Genomic analysis of a phage and prophage from a *Bacillus thuringiensis* strain

Yihui Yuan, Meiying Gao, Qin Peng, Dandan Wu, Pengming Liu and Yan Wu

Key Laboratory of Agricultural and Environmental Microbiology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, PR China

INTRODUCTION

Bacteriophages (phages) are the most abundant biological entities on Earth, with an estimated abundance about 10-fold that of microbes (Hobbie et al., 1977; Rohwer, 2003; Rohwer & Barott, 2013; Weinbauer, 2004). Phages play an important role in the process of global biogeochemistry (Falkowski et al., 2008) and facilitate horizontal gene transfer by moving between different environments (Biers et al., 2008). Many methods have been used to discover and analyse phages and prophages in the biosphere, such as isolation of phages that can infect specific strains and the use of metagenomic methods to analyse non-cultured phages in diverse environments (Bhattacharya et al., 2013; Breitbart et al., 2002, 2003; Edwards & Rohwer, 2005). With progress in whole-genome sequencing technology, more and more phage and prophage genome sequences have been analysed, greatly increasing our understanding of the genetic origin and diversity of phages (El-Arabi et al., 2013; Jalasvuori et al., 2013). To understand the interaction between the phage and its host bacterium, knowledge of the genome sequence of the host strain is also important (Bishop-Lilly et al., 2012).

Microbes of the *Bacillus cereus* group, which is a subdivision of the *Bacillus* genus, include species of *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. weihenstephanensis*, *B. mycoides* and *B. pseudomycooides* (Vilas-Bôas et al., 2007). Several species of the *B. cereus* group are considered pathogens of humans and insects, such as *B. anthracis* which causes anthrax and *B. cereus* which is a pathogen in food contaminants (Jensen et al., 2003; Schnefp et al., 1998). As a new strategy to control pathogenic bacteria by using phages and the phage-encoded endolysins (Fischetti, 2008; Young et al., 2000), numerous phages isolated from the *B. cereus* group have been studied in detail and most of these phages were found to be members of the family *Siphoviridae* (El-Arabi et al., 2013; Klumpp et al., 2010; Liao et al., 2008; Minakhin et al., 2005; Schuch & Fischetti, 2006; Swanson et al., 2012). These studies were mainly...
focused on the genome of these phages, but a few studies on the genetic diversity and the distribution of the Bacillus phages have been reported to date (Jalasvuori et al., 2013).

Among the phages infecting strains of the B. cereus group, eight phages exhibiting high genetic similarity have been reported. These are the phages Gamma, Cherry, Wβ and Fah, which infect strains of B. anthracis and B. cereus (Fouts et al., 2006; Minakhin et al., 2005; Schuch & Fischetti, 2006); phiS3501 and BtCS33, which infect strains of B. thuringiensis (Moumen et al., 2012; Yuan et al., 2012b); and BceA1 and SpaA1, which infect strains of B. thuringiensis and B. cereus. Phage SpaA1 can also infect strains of Staphylococcus pasteuri (Swanson et al., 2012). According to the studies of Swanson et al. (2012), phage BcSA1 and SpaA1 harboured the whole genome of the B. thuringiensis phage MZTP02 and partially resembled the prophage in B. thuringiensis and B. cereus (Liao et al., 2008; Swanson et al., 2012). The results of these studies suggested that Bacillus phages with similar genome sequences might exist in a variety of environments.

B. thuringiensis strain YM-03 showed high toxicity to insects in the order Coleoptera (Gao et al., 2008). In this study, we characterized the genomes of a family Siphoviridae phage phiCM3 from the B. thuringiensis strain YM-03 and a prophage proCM3 in the YM-03 genome. The insertion site of proCM3 and the putative core sequence of the attachment site were also determined.

RESULTS

Isolation and host range of phage phiCM3

A phage designated phiCM3 was isolated from a plate of B. thuringiensis subsp. morrisoni strain YM-03 grown in our laboratory and transmission electron microscopy observation showed that phiCM3 was a family Siphoviridae phage (data not shown). The host range of phiCM3 was determined on 50 B. thuringiensis strains, two B. anthracis strains, one B. cereus strain, one Bacillus subtilis strain and one Bacillus pumilus (Peng et al., 2013) strain (Table S1, available in the online Supplementary Material). Seven B. thuringiensis among the 54 tested strains were infected by phiCM3. The seven B. thuringiensis strains belonged to serotypes of morrisonis, israelensis, huazhongensis, chanpaisis, zhaodongensis, pulsiensis and graciosoensis, respectively. However, B. thuringiensis strain CS33, which was the host strain of phage BtCS33 isolated by our laboratory (Yuan et al., 2012b), was not infected by phiCM3. The infective activity of phage BtCS33 was also tested on YM-03 and no visible plaques were observed. These data suggest that phiCM3 and BtCS33 have completely different host ranges.

Overview of the phiCM3 genome

The genome of phiCM3 was linear with 3′-cohesive ends. The genome size of phiCM3 was 38.8 kb in length and the GC content was 35.46 %, about the same as that of the genome of the B. thuringiensis strain BMB171 (He et al., 2010). Overall, 56 putative coding sequences (CDSs) were found and the functions of the proteins encoded by 30 CDSs were predicted by searching against the NR database in GenBank, the Pfam database and the CDD database, while the functions of the other 26 CDSs are unknown (Fig. 1a; Table S2). Most of the CDSs were transcribed in the forward direction, but three of the CDSs were transcribed in reverse (Fig. 1a). Ten of the 56 CDSs overlapped with the corresponding preceding CDSs. The CDSs in the phiCM3 genome formed a modular genomic structure, composed of the late region (which included genes encoding the structural, host lysis and terminase proteins), the lysogeny-lysis control region (which included the transcription regulator encoding genes) and the early region (which included the DNA replication protein- and integrase-encoding genes). Two genes in the phiCM3 genome were predicted to encode the cell division FtsK/ SpoIIIIE ATPase (phiCM3_26) and the σ70 family sigma factor (phiCM3_40), which were also found in the genome of the other phages with high similarity (Schuch & Fischetti, 2009; Yuan et al., 2012b). The host lysis gene cluster of phiCM3 contained three genes, phiCM3_20 and phiCM3_21, encoding holin, and phiCM3_22, encoding endolysin.

Overview of the proCM3 genome

By draft genome sequencing of the B. thuringiensis strain YM-03, a prophage, designated proCM3, of about 37 kb in genome length with 58 putative CDSs was detected in the genome of YM-03. Among the 58 predicted CDSs in the proCM3 genome, 52 CDSs were similar to the genes of the Bacillus phages TP21-L and BMBtp2 (Fig. 1b; Table S3) (Dong et al., 2013; Klumpp et al., 2010). CDSs encoding the structural, DNA replication, host lysis and regulator proteins were found in the proCM3 genome. These CDSs formed a modular genomic structure and were inserted into the chromosome of B. thuringiensis YM-03. The gene proCM3_1 (Table S3) in the upstream region of the proCM3 genome encoded an ABC transporter permease, which exhibited 99 % similarity to the ABC transporter permease of B. cereus MSX-A1 (GenBank accession number WP000453876). Gene proCM3_57 in the proCM3 genome also encoded an ABC transporter permease, with 98 % similarity to another ABC transporter permease of B. cereus MSX-A1 (GenBank accession number EJ905053). A mutation in the gene encoding the phage minor structural protein might have caused a nonsense mutation, generating two genes, proCM3_55 and proCM3_56 (Fig. 1b). To induce prophage proCM3, strain YM-03 was grown in Luria broth (LB) containing different concentrations of mitomycin C with final concentrations ranging from 0.1 μg ml⁻¹ to 1.0 μg ml⁻¹, at intervals of 0.1 μg ml⁻¹. No inductive phage was detected (data not shown). The gene proCM3_3 in the proCM3 genome was predicted to encode a resolvase, which is a site-specific recombinase (Yang & Steitz, 1995) that might play roles in the integration of the proCM3 genome into the YM-03 genome.
Comparative genome and phylogenetic analysis of the phiCM3 genome

Comparative genome analysis showed that the genome of phiCM3 exhibited high similarity to the genomes of eight phages from the *Bacillus* genus (Fig. 2; Table 1). Among these phages, phages Gamma, Fah, Cherry and Wb had highly similar genome sequences and could infect strains of *B. anthracis* and *B. cereus* (Fouts et al., 2006; Minakhin et al., 2005; Schuch & Fischetti, 2006), while phages BceA1 and SpaA1 exhibited nearly identical genome sequences and infected *B. thuringiensis* strains (Swanson et al., 2012).

Phage phIS3501 is an inducible prophage of *B. thuringiensis*, which is induced by mitomycin C from the latency to lytic stage (Moumen et al., 2012). These nine phage genomes had highly similar genome sequences and collinear genome organizations (Fig. 2). Phages Gamma, Fah, Cherry, Wb and BtCS33 had similar genome structures and phages BceA1, SpaA1 and phiCM3 had almost identical genome structures. The genome structure of phIS3501 was different from those of the other eight phages. Though these phage genomes were highly similar overall, several genes exhibited high diversity, and highly variable regions in the genomes were observed, such as the genes encoding proteins with functions in host lysis and DNA replication.

The proteomes of the nine phages were analysed using CoreGenes and 11 proteins were found to be common in all nine phage genomes. These 11 proteins included phage structural protein, DNA replication protein, regulator protein and cell division FtsK/SpoIIIE ATPase. In addition to the similarity with these eight phages, phiCM3 also showed significant sequence identity with genome fragments from *B. thuringiensis* serovar kurstaki str. HD73 (GenBank accession number CP004069.1), *B. thuringiensis* HD-789 (GenBank accession number CP003763.1), *B. cereus* B4264 (GenBank accession number CP001176.1) and *B. thuringiensis* HD-771 (GenBank accession number CP003752.1). The genome of phiCM3 exhibited clear synteny relationship with these genome fragments from the *Bacillus* species strains.

To analyse the evolutionary relationship of the nine phages, phylogenetic trees based on the whole-genome sequence and the terminase protein were reconstructed (Fig. 3a, b). The phylogenetic tree based on the whole genome (Fig. 3a) showed that phages Gamma, Fah, Cherry and Wb were in the same cluster, while phages phiCM3, BtCS33, SpaA1 and BceA1 had a much closer evolutionary relationship. Phage phIS3501 was far from the other eight phages in evolution, which might be due to its distinct genome structure. The phylogenetic tree based on the amino acid sequences of the terminase larger subunits (Fig. 3b) corresponded with
the phylogenetic tree of the whole-genome sequences, except for the evolutionary position of phage phiS3501.

**Differential evolution of phage genes**

In previous reports, different genes in phage genomes exhibited different rates of evolution in accordance with the changes in their host ranges (Klumpp et al., 2008; Vale et al., 2012). Comparative genome analysis of the nine phages showed that different regions of the phage genome exhibited different levels of similarity. Some genes were highly conserved, while other genes exhibited a high degree of genetic diversity (Fig. 3). To analyse the differential evolution of phage genes, five proteins, encoded by different regions of the phage genomes, were used to reconstruct phylogenetic trees (Fig. 3c). The phylogenetic trees based on the five different phage proteins exhibited different structures. The trees based on the antirepressor and the tail fibre protein showed a shorter branch length, while the trees based on the major capsid protein, DNA replication protein and RNA polymerase σ factor maintained a longer branch length between the phages. These data indicate that the antirepressor and tail fibre proteins were more highly conserved than the major capsid protein, DNA replication protein and RNA polymerase σ factor. Among these five proteins, the tail fibre protein, which was reported to be essential for determination of host specificity, was the most conserved. Lower similarities were observed on the DnaD domain protein, which is involved in phage genome synthesis (Ioannou et al., 2006; Zhang et al., 2006) and the RNA polymerase σ factor, which might regulate gene expression (Schuch & Fischetti, 2009).

**Comparative genome analysis of proCM3**

The genome of prophage proCM3 was found to be highly similar to the genomes of two phages isolated from *B. cereus* (phage TP21-L) and *B. thuringiensis* (phage BM3 Btp2) (Dong et al., 2013; Klumpp et al., 2010; Loessner et al., 1997). Though these three phages exhibited highly
similar genome sequences, their genome structures were quite different (Fig. 4a). Based on the analysis using Mauve, the genomes of these three phages were found to be comprised of three regions in different orders, suggesting that genome rearrangement might have occurred. The genome of prophage proCM3 was more similar to that of phage BMBltp2, an inducible phage of B. thuringiensis (Dong et al., 2013). Phage TP21-L is a virulent phage of B. cereus and the similarity between TP21-L and proCM3 was lower (Klumpp et al., 2010).

Insertion site of prophage proCM3 in the host genome
The genome of proCM3 was integrated into the genome of the host, Bt strain YM-03, downstream of an ABC transporter permease-encoding gene (proCM3_1) and upstream of another ABC transporter permease-encoding gene (proCM3_60) (Fig. 4). A phage gene (proCM_2) encoding the site-specific recombinase was adjacent to the ABC transporter permease gene, with 6 bp overlaps at the ends of the two genes (Fig. 4b; Table S3). According to previous reports, phages integrate into the host genome by recognizing specific attachment sites, which exist at both termini of the phage genome (Kaneko et al., 1998; Karlsson et al., 2006; Zecchi et al., 2012). A 55 bp DNA fragment (corresponding to nt 1895–1949 in the proCM3 genome) at the 3′ terminus of proCM3_1 was found to be identical to the corresponding sequence in the genome of phages TP21-L and BMBltp2 (Fig. 4b). By searching for repeat sequences in the proCM3 genome, a repeat sequence, ‘TTCAAG’, of 6 bp was found at the beginning of the 55 bp overlap fragment (from 1895–1900 bp) and in the sequence between proCM3_59 and proCM3_60 (from 38746–38751 bp).

**DISCUSSION**

In this study, we characterized the genomes of a newly isolated B. thuringiensis phage, phiCM3, and a prophage, proCM3, from the same host strain. Though phiCM3 and proCM3 had a common host, B. thuringiensis strain YM-03, their sequences exhibited few identifiable similarities. Our data corresponded to those of previous reports, which indicated that dissimilarities between phages and prophages in the same host strain were one of the reasons for phage diversity (Hendrix et al., 1999; Kwan et al., 2006).

In the USA and the former Soviet Union, phages Gamma and Fah have been used widely in rapid clinical diagnosis of B. anthracis, which causes the widespread distribution of these two phages (Abshire et al., 2005; Minakhin et al., 2005). Because of the wide distribution of phage Gamma and the high genetic instability of most viral forms, several variants of Gamma, such as γL and γU (GenBank accession numbers DQ222853 and DQ222855), were isolated (Schuch & Fischetti, 2006). The phylogenetic tree (Fig. 5) shows that seven phages infecting the strains of *Bacillus*, together with the variants of these phages, could be classified as the Wβ-group phages. The Wβ-group phages can be defined as phages that can infect the strains of *Bacillus* and harbour genomes similar in genome sequence and structure to that of phage Wβ. In a review by Breitbart & Rohwer (2005), they inferred that the diversity of phages might be high on a local scale while being relatively limited globally. The Wβ-group phages have a wide biogeographical distribution, and are found in the USA, Russia and China. Moreover, they have highly conserved genome sequences, which suggests that the diversity of the *Bacillus* phage might also be diverse at the local level and relatively limited globally.

---

**Table 1. Phages referred to in this study**

<table>
<thead>
<tr>
<th>Phage name</th>
<th>Year of isolation</th>
<th>Place of isolation</th>
<th>Host range</th>
<th>Genome size (bp)</th>
<th>GenBank accession no.</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wβ</td>
<td>1951</td>
<td>USA</td>
<td>B. anthracis, B. cereus</td>
<td>40864</td>
<td>NC_007734</td>
<td>Schuch &amp; Fischetti (2006)</td>
</tr>
<tr>
<td>Cherry</td>
<td>1955</td>
<td>USA</td>
<td>B. anthracis, B. cereus</td>
<td>36615</td>
<td>NC_007457</td>
<td>Fouts et al. (2006)</td>
</tr>
<tr>
<td>Fah</td>
<td>1995</td>
<td>Russia</td>
<td>B. anthracis, B. cereus</td>
<td>37974</td>
<td>NC_007814</td>
<td>Minakhin et al. (2005)</td>
</tr>
<tr>
<td>phiCM3</td>
<td>2011</td>
<td>China</td>
<td>B. thuringiensis</td>
<td>41992</td>
<td>NC_018085</td>
<td>Yuan et al. (2012b)</td>
</tr>
<tr>
<td>SpaA1</td>
<td>2012</td>
<td>Antarctica</td>
<td>B. cereus, B. thuringiensis, S. pasteuri</td>
<td>42784</td>
<td>NC_018277</td>
<td>Moumen et al. (2012)</td>
</tr>
<tr>
<td>BceA1</td>
<td>2012</td>
<td>Antarctica</td>
<td>B. cereus, B. thuringiensis</td>
<td>42932</td>
<td>HE614282</td>
<td>Swanson et al. (2012)</td>
</tr>
<tr>
<td>phiCM3</td>
<td>2012</td>
<td>China</td>
<td>B. thuringiensis</td>
<td>38772</td>
<td>KF296718</td>
<td>This study</td>
</tr>
</tbody>
</table>
Though the genome similarity between the Wβ-group phages was high, they exhibited different host ranges and differential evolution was observed on some regions of the phage genome. In this study, we found that phage phiCM3 and BtCS33 could not infect the host strains of the other. The phages appear to thrive in different environments and hosts by co-evolving with the host and the environment (Berngruber et al., 2010; Golais et al., 2013; Weitz et al., 2013). DNA polymerase is essential for replication of the phage genome and RNA polymerase σ factor is thought to regulate gene transcription of the host bacteria and may have functions in phage reproduction (Schuch & Fischetti, 2006, 2009). As shown in the phylogenetic tree of DNA polymerases and RNA polymerase σ factors, the phages that could infect the B. anthracis and B. cereus strains were clustered into one branch, while the phages that could infect the strains of B. thuringiensis and/or B. cereus were clustered into another branch. The relatively high divergence of the sequences of these two genes might benefit the phages, enabling them to infect different host bacteria. Compared with the rapid evolution of the DNA polymerase and RNA polymerase σ factor, the tail fibre protein and the antirepressor were more highly conserved. The tail fibre protein of Gamma plays roles in host recognition (Schuch & Fischetti, 2006) and mutation of the tail fibre protein modifies host specificity (Lucchini et al., 1999). Though the mutations in the tail fibre protein were not as numerous as those in some other genes, there were indeed some mutations in the tail fibre protein, which might be the hot spot that determines host specificity of the phages (Dai et al., 2010; Doulatov et al., 2004). The antirepressor maintains the lytic form of the phage by inactivating the CI repressor (Fogg et al., 2011; Susskind & Botstein, 1975). Because the gene of the lysogeny control region was conserved in the Wβ-group phages, the antirepressor was also highly conserved. The high mutation frequency of the major capsid protein, which is the main component of the phage capsid that packaged the phage genome, is of particular interest from the perspective of differential evolution of phage genes. According to previous reports, different phages of Wβ group have different sizes of head and different genome lengths (Fouts

Fig. 3. Phylogenetic trees of phiCM3 and related phages. (a) The whole-genome sequences of the nine phages were used to reconstruct the tree. The biogeographical locations of the phages are indicated. Bar, 0.1 substitutions per nucleotide position. (b) The tree reconstructed based on the amino acid sequences of the terminase large subunits is shown. Bar, 0.1 substitutions per amino acid position. (c) The amino acid sequences of the antirepressors, major capsid proteins, tail fibre proteins, DNA replication proteins and RNA polymerase σ factors were used to reconstruct a phylogenetic tree. Bar, 0.2 substitutions per amino acid position. All trees were reconstructed using MEGA 5.
et al., 2006; Minakhin et al., 2005; Schuch & Fischetti, 2006; Swanson et al., 2012; Yuan et al., 2012b). The mutation in the major capsid protein might be associated with the different lengths of the phage genomes.

Phage genomes maintain a mosaic structure and gene exchange might occur by horizontal gene transfer (Dekel-Bird et al., 2013; El Haddad & Moineau, 2013; Labrie et al., 2013), which could then cause the diversity in the phage genome structure. The host lysis genes of the Wβ-group phages were quite different either in the amino acid sequence or in the type of the endolysin (Kikkawa et al., 2008; Yuan et al., 2012a). The genome of phage phiCM3 contained two putative holin and one putative endolysin, while the other phages contained only one putative holin and one putative endolysin. To our knowledge, this was the first report of a phage host cell lysis cassette that contains three genes, in contrast to previous reports of the two-component cell lysis cassette of phages (Sheehan et al., 1999; Wang et al., 2008). The function of the additional holin and the mechanism of action of the phiCM3 host lysis gene cassette need to be studied further.

In previous reports, the genome of the sequenced microbe usually contained one or more prophage genomes and the prophage genomes would lead to the diversity of the bacterial genome (Brüssow et al., 2004; Ivanova et al., 2003). By draft genome sequencing of the host strain YM-03 of phiCM3, the genome of a prophage (designated proCM3) was discovered. This prophage was inserted downstream of an ABC transporter permease-encoding gene with a 55 bp overlap and upstream of another ABC transporter permease-encoding gene. The genome of proCM3 was similar with the genomes of two lytic phages (TP21-L and BMBtp2) of Bacillus and the genomes of these three phages were comprised of three regions in different orders. According to the modular theory of phage evolution (Brüssow et al., 2004), we speculated that these three phages might have a common ancestor with a circular genome and that the genome of the ancestor phage inserted
into different sites of the bacterial genome during evolution. The different insertion sites of phages would result in the rearrangement of the phage genome, thereby generating the three different phages. A core repeat sequence, 'TTCAAG', of 6 bp was found at both termini of the proCM3 genome, which might serve as the attachment sites of proCM3 and the recognition sites of the site-specific recombinase (encoded by proCM3_2, annotated as resolvase).

The Bacillus phages have been isolated rapidly (Ackermann et al., 1994) and the genomes of more than 50 Bacillus phages had been sequenced (http://www.ncbi.nlm.nih.gov/genbank) by 30 May 2013. Because of their application in detecting and controlling pathogenic bacteria of Bacillus in the food chain, for treatment of human infections, for bacteria typing and as biodefence measures (Loessner et al., 1997; Schuch et al., 2002), a better understanding of the genetic nature and evolutionary history of the Bacillus phage is of great importance. Phage contamination can cause serious loss in industrial production of Bt insecticides (Liao et al., 2008). Studying the B. thuringiensis phage genomes will be helpful for isolating phage-resistant B. thuringiensis strains and controlling phage contamination. Besides, phage and phage-producing endolysin could lyse bacteria from outside and exhibited potential as novel antibacterial agents (Porter et al., 2007; Schuch et al., 2002).

The findings of this study enrich our current knowledge of Bacillus phage diversity and evolution, especially for the Wβ-group phages and TP21-L-like phages, and are helpful in developing further practical uses of Bacillus phages.

**METHODS**

**Bacterial strains and phages.** Bacillus thuringiensis strain YM-03 was isolated in our laboratory and exhibited high toxicity to insect larvae of the order Coleoptera. The bacteriophage was isolated and propagated by methods described previously (Carey-Smith et al., 2006; Yuan et al., 2012b), using YM-03 as the host strain. All strains used in this study were grown in LB at 30 °C.

**Phage purification.** The phages propagated on LB-agar plates were washed off with SM buffer [0.58% (w/v) NaCl, 0.2% (w/v) MgSO4]...
These strains were all stored in our laboratory. The _Bacillus thuringiensis_ in this study were tested using the method described previously (Yuan et al., 2012b). In all, 54 strains belonging to the species of _B. thuringiensis_, _B. anthracis_, _B. pumilus_ and _B. subtilis_ were tested. These strains were all stored in our laboratory.

**Host range determination.** The host ranges of the phage isolated in this study were tested using the method described previously (Yuan et al., 2012b). In all, 54 strains belonging to the species of _B. thuringiensis_, _B. anthracis_, _B. pumilus_ and _B. subtilis_ were tested. These strains were all stored in our laboratory.

**Genome sequence and bioinformatics analysis.** Genomes of the phage and strain YM-03 were extracted as previously described (Gao et al., 2008; Yuan et al., 2012b). The end of the phage genome was analysed as described previously (García et al., 2009). The genomes were sequenced using a GS FLX system (Roche) and assembled into contigs using GS De novo Assembler (Roche). The remaining gaps between the contigs of the phage genomes were filled by primer walking. The coding sequences (CDSs) of the phage and prophage genomes were predicted with the FGENE SV software (http://linux1.softberry.com/berry.phtml?topic=genolist&group=programs&subgroup=gfndv) and by visual inspection. The putative function of each gene was predicted by searching against the NR, CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and Pfam (http://pfam.sanger.ac.uk/search/) databases using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The genes encoding the putative tRNAs were analysed using RNAScan (Schattner et al., 2005). Comparative genome analysis of the phages was carried out using Mauve 2.3.1 (Darling et al., 2004), EMBOSS Stretcher (http://emboss.bioinformatics.nl/cgi-bin/emboss/stretcher) and CoreGenes 3.0 (Zafar et al., 2002). The phylogenetic trees in this study were all reconstructed with MEGA 5 (Tamura et al., 2011).

**Nucleotide sequence accession numbers.** The genomes of phage phiCM3 and prophage proCM3 were submitted to GenBank under the accession numbers KF296718 and KF296717, respectively.

**ACKNOWLEDGEMENTS**

This study was supported by the National Natural Science Foundation of China (No. 31170123), the National Project (2009ZX08009-056B) and the projects of the Chinese Academy of Sciences (KSCX2-EW-G-16).

**REFERENCES**


Klumpp, J., Calendar, R. & Loessner, M. J. (2010). Complete nucleotide sequence and molecular characterization of *Bacillus* phage TP21 and its relatedness to other phages with the same name. *Viruses* 2, 961–971.


