

# Genome sequence of a cluster A13 mycobacteriophage detected in *Mycobacterium phlei* over a half century ago

Szilvia Marton<sup>1</sup> · Enikő Fehér<sup>1</sup> · Balázs Horváth<sup>2</sup> · Katalin Háber<sup>3</sup> · Pál Somogyi<sup>3</sup> · János Minárovits<sup>3,4</sup> · Krisztián Bányai<sup>1</sup>

Received: 17 March 2015 / Accepted: 8 September 2015  
© Springer-Verlag Wien 2015

**Abstract** A phage infecting *Mycobacterium phlei* was isolated in 1958 from a soil sample in Hungary. Some physicochemical and biological properties of the virus were described in independent studies over the years. Here, we report the genome sequence of this early mycobacteriophage isolate. The Phlei phage genome measured 50,418 bp, had a GC content of 60.1 % and was predicted to encode 81 proteins and three tRNAs. Phylogeny of the tape measure protein revealed genetic relatedness to other early isolates of mycobacteriophages within subcluster A2. The genomic organization and genetic relationships to other strains showed that the Phlei phage belongs to a novel genetic cluster, designated A13.

## Introduction

Mycobacteriophages belong to the most intensively studied groups of bacteriophages. Several of these viruses have been isolated from environmental and clinical samples and characterized since the middle of the last century [5–8, 10; <http://phagesdb.org/>]. However, the complete sequence of most mycobacteriophage genomes was determined only recently, revealing extensive genetic variability [6, 7; <http://phagesdb.org/>]. In this study, we determined the complete genome sequence of *Mycobacterium* phage Phlei (Phlei phage). The Phlei phage was isolated in 1958 at the National “Koranyi” Institute for Tuberculosis, Budapest, Hungary, from a soil sample, using *Mycobacterium phlei* as a host organism [14]. Although morphological, biophysical and certain biochemical properties of the Phlei phage were investigated previously [10–14], sequencing of its genome was carried out more than fifty years later.

Liquid cultures of *Mycobacterium phlei* were infected with mycobacterium phage Phlei and the phage was harvested as described earlier [10]. Phage particles were concentrated by CsCl equilibrium gradient centrifugation, and the phage DNA was extracted by the proteinase K-phenol-chloroform method. The purified genomic DNA was processed to obtain a barcoded library compatible with the Ion Torrent PGM semiconductor sequencing system [4]. *De novo* genome assembly was executed with MIRA 4.0. GeneMarkS, FGENESV [<http://www.softberry.com/>], Glimmer 3.02 and DNA Master 5.22.9 software were applied for genome annotation and gene prediction [1–3, 15; <http://phagesdb.org/>]. Protein homology was searched by BLASTP analysis (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Next-generation sequencing generated a total of 359,466 Phlei-phage-specific sequence reads. Of these, 346,514 reads mapped on the *de novo* assembled genome at an

---

S. Marton and E. Fehér contributed equally to this work.

✉ Krisztián Bányai  
bkrota@hotmail.com

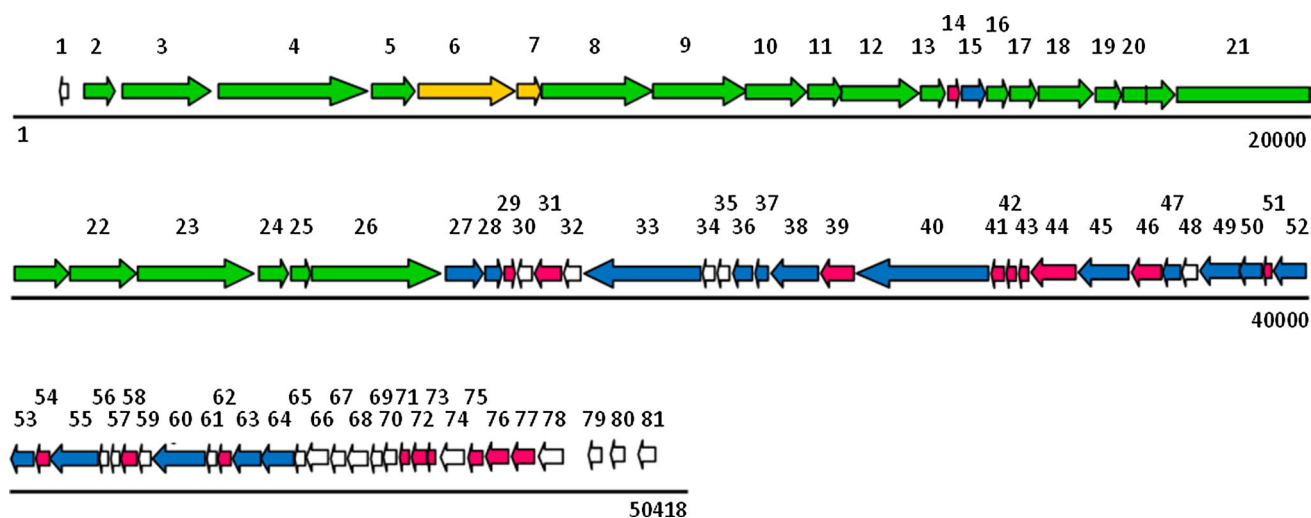
- <sup>1</sup> Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Hungaria krt 21, 1143 Budapest, Hungary
- <sup>2</sup> Biological Research Centre, Hungarian Academy of Sciences, Temesvari krt. 62, 6726 Szeged, Hungary
- <sup>3</sup> Microbiological Research Group, National Center for Epidemiology, Pihenó u. 1, 1529 Budapest, Hungary
- <sup>4</sup> Present Address: Department of Oral Biology and Experimental Dental Research, Faculty of Dentistry, University of Szeged, Tisza Lajos krt. 64., 6720 Szeged, Hungary

average coverage of 680 $\times$ . The assembled genome had a length of 50,418 base pairs with a GC content of 60.1 %. Also, the genome had a 3' sticky overhang. The annotated genome sequence of Phlei phage has been deposited in the GenBank database (accession no. KT206225), and also in the largest mycobacteriophage genome sequence database (<http://phagesdb.org/phages/Phlei/>). Based on BLAST analysis, mycobacteriophages of the subcluster A2 were the closest relatives of the Phlei phage (with  $\leq 52$  % query coverage and  $\leq 77$  % identity). Bioinformatics analysis of the Phlei phage genome predicted 81 open reading frames (ORF) and three tRNAs. Of these, 41 ORFs matched genes with assigned function in GenBank, while the remaining ORFs showed similarity to mycobacteriophage hypothetical genes or did not match any GenBank sequence records. The deduced aa sequences showed 39–84 % identity when compared with available mycobacteriophage sequences using BLAST X. However, some ORFs encoding the putative lysine A, holin and signal peptidase proteins were highly similar to *Mycobacterium hassiacum* sequences (100 %, 97 % and 92 % aa identity, respectively), which may indicate relatively recent gene-transfer events.

The BLAST engine was utilized to assign and compare putative ORFs of the Phlei phage genome to reference phage genes with known functions. The results suggested that the ORFs of the Phlei phage are arranged into functional modules, as seen for other phages (Fig. 1) [7]. The first module might include ORF 2–5, ORF 8–13 and ORF 16–26, encoding the capsid (ORF 12) and tail proteins (ORF 2–4, ORF 18, ORF 22–24) and those responsible for phage assembly, packaging and maturation, including head-to-tail connector proteins (ORF 16–17), a scaffold

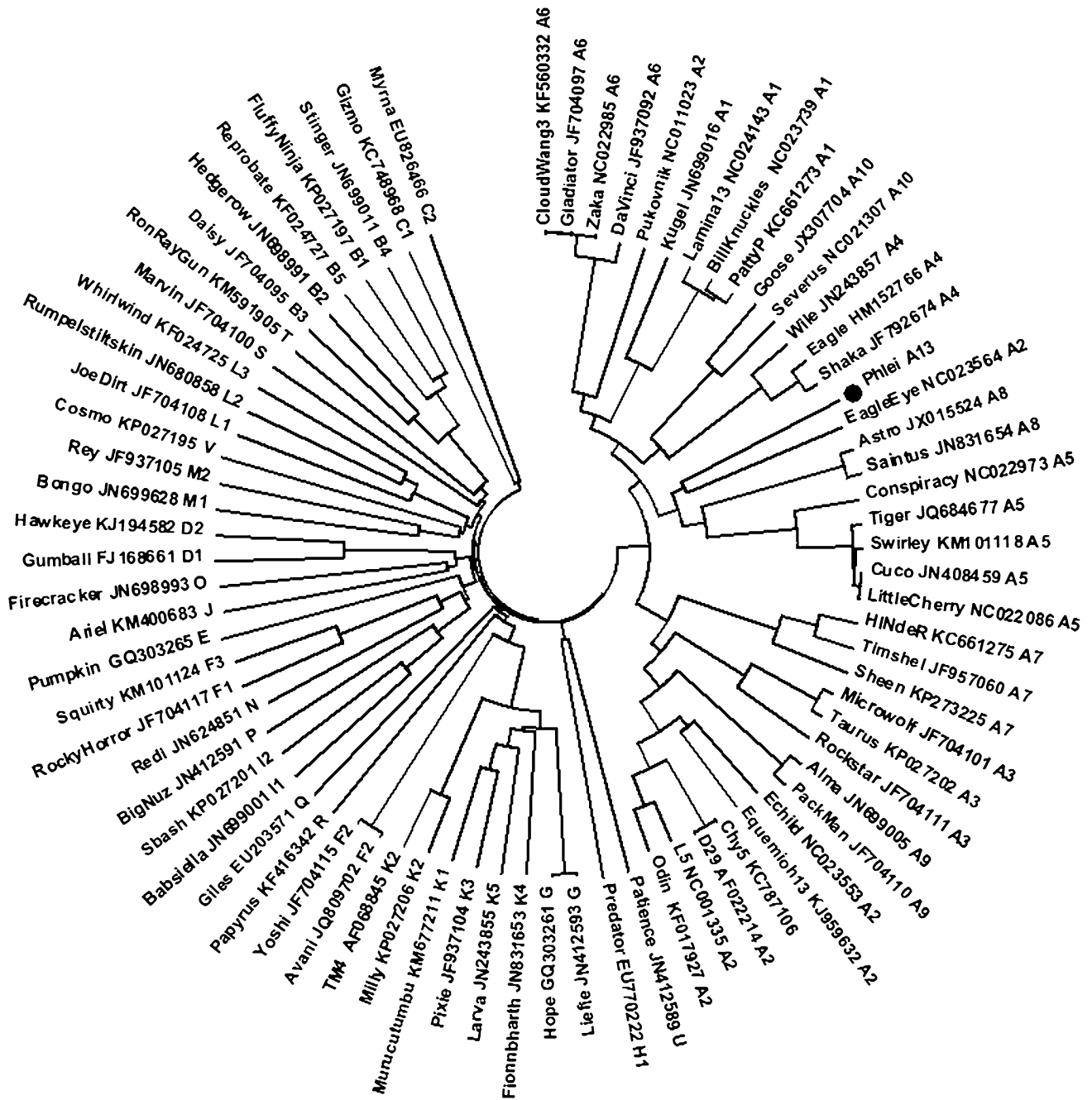
protein (ORF 11), a terminase (ORF 8), a portal protein (ORF 9), a capsid protease (ORF 10), the tail assembly proteins (ORF 19–20, produced by frameshift), the tape measure protein (TMP, ORF 21) and some other hypothetical proteins with potentially similar functions. ORF 6 and ORF 7 encoding the lysine A and holin proteins, respectively, may form a second, 'lysis module' involved in bacterial cell lysis. A third functional module might comprise genes aiding replication, recombination and metabolism. This cluster of ORFs had homologies with mycobacteriophage sequences encoding helicase (ORF 15), primase/helicase (ORF 49), primase (ORF 50), DNA-B-like helicase (ORF 55), ParA and ParB proteins (ORF 27–28), DNA polymerase (ORF 33), endonuclease VII (ORF 52), RecB exonuclease (ORF 60) DNA methylase (ORF 63–64), HTH (helix-turn-helix) DNA binding protein (ORF 36), Hin protein (ORF 37), thymidilate synthase (ORF 38), ribonucleotide reductase (ORF 40), metallophosphoesterase (ORF 45), nrdH glutaredoxin (ORF 47) and MazG nucleotide pyrophosphohydrolase (ORF 53).

Smith et al. [9] suggested that sequence comparison and phylogenetic analysis of the TMP may be useful for distinguishing mycobacteriophages belonging to various subclusters. In our analyses, the TMP of Phlei phage shared the greatest, but still very low, amino acid similarity (41 % identity), with a subcluster A2 mycobacterium phage (EagleEye, accession no. NC023564) and formed a distinct branch on the phylogenetic trees generated by either the neighbor-joining or the maximum-likelihood method (Fig. 2). A more complex and multi-faceted approach to distinguish mycobacteriophages was elaborated by Hatfull et al. in 2010. This approach classifies mycobacteriophages



**Fig. 1** Schematic representation of the *Mycobacterium* phage Phlei genome. Potential phage ORFs are denoted by colored arrows according to gene function. The arrow orientation indicates the reading direction of the putative gene. Green arrows, structure and assembly module genes; yellow, lysis module genes; blue arrows,

replication, recombination, metabolism module genes; pink, hypothetical mycobacteriophage genes without assigned function; colorless arrows, putative ORFs not matching any phage-origin genes when compared to the GenBank references (color figure online)



**Fig. 2** Phylogenetic analysis of mycobacteriophage tape measure protein amino acid sequences representing the main clusters and subclusters (<http://phagesdb.org>). The neighbor-joining phylogenetic

tree was generated using p-distance with 1000 replicates as implemented in the MEGA6 software. *Mycobacterium* phage Phlei is indicated by a black dot

into clusters taking into consideration the results of dot-plot genome clustering, pairwise nt identity values, gene content analysis, and the correlation between pairwise alignments and gene locations [7]. By using this complex classification method (<http://phagesdb.org>), the Phlei phage was designated as the first member of subcluster A13.

Together with subcluster A2 mycobacteriophages, which include two isolates, D29 and L5, found during the early 1950s in the USA and Japan, respectively, the Phlei phage characterized in this study is among the earliest mycobacteriophages isolated to date [5, 8; <http://phagesdb.org>]. Whole-genome sequencing of mycobacteriophage sequences collected during the past few decades and more

recently may widen our knowledge of phage evolution and gene function.

**Acknowledgments** The excellent technical assistance of Mrs. Vavrikné Nagy Andrea is highly appreciated. The study group at Institute for Veterinary Medical Research received financial support from the Momentum program (awarded by the Hungarian Academy of Sciences).

## References

1. Besemer J, Alexandre 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 Res* 29(12):2607–2618
2. Chevreur B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Comput Sci Biol Proc German Conf Bioinform (GCB)* 99:45–56
3. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27(23):4636–4641
4. Fehér E, Pazár P, Lengyel G, Phan TG, Bányai K (2014) Sequence and phylogenetic analysis identifies a putative novel gyrovirus 3 genotype in ferret feces. *Virus Genes*. doi:10.1007/s11262-014-1128-y
5. Ford ME, Sarkis GJ, Belanger AE, Hendrix RW, Hatfull GF (1998) Genome structure of mycobacteriophage D29: implications for phage evolution. *J Mol Biol* 279(1):143–164
6. Hatfull GF (2014) Molecular genetics of mycobacteriophages. *Microbiol Spectr* 2(2):1–36
7. 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. *J Mol Biol* 397(1):119–143
8. Hatfull GF, Sarkis GJ (1993) DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. *Mol Microbiol* 7(3):395–405
9. Smith KC, Castro-Nallar E, Fisher JN, Breakwell DP, Grose JH, Burnett SH (2013) Phage cluster relationships identified through single gene analysis. *BMC Genomics* 14:410
10. Somogyi PA, Petrovsky GV, Grigorev VB (1975) Biophysical properties of mycobacteriophages. *J Gen Virol* 29(2):235–238
11. Somogyi PA, Maso Bel M, Földes I (1982) Methylated nucleic acid bases in *Mycobacterium* and mycobacteriophage DNA. *Acta Microbiol Acad Sci Hung* 29(3):181–185
12. Somogyi PA, Földes I (1983) Incorporation of thymine, thymidine, adenine and uracil into nucleic acids of *Mycobacterium phlei* and its phage. *Ann Microbiol (Paris)* 134A(1):19–28
13. Vajda BP, Háber KJ (1976) Concentration and purification of mycobacteriophages with polyethylene glycol 6000. *Am Rev Respir Dis* 114(1):245–248
14. Vandra E (1965) Thesis. Hungarian Academy of Sciences, Budapest
15. Xu J, Hu J, Wang J, Han Y, Hu Y, Wen J, Li Y, Ji J, Ye J, Zhang Z, Wei W, Li S, Wang J, Wang J, Yu J, Yang H (2003) Genome organization of the SARS-CoV. *Genomics Proteomics Bioinform* 1(3):226–235