



LETTER

A novel transposable Mu-like prophage in *Bacillus alcalophilus* CGMCC 1.3604 (ATCC 27647)

Dear Editor,

Transposable phages, which are reproduced by transposition (Harshey, 2012; Taylor, 1963), have been widely applied in the field of biotechnology to manipulate operon/gene fusions, *in vivo* cloning, random ion mutagenesis, and integration of DNA into bacterial genomes (Abalakina et al., 2008; Akhverdyan et al., 2011). One of the best-studied transposable phages is the Mu phage which was discovered in *Escherichia coli* (Harshey, 2012; Saha et al., 2013; Taylor, 1963). Other transposable phages, such as D108 in *E. coli*, D3112 and B3 in *pseudomonas* (Braid et al., 2004; Wang et al., 2004), and KS10 in *Burkholderia cepacia* complex (Goudie et al., 2008), have also been detected or isolated in *Proteobacteria*. Recently, transposable prophages were identified in Firmicute by searching genome sequences. For example, SglyMu-1 was detected in *Syntrophobotulus glycolicus*, and Firmicute transposable prophages were found in Clostridiales (Toussaint, 2013). It is still undecided whether transposable prophages exist in other bacterial groups.

In this letter, we provide evidence for the first transposable prophage BalMu-1 in Bacilli. The transposable prophage (BalMu-1, Genbank No. KP063902 and KP063903) was identified in *Bacillus alcalophilus* CGMCC 1.3604 (ATCC 27647) through high throughput genome sequencing and PCR-dideoxy chain-termination (Sanger) sequencing.

We first extracted amino acid sequences of the draft genome *B. alcalophilus* to run BLAST against the ACLAME virus database (<http://aclame.ulb.ac.be/Tools/blast.html>) (Leplae et al., 2010). Several open reading frames (ORFs) were predicted to encode transposases and other related prophage genes, suggesting that the *Bacillus* strain likely contained transposable prophages. Considering the genome is incomplete in *B. alcalophilus* (GenBank ALPT01000029), we performed a high throughput re-sequencing of the *B. alcalophilus* (ATCC 27647) genome to extract complete prophage sequences. *B. alcalophilus* was obtained from CGMCC (China General Microbiological Culture Collection Center) with ID CGMCC 1.3604. The high throughput sequenc-

ing from the Illumina HiSeq 2000 system generated 6,805,654 paired-end reads which were estimated to cover the genome at 220-fold depth. The reads were further assembled into contigs using Velvet 1.2.10 (Zerbino and Birney, 2008). According to the ACLAME database, six contigs (>1kb) showed similarity with prophage sequences. They were then assembled together to generate complete sequences of prophage BalMu-1. To obtain an accurate result, PCR-dideoxy sequencing was conducted to adjust low quality regions. The merged sequences were manually checked with Illumina reads and dideoxy sequencing results, using BWA 0.7.10 (Li and Durbin, 2010) for reads mapping and Tablet 1.14.4.10 (Milne et al., 2010) for visualization. The final sequences of BalMu-1 were deposited in GenBank under accession numbers KP063902 and KP063903.

The coverage depth of prophage fragments ($\sim 450\times$) is roughly double that of the genome fragments ($\sim 220\times$), suggesting that BalMu-1 may have two copies in the *B. alcalophilus* CGMCC 1.3604 genome. Consistent with this, analysis of Illumina reads and dideoxy sequencing gave rise to two genome sequences of BalMu-1. The lengths of their genomes were determined to be 39,873 and 39,861 bp (Figure 1), which is longer than the Mu-like prophages SglyMu-1 in *S. glycolicus* (33,611 bp) and transposable phages infecting *Proteobacteria* such as Mu phage (37,000 bp) (Toussaint 2013). The GC content of prophage BalMu-1 was 42.8%, a little higher than the average GC content (37.2%) of the whole genome in *B. alcalophilus* CGMCC 1.3604. Both copies of BalMu-1 contain conserved terminal CA-TGs sequences as Mu and many other transposable prophages (Harshey, 2012). Meanwhile, the two BalMu-1 copies were different in the border regions between the prophage and host genome (Figure 1). Only one copy of BalMu-1 is flanked by a 5 bp direct repeat (GCAAG) and the other copy lacks the repeat sequence, which is assumed to be result from recombination or mutation.

The ORFs of BalMu-1 were predicted by the RAST server (<http://rast.nmpdr.org/>) (Overbeek et al., 2014) and annotated by the ACLAME virus database. Superfamilies and conserved domains of ORFs were analyzed by NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>)

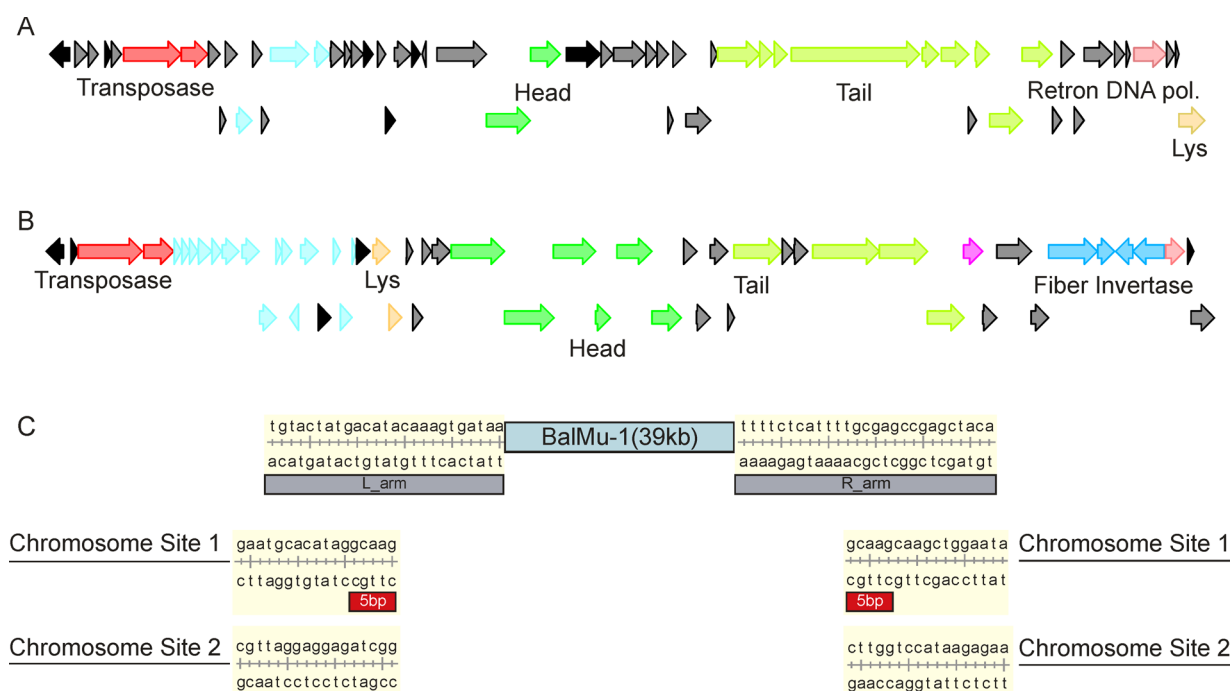


Figure 1. (A, B) Genetic organizations of BalMu-1 (A) and Mu (B). Arrows indicate translated open reading frames (ORFs). Red: Trans-positional replication proteins A and B; black: regulatory proteins; cyan: semi-essential proteins; green: head proteins; light green: tail proteins; blue: fibers region; light red: invertase and reverse transcriptase; yellow: lysis protein; grey: proteins with other functions or no identifiable function. (C) Two copies of BalMu-1 prophages in the *Bacillus alcalophilus* CGMCC 1.3604 (ATCC 27647) chromosome. The ends of each copy of BalMu-1 were identified by self-alignment Blast. The 5 base-pairs flanking the one copy of the prophage were analyzed manually and are indicated by the red box.

and SUPFAM (<http://supfam.org/SUPERFAMILY/>) (Gough et al., 2001) servers (supplementary information, Table S1, based on the sequence of 39,873 bp copy). The ORFs of BalMu-1 are preferentially organized similar to other transposable phages (Figure 1). Both copies of BalMu-1 are predicted to have 56 ORFs. ORF 6 and ORF 7 encode transposases A and B which are involved in the mediation of the transposition process. A similar organization of transposases is observed in SglyMu-1 (Toussaint, 2013). ORF 14 is annotated as DNA polymerase I, but meanwhile phage Mu contains a putative DNA polymerase II large subunit (gp13), making BalMu-1 appear special and indicating that BalMu-1 may have a different means of replication way than other transposable phages. Several ORFs, such as ORF 26, ORF27 and ORFs 37~42, are annotated as phage head and tail proteins. ORF 53 encodes a retron-type RNA-directed DNA polymerase (reverse transcriptase). Although reverse transcriptases are present in retroviruses, ORF 53 displays a greater similarity to reverse transcriptases in SglyMu-1 and other prophages (Toussaint, 2013). Reverse transcriptase is a part of the diversity generating retroelements (DGR) system in transposable

phages. Consequently, self-alignment of SglyMu-1 identified 120 bp TR (template region)/VR (variable region) region of the DGR system. The last ORF (ORF 56) is annotated as N-acetylmuramoyl-L-alanine amidase, which may play a role in phage release through cleaving the linkage between N-acetylmuramoyl and L-amino acid residues in cell-wall glycopeptides. This is different from the phage Mu, in which the genes involved in phage release are located in the middle part of the genome.

Taken together, BalMu-1 is a typical transposable prophage in terms of sequence structure, gene organization, and annotated gene functions. While waiting for further insights into the transposition mechanism of BalMu-1, our results demonstrate the existence of transposable prophage in Bacilli, which offers a potential for developing genetic tools for the molecular manipulation of *Bacillus* and related strains.

FOOTNOTES

This work was supported by the National Basic Research Program (973 Program) of China (2014CB745100) and the National Key Technologies R&D Program of China (2012AA022101).

We thank Dr. Ariane Toussaint for thoughtful discussions. We also thank Prof. Benjamin S. Glick for providing the SnapGene software for producing maps. Finally, we thank Dr. Peng Xu for his help during the revision of the manuscript.

All the authors have declared that no competing interests exist. This article does not contain any studies involving human participants or animals performed by any of the authors.

Supplementary materials are available on the website of *Virologica Sinica*: <http://www.virosin.org>.

Junjie Yang^{1,2,4#}, Yimeng Kong^{1,3#}, Xuan Li¹, Sheng Yang^{1,2,4}✉

1. Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

2. Shanghai Research and Development Center of Industrial Biotechnology, Shanghai 201201, China

3. University of the Chinese Academy of Sciences, Beijing 100080, China

4. Shanghai Collaborative Innovation Center for Biomanufacturing Technology, Shanghai 200237, China

✉Correspondence:

Phone: +86-21-54924173, Fax: +86-21-54924015,

Email: syang@sibs.ac.cn

ORCID: 0000-0003-3742-9989

#These authors contributed equally to this work.

Published online: 22 January 2015

REFERENCES

- Abalakina EG, Tokmakova IL, Gorshkova NV, et al. 2008. Appl Microbiol Biotechnol, 81: 191–200.
- Akhverdyan VZ, Gak ER, Tokmakova IL, et al. 2011. Appl Microbiol Biotechnol, 91: 857–871.
- Braid MD, Silhavy JL, Kitts CL, et al. 2004. J Bacteriol, 186: 6560–6574.
- Goudie AD, Lynch KH, Seed KD, et al. 2008. BMC Genomics, 9: 615.
- Gough J, Karplus K, Hughey R, et al. 2001. J Mol Biol, 313: 903–919.
- Harshey RM. 2012. Mob DNA, 3: 21.
- Leplae R, Lima-Mendez G, Toussaint A. 2010. Nucleic Acids Res, 38: D57–D61.
- Li H, Durbin R. 2010. Bioinformatics, 26: 589–595.
- Milne I, Bayer M, Cardle L, et al. 2010. Bioinformatics, 26: 401–402.
- Overbeek R, Olson R, Pusch GD, et al. 2014. Nucleic Acids Res, 42: D206–D214.
- Saha RP, Lou Z, Meng L, et al. 2013. PLoS Genet, 9: e1003902.
- Taylor AL. 1963. Proc Natl Acad Sci U S A, 50: 1043–1051.
- Toussaint A. 2013. Res Microbiol, 164: 281–287.
- Wang PW, Chu L, Guttman DS. 2004. J Bacteriol, 186: 400–410.
- Zerbino DR, Birney E. 2008. Genome Res, 18: 821–829.