Available online at www.sciencedirect.com
SCIENCE DIRECT
Virology 348 (2006) 57–71
VIROLOGY
www.elsevier.com/locate/yviro

Complete nucleotide sequence and genome analysis of bacteriophage BFK20—A lytic phage of the industrial producer *Brevibacterium flavum*

Gabriela Bukovska [a,*,1], Lubos Klucar [a,1], Cestmir Vlcek [b], Jan Adamovic [a], Jan Turna [c], Jozef Timko [a]

[a] *Institute of Molecular Biology, Centre of Excellence for Molecular Medicine, Slovak Academy of Sciences, Dubravska cesta 21, 845 51 Bratislava, Slovakia*
[b] *Institute of Molecular Genetics, Academy of Sciences of the Czech Republic and Centre for Applied Genomics, Flemingovo nam. 2, Prague 6, CZ-16637, Czech Republic*
[c] *Comenius University, Faculty of Natural Sciences, Department of Molecular Biology, Mlynska dolina B-2, 842 15 Bratislava 4, Slovakia*

Received 22 July 2005; returned to author for revision 14 November 2005; accepted 11 December 2005
Available online 2 February 2006

Abstract

The entire double-stranded DNA genome of bacteriophage BFK20, a lytic phage of the *Brevibacterium flavum* CCM 251 – industrial producer of L-lysine – was sequenced and analyzed. It consists of 42,968 base pairs with an overall molar G + C content of 56.2%. Fifty-five potential open reading frames were identified and annotated using various bioinformatics tools. Clusters of functionally related putative genes were defined (structural, lytic, replication and regulatory). To verify the annotation of structural proteins, they were resolved by 2D gel electrophoresis and were submitted to N-terminal amino acid sequencing. Structural proteins identified included the portal and major and minor tail proteins. Based on the overall genome sequence comparison, similarities with other known bacteriophage genomes include primarily bacteriophages from *Mycobacterium* spp. and some regions of *Corynebacterium* spp. genomes – possible prophages. Our results support the theory that phage genomes are mosaics with respect to each other.

Keywords: Corynebacterium; Siphoviridae; Bacteriophage; BFK20 genome; Lytic phage; Complete genome sequence; Sequence analysis; 2D gel electrophoresis

Introduction

The genera *Corynebacterium* and *Brevibacterium* are Gram-positive bacteria of high G + C content, close to *Nocardia* and *Mycobacteria* and not very distant from *Streptomyces* (Park et al., 1987; Woese, 1987). The non-pathogenic members of genus *Corynebacterium* are widely used as industrial producers in many biotechnological processes (Batt et al., 1985; Krämer, 1996; Nampoothiri and Pandey, 1998). Applications range from feed to food and pharmaceutical products. *C. glutamicum* and *Brevibacterium flavum* play a major nearly exclusive role in the production of economically important amino acids (Hermann, 2003), including L-glutamate, threonine, L-lysine, phenylalanine, glutamine, arginine, tryptophan, isoleucine and histidine.

A serious and still common problem in many biotechnological laboratories and factories is infections of bacterial cultures by bacteriophages. The existence of lytic and temperate phages in corynebacteria has been known for many years (Kato et al., 1984, Patek et al., 1985). Most characterized corynephages are temperate and were isolated after UV irradiation or mitomycin induction (Moreau et al., 1995), but there are not many lytic corynephages defined in detail to date.

Corynephages usually have a polyhedral head (40 × 50–120 × 50 nm) attached to a long, flexible and non-contractile tail (80–250 nm) (Nampoothiri and Pandey, 1998). They contain double-stranded DNA with a G + C content of 54–57%.

* Corresponding author. Fax: +421 2 59307416.
E-mail addresses: Gabriela.Bukovska@savba.sk (G. Bukovska), Lubos.Klucar@savba.sk (L. Klucar), vlcek@img.cas.cz (C. Vlcek), Jan.Adamovic@savba.sk (J. Adamovic), turna@fns.uniba.sk (J. Turna), Jozef.Timko@savba.sk (J. Timko).
[1] Both authors contributed equally.

0042-6822/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2005.12.010

The host range, virion morphology, DNA and protein profiles of several corynephages have been reported (Trautwetter and Blanco, 1988; Trautwetter et al., 1987a). Phages characterized in detail include phages Φ304L, Φ304S, Φ15 and Φ16, which were induced from different *C. glutamicum* ATCC derivatives (Moreau et al., 1995, 1999), ΦGA1 – *B. flavum* phage (Sonnen et al., 1990a), β-converting and γ-nonconverting corynebacteriophage of the pathogenic species *C. diphtheriae* (Buck and Groman, 1981) as well as ΦAAU2 infecting *"Arthrobacter aureus"-C70* (Le Marrec et al., 1996). Bacteriophage CL31 infects *C. lilium* (Trautwetter et al., 1987b), and Cog phage is a virulent phage of *C. glutamicum* (Sonnen et al., 1990b). According to Ackermann (2003), over 5300 phages have been examined by electron microscopy. In the EMBL database, we could find DNA segments from around 1000 different bacteriophages, giving the approximate number of known bacteriophages studied at a molecular level. There are more than 250 completely known genomes of bacteriophages, but no complete nucleotide sequence of any corynephage has yet been reported.

We have previously isolated and characterized corynephage BFK20, which causes lysis of *B. flavum* CCM 251, the industrial producer of L-lysine (Koptides et al., 1992, 1994). Previous study of the host range of BFK20 (Koptides et al., 1992) revealed only one host strain lysed after infection. No lysis of other *B. flavum* ATCC strains, *B. lactofermentum* BLOB and *C. glutamicum* ATCC 13032 was observed. Defense mechanisms of corynebacteria strains against bacteriophage BFK20 were recently investigated in detail (Halgasova et al., 2005).

B. flavum is considered to be the same species as *C. glutamicum* (Liebl et al., 1991). Thus, known genomic sequence of *C. glutamicum* ATCC 13032 (Ikeda and Nakagawa, 2003; Kalinowski et al., 2003) was a great benefit for the BFK20 bioinformatic analysis presented in this study. Bacteriophage BFK20 morphologically belongs to a taxonomical group of unclassified Siphoviridae (Koptides et al., 1992). The phage particle is composed of 50 nm polyhedral head and 200 × 10 nm non-contractible tail. The genome of BFK20 contains a linear double-stranded DNA molecule with 3′ cohesive ends and a GC content of 56.2% (EMBL, accession no. AJ278322). The BFK20 genome consists of 42,968 bp, and it is the first corynephage to be completely sequenced. Using bioinformatics analysis, we have identified 55 putative open reading frames (ORFs) coding for proteins varying in molecular weight (Mw) from 5 to 170 kDa.

Based on the overall genome sequence comparison, similarities with other known bacteriophage genomes were found. Those primarily include bacteriophages from *Mycobacterium* spp. host group and some regions of *Corynebacterium* spp. genomes – possible prophages. Protein-based homology search revealed similarities between BFK20 proteins and other phage and bacterial proteins, which helped predict their functions. Analyses of both DNA and protein support the theory that phage genomes are mosaics with respect to each other. This makes phylogenetic relations of phages more non-linear in comparison to other organisms.

To date, most of the completed genome sequences exist for the phages that infect Gram-negative eubacteria or A + T-rich Gram-positive bacteria such as *Lactococci*, *Streptococci* or *Bacilli*. In this study, we present the complete annotated sequence of the BFK20 genome – a lytic phage of the Gram-positive industrial producer *B. flavum* – and analyze the predicted virion proteins.

Results and discussion

Determination of DNA sequence

The nucleotide sequence of the BFK20 genome was determined by two approaches. Initially, genomic clones containing defined fragments were prepared according to the restriction map (Koptides et al., 1992). The complete nucleotide sequences of these clones were determined step by step using automatic DNA sequencers. Final localization and orientation of sequences on the phage genome were completed by primer walking. Secondly, the whole BFK20 genome was cloned by shotgun, and individual clones were sequenced. The whole nucleotide sequence of BFK20 DNA was assembled (EMBL Acc. No.: AJ278322) using Gap4 program from Staden Package (Staden, 1996). The genome size was estimated as 42,968 bp. Each nucleotide was determined at least twice and up to 10 times in both directions. The overall G + C content of the BFK20 genome is 56.2%. G + C content of *C. glutamicum*, which is considered to be the same species as BFK20 host *B. flavum* (Liebl et al., 1991), is 53.8%.

The presence of *cos* sites is characteristic for phages with a non-headful packaging mechanism. We suggested cohesive ends for the BFK20 genome according to previous results (Koptides et al., 1992). The exact *cos* site sequence was determined by sequencing on genomic DNA in a sequence run-off experiment using the oligonucleotides COS1 and COS2. Primers were situated 277 bp distal to either end of the region containing the expected *cos* site. By comparing these sequences with the previously assembled whole genome (ligated DNA fragments from the shotgun strategy), the *cos* site was assigned to a 13 bp sequence 3′-ACTTCCCCCGCTT and TGAAGGGGGCGAA-3′. Thus, the BFK20 *cos* site possesses a single-stranded 3′ overhang, like many other phages infecting Gram-positive organisms (Brüssow and Desiere, 2001, Chandry et al., 1994; Ganyu et al., 2005; Kaneko et al., 1998; Lillehaug et al., 1993; Mahanivong et al., 2001; van Sinderen et al., 1996).

ORFs prediction and genomic organization of bacteriophage BFK20

Identification of coding regions, located on the BFK20 genome, was based on the application of various methods. These include base preferences, codon preferences, author test, base bias and the presence of START and STOP codons (Staden, 1996). A DNA region was considered as coding if START and STOP codons were present, at least one of the base composition methods was positive and the length of ORF was at least 100 bp. All ORFs begin with ATG/GTG, except for ORFs 14 and 36 which use TTG. As summarized in Table 1, we identified 55 ORFs, predicted as coding regions. These 55