Characterization and comparative genomic analysis of virulent and temperate *Bacillus megaterium* bacteriophages

Abdoallah Sharaf[1,2], Miroslav Oborník[2,3], Adel Hammad[4], Sohair El-Afifi[5] and Eman Marei[5]

[1] Genetic Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt
[2] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic
[3] Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
[4] Department of Microbiology, Faculty of Agriculture, Minia University, Minia, Egypt
[5] Department of Agricultural Microbiology, Virology Laboratory, Ain Shams University, Cairo, Egypt

Corresponding author
Abdoallah Sharaf,
abdoallah.sharaf@agr.asu.edu.eg,
sharaf@paru.cas.cz

ABSTRACT

Next-Generation Sequencing (NGS) technologies provide unique possibilities for the comprehensive assessment of the environmental diversity of bacteriophages. Several *Bacillus* bacteriophages have been isolated, but very few *Bacillus megaterium* bacteriophages have been characterized. In this study, we describe the biological characteristics, whole genome sequences, and annotations for two new isolates of the *B. megaterium* bacteriophages (BM5 and BM10), which were isolated from Egyptian soil samples. Growth analyses indicated that the phages BM5 and BM10 have a shorter latent period (25 and 30 min, respectively) and a smaller burst size (103 and 117 PFU, respectively), in comparison to what is typical for *Bacillus* phages. The genome sizes of the phages BM5 and BM10 were 165,031 bp and 165,213 bp, respectively, with modular organization. Bioinformatic analyses of these genomes enabled the assignment of putative functions to 97 and 65 putative ORFs, respectively. Comparative analysis of the BM5 and BM10 genome structures, in conjunction with other *B. megaterium* bacteriophages, revealed relatively high levels of sequence and organizational identity. Both genomic comparisons and phylogenetic analyses support the conclusion that the sequenced phages (BM5 and BM10) belong to different sub-clusters (L5 and L7, respectively), within the L-cluster, and display different lifestyles (lysogenic and lytic, respectively). Moreover, sequenced phages encode proteins associated with *Bacillus* pathogenesis. In addition, BM5 does not contain any tRNA sequences, whereas BM10 genome codes for 17 tRNAs.

INTRODUCTION

Bacillus megaterium (Firmicutes) is ubiquitous in nature. It is recognized as an endophyte and a potential biocontrol agent for plant diseases (*Kildea et al., 2008*). Furthermore, it is known to produce penicillin, amidase, various baking amylases and is used for the

biotechnological production of pyruvate, vitamin B12, and so on (*Bunk et al., 2009*). In alkaline soils, *B. megaterium* plays a significant role in making insoluble phosphorus accessible to plants via the production of CO_2 and organic acids, which increases soil acidity and the fraction of solubilized phosphates. Certain *B. megaterium* strains even exhibit nitrogen fixation capability (*Bergey, 2009*). Bacteriophages (phages) are the most abundant biological entities present on the planet (*Hambly & Suttle, 2005*). Their capacity to infect and kill their hosts makes them an important factor driving the bacterial evolution and preservation of the ecological balance (*Chan, Abedon & Loc-Carrillo, 2013*; *Singh, Poshtiban & Evoy, 2013*). Phages can be utilized for various purposes, such as therapy for bacterial infections, for biocontrol of pathogens as protein exposure systems, and for bacterial typing (*Haq et al., 2012*). A number of phages infecting the genus *Bacillus* have been identified so far (*Zeigler, 1999*; *Lee et al., 2011*). *Bacillus*-specific phages demonstrate great diversity with regard to morphology, genome sequence length, gene content, and host range, exhibiting high variability among and within the species of this genus. Moreover, numerous bacteriophages that infect and lyse *B. megaterium* have been studied (*Cooney, Jacob & Slepecky, 1975*; *Carvalho & Vary, 1977*; *Vary & Halsey, 1980*; *Van Elsas & Penido, 1982*; *Serwer et al., 2007*; *Eppinger et al., 2011*). These phages possess double-stranded, linear DNA genomes (*Van Elsas, Linhares & Penido, 1982*).

Furthermore, the expansion of next generation sequencing (NGS) technologies (*Eisenstein, 2012*; *Klumpp, Fouts & Sozhamannan, 2012*), in addition to the possibility of sequencing entire genomes or transcriptomes more efficiently and economically, rather than using the Sanger sequencing strategy, has allowed for obtaining full genomic sequences for a wide range of species. Therefore, the advent of NGS technology provides new opportunities for sequencing of a broad range of organisms, including bacteriophages, quickly, reliably, and considerably inexpensively (*Hatfull, 2008*).

Here, we shed light on the biological features, genome sequence and annotation of temperate and virulent *B. megaterium* bacteriophages, representing two different groups according to their host range and amplified fragment length polymorphism (AFLP) profile (*Elmaghraby et al., 2015*). Both phages display different thermal inactivation points (82 and 78 °C) and pH tolerant range (5–9.2 and 5–8.4 pH) while having the same longevity *in vitro* (192 h) (*Elmaghraby et al., 2015*). Electron microscopy proved that both phages belonged to the *Myoviridiae* family (*Elmaghraby et al., 2015*). Furthermore, in this paper, we present an updated phylogenetic analysis of *Bacillus* bacteriophages based on the amino acid sequence of terminases.

MATERIALS AND METHOD

Bacteriophages

Two isolates of *B. megaterium* phages, namely BM5 and BM10, were obtained from soil. Their biological and morphological properties were reported previously (*Elmaghraby et al., 2015*). The high titer phage suspension of each isolate was prepared using a liquid culture enrichment technique (*Marie, 2013*). Fifteen ml of the same high titer phage suspension for each phage was ultra-centrifuged at 30,000 rpm for 90 min at 4 °C in a Beckman

L7-35 ultracentrifuge. The pellet was gently re-suspended in 0.5 ml of a 0.2 M potassium phosphate buffer, with a pH level of 7.2 (*Lanningan & Williams, 1982*). A single-step growth experiment was performed with minor modifications, as reported previously, to determine the latent and rise period and the phage burst size (*Pajunen, Kiljunen & Skurnik, 2000*; *Sillankorva, Neubauer & Azeredo, 2008*). Furthermore, 1 ml of the phage suspension (10^{10} pfu/ml) was mixed with 1 ml of exponential phase culture of *B. megaterium* (10^8 cfu/ml) and incubated at 30 °C for five minutes for phage adsorption. Subsequently, the mixture was centrifuged at 10,000 rpm for 10 min to remove free phage particles. The pellet was re-suspended in 60 ml of a nutrient broth, and the culture was continuously incubated at 30 °C. Phages were sampled at intervals of 5 to 60 min, and their titers were determined. The burst sizes of the phages were estimated by dividing the bacteriophage titers at the plateau phase using the initial phage titers.

DNA isolation, library preparation, and whole genome sequencing

Genomic DNA was extracted from the bacteriophages, as described by *Irshad, Waqas & Saadia (2012)*, with minor adjustments. The purity and concentration of the DNA was evaluated using a Nanodrop Bioanalyzer N1000 (Thermo Scientific, Waltham, MA, USA). In addition, sequencing libraries were prepared by shearing 1 μg of the phage DNA, to generate blunt-ended fragments, after which the Ion adapters were linked using an Ion Xpress^{TM} Fragment Library Kit (Life Technologies, Carlsbad, CA, USA), in accordance with the manufacturer's instructions. The produced fragments were amplified employing the Ion OneTouch 200 Template kit (Life Technologies, Carlsbad, CA, USA). Furthermore, libraries were sequenced on an Ion Torrent PGM semiconductor sequencer using the Ion Torrent 314 chip, by applying the standard protocol (Life Technologies, Carlsbad, CA, USA).

Sequence assembly and annotation

Raw sequencing reads were trimmed and masked using the FASTQ Trimmer and the FASTQ Masker Galaxy Tools Version 1.0.0 (*Blankenberg, Gordon & Kuster, 2010*), which were assembled by the 454 Newbler Assembler software (Roche Applied Science, IN). The collected contigs were visualized and validated by *Consed* (*Gordon & Green, 2013*), resulting in the presence of contigs in each phage, which demonstrated a 60-fold sequence read coverage approximately. In addition, sequence homology analysis and assignment to the phage clusters were performed using BLASTn against the *Bacillus* phage database (http://bacillus.phagesdb.org/: 09/2016) (*Altschul et al., 1990*). The genome sequences of each phage were annotated using the NCBI Prokaryotic Annotation Pipeline. In addition, protein alignment was used for gene predictions to guarantee the consistency of the annotation for closely related genomes. ProSplign (protein aligner software) was utilized to align all the protein sequences. For further refinement, frameshifted alignments and partial alignments were processed using a gene prediction program (GeneMarkS+), and final annotations were established by searching against bacterial and bacteriophage proteins in the SwissProt database (*Besemer, Lomsadze & Borodovsky, 2001*). The gene predictions were verified using the CPT (Center for Phage Technology) Galaxy (https://cpt.tamu.edu/galaxy-pub/). Furthermore, for genome-wide

analyses and other searches, a Phamerator software was employed (*Cresawn et al., 2011*). Subsequently, conserved domains relationships within genes were determined. Proteins were assorted into "Phamilies" (generally referred to as "Phams"), which are groups of proteins that are largely similar to one another. BLASTp and ClustalW were used to determine the pairwise alignment scores (*Cresawn et al., 2011*). Phage RNA Polymerase and RNP-recognized promoters were located using the extractupstreamDNA software (https://github.com/ajvilleg/extractUpStreamDNA), after which the MEME program (http://meme-suite.org) was used for motifs elicitation (*Bailey & Elkan, 1994*; *Bailey et al., 2009*).

Genomic similarity and phylogenetic analysis

Proteome similarity and average nucleotide identity (ANI) values of 16 genome sequences of *Bacillus* phages, which were retrieved from GenBank, were determined using JSpecies 1.2.1 (http://imedea.uib-csic.es/jspecies/download.html) at the default BLASTp threshold of score 75 (*Richter & Rosselló-Móra, 2009*). Gepard was used to create dot plots (*Krumsiek, Arnold & Rattei, 2007*). To simplify the dot plot analyses, certain phage genomes were reverse complemented, and new sequences were generated to re-orient on the basis of the majority of the phages. A maximum likelihood phylogenetic tree was computed by the RAxML program and by utilizing the lg model (*Stamatakis, 2014*). The Bayesian tree was calculated by the PhyloBayes program, with both the lg and the C20 models (*Lartillot, Lepage & Blanquart, 2009*) based on the terminase protein sequence. The supporting values from both methods were merged into a one-rooted tree while the Alloherpesviridae sequences (Anguillid and Cyprinid herpesvirus) were used as the out-group.

Nucleotide sequence accession numbers

The complete genome sequences of phages BM5 and BM10 were deposited in the NCBI GenBank database under accession numbers KT995479 and KT995480, respectively. Moreover, the raw sequencing data were published in the Sequence Read Archive (SRA) database under the accession number SRP136802.

RESULTS

Single-step growth curve

To assess the ability of both phages (BM5 and BM10) to lyse *B. megaterium*, the latent and rise periods and the burst sizes of each phage were determined by a single-step growth curve analysis (Fig. 1). The latent periods for BM5 and BM10 phages were estimated to be 25 min and 30 min, respectively. The rise period for both phages was 45 min. The calculated burst size was 103 pfu/cell and 117 pfu/cell, for the BM5 and BM10 phages, respectively.

Genome organization

The genome sequencing generated 267,273 reads for BM5 phage and 288,254 reads for BM10 phage, forming an approximately 60-fold coverage. Assemblies yielded 165 kbp and 165.2 kbp genomes for BM5 and BM10, respectively. A comparison of the translated BLAST (tblastx) data (*Altschul et al., 1990*) from the two genome sequences revealed that they shared a high level of overall similarity; conserved regions reached 91% identity,
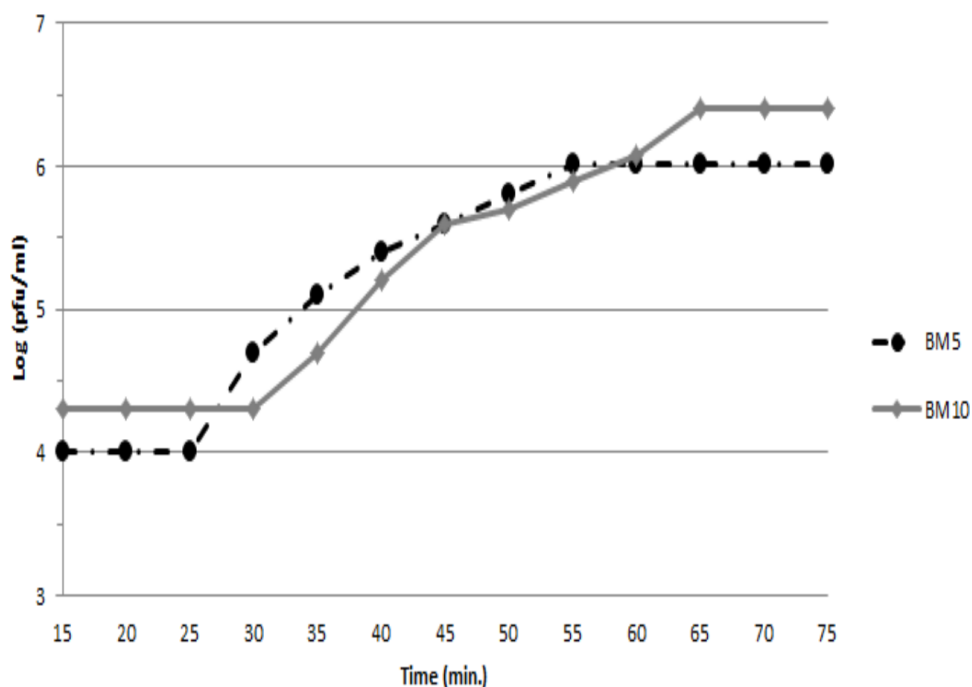
Figure 1 Single-step growth curve of BM5 and BM10 bacteriophages.

distributed within the variable regions and with variability ranging from 23% to 88%. Both phage genomes show the common bacteriophages genome organization, with genes of a similar function clustering together (Fig. 2). The last 30 kb of DNA genome sequences of BM5 encoding, for helicases and helicase related proteins, were highly similar to the DNA genome sequences of BM10 (Fig. 2). Annotation of the BM5 genome revealed 225 ORFs (Open Reading Frames), while the BM10 genome encodes revealed 283 ORFs. A total of 97 of the 225 ORFs (43%) in the BM5 genome and 65 of the 283 (23%) in the BM10 genome were assigned functions in comparison with the identified conserved domains (*Marchler-Bauer, Derbyshire & Gonzales, 2015*) (Tables S1 and S2). In addition, that we identified 17 genes' coding for tRNAs in the BM10 genome (Table 1).

Phage structure and assembly genes

Genes Bm5_9 and Bm5_62 were identified as encoding for the major capsid protein (TIGR04387) (*Choi, McPartland & Kaganman, 2008*) and the phage capsid protein (pfam05065) (*Sutter, Boehringer & Gutmann, 2008*), respectively. They were concurrently thought to be involved in the stabilization of the condensed form of the DNA in phage heads. Genes Bm5_61 and Bm5_184 encode for phage head maturation proteases (Peptidase_U35) (pfam04586), which are also involved in phage head development (*Duda, Martincic & Hendrix, 1995*). Some genes that are essential for the tail development, such as Bm5_66 (TIGR01725), Bm5_68 (cl21657), and Bm5_76 (pfam05709), were also found. Genes involved in head and tail development, such as Bm10_67 encoding for *Caudovirus* prohead
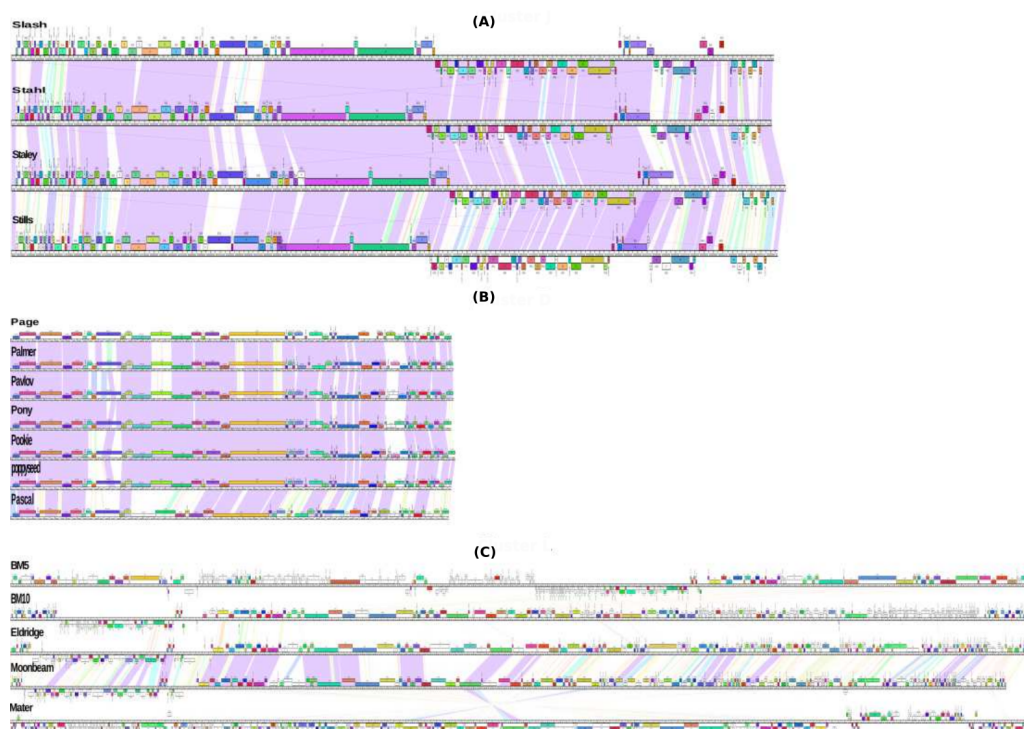
Figure 2 **Genomic organization of Bacillus megaterium phages.** Phages mapped using Phamerator (*Richter & Rosselló-Móra, 2009*); the purple lines between phages underline the regions with high similarity, while the ruler corresponds to genome base pairs. The predicted genes are presented as boxes either above or below the genome (line), depending on whether are rightwards or leftwards transcribed, respectively. Gene numbers are displayed within each box. Genes colored according to their function categories "phams". The unfilled boxes assigned to genes that show low similarity (orphams). *B. megaterium* phages (A) J cluster and (B) D cluster show a high similarity between its members because all members belong to the same sub-clusters (J1 and D1, respectively). Also (C) L cluster phage members belonging to different sub-clusters show a low similarity.

protease (cl01521) (*Duda, Martincic & Hendrix, 1995*), were identified in both BM5 and BM10 genomes (Tables S1 and S2).

Cell lysis genes

A number of genes involved in lytic activities, such as the degradation of a bacterial cell wall during host infection, was also identified. Bm5_11, Bm5_82, and Bm5_198, coding for Lysophospholipase (EC 3.1.1.5) (cd00229), Peptidoglycan-binding (PGRP) domain (COG3409), and Pectate lyase 3 (cl19188), respectively, are known to hydrolyze lipids, Peptidoglycan and Pectate (*Kamen & Woody, 2002*; *Akoh, Lee & Liaw, 2004*). The Bm5_25 and Bm5_77 ORFs were identified as putative endopeptidases and are most probably D-alanyl-D-alanine carboxypeptidase (Peptidase_M) (EC 3.4.16.4) (pfam13539). They belong to the Peptidase M15 family and Prophage tail endopeptidases (pfam06605), required for the infection of the stationary phase bacterial cells (*Lee et al., 2001*; *Stephen et al., 2013*). We also identified Bm10_139 encoding Holin (cl2306), and Bm10_51 coding for Acetylmuramoyl-L-alanine amidase (MurNAc-LAA) (cd02696) (EC 3.5.1.28). The

Table 1 Properties of BM10 phage tRNAs.

ORF no.	Strand	Begin	End	tRNA	Anti-codon
ORF 62	+	29,057	29,148	tRNA-Ser	GCT
ORF 63	+	29,156	29,232	tRNA-Cys	GCA
ORF 64	+	29,274	29,351	tRNA-Asn	GTT
ORF 65	+	29,354	29,428	tRNA-Gly	GCC
ORF 66	+	29,434	29,512	tRNA-Thr	TGT
ORF 67	+	29,516	29,591	tRNA-Glu	TTC
ORF 68	+	29,599	29,677	tRNA-Glu	TTC
ORF 69	+	29,769	29,847	tRNA-Thr	CGT
ORF 70	+	29,849	29,938	tRNA-Ser	TGA
ORF 71	+	30,015	30,093	tRNA-Gln	CAA
ORF 73	+	30,168	30,244	tRNA-Gln	CTG
ORF 74	+	30,339	30,413	tRNA-Gln	TTG
ORF 75	+	30,512	30,588	tRNA-Phe	AAA
ORF 76	+	30,677	30,765	tRNA-Leu	TAA
ORF 77	+	30,765	30,841	tRNA-Ile	GAT
ORF 78	+	30,849	30,935	tRNA-Leu	TAG
ORF 79	+	31,029	31,105	tRNA-Met	CAT

Peptidoglycan endolysin, which lyses the bacterial cell wall during the phage release, was also identified (*Vollmer et al., 2008*) (Tables S1 and S2).

DNA replication and metabolic genes

We identified several genes encoding endonucleases as well. Bm5_55 encodes HNH_3 endonuclease (pfam13392) and both Bm5_52 and Bm5_113 code for a restriction endonuclease (cd00085) (*Makarova, Grishin & Shabalina, 2006*). ORFs with putative roles in DNA replication, repair, and modification were identified in the BM10 phage genome. For instance, Bm10_126 encodes a type II DNA-binding protein (DNABII) (cl00257), which serves as the architectural factor in a variety of cellular processes (*Prieto, Kahramanoglou & Ali, 2012*). Bm10_131 encodes the 5′-3′ polymerase domain of the DNA polymerase (cl02626) (*Lee, Kennedy & Yin, 2009*) and Bm10_136 codes for a RecA recombinase (COG0468). Other ORFs, potentially involved in transcription and translation, include Bm10_189 (RNA polymerase sigma factor (RpoS)) (TIGR02980) and Bm10_24 (tRNA-splicing ligase (RtcB) (pfam01139) (EC:6.5.1.3)), which catalyzes the splicing of tRNA, and it may also participate in tRNA repair and recovery from stress-induced RNA damage (*Tanaka & Shuman, 2011*) (Tables S1 and S2).

Regulatory genes

Several genes that encode proteins that are likely to regulate host functions, are conserved in the phage genomes derived from the *Bacillus* species. The protein encoded by the Bm5_168 gene, containing a bacterial SH3like domain (cl17036), can be found in 28 of the 93 *Bacillus* phages, including phages obtained from cluster D, F, G, and L (*Grose et al., 2014*). We found Sigma-70 region 2 (pfam04542) encoded by the Bm10_188 gene, which

represents the most conserved region of the entire RNA polymerase sigma-G factor protein and is known to interact with the clamp domain of the largest beta polymerase subunit (*Jones et al., 2001*) (Tables S1 and S2).

Genes involved in pathogenesis

The Trimeric dUTP diphosphatase (cl00493) and NTP-Ppase (cd11533) proteins that may show the host pathogenesis are encoded by the Bm10_110 and Bm5_108 genes, respectively. These proteins are commonly present in the bacteriophage genomes and have been made to function as G protein-like regulators, which are required for the transfer of staphylococcal virulence factors (*Tormo-Más, Mir & Shrestha, 2010*; *Tormo-Más et al., 2013*). Bm5_70 encoding Sialidase (cd15482) plays a vital role in the phage pathogenesis (*Telford et al., 2011*) (Tables S1 and S2).

Furthermore, 224 and 266 phage-RNA polymerase (RNP) promoters were identified in the BM5 and BM10 phages, respectively, which were all located between the genes (Tables S3 and S4). Motif analysis conducted in this regard reveals 213 conserved motifs (27 nucleotides) for the BM5 phage, while a shorter conserved motif (21 nucleotides) and with the same motif logo was identified in 254 of the BM10 phage-RNA polymerase promoters (Fig. S1). Moreover, additional eight conserved motifs (27 nucleotides) were identified in the BM10 phage with specific DNA recognition sites for the host (*Bacillus*) transcription. So far, none of the downstream ORFs of this host promoter have been functionally assigned (Fig. S1).

Temperate lifestyle genes

Only the BM5 encodes genes that are suggestive of a temperate lifestyle, such as Integrase_AP2 (pfam14657), were found in a variety of phage integrase proteins, including the ICEBs1 integrase from the *Bacillus subtilis* (*Lee et al., 2007*). Furthermore, the BM5 encodes a putative site-specific tyrosine recombinase (cd01189) with a conserved C-terminal catalytic domain of other phage integrases, such as the P1 Cre and the lambda Int (*Argos et al., 1986*) (Tables S1 and S2).

tRNA genes and codon usage

Phage genomes usually encode a few or no tRNA genes, because they use the host machinery for the synthesis of proteins encoded in the phage genomes. Seventeen tRNA genes were identified in the BM10 genome, while no tRNA genes were identified in the BM5 genome. The 17 BM10 tRNA genes ranged from 77 to 92 nucleotides in size, with one copy each for tRNA-Cys, tRNA-Asn, tRNA-Gly, tRNA-Phe, tRNA-Ile, tRNA-Met, two copies of tRNA-Ser, tRNA-Thr, tRNA-Glu, and tRNA-leu, and three copies of tRNA-Gln (Table 1). All the tRNAs are located in the region between the HTH_XRE (HTH: Helix-turn-Helix) protein-encoding gene and a phage portal protein gene. The presence of tRNAs have been reported in several other phage genomes and are likely to be supporting the phage protein translation efficiency (*Grose et al., 2014*) and virulence (*Bailly-Bechet, Vergassola & Rocha, 2007*).

We compared the codon usage pattern of the BM10 tRNAs with that of its host (*B. megaterium*) using the tRNAscan-SE program (*Chan & Lowe, 2016*; *Lowe & Eddy, 1997*).
We saw that in 10 cases the phage-encoded tRNAs may significantly enhance translation in the phage (Table S5), based on the amino acid usage and/or the codon usage. The products of two genes, Bm10_123 and Bm10_154, appear to encode the nucleotidyltransferase (*El-Arabi et al., 2013*) and the tRNAHis guanylyltransferase (*Hyde et al., 2010*; *El-Arabi et al., 2013*), which may modify the phage or the host tRNAs. The BM10 phage also includes putative tRNAs (tRNA-Thr^{CGT}, tRNA-Gln^{CTG} and tRNA-Phe^{AAA}), which could be utilized during the infection of an alternative host (*Delesalle et al., 2016*).

The genomic relationship between BM5 and BM10

Analysis of the conserved predicted proteins, encoded in the genomes of BM5 and BM10 phages, revealed the huge diversity of these phages, with a total of 407 protein families (phams) of which only 52 (12.8%) were shared by both phages and 355 (87.2%) phams were phage specific or orphams (phams containing a gene product from a single phage). Twenty-one (40.4%) of the 52 shared phams are associated with an identified function (Fig. 3), while the remaining are not.

Bm10_41 and Bm5_159 are members of the pham no.12, assigned to Thymidylate synthase (TS), a well-conserved nucleotide metabolism enzyme identified in 28 *Bacillus* phages (Fig. 3). This enzyme catalyzes the alkylation of the C5 carbon of the pyrimidine nucleotides, known to induce protection against the host restriction system (*Sotelo-Mundo et al., 2006*). Bm10_121, Bm5_222, and Bm5_223 are members of pham no.58, composed of the Exonuclease subunit SbcD, one of the DNA replicator proteins. Genes encoding structural and assembly proteins were also conserved within the *Bacillus* phages (Fig. 3). These proteins show a high level of sequence conservation across the *Myoviridae*, *Siphoviridae*, *Podoviridae,* and *Tectiviridae* phage families. Identified phams include the phage tail sheath protein, encoded by pham n.35 (Bm10_94 and Bm5_193) and the Baseplate J-like protein, encoded by pham n.51 (Bm10_112 and Bm5_214). These genes have homologs in 33 of the 93 recognized *Bacillus* phages (35%). In addition, Phage terminase, encoded by pham n.18 (Bm10_52 and Bm5_170), and *Caudovirus* prohead protease, encoded by pham n.27 (Bm10_84 and Bm5_184), have homologs in 27 of the 93 investigated *Bacillus* phages (Fig. 3).

Several proteins that regulate host metabolism (including pathogenesis) are also conserved in *Bacillus* phages (Fig. 3). The 235 kDa-fam (TIGR01612) protein, encoded by pham n.43 (Bm10_102, Bm10_104, and Bm5_206), represents a Reticulocyte-binding protein which is localized on the host cell surface and is required in the process of invasion of the parasite (*Khan, Jarra & Preiser, 2001*). The carbohydrate-binding domain (cl19911), encoded by pham n.52 (Bm10_113 and Bm5_215), is necessary for the carbohydrate-protein interaction (*Johnson, Joshi & Tomme, 1996*). Two protein phams involved in lysogenic pathways were identified (Fig. 3): Bm10_96 and Bm5_199, members of pham n.37, were identified in three conserved aspartate residues (3D) domain-containing proteins (cd14667), typically found in conjunction with a membrane-bound lytic transglycosylase. Bm10_100 and Bm5_203, members of pham n.41, were assigned to Lysozyme (EC 3.2.1.17) (Fig. 3).
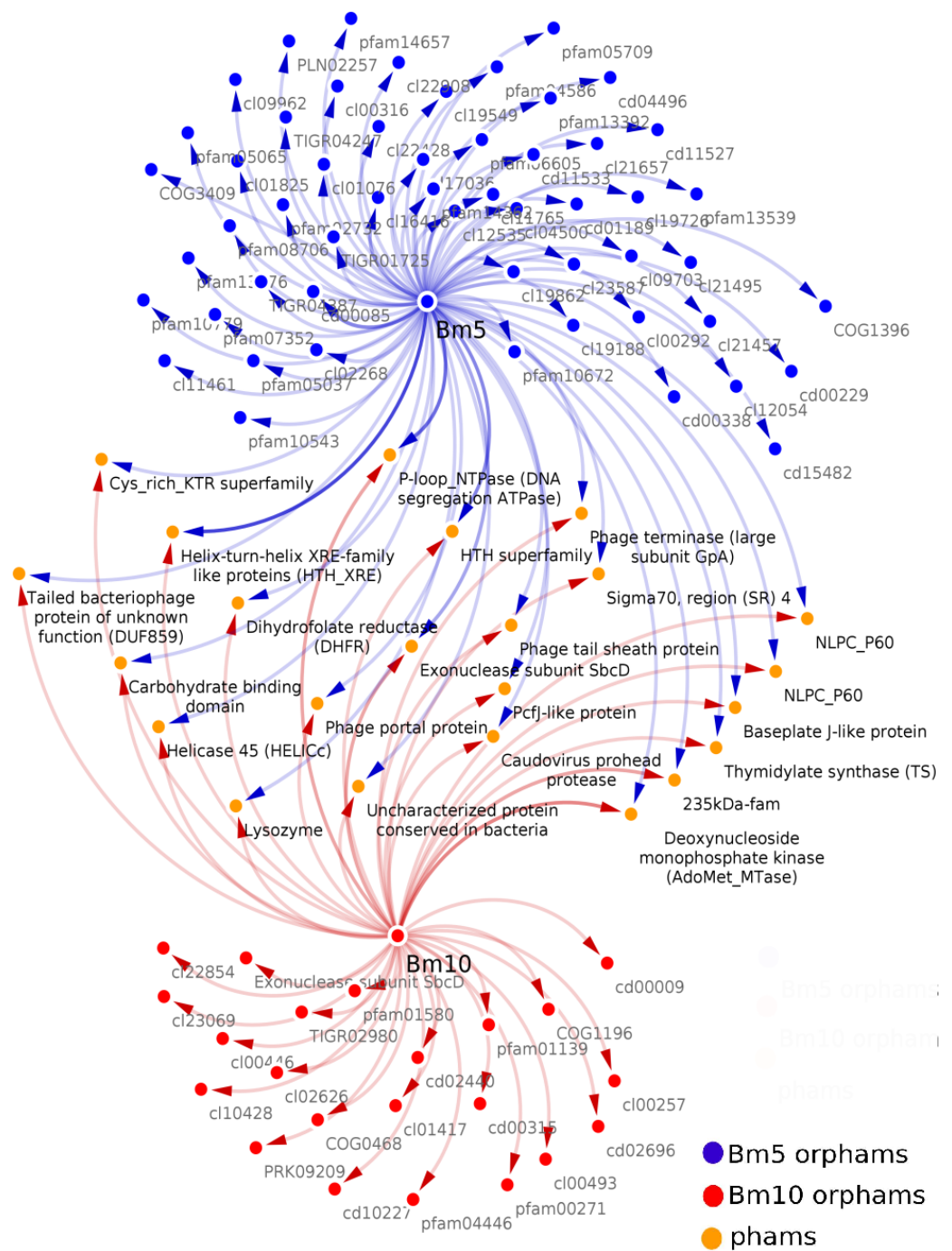
Figure 3 Related Conserved Domains (CD) to the detected Phamilies (phams) using Phamerator. Both BM5 and BM10 orphams (phams containing a gene product from a single phage) presented as blue and red nodes, respectively. While shared phams (orange nodes), located between them. All nodes connected by blue and red arrows represent genes encoding for each pham.

Table 2 Characteristics of reported *Bacillus megaterium* phages with complete genome sequences.

Cluster	Phage name	Size (bp)	GC%	ORFS	tRNA	Integrase presence	Accession number	Family
J1	Slash	80,382	35.23	111	0	Yes	NC_022774.1	*Siphoviridae*
–	Stahl	80,148	35.27	110	0	Yes	NC_028856.1	*Siphoviridae*
J1	Staley	81,656	35.35	113	0	Yes	NC_022767.1	*Siphoviridae*
–	Stills	80,798	35.49	110	0	Yes	NC_028777.1	*Siphoviridae*
D1	Page	39,874	40.71	50	0	No	NC_022764.1	*Podoviridae*
–	Palmer	40,000	40.7	50	0	No	NC_028926.1	*Podoviridae*
–	Pavlov	40,020	40.6	50	0	No	NC_028782.1	–
D1	Pony	39,844	40.7	48	0	No	NC_022770.1	*Podoviridae*
–	Pookie	40,214	40.57	52	0	No	NC_027394.1	*Podoviridae*
D1	poppyseed	39,874	40.71	50	0	No	KF669657.1	–
–	Pascal	39,639	39.9	50	0	No	NC_027372.1	*Podoviridae*
–	BM5	165,030	38.1	224	0	Yes	NC_029069.1	*Myoviridae*
–	BM10	165,213	39.6	283	17	No	KT995480.1	*Myoviridae*
L3	Eldridge	1,650,870	40.4	238	0	No	NC_030920.1	–
–	Mater	164,302	39.5	228	5	No	NC_027366.1	–
–	Moonbeam	161,239	40.2	231	3	No	NC_027374.1	–
–	Silence	40,001	38.3	66	0	No	KT001912.1	–
Singleton	G	497,513	29.93	675	18	Yes	NC_023719.1	*Myoviridae*

Comparative genomics of the *Bacillus megaterium* phages

To determine the evolutionary relationship of the 18 sequenced *B. megaterium* phages available in NCBI database, we examined the described phage genomes using complete genome dot plot analysis and pairwise average nucleotide identities (ANI), as previously reported (*Hatfull et al., 2010*; *Pope et al., 2011*; *Grose et al., 2014*). The properties of the 18 sequenced *Bacillus* phages are presented in Table 2.

Each cluster was analyzed by the nucleotide dot plot analysis to reveal groups of high similarity or sub-clusters (Fig. S2). These sub-clusters were selected based on natural divisions in phage similarity seen in the dot plot analysis and defined by ANI similarity values of at least 66% between every two phages within each sub-cluster. Such sub-clustering either represents evolutionary forces that constrain certain types of phages or is an artifact of phage isolation (*Grose et al., 2014*).

As nucleotide sequences diverge more quickly than amino acid sequences, amino acid dot plot analyses reveal functional conservation. On the other hand, for closely related sequences, the use of nucleotide sequences introduces more bias while building a multiple sequence alignment. Single gene product analysis was performed to determine phage cluster, using a large terminase gene product (TerL), which was identified in all known *Bacillus* phages and has been used for this purpose previously (*Grose et al., 2014*). A dot plot alignment of the Terminase gene product (TerL) validated our basic cluster/sub-cluster assignment, with all phages grouping in agreement with their clusters or sub-clusters, and the G and Silence phages remaining singletons (Fig. 4B).
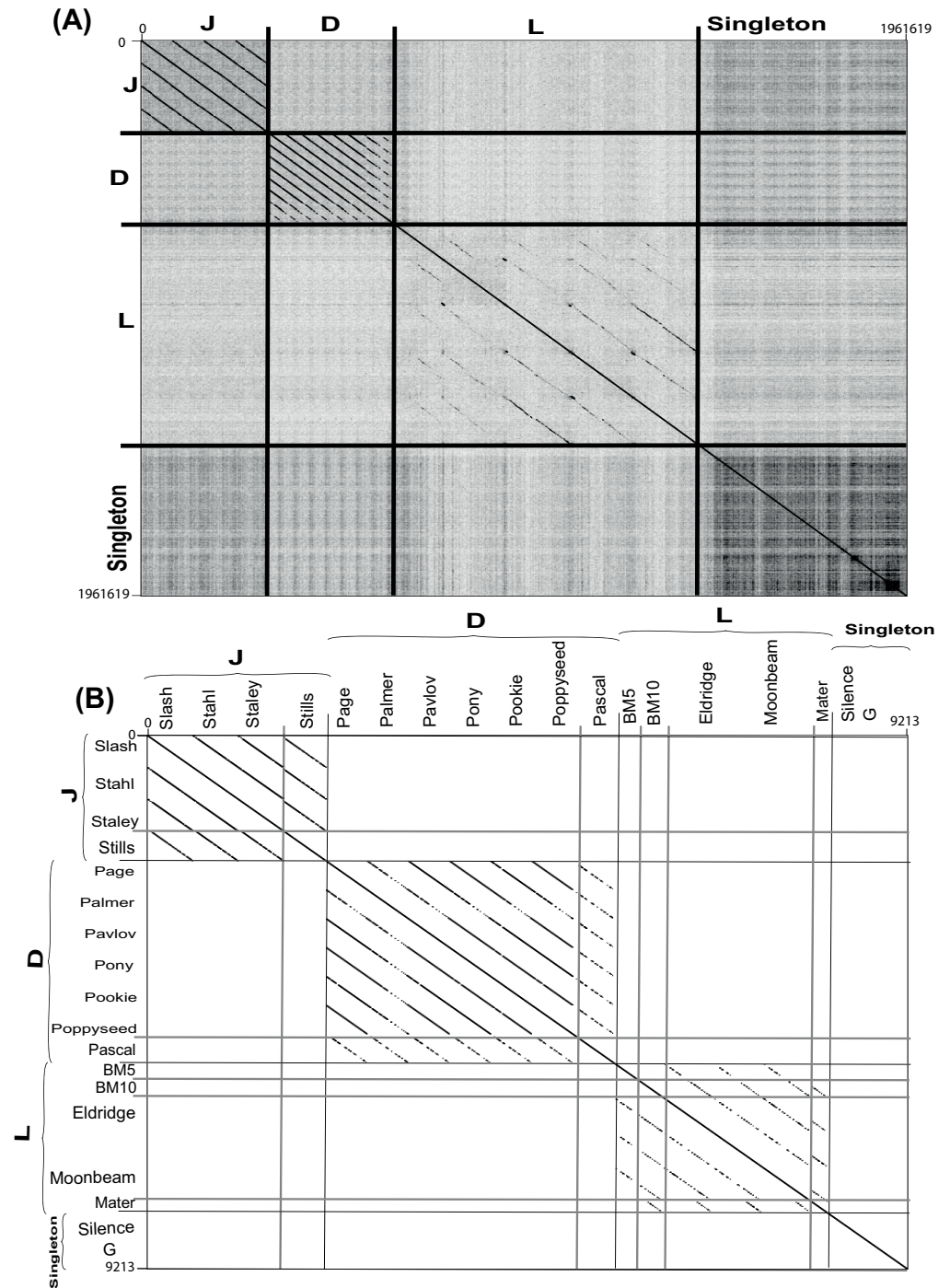
Figure 4 Complete nucleotide genome sequence and single gene amino acid (large terminase) dot plot analysis of *Bacillus megaterium* reveals three clusters (D, L and J) and two singletons. (A) full genome sequence and (B) single gene amino acid. Thick black lines indicate cluster assignments while thick light gray lines indicate cluster assignments, which provided on the *X*- and *Y*-axis. Dot plots created using Gepard (*Krumsiek, Arnold & Rattei, 2007*). Single gene amino acid sequences were chosen from annotated large terminase gene products.

Phylogenetic analysis

Analysis of terminase (TerL) diversity in BM5, BM10 and the other 87 *Bacillus* phages (Fig. 5) provides a robust assignment of their phylogeny (*Casjens, 2011*) and can be used to interpret the phage replication strategy, (*Grose et al., 2014*) and to investigate the hybrid generations between phage genera (*Canchaya, Fournous & Chibani-Chennoufi, 2003*) or families (*Recktenwald & Schmidt, 2002*). The 89 *Bacillus* phages were grouped into five main clades (Fig. 5).

On the phylogenetic tree, we have indicated the numbers of tRNA genes encoded in each phage order to describe its distribution among the *Bacillus* phages. An association between sub-cluster and number of tRNA genes was been identified. Phages without tRNAs were observed in all sub-clusters—L8, L5, H1, J1, F1, K2, and D1. Phages encoding similar numbers of tRNAs were identified in sub-clusters L1 and L4; the same pattern was observed in *Mycobacteriophages* (*Delesalle et al., 2016*). An exception was found in the F2 sub-cluster, where only one member (phiIS3501) contained one tRNA gene. It was described as an unusual pro-phage, different from the rest of its family, containing the tRNA-Met gene located upstream of the lysogenic integrase gene (*Moumen, Nguen-The & Sorokin, 2012*). Also, for identification of the lytic and temperate clusters within the *Bacillus* phage, the presence and absence of integrase genes in each phage was exhibited in the tree. We identified seven lytic sub-clusters (L8, L4, L5, K2, H1, f1 and D1), with all their members lacking the integrase gene, and two temperate sub-clusters (J1 and F2). Interestingly, our sequenced BM5 phage was the only phage containing the lysogenic integrase gene within the L cluster and even within all long DTR *Myoviridae Bacillus* phages. Other temperate phages were previously observed within the lytic *Mycobacteriophages* cluster (*Delesalle et al., 2016*). To confirm clustering of the BM5 and BM10 phages, a whole genome pairwise ANI analysis was performed using JSpecies 1.2.1 between the BM5 and BM10 phages and other phages clustered in the same clade (Table S6). The ANI percentages between the BM5 phages and the L5 cluster phages (Phrodo, Riley, AvesoBmore, BigBertha, and Troll) ranged from 94% to 98%, while it ranged from 86% to 93% between the BM10 phage and the L7 cluster phages (vB_BceM_Bc431v3 and Bcp1).

DISCUSSION

Our isolated phages (BM5 and BM10) show low latent period, which is lower than expected for *B. megaterium* (45–60 min) and other *Bacillus* phages (34–145 min) (*Brodetsky & Romig, 1965*; *Colasito & Rogoff, 1969a*; *Colasito & Rogoff, 1969b*; *Cooney, Jacob & Slepecky, 1975*; *El-Arabi et al., 2013*; *Krasowska, Biegalska & Augustyniak, 2015*). This is probably because of the strong activity of endolysins in the BM5 and BM10 phages, which is likely due to mutations in their genes encoding of holin (*Stephen et al., 2013*) and large burst size, which, in turn, might be caused due to different factors, such as host cell size, culture age, and/or the host physiological conditions (*Abedon, Hyman & Thomas, 2003*; *Choi, Kuatsjah & Wu, 2010*; *Bolger-Munro, Cheung & Fang, 2013*). Moreover, the larger burst size and shorter latency period for both phages reflect their active propagation in the host (*Delesalle et al., 2016*; *Wilson, 1973*).
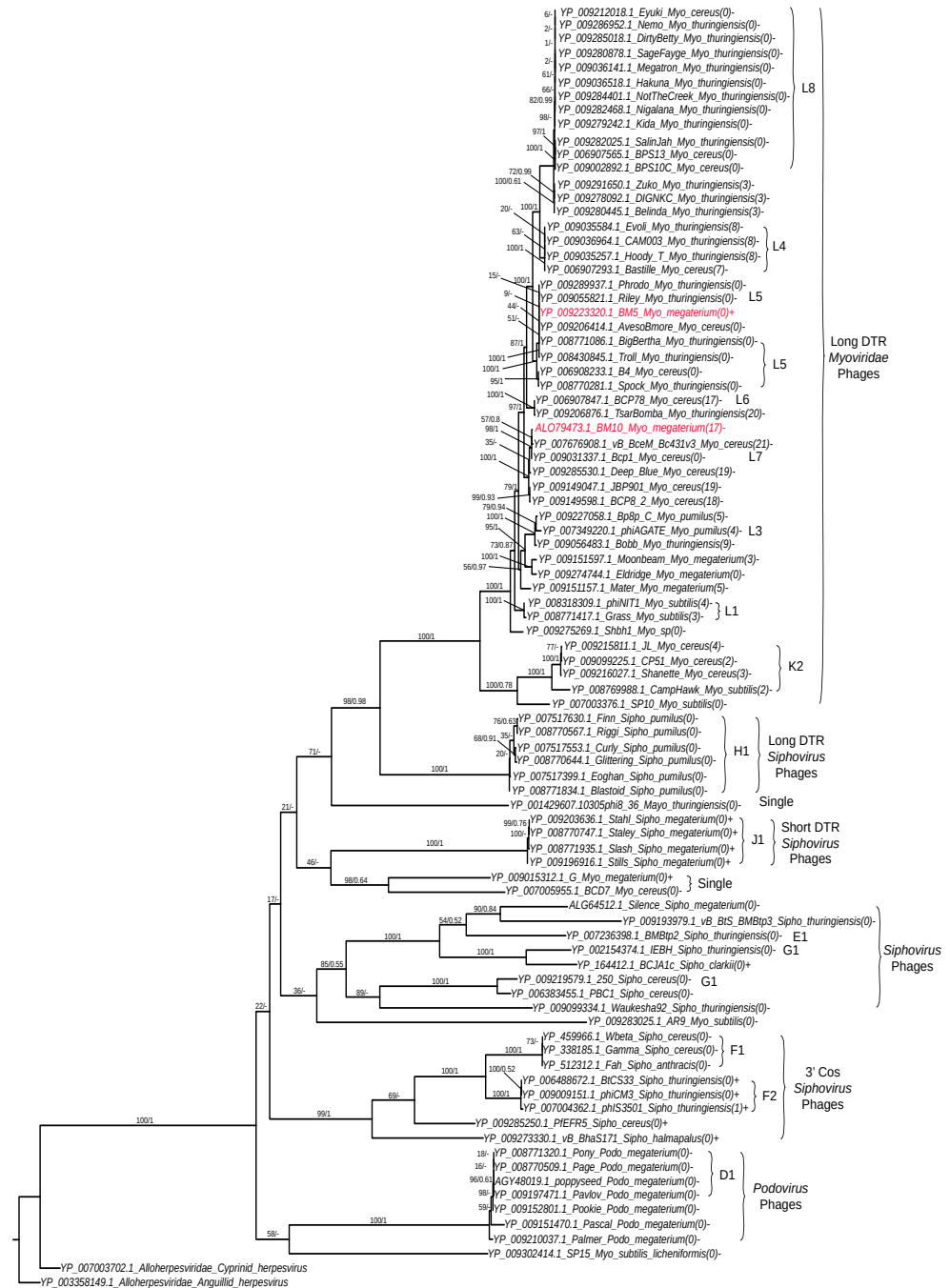
Figure 5 **Phylogenetic tree analysis of the *Bacillus* terminase reflects complete genome cluster.** Sequenced phages, colored in red. Whole genome sub-cluster assignment and phage families, indicated on the right. Replication approaches for phages also, indicated as a direct terminal repeats (DTR) and cohesive ends (cos). Moreover, the presence and absence of integrase genes as (+) or (−), respectively and number of tRNA genes for each phage were indicated as the number between brackets. The phylogenetic tree were constructed using a Maximum Likelihood distance matrix (*Stamatakis, 2014*) and Bayesian Monte Carlo Markov Chain (MCMC) matrix (*Lartillot, Lepage & Blanquart, 2009*). The bootstraps from both methods merged in one rooted tree using the out-group Alloherpesviridae sequences (Anguillid and Cyprinid herpesvirus). Bootstrap values indicate the number of times a node was supported in 2000 resampling replications.

Bioinformatic analyses of the BM5 and BM10 genomes reveal that the genome organization and annotation are in general agreement with other studies of bacteriophages (*Grande et al., 2014*; *Leon-Velarde et al., 2014*). A similar mosaic genome structure was observed in most other phage genomes, indicating the extensive horizontal genetic exchange within the phage communities (*Rohwer & Edwards, 2002*; *Van Dessel et al., 2005*). Moreover, bacteriophage proteins represent the majority of the members of the identified protein families. The obtained identified genes were involved in phage structuring, DNA replication, nucleotide metabolism, lysis, and the repairing of proteins. The presence of the 17th tRNAs in the BM10 phage genome, which has a larger burst size (117 pfu/cell) than the BM5 phage (103 pfu/cell), along with the absence of tRNAs support the hypothesis that bacteriophages with tRNA genes have a larger burst size as a result of their fast propagation in their hosts (*Delesalle et al., 2016*; *Wilson, 1973*). This high number of tRNAs within phage genomes is associated with phage virulence and higher codon usage bias (*Bailly-Bechet, Vergassola & Rocha, 2007*).

The analysis of the genome sequences of the BM5 and BM10 phages revealed high diversity (66.5% ANI) in these phages. Some of the identified phams can be used in evolutionary and phylogenetic studies. As reported for *Bacillus* phages (*Grose et al., 2014*), single, ubiquitous, and semi-conserved genes can be utilized for the cluster prediction, particularly, when the whole genome sequence is unavailable. A close evolutionary relationship between the two isolated phages was emphasized by using the 21 identified phams, and it provided information that might prove useful for the comparing the evolution of the gene complement in *B. megaterium* phages and the phylogenetically related *Bacillus* bacteriophages.

Whole genome comparisons of the *B. megaterium* phages generated three clusters (D, L, and J) with genome similarity values of over 50% (Fig. 4A). Two singletons (G and Silence phage) were identified. The ANI values were also estimated within each cluster and determined to have at least 63% similarity of phages within a single cluster. In addition to showing strong evolutionary relationships, whole genome nucleotide dot plots also showed smaller regions of homology (<50% span length) between phages of different clusters. These local homology regions are supposed to be sites of recombination (Fig. 4A). Likewise, the analysis of the proteins conserved between each cluster phage revealed that cluster J, which contained only temperate phages, contained the highest number of conserved proteins (78 phams), whereas 41 and 13 phams were found in the cluster L and D phages, respectively. Interestingly, 20 phams were identified between the BM5 phage and the cluster D phages (Fig. 2). These results are comparable with those described for the 98 *Bacillus* phages by using the same gene product (TerL) (*Grose et al., 2014*).

The TerL-based phylogenetic analysis supported the clustering and sub-clustering of the genome comparison analysis, and it can be used for conducting further studies of the evolutionary relationships between phage families and phage-packaging strategies. Most *Bacillus* phages that belong to the *Myoviridae* family (specifically, those that belong to clusters H1, K2, L1, L3, L4, L5, L6, L7, and L8) pack their DNA in a similar concatemer structure, thereby resulting in DNAs with long and direct terminal repeats (DTRs). Clusters F1 and F2 (*Siphoviridae Bacillus* phages) are packed with 3′ cos ends. Cluster

J of *Bacillus* phages, which belongs to the *Siphoviridae* family, has short DTRs. This type of prediction can be used as a guide to facilitate the experimental determination of tailed-phage chromosome's end structure. Based on our phylogenetic analysis, we rejected the hypothesis that the lytic phage clusters are more likely to contain tRNA genes than temperate phage clusters that are similar to *Mycobacteriophages* (*Bailly-Bechet, Vergassola & Rocha, 2007*).

CONCLUSION

The present study reports the biological and genome properties of virulent and temperate *B. megaterium* phages. Both the BM5 and BM10 phages displayed high lytic activity to *B. megaterium*. Moreover, both phages lacked repressor determinants to maintain lysogeny by down-regulating lytic promoters and confer superinfection immunity. This increases the potential risk of using *B. megaterium* as a biocontrol and a biofertilizer agent. Moreover, putative tRNAs were identified, thereby revealing the ability of the BM10 phage to infect the other hosts. The first comparative whole genome nucleotide sequences analysis and a large terminase (TerL) protein phylogeny, which reveal the clustering and sub-clustering of *B. megaterium* phages, are presented.

Finally, the study attempts to present the most comprehensive phylogenetic analysis of the *Bacillus* phages, based on terminase amino acids sequences, thereby revealing a robust relationship between the phage families and packing strategies and supporting the fact that the distribution of tRNA genes in the *Bacillus* phage is specific to sub-clusters. On the other hand, our screening of the 87 *Bacillus* phages genomes for integrases and tRNA genes revealed that 66.7% of the 66 lytic *Bacillus* phages lacked tRNA genes while 25% of the 20 temperate *Bacillus* phages contained tRNA genes (Tables S3 and S4). Hence, we do not agree with the assumption that lytic bacteriophages are more likely to contain tRNA genes than temperate bacteriophages.

ACKNOWLEDGEMENTS

The authors would like to thank the Virology Laboratory, Agricultural Microbiology Department, Faculty of Agriculture, Minia University, Minia, Egypt. for providing the studied phages; Center for Research in Agricultural Genomics (CRAG) service laboratory, Barcelona, Spain for providing the sequencing instruments and reagents used in the study; MetaCentrum/CERIT-SC for providing the computing cloud to perform the bioinformatic analyses and Prof. Don Cowan, Director of UP Genomics Research Institute, University of Pretoria for assistance with language editing and scientific reviewing.

Funding

This work was supported by the project Centre for Research of Pathogenicity and Virulence of Parasites (No. CZ.02.1.01/0.0/0.0/16_019/0000759) funded by the European Regional Development Fund (ERDF), and the Ministry of Education, Youth and Sport (MEYS). The

funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Centre for Research of Pathogenicity and Virulence of Parasites: CZ.02.1.01/0.0/0.0/16_019/0000759.
European Regional Development Fund (ERDF), Ministry of Education, Youth and Sport (MEYS).

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Abdoallah Sharaf conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Miroslav Oborník conceived and designed the experiments, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Adel Hammad and Eman Marei conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Sohair El-Afifi conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The complete genome sequences of the phages BM5 and BM10 are available at NCBI GenBank: KT995479 and KT995480. The raw sequencing data are available at Sequence Read Archive (SRA): SRP136802.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5687#supplemental-information.

REFERENCES

Abedon ST, Hyman P, Thomas C. 2003. Experimental examination of bacteriophage latent-period evaluation as response to bacterial availability. *Applied and Environmental Microbiology* **69**:7499–7506 DOI 10.1128/AEM.69.12.7499-7506.2003.

Akoh CC, Lee GC, Liaw YC. 2004. GDSL family of serine esterases/lipases. *Progress in Lipid Research* **43**:534–552 DOI 10.1016/j.plipres.2004.09.002.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Biologia* **215**:403–410.

Argos P, Landy A, Abremski K, Egan JB, Haggard-Ljungquist E, Hoess RH, Kahn ML, Kalionis B, Narayana SV, Pierson LS. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *The EMBO Journal* **5**:433–440 DOI 10.1002/j.1460-2075.1986.tb04229.x.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**:202–208 DOI 10.1093/nar/gkp335.

Bailey TL, Elkan C. 1994. *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Menlo Park: AAAI Press, 28–36.

Bailly-Bechet M, Vergassola M, Rocha E. 2007. Causes for the intriguing presence of tRNAs in phages. *Genome Research* **17**:1486–1495 DOI 10.1101/gr.6649807.

Bergey DH. 2009. *Bergey's manual of systematic bacteriology—vol 3: the firmicutes.* New York Inc.: Springer-Verlag DOI 10.1007/b92997.

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research* **29**:2607–2618.

Blankenberg D, Gordon A, Kuster G. 2010. Von Galaxy Team. *Manipulation of FASTQ Data with Galaxy Bioinformatics* **26**:1783–1785.

Bolger-Munro M, Cheung K, Fang A. 2013. T4 bacteriophage average burst size varies with *Escherichia coli* B23 cell culture age. *Journal of Experimental Microbiology and Immunology* **17**:115–119.

Brodetsky AM, Romig WR. 1965. Characterization of Bacillus subtilis bacteriophages. *Journal of Bacteriology* **90**:1655–1663.

Bunk B, Biedendieck R, Jahn D, Vary PS, Flickinger MC. 2009. *Encyclopedia of Industrial Biotechnology: Bacillus Megaterium and other Bacilli: industrial applications.* Hoboken: John Wiley & Sons, Inc. DOI 10.1002/9780470054581.eib063.

Canchaya C, Fournous G, Chibani-Chennoufi S. 2003. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* **6**:417–424 DOI 10.1016/S1369-5274(03)00086-9.

Carvalho PM, Vary JC. 1977. Isolation and characterization of a Bacillus megaterium QMB1551 bacteriophage. *The Journal of General Virology* **36**:547–550 DOI 10.1099/0022-1317-36-3-547.

Casjens SR. 2011. The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews Microbiology* **9**:647–657 DOI 10.1038/nrmicro2632.

Chan BK, Abedon ST, Loc-Carrillo C. 2013. Phage cocktails and the future of phage therapy. *Future Microbiology* **8**:769–783 DOI 10.2217/fmb.13.47.

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research* **44**:D184–D189 DOI 10.1093/nar/gkv1309.

Choi C, Kuatsjah E, Wu E. 2010. The effect of cell size on the burst size of T4 bacteriophage infections of *Escherichia coli* B23. *Journal of Experimental Microbiology and Immunology* **14**:85–91.

Choi KH, McPartland J, Kaganman I. 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *Journal of Molecular Biology* **378**:726–736 DOI 10.1016/j.jmb.2008.02.059.

Colasito DJ, Rogoff MH. 1969a. Characterization of lytic bacteriophages of Bacillus thuringiensis. *Journal of General Virology* **5**:267–274 DOI 10.1099/0022-1317-5-2-267.

Colasito DJ, Rogoff MH. 1969b. Characterization of temperate bacteriophages of Bacillus thuringiensis. *Journal of General Virology* **5**:275–281 DOI 10.1099/0022-1317-5-2-275.

Cooney PH, Jacob RJ, Slepecky RA. 1975. Characteristics of a Bacillus megaterium bacteriophage. *Journal of General Virology* **26**:131–134 DOI 10.1099/0022-1317-26-1-131.

Cresawn SG, Bogel M, Day N, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* **12**:395 DOI 10.1186/1471-2105-12-395.

Delesalle VA, Tanke NT, Vill AC, Krukonis GP. 2016. Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes. *Bacteriophage* **6**:e1219441 DOI 10.1080/21597081.2016.1219441.

Duda RL, Martincic K, Hendrix RW. 1995. Genetic basis of bacteriophage HK97 prohead assembly. *Journal of Molecular Biology* **247**:636–647 DOI 10.1016/S0022-2836(05)80144-5.

Eisenstein M. 2012. Oxford nanopore announcement sets sequencing sector abuzz. *Nature Biotechnology* **30**:295–296 DOI 10.1038/nbt0412-295.

El-Arabi TF, Griffiths MW, She YM, Villegas A, Lingohr EJ, Kropinski AM. 2013. Genome sequence and analysis of a broad-host range lytic bacteriophage that infects the Bacillus cereus group. *Virology Journal* **10**:48 DOI 10.1186/1743-422X-10-48.

Elmaghraby I, Carimi F, Sharaf A, Marei EM, Hammad AMM. 2015. Isolation and identification of Bacillus megaterium Bacteriophages via AFLP technique. *Current Research in Bacteriology* **8**:77–89 DOI 10.3923/crb.2015.77.89.

Eppinger M, Bunk B, Johns MA, Edirisinghe JN, Kutumbaka KK, Koenig SSK, Creasy HH, Rosovitz MJ, Riley DR, Daugherty S, Martin M, Elbourne LDH, Paulsen I, Biedendieck R, Braun C, Grayburn S, Dhingra S, Lukyanchuk V, Ball B, Ul-Qamar R, Seibe J, Bremer E, Jahn D, Ravel J, Vary PS. 2011. Genome sequences of the biotechnologically important Bacillus megaterium strains QM B1551 and DSM319. *Journal of Bacteriology* **193**:4199–4213 DOI 10.1128/JB.00449-11.

Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. *Bioinformatics* **29**(22):2936–2937 DOI 10.1093/bioinformatics/btt515.

Grande L, Michelacci V, Tozzoli R, Ranieri P, Maugliani A, Caprioli A, Morabito S. 2014. Whole genome sequence comparison of vtx2-converting phages from Enteroaggregative Haemorrhagic *Escherichia coli* strains. *BMC Genomics* **15**:574 DOI 10.1186/1471-2164-15-574.

Grose JH, Jensen GL, Burnett SH, Breakwell DP. 2014. Correction: genomic comparison of 93 Bacillus phages reveals 12 clusters, 14 singletons and remarkable diversity. *BMC Genomics* **15**:1184 DOI 10.1186/1471-2164-15-1184.

Hambly E, Suttle CA. 2005. The viriosphere, diversity, and genetic exchange within phage communities. *Current Opinion in Microbiology* **8**:444–450 DOI 10.1016/j.mib.2005.06.005.

Haq IU, Chaudhry WN, Akhtar MN, Andleeb S, Qadri I. 2012. Bacteriophages and their implications on future biotechnology: a review. *Virology Journal* **9**:9 DOI 10.1186/1743-422X-9-9.

Hatfull GF. 2008. Bacteriophage genomics. *Current Opinion in Microbiology* **11**:447–453 DOI 10.1016/j.mib.2008.09.004.

Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH, Hoyte NN, Bowman CA, Tantoco AT, Paladin EC, Myers MS, Smith AL, Grace MS, Pham TT, O'Brien MB, Vogelsberger AM, Hryckowian AJ, Wynalek JL, Donis-Keller H, Bogel MW, Peebles CL, Cresawn SG, Hendrix RW. 2010. Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *Journal of Molecular Biology* **397**:119–143.

Hyde SJ, Eckenroth BE, Smith BA, Eberley WA, Heintz NH, Jackman JE, Doublié S. 2010. tRNA(His) guanylyltransferase (THG1), a unique 3′-5′ nucleotidyl transferase, shares unexpected structural homology with canonical 5′-3′ DNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **107**:20305–20310 DOI 10.1073/pnas.1010436107.

Irshad LH, Waqas NC, Saadia A. 2012. Isolation and partial characterization of a virulent Bacteriophage IHQ1 specific for Aeromonas punctata from stream water. *Microbial Ecology* **63**:954–963 DOI 10.1007/s00248-011-9944-2.

Johnson PE, Joshi MD, Tomme P. 1996. Structure of the N-terminal cellulose-binding domain of Cellulomonas fimi CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **35**:14381–14394 DOI 10.1021/bi961612s.

Jones CE, Mueser TC, Dudas KC, Kreuzer KN, Nossal NG. 2001. Bacteriophage T4 gene 41 helicase and gene 59 helicase-loading protein: a versatile couple with roles in replication and recombination. *Proceedings of the National Academy of Sciences of the United States of America* **98**:8312–8318 DOI 10.1073/pnas.121009398.

Kamen DE, Woody RW. 2002. Folding kinetics of the protein pectate lyase C reveal fast-forming intermediates and slow proline isomerization. *Biochemistry* **41**:4713–4723 DOI 10.1021/bi0115129.

Khan SM, Jarra W, Preiser PR. 2001. The 235 kDa rhoptry protein of Plasmodium (yoelii) yoelii: function at the junction. *Molecular and Biochemical Parasitology* **117**:1–10 DOI 10.1016/S0166-6851(01)00333-4.

Kildea S, Ransbotyn V, Khan MR, Fagan B, Leonard G, Mullins E, Doohan FM. 2008. Bacillus megaterium shows potential for the biocontrol of septoria tritici blotch of wheat. *Biological Control* **47**:37–45 DOI 10.1016/j.biocontrol.2008.07.001.

Klumpp J, Fouts DE, Sozhamannan S. 2012. Next generation sequencing technologies and the changing landscape of phage genomics. *Bacteriophage* **2**:190–199 DOI 10.4161/bact.22111.

Krasowska A, Biegalska A, Augustyniak D, Łoś M, Richert M, Łukaszewicz M. 2015. Isolation and characterization of phages infecting Bacillus subtilis. *BioMed Research International* **2015**:1–10 DOI 10.1155/2015/179597.

Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**:1026–1028 DOI 10.1093/bioinformatics/btm039.

Lanningan S, Williams DST. 1982. Methods for the direct ssolation and enumeration of Actinophages in soil. *Journal of General Microbiology* **128**:2063–2071 DOI 10.1099/00221287-128-9-2063.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**:2286–2288 DOI 10.1093/bioinformatics/btp368.

Lee CA, Auchtung JM, Monson RE, Grossman AD. 2007. Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of Bacillus subtilis. *Molecular Microbiology* **66**:1356–1369 DOI 10.1111/j.1365-2958.2007.06000.x.

Lee WJ, Billington C, Hudson JA, Heinemann JA. 2011. Isolation and characterization of phages infecting Bacillus cereus. *Letters in Applied Microbiology* **52**:456–464 DOI 10.1111/j.1472-765X.2011.03023.x.

Lee YS, Kennedy WD, Yin YW. 2009. Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* **139**:312–324 DOI 10.1016/j.cell.2009.07.050.

Lee W, McDonough MA, Kotra L, Li Z-H, Silvaggi NR, Takeda Y, Kelly JA, Mobashery S. 2001. A snapshot of the final step of bacterial cell wall biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **98**:1427–1431 DOI 10.1073/pnas.98.4.1427.

Leon-Velarde CG, Kropinski AM, Chen S, Abbasifar A, Griffiths MW, Odumeru JA. 2014. Complete genome sequence of bacteriophage vB_YenP_AP5 which infects Yersinia enterocolitica of serotype O:3. *Virology Journal* **11**:188 DOI 10.1186/1743-422X-11-188.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transferRNA genes in genomic sequence. *Nucleic Acids Research* **25**:955–964 DOI 10.1093/nar/25.5.0955.

Makarova KS, Grishin NV, Shabalina SA. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1**:7 DOI 10.1186/1745-6150-1-7.

Marchler-Bauer A, Derbyshire MK, Gonzales NR. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research* **43**:222–226 DOI 10.1093/nar/gku1221.

Marie EM. 2013. Isolation and characterization of Bacillus subtilis phage from soil cultivated with Liquorices root. *Microbiological Research* **4**:43–49 DOI 10.5829/idosi.ijmr.2013.4.1.7231.

Moumen B, Nguen-The C, Sorokin A. 2012. Sequence analysis of inducible prophage phIS3501 integrated into the Haemolysin II gene of *Bacillus thuringiensis* var israelensis ATCC35646. *Genetics Research International* **2012**:e543286 DOI 10.1155/2012/543286.

Pajunen M, Kiljunen S, Skurnik M. 2000. Bacteriophage Φ Ye, specific for Yersinia enterocolitica serotype 0: 3; is related to coliphages T3 and T7. *Journal of Bacteriology* **182**:5114–5120 DOI 10.1128/JB.182.18.5114-5120.2000.

Pope WH, Jacobs-Sera D, Russell DA, Peebles CL, Al-Atrache Z, Alcoser TA, Alexander LM, Alfano MB, Alford ST, Amy NE, Anderson MD, Anderson AG, Ang AA, Ares Jr M, Barber AJ, Barker LP, Barrett JM, Barshop WD, Bauerle CM, Bayles IM, Belfield KL, Best AA, Borjon Jr A, Bowman CA, Boyer CA, Bradley KW, Bradley VA, Broadway LN, Budwal K, Busby KN, Campbell IW, Campbell AM, Carey A, Caruso SM, Chew RD, Cockburn CL, Cohen LB, Corajod JM, Cresawn SG, Davis KR, Deng L, Denver DR, Dixon BR, Ekram S, Elgin SC, Engelsen AE, English BE, Erb ML, Estrada C, Filliger LZ, Findley AM, Forbes L, Forsyth MH, Fox TM, Fritz MJ, Garcia R, George ZD, Georges AE, Gissendanner CR, Goff S, Goldstein R, Gordon KC, Green RD, Guerra SL, Guiney-Olsen KR, Guiza BG, Haghighat L, Hagopian GV, Harmon CJ, Harmson JS, Hartzog GA, Harvey SE, He S, He KJ, Healy KE, Higinbotham ER, Hildebrandt EN, Ho JH, Hogan GM, Hohenstein VG, Holz NA, Huang VJ, Hufford EL, Hynes PM, Jackson AS, Jansen EC, Jarvik J, Jasinto PG, Jordan TC, Kasza T, Katelyn MA, Kelsey JS, Kerrigan LA, Khaw D, Kim J, Knutter JZ, Ko CC, Larkin GV, Laroche JR, Latif A, Leuba KD, Leuba SI, Lewis LO, Loesser-Casey KE, Long CA, Lopez AJ, Lowery N, Lu TQ, Mac V, Masters IR, McCloud JJ, McDonough MJ, Medenbach AJ, Menon A, Miller R, Morgan BK, Ng PC, Nguyen E, Nguyen KT, Nguyen ET, Nicholson KM, Parnell LA, Peirce CE, Perz AM, Peterson LJ, Pferdehirt RE, Philip SV, Pogliano K, Pogliano J, Polley T, Puopolo EJ, Rabinowitz HS, Resiss MJ, Rhyan CN, Robinson YM, Rodriguez LL, Rose AC, Rubin JD, Ruby JA, Saha MS, Sandoz JW, Savitskaya J, Schipper DJ, Schnitzler CE, Schott AR, Segal JB, Shaffer CD, Sheldon KE, Shepard EM, Shepardson JW, Shroff MK, Simmons JM, Simms EF, Simpson BM, Sinclair KM, Sjoholm RL, Slette IJ, Spaulding BC, Straub CL, Stukey J, Sughrue T, Tang TY, Tatyana LM, Taylor SB, Taylor BJ, Temple LM, Thompson JV, Tokarz MP, Trapani SE, Troum AP, Tsay J, Tubbs AT, Walton JM, Wang DH, Wang H, Warner JR, Weisser EG, Wendler SC, Weston-Hafer KA, Whelan HM, Williamson KE, Willis AN, Wirtshafter HS, Wong TW, Wu P, Yang Yj, Yee BC, Zaidins DA, Zhang B, Züniga MY, Hendrix RW, Hatfull GF. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. *PLOS ONE* **6**:e16329 DOI 10.1371/journal.pone.0016329.

Prieto AI, Kahramanoglou C, Ali RM. 2012. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Research* **40**:3524–3537 DOI 10.1093/nar/gkr1236.

Recktenwald J, Schmidt H. 2002. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. *Infection and Immunity* **70**:1896–1908 DOI 10.1128/IAI.70.4.1896-1908.2002.

Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**:19126–19131 DOI 10.1073/pnas.0906412106.

Rohwer F, Edwards R. 2002. The phage proteomic tree: a genome-based taxonomy for phage. *Journal of Bacteriology* **184**:4529–4535 DOI 10.1128/JB.184.16.4529-4535.2002.

Serwer P, Hayes SJ, Thomas JA, Hardies SC. 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. *Virology journal* **4**:21 DOI 10.1186/1743-422X-4-21.

Sillankorva S, Neubauer P, Azeredo J. 2008. Isolation and characterization of a T7-like lytic phage for Pseudomonas fluorescens. *BioMed Central Biotechnology* **8**:1–11.

Singh A, Poshtiban S, Evoy S. 2013. Recent advances in bacteriophage based biosensors for food-borne pathogen detection. *Sensors* **13**:1763–1786 DOI 10.3390/s130201763.

Sotelo-Mundo RR, Changchien L, Maley F, Montfort WR. 2006. Crystal structures of thymidylate synthase mutant R166Q: structural basis for the nearly complete loss of catalytic activity. *Journal of Biochemical and Molecular Toxicology* **20**:88–92 DOI 10.1002/jbt.20122.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313 DOI 10.1093/bioinformatics/btu033.

Stephen RS, Mahony J, Courtin P, Chapot-Chartier MP, Van Pijkeren JP, Britton RA, Neve H, Heller KJ, Aideh B, Vogensen FK, Van Sinderen D. 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. *Journal of Biological Chemistry* **288**:5581–5590 DOI 10.1074/jbc.M112.444901.

Sutter M, Boehringer D, Gutmann S. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural & Molecular Biology* **15**:939–947 DOI 10.1038/nsmb.1473.

Tanaka N, Shuman S. 2011. RtcB is the RNA ligase component of an *Escherichia coli* RNA repair operon. *Chemistry* **286**:7727–7731.

Telford JC, Yeung JH, Xu G, Kiefel MJ, Watts AG, Hader S, Chan J, Bennet AJ, Moore MM, Taylor GL. 2011. The Aspergillus fumigatus sialidase is a 3-deoxy-D-glycero-D-galacto-2-nonulosonic acid hydrolase (KDNase): structural and mechanistic insights. *Chemistry* **286**:10783–10792.

Tormo-Más MA, Donderis J, García-Caballer M, Alt A, Mir-Sanchis I, Marina A, Penadés JR. 2013. Phage dUTPases control transfer of virulence genes by

a proto-oncogenic G protein-like mechanism. *Molecular Cell* **49**:947–958 DOI 10.1016/j.molcel.2012.12.013.

Tormo-Más MA, Mir I, Shrestha A. 2010. Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *NATURE* **465**:779–782 DOI 10.1038/nature09065.

Van Dessel W, Van Mellaert L, Liesegang H, Raasch C, De Keersmaeker S, Geukens N, Lammertyn E, Streit W, Ann J. 2005. Complete genomic nucleotide sequence and analysis of the temperate bacteriophage VWB. *Virology* **331**:325–337 DOI 10.1016/j.virol.2004.10.028.

Van Elsas JD, Linhares MM, Penido EGC. 1982. Properties of a new Bacillus megaterium bacteriophage with elongated head [isolated from the soil]. *Revista de Microbiologia* **13**.

Van Elsas JD, Penido EG. 1982. Characterization of a new Bacillus megaterium bacteriophage, MJ-1, from tropical soil. *Antonie van Leeuwenhoek* **48**:365–371 DOI 10.1007/BF00418289.

Vary PS, Halsey WF. 1980. Host-range and partial characterization of several new bacteriophages for Bacillus megaterium QM b1551. *Journal of General Virology* **51**:137–146 DOI 10.1099/0022-1317-51-1-137.

Vollmer W, Joris B, Charlier P, Foster S. 2008. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiology Reviews* **32**:259–286 DOI 10.1111/j.1574-6976.2007.00099.x.

Wilson JH. 1973. Function of the bacteriophage T4 transfer RNA's. *Journal of Molecular Biology* **74**:753–757 DOI 10.1016/0022-2836(73)90065-X.

Zeigler DR. 1999. *Bacillus Genetic Stock Center catalog of strains. Part 2: Bacillus thuringiensis and Bacillus cereus.* Columbus: The Bacillus Genetic Stock Center, The Ohio State University.