base pairs (bp) was input in a FASTA format and 11 (Bacterial code) under Genetic codes was selected. Nucleic acid and protein sequence alignments for predicted genes were performed individually using the BLAST program (Altschul et al., 1997). Coding sequence analysis was performed using the GeneMark program (Besemer and Borodovsky, 1999) available on the Mark Borodovsky Laboratory website (en.wikipedia.org/wiki/Mark_

Borodovsky) and the results were validated using the Glimmer program. The average length of putative genes and the density of genes in the phage AB3 genome were analyzed using the EditSeq program.

Phylogenetic analysis of the phage AB3 RNA polymerase gene

The phage AB3 RNA polymerase gene was aligned with its homologous genes. Multiple sequence alignment for the RNA polymerase gene was performed using the ClustalX (Jeanmougin et al., 1998) and a phylogenetic tree was constructed using Phylip (Retief, 2000).

CoreGenes analysis of phage AB3 and phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2

CoreGenes (Zafar et al., 2002) is a software tool designed for detecting common core genes across 2-5 small genomes (< 300 kb). This program primarily uses BLASTP to analyze the results. Using phage AB3 as the reference, phages AB1 (European Association for the Study of the Liver, 2012), AP22, YMC/09/02/B1251 ABA BP (Jeon et al., 2012), and IME-AB2 were queried.

Genomic alignment of phage AB3 with phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2

Mauve (Darling et al., 2004) is a software program used to align conservative genome sequences. Using Mauve, the genomic sequence of phage AB3 was aligned with the genomes of phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2.

RESULTS

Sequence assembly and analysis of the phage AB3 genome

Phage AB3 (Figure 1) was amenable to DNA restriction endonuclease digestion in pilot tests, as it could be completely digested with DNase I; thus, the phage AB3 genome was found to be a double-stranded DNA. The software assembly of the sequencing results revealed that the total genome length was 31,185 bp with a G+C mol% of 39.18%. The CpGPlot program was used to determine G+C content using the default settings of EMBOSS from the European Bioinformatics Institute. The overall G+C content was relatively uniform, without GC islands or notably high G+C content areas greater than 200 bp. Typically, genes with significant higher-than-average G+C content may be exogenous genes produced through horizontal transfer. No tRNA gene was detected using tRNAscan-SE and no tandem repeat was detected using Tandem Repeat Finder.

Predicted encoding genes of phage AB3

Six-frame analysis using the ORF Finder program revealed a total of 97 ORFs greater than 100 bp in length, with most located on the negative strand; this finding was consistent with those of coding sequence analysis. However, gene-coding sequence characteristics alone

do not verify the presence of a real gene. Analysis of the phage AB3 genome using GeneMark identified a total of 28 possible coding sequences, corroborating the results obtained using the Glimmer program.

Analysis of phage AB3 gene length and gene density

The longest gene in the genome was orf5 (3099 bp) and the shortest gene was orf23 (120 bp). There were 6 genes greater than 2000 bp in length. No encoding gene was identified in a fragment spanning 1477 bp at the 5' end of the genome. The analysis showed that the entire coding sequence was located on the negative strand, representing 90.8% of the whole genomic sequence.

Analysis of the phage AB3 promoter and transcription terminator

Using the FindTerm program to predict the genome transcription terminator, we identified only 1 potential transcriptional terminator site independent of the ρ-factor. The BPROM program was used to predict promoters in the genome and, given the positions of gene sequences in the phage, 4 possible promoter sequences were found. Promoter positions were 2875, 15226, 29591, and 30501, which may correspond to genes orf03, orf09, orf26, and orf27, respectively.

Phylogenetic analysis of the phage AB3 RNA polymerase gene

The RNA polymerase gene of phage AB3 was aligned with its 10 homologous genes and a phylogenetic tree was created using Phylip. The RNA polymerase gene of phage AB3 was the most similar to that of phage phiAB1 (i.e., phage AB1) (Figure 2).

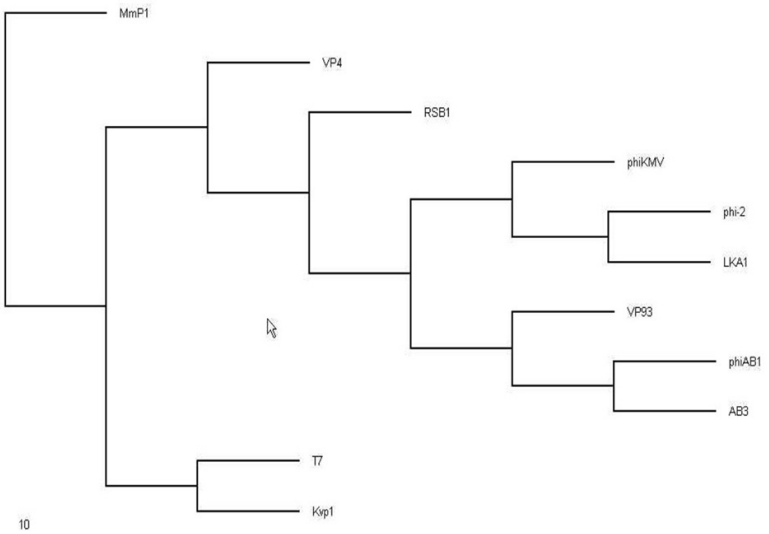

Figure 2. Phylogenetic tree of phage AB3 RNA polymerase gene.

CoreGenes analysis of phage AB3 and phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2

When the BLASTP threshold score was set at 75 (high homology), there are only 1-2 core genes across the several genomes. This finding illustrates the diversity of individual phages and shows that considerable divergence in homology has occurred during evolution, although they are the same-host phages. However, when the score was set to 30 (low homology), different results were obtained. Phage AB3 shares 21-24 common genes with phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2. Although the score was relatively low, this finding also demonstrates that these phages have a high degree of similarity and may belong to the same family classification and have a common evolutionary ancestor. Phage diversity is also manifested in *A. baumannii* phages (Figure 3).

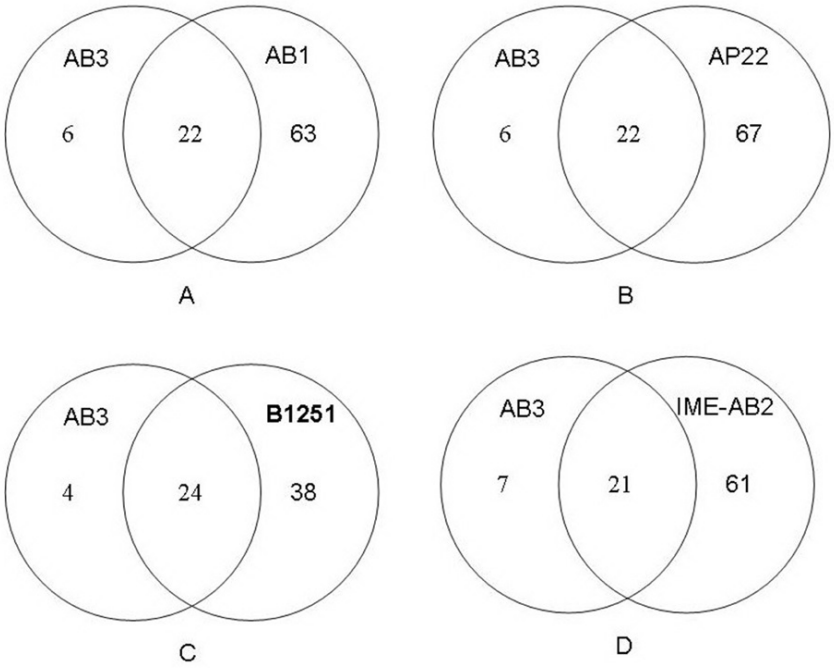

Figure 3. CoreGenes analysis of the phage AB3 genome.

Genomic alignment of phage AB3 with phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2

Phage AB3 has the shortest genomic sequence among the phages examined. Highly homologous sequences in phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2 are located at the 5' end of the AB3 and AB1 genomes and at the 3' end in the phages AP22, YMC/09/02/B1251 ABA BP, and IME-AB2 genomes. Homology alignment of 5 *A. baumannii* phages revealed few similarities, suggesting great diversity among the 5 phages although they all target the same host (Figure 4).

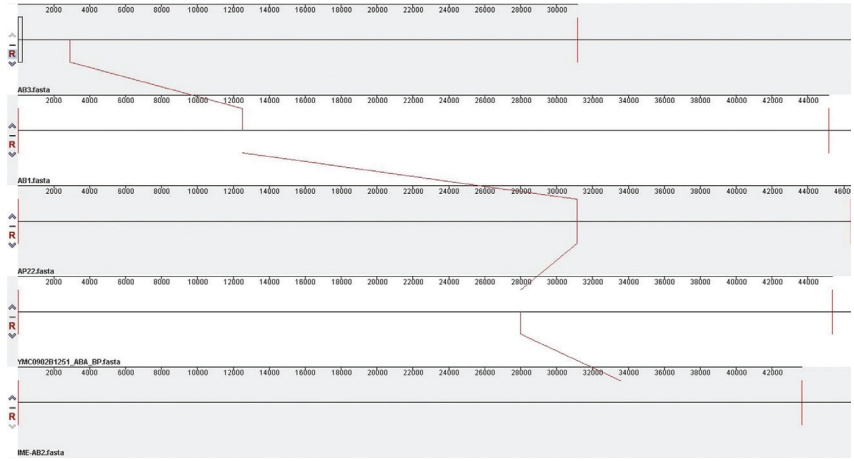

Figure 4. Genomic alignment of phage AB3 with phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2.

Phage AB3 was further aligned with its closest match, phage AB1. The figure suggests that 2 phages have many similarities. However, the size of genomic regions and nucleotide arrangement in the genomic loci with shared similarity differ in the 2 phages. This finding shows that the 2 phage genomes share high similarity and may be derived from a common ancestor (Figure 5).

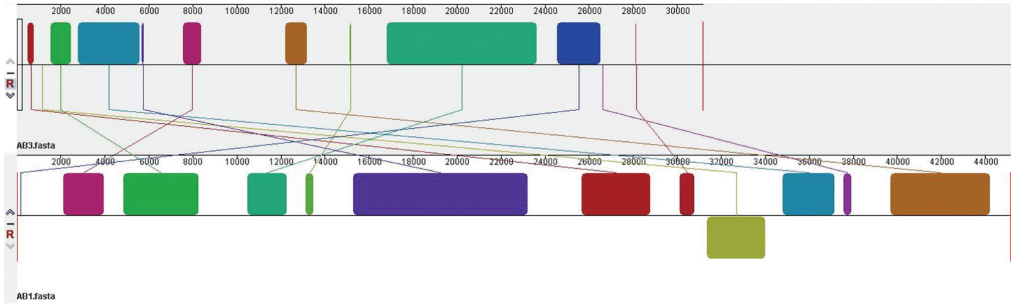

Figure 5. Genomic alignment of phage AB3 and phage AB1.

DISCUSSION

The phage AB3 genome consists of 31,185 bp, which is smaller than the 3 strains of *A. baumannii* phages that were sequenced previously (European Association for the Study of the Liver, 2012). The phage AB3 genome contains a large number of sequences that are similar to phage AB1. The G+C mol% is 39.18%, which is higher than in the 3 strains of *A. baumannii* phages and is most similar to that of phage AB1. Comparisons of core genes between phage AB3 and phages AB1, AP22, YMC/09/02/B1251 ABA BP, and IME-AB2 suggest that these phages belong to the same family, but diverged during evolution. For example, only 1 tRNA gene was identified in phage AB1, while no tRNA gene was detected in phage AB3. In addition, there was a single negative-strand RNA virus bacteriophage AP205 of *A. baumannii* (Klovins et al., 2002). The genome size and structure of phage AP205 differ dramatically from

the above strains of phages, suggesting rich diversity in the phages.

An ORF is a prerequisite for a DNA fragment to be considered a gene-coding sequence. Six-frame analyses of phage AB3 with ORF Finder (Rombel et al., 2002) revealed a total of 97 ORFs greater than 100 bp, most of which were located on the negative strand. This finding is consistent with those of coding sequence analysis. However, gene sequence characteristics do not necessarily verify that an ORF is a true gene. We then used GeneMark for genomic analysis of phage AB3 and identified 28 potential coding sequences, corroborating the results of Glimmer analysis. Phage AB3 has a relatively short genomic sequence, resulting in significantly fewer encoding genes than the other 4 strains of phages targeting the same host. Same-host bacteria can have phages of different sizes, which is commonly observed in *Escherichia coli* and *Pseudomonas aeruginosa* phages.

The length of genes in the phage AB3 genome were found to range from 120 bp (orf23) to 3099 bp (orf5), with an average length of 1011 bp, which is largely consistent with the other 4 phages. Of the 28 predicted genes, 24 share homology with phage AB1, including 20 with function-assigned genes. This finding suggests that phage AB3 may have a common ancestor with phage AB1, or may be derived from the latter. Both phage AB3 and phage AB1 are in the same taxonomic group. This view is further supported by phylogenetic analysis of the RNA polymerase gene. Regarding gene length distribution, 6 genes were greater than 2000 bp in the phage AB3 genome. These 6 genes encode RNA polymerase, DNA polymerase, and structural proteins.

The FindTerm program was used to identify potential transcription terminators in the phage AB3 genome and identified only 1 transcriptional terminator site independent of the ρ-factor. Phage AB3 genes are arranged in an orderly fashion, with early, intermediate, and late genes largely arranged in succession. The putative transcription terminator is located in a critical position, which is the point of division between intermediate genes and late genes in phage AB3. The 5' end of the transcription terminator contains the late genes, most of which are phage structural proteins. The 3' end of the transcription terminator contains early and intermediate genes, which encode non-structural phage proteins. Prediction of promoters is much more difficult than that of transcription terminators. Using the BPROM program to identify potential promoters in the phage AB3 genome, 4 possible promoter sequences were identified, accounting for only part of the promoter repertoire of phage AB3. Further studies will be conducted to identify additional promoters in the phage.

Homology analysis revealed that phage AB1 is highly correlated with phage AB3. However, the size of the genomic regions and nucleotide arrangement in the genomic loci with shared similarity differed for the 2 phages. This finding indicates that the 2 phage genomes share high similarity and may be derived from a common ancestor, but diverged considerably during evolution. The RNA polymerase gene is the gold standard for classification of the T7 phage family. We aligned the RNA polymerase gene of phage AB3 with its 10 correlated homologous genes using ClustalX and constructed a phylogenetic tree using Phylip based on maximum likelihood distances. The results showed that the RNA polymerase of phage AB3 is highly homologous to that of phage AB1 and that phage AB3 may taxonomically be a phiKMV-like virus in the T7 phage family (Lee et al., 2004).

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of

protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
Besemer J and Borodovsky M (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res.* 27: 3911-3920.
Darling AC, Mau B, Blattner FR and Perna NT (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14: 1394-1403.
Dong H, Lin W, Zhang CK, Xiong H, et al. (2001). Genomic sequence and expression analyses of human chromatin assembly factor 1 p150 gene. *Gene* 264: 187-196.
European Association for the Study of the Liver (2012). EASL clinical practice guidelines: management of chronic hepatitis B virus infection. *J. Hepatol.* 57: 167-185.
Gupta R and Prasad Y (2011). Efficacy of polyvalent bacteriophage P-27/HP to control multidrug resistant *Staphylococcus aureus* associated with human infections. *Curr. Microbiol.* 62: 255-260.
Jeanmougin F, Thompson JD, Gouy M, Higgins DG, et al. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23: 403-405.
Jeon J, Kim JW, Yong D, Lee K, et al. (2012). Complete genome sequence of the podoviral bacteriophage YMC/09/02/B1251 ABA BP, which causes the lysis of an OXA-23-producing carbapenem-resistant *Acinetobacter baumannii* isolate from a septic patient. *J. Virol.* 86: 12437-12438.
Klovins J, Overbeek GP, van den Worm SH, Ackermann HW, et al. (2002). Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J. Gen. Virol.* 83: 1523-1533.
Lee K, Kim YA, Park YJ, Lee HS, et al. (2004). Increasing prevalence of vancomycin-resistant enterococci, and cefoxitin-, imipenem- and fluoroquinolone-resistant gram-negative bacilli: a KONSAR study in 2002. *Yonsei Med. J.* 45: 598-608.
Lowe TM and Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955-964.
Retief JD (2000). Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* 132: 243-258.
Rice P, Longden I and Bleasby A (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16: 276-277.
Rombel IT, Sykes KF, Rayner S and Johnston SA (2002). ORF-FINDER: a vector for high-throughput gene identification. *Gene* 282: 33-41.
Zafar N, Mazumder R and Seto D (2002). CoreGenes: a computational tool for identifying and cataloging "core" genes in a set of small genomes. *BMC Bioinformatics* 3: 12.