Distribution and function of prophage phiRv1 and phiRv2 among *Mycobacterium tuberculosis* complex

Xiangyu Fan\(^a,b\), Abu Algasim Elgaili Abd Alla\(^a\) and Jianping Xie\(^a*\)

\(^a\)State Key Laboratory Breeding Base of Eco-Environment and Bio-Resource of the Three Gorges Area, Key Laboratory of Ministry of Education Eco-Environment of the Three Gorges Reservoir Region, Institute of Modern Biopharmaceuticals, School of Life Sciences, Southwest University, Beibei, Chongqing 400700, China; \(^b\)School of Biological Science and Technology, University of Jinan, Shandong 250022, China

Communicated by Ramaswamy H. Sarma

(Received 30 October 2014; accepted 2 February 2015)

*Mycobacterium tuberculosis* complex (MTBC) is notorious for causing diseases, such as tuberculosis. Tuberculosis caused by *M. tuberculosis* remains a global public health concern. Two prophages, phiRv1 and phiRv2, can be found among most MTBC genomes. However, no precise functions have been assigned for the two prophages. In this paper, to find out the function of these two prophages, the distribution and function of phiRv1 and phiRv2 in MTBC genomes were analyzed from multiple omics data. We found that complex insertion, deletion, and reorganization appeared on the locus of two prophages in MTBC genomes; some genes of the two prophages can be translated and are functional from proteomic data; the expression of other prophage genes, such as Rv1577c, Rv2650c, Rv2652c, Rv2659c, and Rv2658c, can vary with environmental stresses and might enhance the fitness of MTBC. These data will facilitate our in-depth understanding of their function.

**Keywords:** *Mycobacterium tuberculosis* complex; prophage; genome; proteome; transcriptomic

1. Introduction

Prophages are temperate phages or their components, which are integrated into the host bacterial genome during evolution. The full length of prophages or prophage-like elements can be found in the bacterial genomes (Fan, Xie, Li, & Xie, 2014). Full-length prophages have complete excision and integration cassette, lysis cassette, and bacteriophage structure proteins. They can detach from the host genome, replicate, and assemble into phage particles. In contrast, prophage-like elements cannot shed from the host genome due to various deficiencies. The functions of prophages are very intriguing. Their impact on the host genome diversity is widely noted (Lang, Zhaxybayeva, & Beatty, 2012). Some prophages can encode bacterial virulence factors, thereby involving in bacterial pathogenesis (Brussow, Canchaya, & Hardt, 2004; Rabinovich, Sigal, Borovok, Nir-Paz, & Herskovits, 2012; Varani, Monteiro-Vitorello, Nakaya, & Van Sluys, 2013).

*Mycobacterium tuberculosis* complex (MTBC) refers to a genetically related group of *Mycobacterium* species that can render tuberculosis in humans or other organisms. Tuberculosis, though largely treatable, claims more people globally, and particularly, in China, than any other single infectious disease. The knowledge gap about its pathogenesis hinders better control (Fan, Tang, Yan, & Xie, 2014). Two prophage-like elements, phiRv1 and phiRv2, are found within the *M. tuberculosis* H37Rv genome (Bibb & Hatfull, 2002; Cole et al., 1998; Hendrix, Smith, Burns, Ford, & Hatfull, 1999), with few functionality data (Hatfull, 2010; Pedulla et al., 2003). The role of prophages in the pathogenesis of MTBC remains largely untapped. In this study, genome, proteome, and transcriptomic data were mined to survey the distribution of phiRv1 and phiRv2 among MTBC genomes to explore their function in the context of MTBC pathogenicity.

2. Materials and methods

2.1. Data collection

All genome sequences (Table S1) of bacteria for analysis were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/genbank). All transcriptomic and proteomic data used here were freely downloaded from the internet (Betts, Lukey, Robb, McAdam, & Duncan, 2002; Boshoff et al., 2004; Butcher, 2004; Fontan, Aris, Ghanny, Soteropoulos, & Smith, 2008; Keren, Minami, Rubin, & Lewis, 2011; Muttucumaru, Roberts, Hinds, Stabler, & Parish, 2004; Schnappinger et al., 2003). GEO accession numbers are not shown.

*Corresponding author. Email: georgex@swu.edu.cn

© 2015 Taylor & Francis
2.2. Evaluation methods

Comparative genomic analyses of prophages were performed at the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). As we know, twofold difference is a widely accepted cut-off for biological differences. For up- and down-regulated prophage genes, a regulation of at least twofold was used as a criterion. And probabilistic corrections (p-value or false discovery rate) of all transcriptomic data used here have been considered in the primary paper.

3. Results and discussion

3.1. Distribution of prophages in MTBC genomes

All retrievable MTBC genomes sequenced before 18 September 2014 were included in this study. Based on BLAST-N, we obtained the prophage information in different MTBC genomes (Table S1). Data showed that all strains of *M. tuberculosis* have at least one prophage (either phiRv1 or phiRv2); *M. bovis* AF2122/97 contains just one prophage, phiRv1; all strains of *M. canetti*, *M. africanum*, and *M. bovis* BCG have neither phiRv1 nor phiRv2, but some strains of *M. canetti* harbor other prophages (Fan, Xie, et al., 2014).

We investigated the distribution of phiRv1 and phiRv2 in different strains of *M. tuberculosis*. There are many substrains in *M. tuberculosis*, such as EAI (East-African-Indian), CAS (Central-Asian), Beijing, ‘X’, Haarlem, and LAM (Latin-American-Mediterranean) (Comas, Homolka, Niemann, & Gagneux, 2009). Data showed that all strains of EAI, CAS, and Haarlem strain families and a minority of ‘Beijing’ strains had both phiRv1 and phiRv2; all strains of ‘X’ and LAM and a majority of ‘Beijing’ strains carry just one prophage, phiRv2. Two strains isolated from India, *M. tuberculosis* RGTB423 and *M. tuberculosis* RGTB327, harbor phiRv1 and phiRv2, respectively. And RGTB423 is the only *M. tuberculosis* strain carrying just phiRv1. In fact, prophage deletion had been previously reported in the MTBC genome. Two prophages phiRv1 and phiRv2 of *M. tuberculosis* H37Rv corresponding to regions RD3 and RD11 were absent in BCG (Brosch et al., 2002). Our data showed that phiRv1 or phiRv2 deletion exists in different *M. tuberculosis* strains.

We systematically analyzed the integrating sites of prophages phiRv1 and phiRv2 in different strains. In *M. tuberculosis* strains H37Rv, H37Ra, CAS/NITR204, EA15, EA15/NITR206, str. Beijing/NITR203, RGTB423, RGTB327, str. Haarlem/NITR202, and *M. bovis* AF2122/97, phiRv1 was integrated into the region between *bioD* (dethiobiotin synthetase D) and *bioB* (biotin synthetase B); in *M. tuberculosis* strains 7199-99, str. Erdman = ATCC 35801, and str. Haarlem, phiRv1 was integrated into the region between *mhpE* (4-hydroxy-2-oxovalerate aldolase MhpE) and *rmlC* (dTDP-4-dehydrorhamnose 3,5-epimerase RmlC); in *M. tuberculosis* strains CDC1551, phiRv1 was integrated into the region between *bphl* (4-hydroxy-2-oxovalerate aldolase) and *strM* (dTDP-4-dehydrorhamnose 3,5-epimerase) (Figure S1). Prophage phiRv1 insertion sites lie within the REP13E12 repetitive element (Bibb & Hatfull, 2002). In all strains containing phiRv2, the element was integrated into the flank of *arsR* (ArsR-type transcription regulator), *arsC* (arsenical resistance protein), and four tRNA genes (Figure S2). As we know, insertion sites frequently lie within tRNA genes. Four tRNA genes (tRNA-val, -gly, -cys, and -val) were identified as integration sites of phiRv2. However, the four tRNA genes were found to be deleted in *M. tuberculosis* CDC1551. Data showed that phiRv1 rearrangement is common among *M. tuberculosis* strains, and phiRv2 was more stable (Fleischmann et al., 2002).

The same prophage genomes underwent evolutionary change such as recombination, deletion, or insertion. Some phiRv2 genomes integrated into *M. tuberculosis* strains H37Rv, H37Ra, str. Beijing/NITR203, RGTB327, CAS/NITR204, and KZN 4207 shared strong sequence similarities (identity > 99%, 100% coverage). However, other phiRv2 genomes shared only about 80% sequence (identity > 99%) with them (Figure 1(A) and Figure S3), implicating recombination events at prophage phiRv2 loci. PhiRv2 genomes of *M. tuberculosis* strains H37Rv and CDC1551 can serve as example. Figure 1(B) shows that two phiRv2 genes (Rv2648 and Rv2649) are lost in *M. tuberculosis* strains CDC1551, replaced by MT2726. Most phiRv1 genomes are very similar. However, the *M. tuberculosis* RGTB327 genome contains only a short fragment of phiRv1, indicating deletion in prophage phiRv1 loci (Figure S1). Figure 2 shows that three segments of genomic are absent in phiRv1 genomes of *M. tuberculosis* RGTB327 compared with *M. tuberculosis* H37Rv. And some prophage genes (for example Rv1574, Rv1579c, Rv1580c, Rv1581c, Rv1582c, Rv1583c, Rv1584c, and Rv1585c) are lost in the RGTB327 strain.

In short, the deletion or lost of prophages phiRv1 or phiRv2 during evolution can occur in different species of MTBC, and even the diverse strains of the same species; the insertion sites of phiRv1 vary with different *M. tuberculosis* strains, and even those strains belonged to the same strain grouping (such as Haarlem grouping); homologous recombination or gene deletion phenomenon can be found in the genome of prophages phiRv1 or phiRv2, and the similarities among these prophages seemed irrelevant to the kinship of their hosts. Prophages usually can enhance the virulence of pathogens (Tinsley, Bille, & Nassif, 2006; Wagner & Waldor, 2002). The comparative genomic analysis applied in this study corroborates that of previous studies (Hatfull, 2010; Pedulla...
et al., 2003), suggesting that there are no significant associations between MTBC and prophages.

### 3.2. Proteomic data demonstrated that prophages phiRv1 and phiRv2 are functional units

In general, not all prophage ORFs can be translated in the dormant stage (lysogenic). Some prophages exist as dormant DNA within the host genome. Repressor-encoding ORFs (Hendrix, Lawrence, Hatfull, & Casjens, 2000), and proteins endowed bacteria adaption (Wang et al., 2010), such as pathogenic factors, stress resistance proteins, can be actively expressed.

It is interesting to know whether phiRv1 or phiRv2 encodes proteins beneficial to pathogens. To this end, proteome data of *M. tuberculosis* H37Rv were mined. Some phiRv1 genes are lost in RGTB327 strain. They are Rv1574, Rv1579c, Rv1580c, Rv1581c, Rv1582c, Rv1583c, Rv1584c, and Rv1585c.

Figure 1. Comparative genomic analysis of prophages phiRv2. (A) Diagrammatic sketch of phiRv2 genomes in different MTBC strains. (B) Comparative genomic analysis between phiRv2 in *M. tuberculosis* H37Rv genome, and phiRv2 in *M. tuberculosis* CDC1551. Two phiRv2 genes (Rv2648 and Rv2649) are lost in *M. tuberculosis* strains CDC1551, replaced by MT2726.

Figure 2. Comparative genomic analysis between a short fragment of phiRv1 in *M. tuberculosis* RGTB327 genome and phiRv1 in *M. tuberculosis* H37Rv. Three segments of genomic are absent in phiRv1 genomes of *M. tuberculosis* RGTB327. Some phiRv1 genes are lost in RGTB327 strain. They are Rv1574, Rv1579c, Rv1580c, Rv1581c, Rv1582c, Rv1583c, Rv1584c, and Rv1585c.
Phage capsid protein (Rv1576c and Rv2650c), phage terminase (Rv1578c), and phage integrase (Rv1586c and Rv2659c) are present in both proteome data. These proteins are not conserved in most sequenced mycobacteriophages. They may help *M. tuberculosis* H37Rv to adapt to environment and be associated with the pathogenesis of *M. tuberculosis* H37Rv. Since, prophages phiRv1 and phiRv2 are temperate phages, these proteins found in proteomic data may indicate production of phage particles. Furthermore, phage particles may help *M. tuberculosis* to adapt to environmental fluctuations and be involved in virulence.

### 3.3 Transcriptomic data

#### 3.3.1 Expression of prophage genes can respond to environmental stresses

As intracellular pathogen, it is essential for MTBC member to tackle different stresses such as, acidic environment, low oxygen, and nutritional deficiency to survive and thrive (Fontan et al., 2008). Prophages can improve the capability of pathogens to overcome such stresses (Wang et al., 2010). Mining the variety of Transcriptomic profilings of *M. tuberculosis* responses to stresses (Betts et al., 2002; Boshoff et al., 2004; Butcher, 2004; Fontan et al., 2008; Keren et al., 2011; Muttucumaru et al., 2004; Schnappinger et al., 2003) can provide information about prophages.

Six prophage genes (Rv1577c, Rv1585c, Rv2652c, Rv2656c, Rv2658c, and Rv2659c) were up-regulated in
the nutrient starvation model of *M. tuberculosis* (Figure 3 and Table S2) (Betts et al., 2002). Two prophage genes (Rv2655c and Rv2658c) were up-regulated in the oxygen-depleted non-replicating (NRP) model of *M. tuberculosis* (Figure 3, Table S2) (Muttucumaru et al., 2004). Five prophage genes (Rv1584c, Rv2650c, Rv2651c, Rv2652c, and Rv2659c) were up-regulated (Figure 3 and Table S2) in *M. tuberculosis*-infected wild-type mice macrophages (Schnappinger et al., 2003) and human macrophage THP-1 cells (Fontan et al., 2008).

Four prophage genes (Rv1577c, Rv1581c, Rv2650c, and Rv2651c) were up-regulated (Figure 3 and Table S2) in the *M. tuberculosis* persisters model (Keren et al., 2011).

No prophage genes are down-regulated under these conditions.

The overview of prophage gene expression under a variety of biological conditions is presented in Figure 3. Prohead protease of prophage phiRv1, Rv1577c, was up-regulated in the nutrient starvation model and persister model. Capsid protein of prophage phiRv2, Rv2650c, was up-regulated in the macrophage model and persister model. Terminase (Rv2652c) and integrase (Rv2659c) of phiRv2 were up-regulated in the nutrient starvation model and macrophage model. Rv2658c, a function unknown protein of phiRv2, was up-regulated in the nutrient starvation model and oxygen-depleted NRP model. These up-regulated prophage genes might improve the fitness of *M. tuberculosis*.

### 3.3.2. Oxidative phosphorylation inhibitors up-regulate the expression of a large number of prophage genes

Most *M. tuberculosis* H37Rv prophage genes were up-regulated by metabolism inhibitor treatment (Figure 4 and Table S3) (Boshoff et al., 2004; Liang et al., 2011, 2012). Oxidative phosphorylation inhibitors, such as valinomycin, KCN, N,N′-dicyclohexylcarbodiimide (DCCD), and 2,4-dinitrophenol (DNP) promoted the expression of most prophage genes under which the electron transfer chain of *M. tuberculosis* was blocked and the bacteria are supposed to be under hypoxia. The up-regulated prophage genes might implicate in host hypoxia adaptation or set the stage for the prophage to excise from the host chromosome. Since these prophages are functional units, the most parsimonious explanation is the excision of the prophage rather than a putative role in host hypoxia.

### 4. Conclusions

In summary, to shed more light on the poorly defined relation between phiRv1 or phiRv2 and MTBC pathogenesis (Hatfull, 2010; Pedulla et al., 2003), genomic, proteomic, and transcriptomic data of MTBC were mined to find the distribution and function of phiRv1 and phiRv2. At the genomic level, complex insertion, deletion, and reorganization appeared on the locus of two prophages. The proteome data showed that some phiRv1 and phiRv2 genes were translated and had function. The transcriptomic data demonstrated that some prophage genes were induced during many stresses. Moreover, oxidative phosphorylation inhibitors can induce expression of most prophage genes, implicating that phiRv1 and phiRv2 elements can sense oxygen status within the host and respond to the variation accordingly. The diverse phenotypes of phiRv1 and phiRv2 suggest important roles. Deletion experiments are needed to reveal phiRv1 and phiRv2 function.
