

RESEARCH ARTICLE

Open Access

# Prophage-like elements present in *Mycobacterium* genomes

Xiangyu Fan, Longxiang Xie, Wu Li and Jianping Xie\*

## Abstract

**Background:** Prophages, integral components of many bacterial genomes, play significant roles in cognate host bacteria, such as virulence, toxin biosynthesis and secretion, fitness cost, genomic variations, and evolution. Many prophages and prophage-like elements present in sequenced bacterial genomes, such as *Bifidobacteria*, *Lactococcus* and *Streptococcus*, have been described. However, information for the prophage of *Mycobacterium* remains poorly defined.

**Results:** In this study, based on the search of the complete genome database from GenBank, the Whole Genome Shotgun (WGS) databases, and some published literatures, thirty-three prophages were described in detail. Eleven of them were full-length prophages, and others were prophage-like elements. Eleven prophages were firstly revealed. They were phiMAV\_1, phiMAV\_2, phiMmcs\_1, phiMmcs\_2, phiMkms\_1, phiMkms\_2, phiBN42\_1, phiBN44\_1, phiMCAN\_1, phiMyesm\_1, and phiW7S\_1. Their genomes and gene contents were firstly analyzed. Furthermore, comparative genomics analyses among mycobacteriophages showed that full-length prophage phi172\_2 belonged to mycobacteriophage Cluster A and the phiMmcs\_1, phiMkms\_1, phiBN44\_1, and phiMCAN\_1 shared high homology and could be classified into one group.

**Conclusions:** To our knowledge, this is the first systematic characterization of mycobacteriophages, their genomic organization and phylogeny. This information will afford more understanding of the biology of *Mycobacterium*.

**Keywords:** Prophage, Mycobacteriophage, Phylogeny, Comparative genomics

## Background

Phages can be divided into virulent or temperate based on their relationship with the host. Temperate phage inserts and integrates into its host genome upon infection, and can reside as quiescent prophage. Prophage does not infect its host and maintains the dormant state [1]. Whole-genome sequencing reveals that prophage DNAs are widespread among bacterial genomes, even up to 20% of the host genome content [2]. Prophages are important genetic components transferred horizontally that can impart bacterial genome variability, evolution, and virulence [1,3]. Some prophage genes contribute to the adaptation of bacteria to their specific ecological niches [3]. This has been demonstrated in many bacteria [1,4,5], but a little is known for *Mycobacterium* prophages.

There is huge gap between the number of mycobacteriophages isolated and cognate prophages found within mycobacteria. To date, there are 3427 mycobacteriophages isolated and 448 of them with genome sequenced. They can be assembled into 20 clusters (A-T) and seven of them are singletons [6,7]. In contrast with large number of sequenced mycobacteriophages, their cognate prophages are poorly defined. Only the following mycobacteriophage sequences have been described. Two prophage-like elements, phiRv1 and phiRv2, have been detected in *Mycobacterium tuberculosis* H37Rv genome [8]; two prophage-like elements, PhiMU01 and PhiMU02, are found within *M. ulcerans* Agy99 genome [9]; 10 putative prophages, named phiMmar01–10, are found in *M. marinum* M and two of them, phiMmar02 and phiMmar08, are full-length prophages [10]; the *M. abscessus* ATCC 19977 chromosome contains a full-length prophage and three prophage-like elements [11]; prophage Araucaria is found in *M. abscessus* subsp. bolletii BD genome [6]; two prophages are found in pathogen *M. abscessus* Strain 47J26 [12]; a potential prophage in *M. abscessus* M93 is described [13]; *M. massiliense*

\* Correspondence: georgex@swu.edu.cn  
Institute of Modern Biopharmaceuticals, State Key Laboratory breeding base of Three Gorges Eco-environment and Bioresources, Eco-Environment Key Laboratory of the Three Gorges Reservoir Region, Ministry of Education, School of Life Sciences, Southwest University, 400715 Chongqing, China

Strain M172 contains putative mycobacteriophage [14]; a 55-kb region encodes a putative prophage in *M. canettii* STB-I [15]; a 40-kb prophage is predicted in addition to two prophage-like elements also are seen in *M. simiae* strain DSM 44165 [16]. Many *Mycobacterium* prophages remain to be characterized. Knowledge regarding their genomic composition, distribution can facilitate the elucidation of the biology of *Mycobacterium*.

In this study, we screened all available *Mycobacterium* complete genomes sequences from GenBank, shotgun assembly sequences from Whole Genome Shotgun (WGS) databases, and searched for mycobacteriophages in published literatures. Together, 33 prophages were described in detail, and 11 of them were previously undocumented prophages among *Mycobacterium* genomes. The genomes, gene contents, comparative genomics studies and the relationships among them were characterized.

## Results and discussion

### Prophages in *Mycobacterium* genomes

Though the identification of prophages from sequenced bacterial genomes is difficult [1], prophage sequences can be found by several approaches. Integrases are well-recognized diagnostic markers for prophages within bacterial genomes [17-23]. Web servers and programs for prophages identification are available [24-28]. In this study, we used an integrated protocol to streamline the identification. Firstly, PHAST (PHAge Search Tool) was used to search *Mycobacterium* genomes. Secondly, the presence or absence of the integrase genes was tested to exclude negative results. Finally, mycobacteriophage sequences were identified based on the homology between prophage ORFs (open reading frames) and known phage genes. Thirty mycobacterial complete genomes (see Additional file 1) were retrieved. Eleven new prophages were identified. The genomic features of these newly identified mycobacteriophages are described in Table 1.

In the WGS databases, some mycobacteria containing prophages are also reported [12-16]. Since the whole genome sequences of these mycobacteria and the specific information of these prophages are not available, we searched for prophages in five mycobacterial shotgun assembly sequences contigs (see Additional file 1) using the method mentioned above. The results showed that prophages were found in some sequences contigs of *M. abscessus* Strain 47J26, *M. abscessus* M93, and *M. massiliense* M172 (Table 1). Prophages previously reported in the genomes of *M. canettii* CIPT 140070007 and *M. simiae* DSM 44165 cannot be detected in our study. With annotated whole genomic sequence, this puzzle might be solved.

Some mycobacteria harboring prophages have been detailed in previous studies [6,8,10,11], which are included in Table 1. Four of them contained in *M. abscessus* ATCC 19977 chromosome are not designated. We named them

phiMAB\_1, phiMAB\_2, phiMAB\_3, and phiMAB\_4, respectively. We noted that two prophage, PhiMU01 and PhiMU02, mentioned in *M. ulcerans* Agy99 genome, lack specific information and cannot be detected.

Overall, thirty-three prophages were described, and six prophages had been mentioned, but without specific information. Eleven prophages were found from the complete genome database; five prophages were retrieved from the WGS databases; seventeen of them were reported prophages with specific sequence information. Their size range was from 6 kb to 80.5 kb. Based on the length of prophage genome (the length of mycobacteriophage genomes is 41,441 bp – 164,602 bp, <http://phagesdb.org/>), 11 prophages can be considered as full-length prophage. The remaining 22 prophages were prophage-like elements. The result showed that small prophage-like elements were more prevalent than putative full-length prophages. The small prophage-like elements might be more stable due to mutational decay and loss of some genes somehow involved in genome excision. Small prophage-like elements were more stable and can be more easily detected than the full-length prophages. Through the tRNA search tool, 19 prophages were integrated into tRNA genes (Table 1). The frequency of tRNA integration was tRNA-Leu (4/19), tRNA-Arg (4/19), tRNA-Val (2/19), tRNA-Lys (2/19), tRNA-Pro (2/19), tRNA-Met (2/19), tRNA-Phe (1/19), tRNA-Gly (1/19), tRNA-Ala (1/19). The genome of *M.sp.* KMS, *M.sp.* MCS, *M. avium* 104, *M. tuberculosis* H37Rv, *M. marinum* M, *M. abscessus* ATCC 19977, *M. abscessus* Strain 47J26, and *M. massiliense* Strain M172 was polylysogenic.

### New prophages of *Mycobacterium* genomes

#### Full-length prophage phiMAV\_1 in the genome of *M. avium* 104

Prophage phiMAV\_1, spanning from MAV\_0779 (integrase gene) to MAV\_0841 (excisionase DNA binding protein), contains sixty-three ORFs (see Additional file 2), and is flanked by two 20-bp repeats (Table 1) reminiscent of *attL* and *attR* sites. There is no predicted tRNA within the prophage. PhiMAV\_1 cannot be categorized into any known phage clusters and might represent new singleton type [29].

Based on Blast-p, 41 phiMAV\_1 ORFs show more or less amino acid sequence similarity to other known phage genes, and 17 can be assigned functionalities based on homology (see Additional file 2). PhiMAV\_1 genome consists of different functional modules (Figure 1).

The lysis module consists of MAV\_0786 and MAV\_0787, which encode cutinase and glycosyl hydrolase respectively that can lyse bacterium and enable the release of progeny phages. The DNA packaging and structural modules extend from MAV\_0795 to MAV\_0813. MAV\_0795, MAV\_0797, and MAV\_0803 all encode putative tail protein. MAV\_0798 and MAV\_0799 all encode putative structural protein. MAV\_0800, MAV\_0802, and

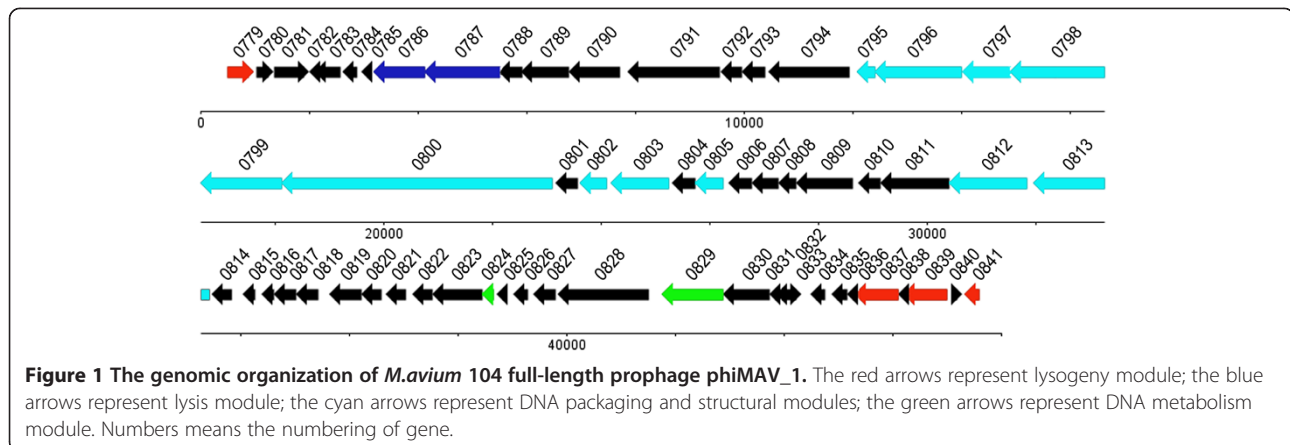
**Table 1 Genomic features of prophages in *Mycobacterium* genomes**

Prophages	Host	Coordinate	Insertion sites	Size	Putative <i>attB</i> regions of prophage-like elements	References
phiMAV_1	<i>M.avium</i> 104	746,437-794,445	-	48.0 kb	GGACCTGCGGATTAAGTC	This study
phiMAV_2		1,446,840-1,463,298	-	16.5 kb	GTGGTCAGCTT	This study
phiMmcs_1	<i>M.sp.</i> MCS	3,067,419-3,080,311	tRNA-Pro	12.9 kb	CCGTTGCCGT	This study
phiMmcs_2		4,054,082-4,063,334	-	9.3 kb	GGTGAGGGCGT	This study
phiMkms_1	<i>M.sp.</i> KMS	3,085,307-3,098,199	-	12.9 kb	CCGTTGCCGT	This study
phiMkms_2		4,088,564-4,097,816	-	9.3 kb	GGTGAGGGCGT	This study
phiBN42_1	<i>M.canettii</i> CIPT 140070010	1,514,379-1,522,784	tRNA-Arg	8.4 kb	GTGCCCCGGCAGGATTCG	This study
phiBN44_1	<i>M.canettii</i> CIPT 140060008	3,433,693-3,444,793	-	11.1 kb	CCGAGAAGAAGTCATGGTTCT	This study
phiMCAN_1	<i>M.canettii</i> CIPT 140010059	1,180,578-1,191,783	-	11.2 kb	GTTTCGAGTCCGACTGGGGCAC	This study
phiMyesm_1	<i>M.smegmatis</i> JS623	4,221,770-4,232,715	-	10.9 kb	GCCGACGACG	This study
phiW7S_1	<i>M.sp.</i> MOTT36Y	968,527-980,490	tRNA-Ala	12.0 kb	GTTTCGATCGAGTAGGTCAGGGTTCGATTCCC	This study
phiRv1 <sup>a</sup>	<i>M.tuberculosis</i> H37Rv	1,779,267-1,788,525	-	9.3 kb	GGTTGGCCGTGG	[8]
phiRv2 <sup>a</sup>		2,970,063-2,980,853	tRNA-Val	10.8 kb	CCGCGCAATAAACGCGCAATA	[8]
phiMU01 <sup>a</sup>	<i>M.ulcerans</i> Agy99	NM	NM	18 kb	NM	[9]
phiMU02 <sup>a</sup>		NM	NM	24 kb	NM	[9]
phiMmar01 <sup>a</sup>	<i>M.marinum</i> M	27,047- 33,095	tRNA-Leu	6.0 kb	NM	[10]
phiMmar02 <sup>a</sup>		4,812,334- 4,869,620	tRNA- Lys	57.8 kb	NM	[10]
phiMmar03 <sup>a</sup>		5,460,072- 5,470,770	tRNA- Leu	10.7 kb	NM	[10]
phiMmar04 <sup>a</sup>		5,628,721- 5,636,467	tRNA- Arg	7.7 kb	NM	[10]
phiMmar05 <sup>a</sup>		5,884,651- 5,904,290	tRNA-Phe	19.6 kb	NM	[10]
phiMmar06 <sup>a</sup>		2,567,503- 2,589,174	-	21.6 kb	NM	[10]
phiMmar07 <sup>a</sup>		3,082,858- 3,100,046	tRNA- Pro	17.2 kb	NM	[10]
phiMmar08 <sup>a</sup>		3,808,513-3,851,917	tRNA-Leu	43.0 kb	NM	[10]
phiMmar09 <sup>a</sup>		688,611-695,966	tRNA-Gly	7.4 kb	NM	[10]
phiMmar10 <sup>a</sup>		4,405,758- 4,430,810	tRNA-Met	25.1 kb	NM	[10]
Prophage Araucaria <sup>a</sup>	<i>M.abscessus</i> subsp. bolletii BD Contig17	NM (1,000-65,113)	-	64.1 kb	NM	[6]
phiMAB_1 <sup>#</sup>	<i>M.abscessus</i> ATCC 19977	1,754,551-1,835,095	tRNA-Met	80.5 kb	NM	[11]
phiMAB_2 <sup>#</sup>		233,621-247,981	tRNA-Arg	14.4 kb	NM	[11]
phiMAB_3 <sup>#</sup>		770,916-778,753	tRNA-Lys	7.8 kb	NM	[11]
phiMAB_4 <sup>#</sup>		4,909,957-4,959,626	-	49.7 kb	NM	[11]
phiMAB47J26_1 <sup>#</sup>	<i>M.abscessus</i> Strain 47 J26 Contig02	NM (55,744-116,295)	NM (tRNA-Leu)	60.5 kb	NM (GCGGACTTAAATCCGCCAAGTGTCTCGGTTTCGAGTCCGACTGGGGCAC)	[12]
phiMAB47J26_2 <sup>#</sup>	<i>M.abscessus</i> Strain 47 J26 Contig03	NM (58,656-105,841)	-	47.2 kb	NM	[12]
phiOUW_1 <sup>#</sup>	<i>M.abscessus</i> M93 Contig09	NM (267,663-342,276)	NM (tRNA-Arg)	74.6 kb	NM (GTGCGCCCGAAGGGATTCTGAACCCCTAACCTTCTG)	[13]

**Table 1 Genomic features of prophages in *Mycobacterium* genomes (Continued)**

phiM172_1 <sup>#</sup>	<i>M.massiliense</i> Strain M172 Contig06	NM (358,576-422,610)	NM (tRNA-Val)	64.0 kb	NM (TTGGTGGGCGCGAGGGTTTCGAACCC)	[14]
phiM172_2 <sup>#</sup>		NM (456,450-513,710)	-	57.3 kb	NM (CAACCAGTCGGCCTGA)	[14]
phiSTB-1 <sup>#</sup>	<i>M.canettii</i> STB-1 (CIPT 140070007)	NM	tRNA-Lys	55 kb	NM	[15]
phiDSM_1 <sup>#</sup>	<i>M.simiae</i> DSM 44165	NM	NM	40 kb	NM	[16]
phiDSM_2 <sup>#</sup>		NM	NM	7 kb	NM	[16]
phiDSM_3 <sup>#</sup>		NM	NM	18 kb	NM	[16]

NM means that these data do not be mentioned; parentheses means that these data is shown in this study; – means these prophages are not integrating into tRNA genes; <sup>a</sup>those prophages has been described and named; <sup>#</sup>those prophages has been described, but did not be named.



MAV\_0805 encode phage tail tape measure protein, tail assembly chaperone, and phage capsid and scaffold protein. MAV\_0812 and MAV\_0813 encode putative portal protein and phage terminase engaged in the phage head morphogenesis. The DNA metabolism module includes MAV\_0824 and MAV\_0829. MAV\_0824 encodes exonuclease and MAV\_0829 encodes recombination and repair protein RecT. The lysogeny module consists of MAV\_0837, MAV\_0839, MAV\_0841 and MAV\_0779. MAV\_0779 and MAV\_0841 encode phage integrase and excisionase DNA binding protein. Both MAV\_0837 and MAV\_0839 encode phage antirepressor protein.

In addition to ORFs similar to other phage genes, two ORFs show unexpected similarity to bacterial key proteins. MAV\_0835 encodes type VI secretion protein IcmF (Intracellular Multiplication F), a core component of type VI secretion system in *Pseudomonas aeruginosa*, *Vibrio cholerae* or other pathogenic bacteria [30-32]. Based on Blast-p, type VI secretion system was not documented in mycobacteria except for *M. avium* 104 and *M. parascrofulaceum*. IcmF is involved in bacterial motility, adherence to epithelial cells, and conjugation frequency [31], and has been reported in an avian pathogenic *Escherichia coli* (APEC) strain [32]. In addition, MAV\_0790 encodes PPE family protein, a widespread *Mycobacterium* unique protein. This implies that MAV\_0835 and MAV\_0790 play a role in the physiology and pathogenicity of *M. avium* 104.

#### Prophage-like elements phiMAV\_2

Prophage phiMAV\_2 (Figure 2), integrated into a hypothetical gene (MAV\_1505) in *M. avium* 104, extends from MAV\_1484 (integrase gene) to MAV\_1504 (Phage terminase) and contains 21 ORFs (see Additional file 3) flanked by an 11-bp repeat (Table 1), indicative of *attL* and *attR* sites. No tRNA is found in the genome of phiMAV\_2. Based on Blast-p, only nine ORFs have sequence similarity to other phage genes at the amino acid sequence level. Six ORFs of the phiMAV\_2 prophage genome can be assigned function based on database search, namely the

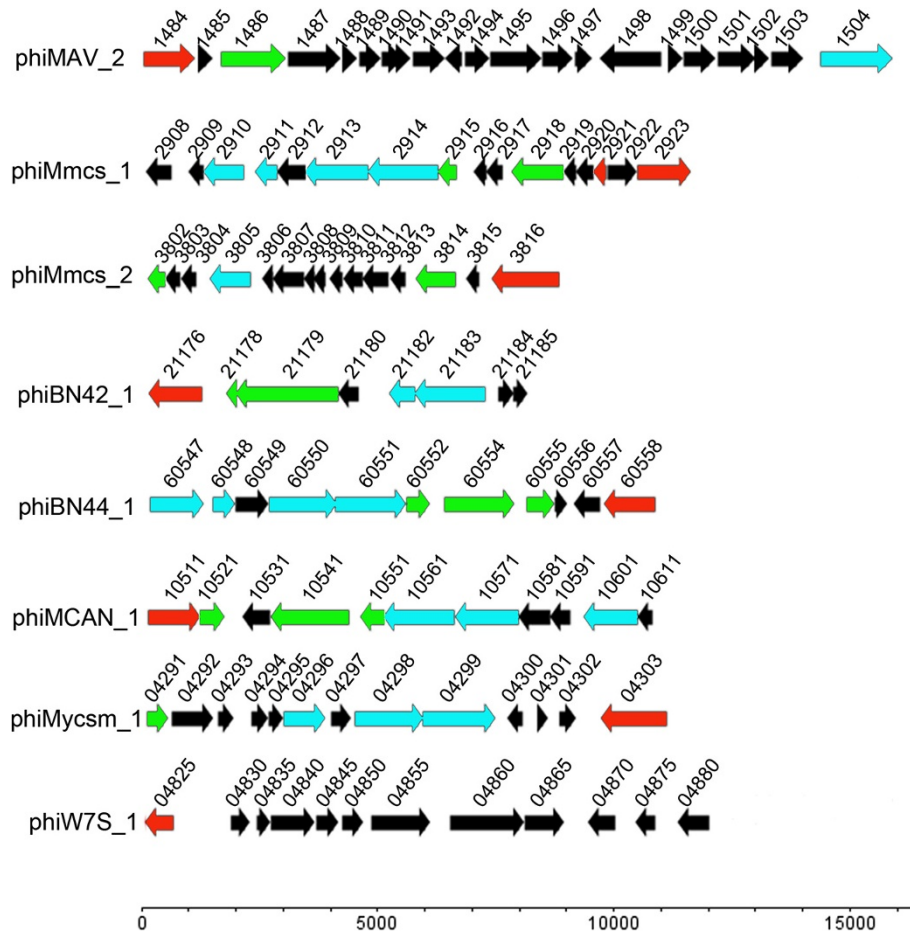
integrase gene (MAV\_1484), response regulator receiver protein (MAV\_1485), DNA primase/polymerase (MAV\_1486), Y4cG protein (MAV\_1493), transposase (MAV\_1498) and phage terminase (MAV\_1504). Other phiMAV\_2 prophage ORFs similar to known bacterial functional proteins are also identified (see Additional file 3).

#### Prophage-like elements phiMmcs\_1, phiMmcs\_2, phiMkms\_1, and phiMkms\_2

There are two prophage-like elements in *M.sp.MCS*, phiMmcs\_1 and phiMmcs\_2. Prophage phiMmcs\_1 (Figure 2), which is integrated into a tRNA-pro (Mmcs\_R0021) in *M.sp.MCS*, extends from Mmcs\_2923 (integrase gene) to Mmcs\_2908 (transglycosylase-like protein) and contains sixteen ORFs (see Additional file 4) flanked by a 10-bp repeat (Table 1), indicative of *attL* and *attR* sites. No tRNA is found in the genome of phiMmcs\_1. Only nine ORFs can be assigned function based on amino acid sequence homology. The prophage phiMmcs\_1 genome contains 4 modules. The lysis module appeared to be limited to Mmcs\_2908, whose protein product has 50% sequence identity to lysin of *Rhodococcus phage* REQ1. The structural module consists of Mmcs\_2910 and Mmcs\_2914. Mmcs\_2910, Mmcs\_2911, Mmcs\_2913, and Mmcs\_2914 encode phage major capsid protein, scaffolding protein, phage portal protein, and phage terminase, respectively. The DNA metabolism module has two genes (Mmcs\_2915 and Mmcs\_2918), whose predicted protein products are HNH endonuclease and DNA repair protein RadA, respectively. The lysogeny module consists of Mmcs\_2921 (putative phage excisionase) and Mmcs\_2923 (phage integrase).

The phiMmcs\_2 prophage remnant inserts between Mmcs\_3803 and Mmcs\_3817. The prophage sequence contains 15 ORFs (see Additional file 5) and is flanked by two 11-bp repeats, indicating the existence of putative *attL* and *attR* sites. Based on Blast-p, only 8 ORFs have sequence similarity to other phage genes at the amino acid sequence level and 4 can be assigned function, namely Mmcs\_3802 (HNH endonuclease), Mmcs\_3805





**Figure 2 Genomic organization of some defective prophage-like elements among mycobacteria.** Numbers means the numbering of gene. The red arrows represent lysogeny module; the blue arrows represent lysis module; the cyan arrows represent DNA packaging and structural modules; the green arrows represent DNA metabolism module.

(phage major capsid protein), Mmcs\_3814 (HNH endonuclease domain-containing protein), and Mmcs\_3816 (phiRv1 integrase).

PhiMkms\_1 and phiMkms\_2 (see Additional files 6 and 7) are prophage-like elements in *M.sp.KMS*. PhiMmcs\_1 is identical to phiMkms\_1 and represents same prophage. They also insert into the same location in host genome. PhiMmcs\_2 and phiMkms\_2 is just the same scenario as phiMkms\_1 and phiMkms\_2.

#### Prophage-like elements phiBN42\_1, phiBN44\_1, and phiMCAN\_1

PhiBN42\_1, phiBN44\_1, and phiMCAN\_1 are found in *M. canettii* CIPT 140070010, *M.canettii* CIPT 140060008, and *M.canettii* CIPT 140010059 respectively. Prophage phiBN42\_1 (Figure 2), which is integrated into a tRNA-arg (BN42\_tRNA41) in *M.canettii* CIPT 140070010, extends from BN42\_21176 (integrase gene) to BN42\_21185 (hypothetical protein) and contains only eight ORFs (see Additional file 8) flanked by a 19-bp repeat (Table 1),

indicative of *attL* and *attR* sites. No tRNA is found in the genome of phiBN42\_1. Only seven genes have sequence similarity to other phage genes, five of which can be assigned function. There are BN42\_21176 (integrase), BN42\_21178 (excisionase), BN42\_21179 (DNA primase), BN42\_21182 (phage prohead protease), and BN42\_21183 (phage major capsid protein).

The phiBN44\_1 prophage remnant is located between BN44\_60546 and BN44\_60559 in *M.canettii* CIPT 140060008, flanked by a 22-bp repeat (Table 1), representing candidates for the *attL* and *attR* sites. There are 11 ORFs in phiBN44\_1 prophage genome (see Additional file 9). Eight are similar to other phage genes and can be assign function. There are BN44\_60547 (phage major capsid protein), BN44\_60548 (scaffolding protein), BN44\_60550 (Phage portal protein), BN44\_60551 (Phage Terminase), BN44\_60552 (HNH endonuclease), BN44\_60554 (DNA primase), BN44\_60557 (XRE family transcriptional regulator), and BN44\_60558 (phage integrase). Additionally, BN44\_60555 encodes protein similar to *Human adenovirus* DNA

polymerase and BN44\_60556 encodes protein similar to K<sup>+</sup> transporter of many bacteria.

Prophage phiMCAN\_1 (Figure 2), which is integrated into between MCAN\_10501 and MCAN\_10621 in *M. canettii* CIPT 140010059, contains only 11 ORFs flanked (see Additional file 10) by a 22-bp repeat (Table 1), indicative of *attL* and *attR* sites. No tRNA is found in the genome of phiMCAN\_1. Only 8 ORFs similar to other phage genes at the amino acid sequence level and seven genes have been assigned function. There are MCAN\_10511 (phage integrase), MCAN\_10521 (DNA-binding protein), MCAN\_10541 (DNA primase), MCAN\_10551 (HNH endonuclease), MCAN\_10561 (phage terminase), MCAN\_10571 (phage portal protein), and MCAN\_10601 (phage major capsid protein).

**Prophage-like elements phiMycsm\_1 and phiW7S\_1**

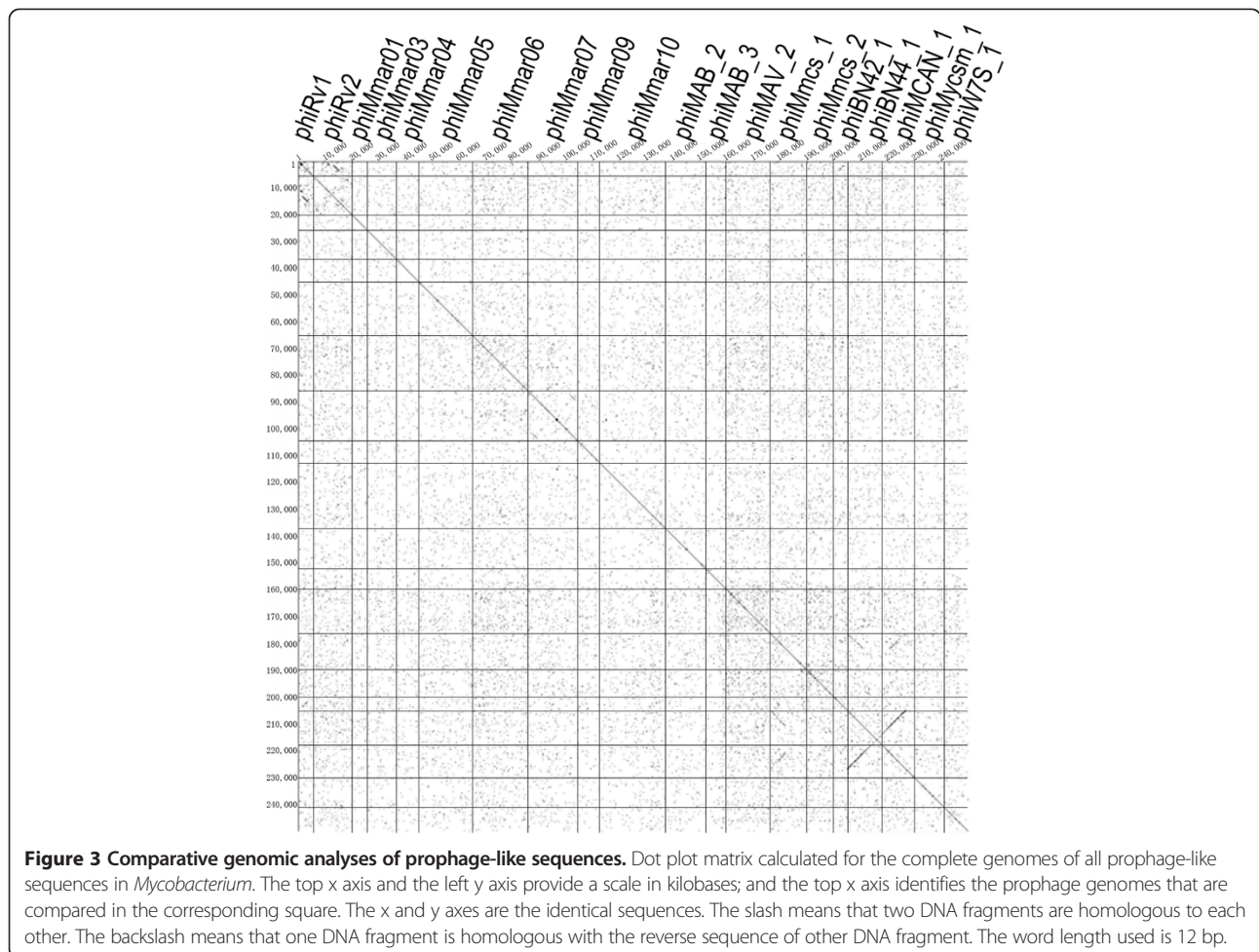
Prophage phiMycsm\_1 (Figure 2), inserted between Mycsm\_04290 and Mycsm\_04304 in *M. smegmatis* JS623, contains 13 ORFs (see Additional file 11) flanked by a 10-bp repeat (Table 1), indicative of *attL* and *attR* sites. No tRNA is found in the genome of phiMycsm\_1. Nine

ORFs show the protein sequence similarity to other phage genes, in which six ORFs have the descriptive function: Mycsm\_04291 (phage integrase), Mycsm\_04296 (DNA-binding protein), Mycsm\_04298 (DNA primase), Mycsm\_04299 (HNH endonuclease), Mycsm\_04302 (phage terminase), and Mycsm\_04303 (phage portal protein). Additionally, Mycsm\_04293, whose protein product is similar to glycerate kinase, is also present in phiBN44\_1.

Prophage phiW7S\_1 (Figure 2) integrated into a tRNA-ala (W7S\_t25871) in *M.sp.* MOTT36Y, extends from W7S\_04825 (integrase gene) to W7S\_04880 (hypothetical protein) and contains 12 ORFs (see Additional file 12) flanked by a 33-bp repeat (Table 1), indicative of *attL* and *attR* sites. No tRNA is found in the genome of phiW7S\_1. Only six genes have sequence similarity to other phage genes and three of them have annotated function, which are W7S\_04825 (integrase), W7S\_04845 (pantothenate kinase), and W7S\_04855 (transposase).

**Grouping of full-length prophages**

We searched all the literatures published so far about full-length mycobacteriophages. Only one prophage



Araucaria is assigned to a Dori-like prophage [6]. BlastN (<http://phagesdb.org/blast/>) and dot plot matrix of the genomes of full-length mycobacteriophages and mycobacteriophage clusters (A-T and singletons) revealed that phi172\_2 shared sequence similarity to cluster A (see Additional file 13); phiMAB\_1 shared an even weaker sequence similarity to subcluster F1 (see Additional file 14); phiMAB47J26\_1 shared an even weak sequence similarity to subcluster F1 and cluster N (see Additional file 15); phiMAB47J26\_2 shared an even weak sequence similarity to cluster P, subcluster F1, and cluster N (see Additional file 16); phi172\_1 shared an even weaker sequence similarity to subcluster F1 and cluster N (see Additional file 17). The remaining full-length prophages had no close relatives to any cluster. We proposed that phi172\_2 was grouped into cluster A, and other full-length mycobacteriophages did not belong to any mycobacteriophage clusters and were 'singletons'.

#### Comparative genomics of prophage-like elements

Dot plot matrix was generated for the complete genomes of 22 mycobacteriophage-like elements in this study (Figure 3). The figure displays that phiMmcs\_1, phiMkms\_1, phiBN44\_1, and phiMCAN\_1 are more closely related to each other than to other mycobacteriophage-like elements, and can be classified as one group. In a simple NCBI 'Align two sequences' comparison, the comparison between phiMmcs\_1 (or phiMkms\_1) and phiBN44\_1 shows that one of the major segments less than 2801 bp has greater than 71% identity, and four

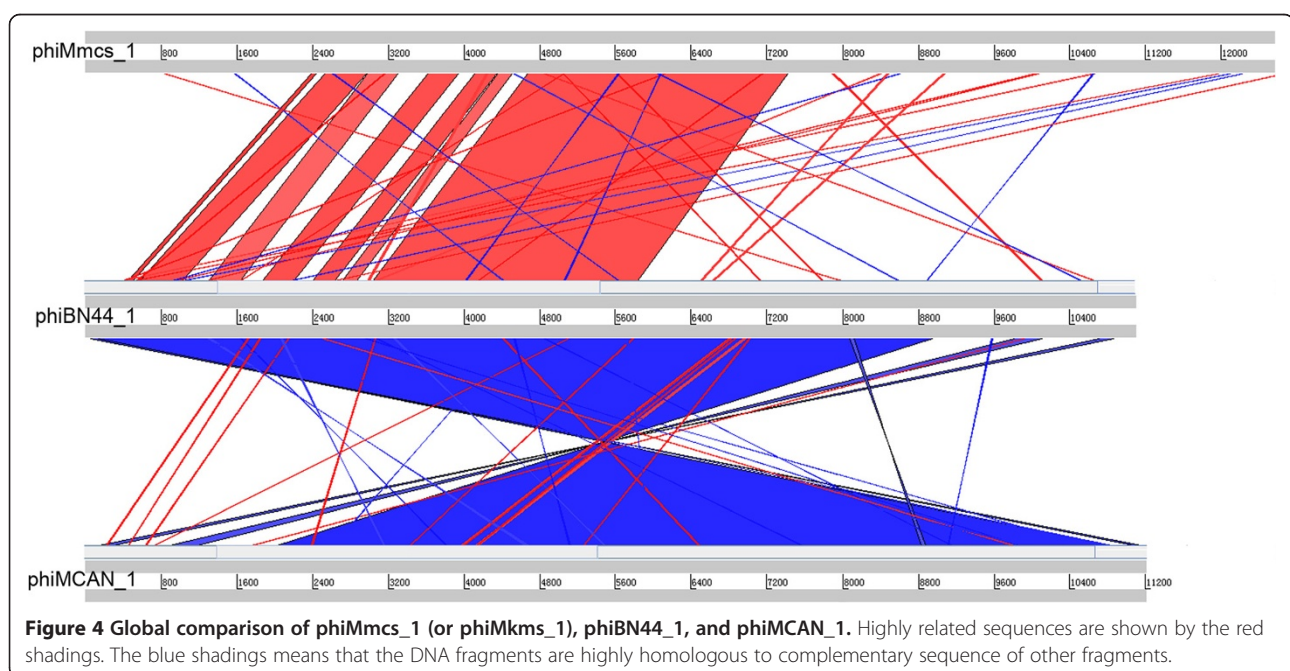
segments less than 200 bp are reported to have 68% identity (Figure 4). The comparison between reverse complementary sequence of phiMCAN\_1 and phiBN44\_1 shows that one of the major segments 8952 bp has greater than 85% identity (Figure 4). Further analysis indicated a lack of homology between the prophage of *M.tuberculosis* H37Rv and other prophage-like elements.

#### Phylogeny of prophage integrases

Integrase can be found in virtually each prophage genome found in this study. And it can serve as good marker for the phylogeny of prophage phiRv1 element encodes a serine site-specific recombinase and phiRv2 encodes a tyrosine recombinase [33]. All integrases fall into the two categories (Figure 5). The serine recombinase division includes phiMymsm\_1, phiMmcs\_2 (phiMkms\_2) and phiRv1. The tyrosine recombinase division includes the remaining prophages and phiRv2. PhiMmcs\_1 (phiMkms\_1), phiBN44\_1, and phiMCAN\_1 belong to the same clade, consistent with the comparative genomic result. The distance between prophages had little relevance to the phylogeny between their hosts, suggestive of independent evolutionary trajectory.

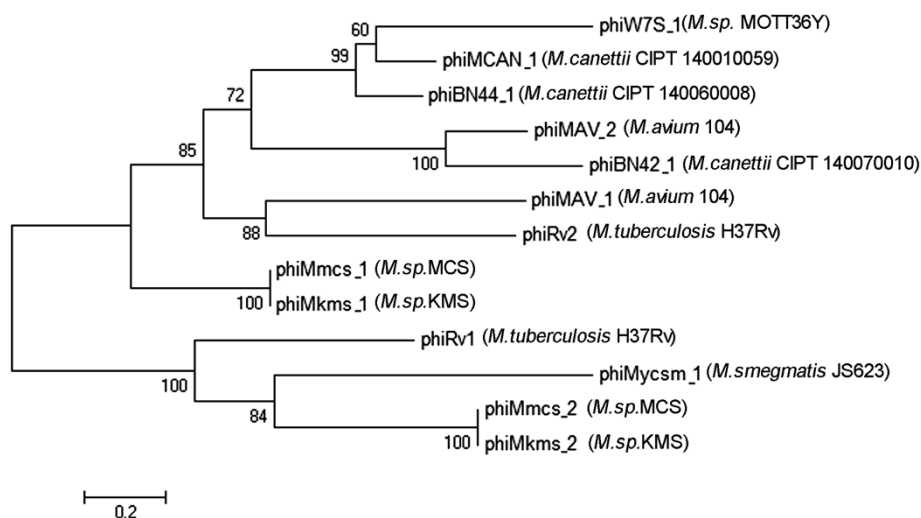
#### Conclusions

In brief, we present here thirty-three mycobacteriophages mined from sequenced mycobacterial genomes, the WGS databases, and some published literatures. Eleven prophages were newly identified prophages from complete genome database; five prophages were from the WGS databases; seventeen prophages were reported with specific



**Figure 4** Global comparison of phiMmcs\_1 (or phiMkms\_1), phiBN44\_1, and phiMCAN\_1. Highly related sequences are shown by the red shadings. The blue shadings means that the DNA fragments are highly homologous to complementary sequence of other fragments.





**Figure 5 Phylogeny of prophage integrases.** Unrooted phylogenetic relationships are represented using NJTree. Bootstrap values from 1,000 reiterations are shown.

sequence information. The genome sequences, gene contents of eleven newly identified prophages were analyzed. Comparative genomic analysis revealed that one full-length mycobacteriophage phi172\_2 belonged to cluster A and one group having recognizable sequence similarity was verified and contained four small prophage-like elements, including the phiMmcs\_1, phiMkms\_1, phiBN44\_1, and phiMCAN\_1. To our knowledge, this represents the first systematic analysis of mycobacteriophages. With more forthcoming *Mycobacterium* genome sequences and thorough mycobacteriophages screening, we can generate a more comprehensive picture of the role of prophages in mycobacterial evolution, adaptations and physiology.

## Methods

### Data collection and mycobacteriophage identification

DNA sequences of bacteria for analysis were downloaded from multiple databases, such as NCBI (the National Center for Biotechnology Information). PHAST (<http://phast.wishartlab.com/index.html>) were firstly used for analyzing bacterial genome to find candidate prophages [24]. An integrase gene was screened from candidate prophage genome for in these results to drop false negative results [17-20]. Finally, prophages were identified on the basis of the presence of significant homology between ORFs (open reading frames) and known phage genes [17].

### Analysis of mycobacteriophage genome sequence

Prophage sequence was annotated using a variety of programs including Glimmer [34]. tRNA and tmRNA genes were identified using tRNA-Scan-SE (<http://lowelab.ucsc.edu/tRNAscan-SE/>) [35] and ARAGORN (<http://mbio-serv2.mbioekol.lu.se/ARAGORN/>) [36]. BLAST analyses were performed remotely at the NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and the phagesdb.org site (<http://phagesdb.org/blast/>).

Some data about mycobacteriophage genomes was downloaded from the phagesdb.org site (<http://phagesdb.org/>). DNAMAN was used to searching the flank of prophage to find *attL* and *attR* sites. Sequences were submitted entries to the GenBank sequence database by Sequin (<http://www.ncbi.nlm.nih.gov/projects/Sequin/index.html>). Comparative genomic analyses of prophage could be carried out by Blast-N for the global comparison of phiMmcs\_1 (or phiMkms\_1), phiBN44\_1, and phiMCAN\_1 and Geneious software for the dotplot of all the mycobacteriophage-like sequences [37]. Multiple sequence alignment and the construct of phylogenetic trees were performed using ClustalW (<http://embnet.vital-it.ch/software/ClustalW.html>) or MEGA4 [38].

## Additional files

**Additional file 1: Table S1.** Mycobacterial genomes retrieved in this study.

**Additional file 2: Table S2.** Database matches for phiMAV\_1.

**Additional file 3: Table S3.** Database matches for phiMAV\_2.

**Additional file 4: Table S4.** Database matches for phiMmcs\_1.

**Additional file 5: Table S5.** Database matches for phiMmcs\_2.

**Additional file 6: Table S6.** Database matches for phiMkms\_1.

**Additional file 7: Table S7.** Database matches for phiMkms\_2.

**Additional file 8: Table S8.** Database matches for phiBN42\_1.

**Additional file 9: Table S9.** Database matches for phiBN44\_1.

**Additional file 10: Table S10.** Database matches for phiMCAN\_1.

**Additional file 11: Table S11.** Database matches for phiMycsm\_1.

**Additional file 12: Table S12.** Database matches for phiW7S\_1.

**Additional file 13: Figure S1-S11.** Comparative genomic analyses of phi172\_2 and cluster A (subcluster A1-A11) mycobacteriophage.

**Additional file 14: Figure S12.** Comparative genomic analyses of phiMAB\_1 and subcluster F1 mycobacteriophage.

**Additional file 15: Figure S13-S14.** Comparative genomic analyses of phiMAB47J26\_1, subcluster F1 and cluster N mycobacteriophage.

**Additional file 16: Figure S15-S17.** Comparative genomic analyses of phiMAB47J26\_2, cluster P, subcluster F1 and cluster N mycobacteriophage.

**Additional file 17: Figure S18-S19.** Comparative genomic analyses of phi172\_1, subcluster F1 and cluster N mycobacteriophage.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

XF participated in the design of the study, analyzed data and wrote the paper. LX and WL helped to modify the manuscript. JX designed the research and wrote the paper. All authors read and approved the final manuscript.

#### Acknowledgements

This work was supported by National Natural Science Foundation [grant numbers 81371851, 81071316, 81271882 and 81301394], New Century Excellent Talents in Universities [grant number NCET-11-0703], National Megaprojects for Key Infectious Diseases [grant numbers 2008ZX10003-006], Excellent PhD thesis fellowship of southwest university [grant numbers kb2010017, ky2011003], the Fundamental Research Funds for the Central Universities [grant numbers XDJK2011D006, XDJK2012D011, XDJK2012D007, XDJK2013D003 and XDJK2014D040], The Chongqing municipal committee of Education for postgraduates excellence program [grant numbers YJG123104], The undergraduates teaching reform program [grant numbers 2011YJ052].

Received: 4 August 2013 Accepted: 24 March 2014

Published: 27 March 2014

#### References

1. Varani AM, Monteiro-Vitorello CB, Nakaya HI, Van Sluys MA: **The role of prophage in plant-pathogenic bacteria.** *Annu Rev Phytopathol* 2013, **51**:429–451.
2. Casjens S: **Prophages and bacterial genomics: what have we learned so far?** *Mol Microbiol* 2003, **49**(2):277–300.
3. Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H: **Prophage genomics.** *Microbiol Mol Biol Rev* 2003, **67**(2):238–276.
4. Zou QH, Li QH, Zhu HY, Feng Y, Li YG, Johnston RN, Liu GR, Liu SL: **SPC-P1: a pathogenicity-associated prophage of *Salmonella paratyphi C*.** *BMC Genomics* 2010, **11**:729.
5. Fortier LC, Sekulovic O: **Importance of prophages to evolution and virulence of bacterial pathogens.** *Virulence* 2013, **4**(5):354–365.
6. Sassi M, Bebeacua C, Drancourt M, Cambillau C: **The first structure of a mycobacteriophage, the *Mycobacterium abscessus* subsp. bolletii phage Araucaria.** *J Virol* 2013, **87**(14):8099–8109.
7. Hatfull GF: **Complete genome sequences of 138 mycobacteriophages.** *J Virol* 2012, **86**(4):2382–2384.
8. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Felwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, *et al*: **Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.** *Nature* 1998, **393**(6685):537–544.
9. Stinear TP, Seemann T, Pidot S, Frigui W, Reyssset G, Garnier T, Meurice G, Simon D, Bouchier C, Ma L, Tichit M, Porter JL, Ryan J, Johnson PD, Davies JK, Jenkin GA, Small PL, Jones LM, Tekaia F, Laval F, Daffe M, Parkhill J, Cole ST: **Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer.** *Genome Res* 2007, **17**(2):192–200.
10. Stinear TP, Seemann T, Harrison PF, Jenkin GA, Davies JK, Johnson PD, Abdellah Z, Arrowsmith C, Chillingworth T, Churcher C, Clarke K, Cronin A, Davis P, Goodhead I, Holroyd N, Jagels K, Lord A, Moule S, Mungall K, Norbertczak H, Quail MA, Rabinowitz E, Walker D, White B, Whitehead S, Small PL, Brosch R, Ramakrishnan L, Fischbach MA, Parkhill J, *et al*: **Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*.** *Genome Res* 2008, **18**(5):729–741.
11. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, Macheras E, Heym B, Herrmann JL, Daffe M, Brosch R, Risler JL, Gaillard JL: **Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*.** *PLoS One* 2009, **4**(6):e5660.
12. Chan J, Halachev M, Yates E, Smith G, Pallen M: **Whole-genome sequence of the emerging pathogen *Mycobacterium abscessus* strain 47J26.** *J Bacteriol* 2012, **194**(2):549.
13. Broussard GW, Oldfield LM, Villanueva VM, Lunt BL, Shine EE, Hatfull GF: **Integration-dependent bacteriophage immunity provides insights into the evolution of genetic switches.** *Mol Cell* 2013, **49**(2):237–248.
14. Choo SW, Yusoff AM, Wong YL, Wee WY, Ong CS, Ng KP, Ngeow YF: **Genome analysis of *Mycobacterium massiliense* strain M172, which contains a putative mycobacteriophage.** *J Bacteriol* 2012, **194**(18):5128.
15. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Loch C, Gutierrez MC, *et al*: **Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*.** *Nat Genet* 2013, **45**(2):172–179.
16. Sassi M, Robert C, Raoult D, Drancourt M: **Non-contiguous genome sequence of *Mycobacterium simiae* strain DSM 44165(T.).** *Stand Genomic Sci* 2013, **8**(2):306–317.
17. Ventura M, Zomer A, Canchaya C, O'Connell-Motherway M, Kuipers O, Turroni F, Ribbera A, Foroni E, Buist G, Wegmann U, Shearman C, Gasson MJ, Fitzgerald GF, Kok J, van Sinderen D: **Comparative analyses of prophage-like elements present in two *Lactococcus lactis* strains.** *Appl Environ Microbiol* 2007, **73**(23):7771–7780.
18. Ventura M, Turroni F, Lima-Mendez G, Foroni E, Zomer A, Duranti S, Giubellini V, Bottacini F, Horvath P, Barrangou R, Sela DA, Mills DA, van Sinderen D: **Comparative analyses of prophage-like elements present in bifidobacterial genomes.** *Appl Environ Microbiol* 2009, **75**(21):6929–6936.
19. Ventura M, Canchaya C, Pridmore D, Berger B, Brüssow H: **Integration and distribution of *Lactobacillus johnsonii* prophages.** *J Bacteriol* 2003, **185**(15):4603–4608.
20. Ventura M, Canchaya C, Kleerebezem M, de Vos WM, Siezen RJ, Brüssow H: **The prophage sequences of *Lactobacillus plantarum* strain WCFS1.** *Virology* 2003, **316**(2):245–255.
21. Ventura M, Turroni F, Foroni E, Duranti S, Giubellini V, Bottacini F, van Sinderen D: **Analyses of bifidobacterial prophage-like sequences.** *Antonie Van Leeuwenhoek* 2010, **98**(1):39–50.
22. Ventura M, Lee JH, Canchaya C, Zink R, Leahy S, Moreno-Munoz JA, O'Connell-Motherway M, Higgins D, Fitzgerald GF, O'Sullivan DJ, van Sinderen D: **Prophage-like elements in bifidobacteria: insights from genomics, transcription, integration, distribution, and phylogenetic analysis.** *Appl Environ Microbiol* 2005, **71**(12):8692–8705.
23. Zhao Y, Wang K, Ackermann HW, Halden RU, Jiao N, Chen F: **Searching for a "hidden" prophage in a marine bacterium.** *Appl Environ Microbiol* 2010, **76**(2):589–595.
24. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS: **PHAST: a fast phage search tool.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W347–W352.
25. Fouts DE: **Phage\_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences.** *Nucleic Acids Res* 2006, **34**(20):5839–5851.
26. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R: **Prophinder: a computational tool for prophage prediction in prokaryotic genomes.** *Bioinformatics* 2008, **24**(6):863–865.
27. Bose M, Barber RD: **Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences.** *In Silico Biol* 2006, **6**(3):223–227.
28. Akhter S, Aziz RK, Edwards RA: **PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies.** *Nucleic Acids Res* 2012, **40**(16):e126.
29. Hatfull GF, Jacobs-Sera D, Lawrence JG, Pope WH, Russell DA, Ko CC, Weber RJ, Patel MC, Germane KL, Edgar RH: **Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size.** *J Mol Biol* 2010, **397**(1):119–143.
30. Silverman JM, Brunet YR, Cascales E, Mougous JD: **Structure and regulation of the type VI secretion system.** *Annu Rev Microbiol* 2012, **66**:453–472.
31. Das S, Chakraborty A, Banerjee R, Chaudhuri K: **Involvement of in vivo induced icmF gene of *Vibrio cholerae* in motility, adherence to epithelial cells, and conjugation frequency.** *Biochem Biophys Res Commun* 2002, **295**(4):922–928.

32. de Pace F, Boldrin de Paiva J, Nakazato G, Lancellotti M, Sircili MP, Guedes Stehling E, Dias da Silveira W, Sperandio V: **Characterization of lcmF of the type VI secretion system in an avian pathogenic *Escherichia coli* (APEC) strain.** *Microbiology* 2011, **157**(Pt 10):2954–2962.
33. Bibb LA, Hatfull GF: **Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, phiRv1.** *Mol Microbiol* 2002, **45**(6):1515–1526.
34. Delcher AL, Bratke KA, Powers EC, Salzberg SL: **Identifying bacterial genes and endosymbiont DNA with Glimmer.** *Bioinformatics* 2007, **23**(6):673–679.
35. Schattner P, Brooks AN, Lowe TM: **The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs.** *Nucleic Acids Res* 2005, **33**(Web Server issue):W686–W689.
36. Laslett D, Canback B: **ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.** *Nucleic Acids Res* 2004, **32**(1):11–16.
37. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 2012, **28**(12):1647–1649.
38. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**(8):1596–1599.

doi:10.1186/1471-2164-15-243

**Cite this article as:** Fan *et al.*: Prophage-like elements present in *Mycobacterium* genomes. *BMC Genomics* 2014 **15**:243.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

