Transcriptional profiling of bacteriophage BFK20: Coexpression interrogated by “guilt-by-association” algorithm

Tomas Majtan ⁎, Nora Halgasova, Gabriela Bukovska, Jozef Timko

Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, 845 51 Bratislava 45, Slovakia

Received 17 May 2006; returned to author for revision 4 September 2006; accepted 19 September 2006

Available online 17 October 2006

Abstract

Global gene expression profiling of bacteriophage BFK20 infecting the industrial L-lysine producer Brevibacterium flavum CCM 251 was performed using DNA microarray. The relative gene expressions were measured in fourteen time samples collected during phage development. Phage genes were classified as early, middle, late or unassigned based on complex expression patterns during infection. Temporal classification of BFK20 genes was in concordance with previous predictions. However, proposed late regulatory genes were reclassified and new functional assignments for ORF55 were strongly suggested. Furthermore, we consider possible functions of other genes and their products regarding coexpression pattern by using “guilt-by-association” algorithm. Microarray results were validated using real-time RT-PCR. The detailed description of phage BFK20 transcriptional profile can answer the basic questions of its life cycle and it also can help to prevent phage contamination during industrial fermentation. In addition, this work presents the first complete microarray time course study of gene expression utilizing loop design.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Bacteriophage BFK20; DNA microarray; Loop design; Transcriptional profile; Siphoviridae; Guilt-by-association

Introduction

Corynephage BFK20 is a lytic phage of the industrial L-lysine producing strain Brevibacterium flavum CCM 251 (Koptides et al., 1992). It is a first corynephage with an entirely sequenced genome that consists of linear double stranded DNA with cos sites (EMBL accession no. AJ278322). Morphologically, the bacteriophage BFK20 belongs to an unclassified Siphoviridae sharing a similar genome organization with other members of this taxonomical group (Brussow and Desiere, 2001). Recent study of the host spectrum revealed that only B. flavum CCM 251 lysed after infection, while abortive infection or adsorption barrier were suggested as defense mechanisms against bacteriophage BFK20 infection in other tested corynebacterial strains (Halgasova et al., 2005). Fifty-five open reading frames (ORFs) were identified on the phage genome using bioinformatics approaches. Phage ORFs were distributed into early or late mainly according to transcription orientation and biological functions for several of them were suggested (Bukovska et al., 2006). Therefore, experimental evidence of the phage BFK20 ORFs temporal classification and biological functions has become inevitable in further detailed characterization of corynephage BFK20.

Traditional methods for gene expression profiling, such as Northern blot analysis, provide only limited impact, while complex relationship between genes and their products remains obscured. The well-established DNA microarray technology has promised whole-genome approach in gene expression studies and transferred this research into the field of functional genomics. The most common application of this high-throughput technique is to detect and measure differential gene expression of identical cell lines exposed to two different conditions. However, in phage research it is much more attractive to study gene expression during the time course of host infection. First complete transcriptional profiles of bacteriophage genomes during host infection were obtained by time-consuming and laborious Northern blot analysis on Streptococcus thermophilus temperate bacteriophage sfi21.
(Ventura et al., 2002) and virulent phage sfi19 (Ventura and Brussow, 2004). To date, microarray analysis of phage gene expression has been performed on coliphage T4 (Luke et al., 2002) and two streptococcal phages DT1 and 2972 (Duplessis et al., 2005).

In this study, we report transcriptional profile of bacteriophage BFK20 during infection of its host B. flavum CCM 251 using custom DNA microarray that was developed to track changes in the phage gene expression. Contrary to bioinformatics results, the BFK20 genes were classified into four distinct groups based on their complex expression behavior. Proposed functional assignments were supplemented by expression data and new functions were proposed for several genes and their products using “guilt-by-association” (GBA) algorithm (Thompson et al., 2002). The GBA has been used to both explain why genes might have correlated expression in the set of experiments and infer what might be a function of the gene coexpressed with the gene of better known function (Wu et al., 2002; Draghici et al., 2003; Lee et al., 2004). Using the GBA approach, it was possible to group non-related or non-homologous genes in the same functional module and vice versa to assign new function for previously characterized genes based on their coexpression motif. Fairly characterized genes may be subsequently used for development of efficient phage defense mechanisms to combat fermentation failures caused by phage contaminations.

Results

Bacteriophage BFK20 growth

Basic characteristics of bacteriophage BFK20 infective cycle such as latent period, the rise period and the burst size were determined. The latent period of BFK20 infecting B. flavum CCM 251 was 60 min (Fig. 1) when propagated in LB medium at 30 °C. The maximal burst was achieved at 110 min indicating the 50-min rise period. The burst size of approximately 90 phage particles per bacterium was calculated as the ratio of the final count of liberated phage particles to the initial count of infected bacterial cells. Latent period was investigated in phage gene expression analysis by using DNA microarray including two samples from rise period and plateau, respectively.

Experimental design

Bacteriophage BFK20 DNA microarrays and its gene expression profiling were made according to the MIAME recommendations (Brazma et al., 2001). Each out of the 55 ORFs of bacteriophage BFK20 DNA was represented by one probe designed from central region of gene coding sequence except for the ORF15 for which three isolated probes were designed: ORF15_1 covering the beginning of the gene, ORF15_2 from the end and ORF15_3 from the central region. The microarrays also included several normalization controls (Supplementary Table). Total RNA samples were obtained from infected cells at fourteen specific time points: first twelve samples covered the latent period and the last two samples were included to monitor the rise period and plateau, respectively. After DNase I treatment, samples were spiked with DsRed RNA and reverse transcribed in the presence of a fluorescent label. Differentially labeled cDNAs for each time were hybridized to the DNA microarrays according to the proposed loop design (Fig. 2). The main loop experiment is represented by the outer 14-unit loop. The verification loop experiment represented by the inner 4-unit loop was performed for those samples, where the most significant changes in expression were detected by the main loop experiment. Thus, one experiment was composed of fourteen or four slides, respectively, corresponding to selected time points of phage development. Both experiments were repeated twice at least. Finally, expression graph of relative abundances of RNA for each ORF over the progress of infection was constructed (Fig. 3a).

The complete information about BFK20 DNA microarray composition and construction was also deposited in...
the NCBI Gene Expression Omnibus (GEO) under the platform accession number GPL2741. MIAME-compliant database also includes the obtained microarray data from loop design experiments under the series accession number GSE3164.

Temporal patterns of BFK20 transcription

Bacteriophage BFK20 genes were classified into early, middle or late or remained unassigned according to the phage BFK20 global expression profiling during host infection. Combination of the template matching algorithm (Pavlidis and Noble, 2001) and K-means clustering (Soukas et al., 2000) was employed for complex temporal classification. A plot of the relative abundance of RNAs for each ORF over the time of infection (Fig. 3a) represented the expression profile of phage genes during the latent period. The obtained gene expression patterns were compared with the known BFK20 genome organization. Finally, the genetic map of the phage, annotated with some ORFs and predicted functional assignments, was constructed (Fig. 3b).

Early genes

The 19 genes comprising early genes are those detected in the first 5 time samples (0, 1, 3, 5 and 7 min). Representative expression curve is shown in Fig. 4.

Transcripts of ORF29, ORF30, ORF31, ORF33, ORF35, ORF39, ORF47 and ORF48 were simultaneously first time detected 1 min post-infection, but they were the most abundant in 7 min (ORF29, ORF30, ORF33, ORF47, ORF48), 10 min (ORF31, ORF35) or 15 min (ORF39) post-infection. RNA expression of the ORF29, ORF30, ORF31, ORF33 and ORF35 was stopped at 30 min after infection, what was in contrast to the remaining ones, which were still detectable at 50 min (ORF48) and 120 min (ORF39 and ORF47). The expression maximum of immediate early transcripts was similar. However, ORF39 transcripts were twice as abundant as the remainder of
Middle genes
homology to other mainly viral genes. The expression patterns within the subcluster and by the lack of unclear function was limited by the variations in gene functional assignments for the genes with unpredicted or regulators. Unfortunately, the data analysis and further possible from different promoters or the activities of two different this feature could be due to the transcription of given sequences detected at 40 min after infection. The possible explanation of the presence of two peaks. The first expression maximum was measured at 15 min post-infection, while the second peak was detected at 40 min after infection. The possible explanation of this feature could be due to the transcription of given sequences from different promoters or the activities of two different regulators. Unfortunately, the data analysis and further possible functional assignments for the genes with unpredicted or unclear function was limited by the variations in gene expression patterns within the subcluster and by the lack of homology to other mainly viral genes.

Middle genes

Expression patterns of ORF37 and ORF50–ORF54 were almost identical particularly concerning the time and the rate of expression (Fig. 4). Their expression began at 7 min post-infection, the maximal transcript abundance was measured at 15 min for ORF37 and ORF50–ORF53 or at 20 min for ORF54 and their transcripts were entirely absent after 40 min. The middle genes cluster represented the most homogeneous group of genes regarding the complex expression behavior and the organization on the BFK20 genome (except for ORF37, which is surrounded by early genes). The almost identical expression profile, the localization on the genome and also the homology results point out to their putative regulatory functions.

Late genes

The 25 ORFs (ORF2–ORF6, ORF8–ORF15 and ORF17–ORF28) were identified as late genes (Fig. 4). Late transcripts were first detected at 20 min post-infection, but they were completely switched on after 40 min. The expression of late genes lasted until 120 min. In general, the microarray data are in agreement with the proposed assignments (Bukovska et al., 2006). However, the predicted subgroup of late genes with regulatory function was excluded from this cluster and now represents the middle genes cluster (Fig. 4). Microarray results also confirmed the model of lysis cassette with the experimentally identified gp24 as lysis and gp26 as holin (Bukovska et al., 2006).

The ORF23 and ORF11 generated the most abundant transcripts. Their relative expression was 79× and 63×, respectively, higher than the expression rate of the least abundant transcript of ORF36. High expression levels were also found for ORF6 and ORF12 coding for major head and major tail protein, respectively. The virion proteins have been analyzed by 2D gel electrophoresis recently (Bukovska et al., 2006). The most intensive spot was identified as the major tail protein gp12, which likely corresponds to the most intensive band on the SDS–PAGE gel of proteins from purified phage particles (Fig. 5). The lack of major capsid protein gp6 in the phage particle protein profile as the one of highly expressed and predominant proteins is most probably due to its chain-mail character typical for the group of phage HK97-like major capsid proteins and was discussed in detail by Bukovska et al. (2006). Presence of small proteins with predicted molecular weight below 14.4 kDa on the gel indicated existence of at least three small structural proteins. Massively transcribed late genes ORF11 and ORF23 might be suitable candidates coding for such small highly expressed structural proteins. Additionally, the time course of phage protein synthesis during host infection revealed the lack of host proteins shut-off (data not shown).

The late cluster also included the largest predicted gene ORF15, for whose detection three different probes were designed. Interestingly, only two marginal probes ORF15_1 and ORF15_2 were clustered together and central probe ORF15_3 remained unassigned. The central probe likely hybridized less efficiently, what was supported by different labeling approaches. The first approach involved preparation of labeled cDNA from 60 min total RNA sample using only ORF15-specific primers. The second approach involved PCR amplification and labeling of the whole ORF15. In all cases, the fluorescent intensities from ORF15 probes copied previous findings (data not shown). This result also indicated that ORF15 is likely coding for only one transcript.
Unassigned genes

A small number of genes did not fit into the early, middle or late gene clusters. The group of unassigned genes included ORF1, ORF7, ORF16, ORF49 and ORF55 (Fig. 4).

Expression profile of ORF1 showed two expression peaks as the early genes ORF40–ORF42, but in reverse order regarding maximal expression. ORF1 first appeared very early post-infection (3 min) and reached the first peak at 10 min at the same average expression level as any other early gene. Subsequently, its transcripts were diminished until 30 min. Finally, the expression led into the second peak at 60 min post-infection. The reasons for such behavior could be the same as for ORF40–ORF42.

Expression patterns of ORF7 and ORF16 were quite similar. Their transcripts were detected at 60 min post-infection for the first time and thus represented the last switched on BFK20 genes. Relative expression of ORF7 and ORF16 was approximately 10× lower than the average relative expression of late genes. Nevertheless, ORF7 and ORF16 could be classified as late genes with altered expression.

Transcripts of ORF49 and ORF55 were first detected at 10 and 15 min post-infection, but maximal expression was reached at 50 and 60 min, respectively. Such a long time gap between the first detection and the maximal expression was not observed for any other gene. Their expression pattern reminded late rather than middle genes (ORF50–ORF54), which are located between them.

Validation of microarray analysis results

Validation of the phage gene expression was carried out using real-time RT-PCR only for five phage genes, whose cluster memberships were clear and whose functional assign-

ments were experimentally confirmed or for which sequence homology search revealed highly significant results (Bukovska et al., 2006). This included late genes ORF2, ORF12 and ORF24 and early genes ORF41 and ORF47.

The deduced relative expression levels of five bacteriophage BFK20 genes during host infection are summarized in semi-log plot of Fig. 6. Gene expression rate measured by real-time RT-PCR was compared with results obtained using DNA microarray. The resulting data clearly confirmed differential expressions of the selected phage genes, their trends of expression in time course of host infection, although PCR assay provided an extended dynamic range than DNA microarray (Chuaqui et al., 2002; Herold et al., 2005). However, it must be mentioned that the input amount of sample into real-time RT-PCR assay was adjusted to maximize dynamic range and not to copy the DNA microarray gene expressions.

Phage DNA replication

Bacteriophage BFK20 DNA replication during time course of host infection was investigated by Southern blot analysis. Total DNAs purified from infected cells during latent period of phage BFK20 infective cycle were hybridized with DNA probe derived from ORF21 coding for tail fiber protein (Fig. 7). We detected specific signal in each selected time point, which intensity was roughly similar for 0 and 10 min sample p.i. and from 20 min constantly grew up until the end of latent period. This indicates the on-going replication of phage BFK20 DNA in host cells during latent period, which is initiated between 10 and 20 min after host infection.

Discussion

We present the complete gene expression profiling of bacteriophage BFK20 during infection of the industrial l-lysine producer B. flavum CCM 251. The DNA microarray was designed and constructed to study phage genes transcription...
during the phage BFK20 life cycle after host infection. The microarray data revealed the relative abundances of each transcript within the complex pool of RNAs and so their expression patterns during the phage development were obtained.

The selection of experimental design represents the guide throughout complete microarray procedure and determines the effectiveness of microarray analysis. The time course study of viral gene expression brings several complications and thus presents unique technical challenge to a conventional microarray arrangement. Not only the mechanisms of rapid transcript production, maintenance and degradation, but also the limited knowledge about individual genes, their relationship and regulation and the lack of a usable phage reference state complicates experimental design and data analysis. We have tried to overcome these shortcomings by using a loop design instead of universal reference design. To date, the reference design was used to track changes in transcript abundances, when labeled sample cDNA was compared to labeled reference cDNA reverse transcribed either from representative pool of transcripts (Luke et al., 2002) or from total RNA of non-infected cells (Duplessis et al., 2005). In these studies, a reference sample provides only universal comparison standard for normalization and thus produces data without any biological relevance to phage gene expression. Moreover, reference sample prepared from total RNA of non-infected cells provides neither constant background nor normalization standard and single-channel DNA microarray experiment could be performed instead of reference design. In contrast, the loop design allowed us to detect small differences in gene expression of closely related states with direct comparison to the preceding and following sample. Furthermore, loop design is at least equally robust as reference design to failed arrays and the precision of differential gene expression measurement is higher (Vinciotti et al., 2005; Tempelman, 2005). Data from main loop were supplemented and also verified with relative expressions obtained from minor validation loop. Finally, the gene expression trends were confirmed by real-time RT-PCR and genetic map of phage BFK20 containing temporal classification of genes with several strongly proposed functions was constructed (Fig. 3b).

Phage BFK20 genes were clustered into four groups defined as early, middle, late and unassigned based on the gene transcription pattern during infection cycle. Genes could be classified based on the time when the transcript first appeared (Luke et al., 2002), when the transcript became the most abundant (Paulose-Murphy et al., 2001) or following the period of time during which the greatest change in transcript abundance occurred (Duplessis et al., 2005). The discrepancies in temporal classifications are clearly visible, when comparison of transcriptional maps of homologous genes is being performed. Transcriptional profile of bacteriophage sfi19 was obtained as early, middle, late and unassigned based on the gene transcription pattern during infection cycle. Genes could be classified based on the time when the transcript first appeared (Luke et al., 2002), when the transcript became the most abundant (Paulose-Murphy et al., 2001) or following the period of time during which the greatest change in transcript abundance occurred (Duplessis et al., 2005). The discrepancies in temporal classifications are clearly visible, when comparison of transcriptional maps of homologous genes is being performed. Transcriptional profile of bacteriophage DT1 provided by DNA microarray (Duplessis et al., 2005). The temporal classification shift of several homologous sfi19 middle genes to DT1 early genes as well as sfi19 late genes to DT1 middle genes was shown. We suggested that our temporal classification considered all parameters of gene expression thus establishing solid base for correct gene clustering.

The transcriptional map of bacteriophage BFK20 shows gene expressions during the phage latent period (Fig. 3a). The plot of the relative transcript abundances at infection time revealed that the early transcripts were not strictly terminated when the switch to middle and subsequently to late genes occurred. The absence
or alteration of the mRNA processing and degradation machinery plays central role in determining the phage transcripts abundance. Presence of mRNA processing and degradation system, such as bacteriophage T4 endoribonuclease regB (Sanson et al., 2000) or host encoded RNAses (Mudd et al., 1988; Cheng and Deutscher, 2005), allows for rapid switch to late genes with complete cut-off regarding early transcription. In the case of bacteriophage BFK20, there is no evidence about phage-encoded RNA degradation enzymes, which is supported by both the bioinformatics analysis (Bukovska et al., 2006) and the DNA microarray transcriptional profiling. Likewise, gene expression slows down very gently after reaching the end of the latent period. The unique mechanism of phage transcript protection and stabilization was demonstrated for bacteriophage T7 gene product 0.7 (Marchand et al., 2001a), which exhibits protein kinase activity. The bacteriophage BFK20 proposed pseudogene ORF4 represents a highly transcribed late gene, whose gene product contains the ATP-binding site (Bukovska et al., 2006) and could be active in similar way to T7 gene product 0.7 (Marchand et al., 2001b). However, further experiments including gp4 over-expression and mutation must be performed to verify this hypothesis.

Our classification revealed several discrepancies compared to the proposed temporal clustering (Bukovska et al., 2006). The region covering ORF50–ORF55 was previously assigned to late genes. We found their expression profiles almost identical to each other (ORF50–ORF54) and according to timing of their expression they created middle gene cluster together with ORF37. We suggest that the middle genes are responsible for retransferring the transcriptional machinery from early to the late promoters. The gene products of both ORF37 and ORF54, possessing lambda-repressor HTH signature, may repress early transcription, while others could initiate late transcription. ORF37 is separated from remaining middle genes and is transcribed from opposite strand regarding the transcription orientation of ORF50–54. We suggest that the ORF37 transcription is regulated mainly from its own promoter. However, proposed transcriptional repressors do not initiate such a strong mRNA decay in early transcripts as was shown for the T4 phage (Luke et al., 2002), which could also be influenced by absence or alteration of mRNA processing machinery (see above).

The ORF40–ORF42 transcription patterns revealed the presence of two expression peaks: at the early stage and also at the beginning of the late stage of phage development. Gene expression of ORF40–ORF42 in the late phase of phage cycle could be due to activity of the experimentally identified F1 promoter, which is located upstream to this gene cluster (GenBank accession no. L13772; Koptides et al., 1992). ORF41 encodes a putative helicase, while ORF40 and ORF42 encode proteins with unknown functions possessing homology to several putative phage proteins (Bukovska et al., 2006). Similar gene expression pattern with two peaks was observed for putative helicases of streptococcal phages DT1 and 2972 (Duplessis et al., 2005). We consider that the potential helicase gp41 is involved both in phage DNA replication process and in phage recombination, repair and regulation of replication origin such as phage T4 helicase uvsW (Carles-Kinch et al., 1997; Dudas and Kreuzer, 2001; Sickmier et al., 2004). Gene products of ORF40 and ORF42 may participate in such function. Moreover, the gene encoding putative helicase gp41 or other gene from ORF40–ORF42 cluster may serve as suitable targets for a development of effective antisense RNA-based phage defense strategy (Sturino and Klaenhammer, 2002, 2004).

Similar feature in expression profile with the presence of two peaks was observed for the unassigned gene ORF1. The gene product of ORF1 is a candidate for the small phage terminase subunit according to bioinformatics analysis, but it may also participate in DNA replication or transcription regulation process. Similar phenomenon in gene expression of the proposed small phage terminase subunit was shown for cos-type phage DT1, but not for pac-type phage 2972 (Duplessis et al., 2005), what could pointed out to the different involvement of small terminase in cos-or pac-type packaging mechanisms. We propose the existence of at least two promoters for this gene. Alternatively, the activity of a single promoter might be dependent on the superhelicity of DNA, which might change during the phage’s life cycle. During the intracellular phage development, when phage DNA is either circularized or in a concatameric structure, promoters’ activity should be partially coupled with expression of other unassigned gene ORF55 as well as late genes located downstream to ORF1. The expression pattern of ORF55 did not copy the transcription of the preceding middle genes, but rather followed late genes. Similar phenomenon was observed for bacteriophages DT1, sf19 and sf21. In these phages the last gene in the group of proposed transcription regulators was classified in a different temporal group to the rest of the region. The reason for such behavior remains unclear. ORF55 encodes a putative HNH endonuclease. According to the results of Crutz-Le Coq et al. (2002), the gene product of ORF55 could be included in the same functional module as terminase (ORF2) and is possibly involved in DNA packaging. If so, then the gene product of ORF55 could act either as a site-specific endonuclease playing analogous role to the phage large terminase subunit, or structure-specific endonuclease clearing branched replicative DNA prior the packaging as bacteriophage T4 endonuclease VII (Golz and Kemper, 1999) or bacteriophage T7 endonuclease I (Picksley et al., 1990).

Our study provides complex view on global gene expression of bacteriophage BFK20 during host infection and on genome organization of the Siphoviridae family in general. Microarrays make possible to follow each stage of bacteriophage development on a gene-by-gene basis in a complex manner. We found the loop design very useful and suitable for the viral time course gene expression studies. The expression patterns of phage genes during host infection are helpful for a detailed characterization and functional assignments of related genes either by sequence homology or on a transcriptional profile basis (“guilt-by-association” algorithm; Thompson et al., 2002; Wolfe et al., 2005). In addition, developed DNA microarray can be used for comparative genome analysis to elucidate corynephages evolution and relationship. Such analysis could reveal well-conserved genes shared between various genomes. Thoroughly described genes and their products could subsequently be utilized in the
development of effective anti-phage mechanisms such as antisense RNAs, whereas phage contaminations represent the still emerging issues during biotechnological fermentations.

Materials and methods

Bacterial strain, bacteriophage, media and culture conditions

Corynebacterium sp. B. flavum CC351 was propagated on L-lysine producer B. flavum CC351 (hse<sup>+</sup>, aec<sup>+</sup>) as an exclusive host. Preparation and purification of phage lysate and phage DNA were performed as described previously (Koptides et al., 1992; Halgasova et al., 2005).

Bacterial cells were grown in LB medium or on solid LB agar plates (Sambrook et al., 1989). Bacterial cultures were incubated at 30 °C.

Bacteriophage propagation and assays

The phage stock and phage titer were performed as reported elsewhere (Halgasova et al., 2005). One-step growth experiment was performed as described by Pajunen et al. (2000) and Chang et al. (2005) with modifications. A B. flavum CC351 culture (OD<sub>570</sub> 0.3) was infected by phage BFK20 giving the MOI of 0.1 and allowed to adsorb for 5 min at RT. After centrifugation at 6 000 × g for 5 min, infected cells were resuspended at their original density in pre-warmed LB medium and incubation continued at 30 °C with vigorous shaking. Samples were taken periodically over 120 min, immediately diluted and plated for phage titration. The complete experiment was done in triplicate. The latent period was defined as time between infection and initial lysis of infected cells. Total time interval from the end of the latent period to the completion of bacterial lysis represented the rise period. The average burst size was calculated as the ratio of the final count of liberated phage particles to the initial count of infected bacterial cells.

Total RNA preparation and purification

The bacterial suspensions for the time course experiment were prepared according to the protocol of phage stock preparation except that MOI was elevated to fifteen to ensure complete infection of host cells. After infection, suspensions were cultivated at 30 °C and aliquot samples were taken at 0, 1, 3, 5, 7, 10, 15, 20, 30, 40, 50, 60, 80 and 120 min. Total RNA was isolated using hot phenol–chloroform extraction protocol according to Kormanec (2001). The contaminating DNA was removed using DNase I treatment. Quality, quantity and purity of extracted total RNAs were determined by electrophoresis in denaturing formamide gel and spectrophotometrically.

Spike control development and preparation

New spiking control system was developed in our laboratory in order to control labeling process and for normalization purposes. It employed red fluorescent protein DsRed nucleotide sequence located downstream to T7 promoter.

Briefly, DsRed coding sequence was cut out from pDsRed expression vector (Clontech, USA) by restriction endonucleases BamHI and NotI. The 703-bp fragment was inserted into pET-26b (+) expression vector (Novagen, USA) digested with BamHI/NotI and the recombinant plasmid pET26bDsRed was isolated. 1.5 μg of pET26bDsRed linearized with Xhol served as a template DNA for in vitro transcription. The reaction was performed by using 150 U of T7 RNA polymerase (NEB, UK) in the presence of 100 U of RNase Inhibitor (Roche, Germany) for 3 h at 37 °C and subsequently stopped with 4 mM EDTA. After DNase I treatment, DsRed RNA was purified using RNAeasy Mini Kit (Qiagen, Germany).

The DsRed RNA spike was subsequently used in every labeling reaction. The spike was tested for signal sensitivity in the concentration range from 5 ng to 200 ng (data not shown). We found that 50 ng of DsRed RNA in a 25-μg total RNA sample provided the best sensitivity and signal saturation. Moreover, the designed microarray probe proved to be highly specific for its target cDNA and also in complex samples no cross-hybridization was observed. For the bacteriophage BFK20 gene expression study, the spiking control served as a positive control of the labeling process and the ratio control.

Subsequently, it turned out to be useful for with-in array normalization in combination with other control elements located on the array.

Labeling

Total RNAs were labeled using standard cDNA protocol (Hegde et al., 2000) with modifications. The reverse transcription was performed in the dark on the UNO-II thermocycler (Biometra, Germany). The reaction consisting of 20 μg purified total RNA, 9 μg of random hexamers (Invitrogen, USA) and 50 ng of DsRed RNA spike was mixed and heated for 10 min at 70 °C and immediately chilled on ice. RNA was reverse transcribed using 400 U of SuperScript II RNaseH<sup>-</sup> (Invitrogen, USA) in the presence of 2 mM dATP, dTTP, dGTP, 1 mM dCTP and 1 mM Cy3-dCTP or Cy5-dCTP (Amersham Biosciences, UK) and 25 U of RNase OUT Ribonuclease Inhibitor (Invitrogen, USA). The reaction was heated for 10 min at 23 °C followed by 3 h incubation at 42 °C. Unreacted dye was removed using QIAquick PCR Purification Kit (Qiagen, Germany).

Microarray design and construction

Primer design

The BFK20 DNA microarrays were generated by deposition of PCR products that were specific to each open reading frame (ORF) of bacteriophage BFK20 genome. The primers used for probe generation and amplification were designed according to known BFK20 genome sequence (GenBank accession no. AJ278322). In addition, constructed DNA microarrays contained several control probes. From host B. flavum CC351 were selected sigA (GenBank accession no. AF320817), sigB
Crosslinked by UV irradiation at 250 mJ (UV Stratalinker 1800, Telechem International, USA) in a thermostated room (22 °C) and humidity (47%). DNA microarrays were dried at 37 °C for 1 h, then stored lyophilized at −80 °C until hybridization.

Microarray printing and processing

BFK20 probes were amplified from purified phage genomic DNA. The control probes were prepared by PCR using genomic DNA from B. flavum CCM 251 or pDsRed and pBSdotA plasmid DNA as the templates. pBSdotA contains 483 bp dotA fragment originated from C. burnetti (GenBank accession no. NC_002971) (gift from Dr. J. Melnicakova, Institute of Virology, Bratislava, Slovakia). PCR was conducted on the UNO-II thermocycler (Biometra, Germany) in total volume of 50 μl using 2.5 U of Pfu DNA polymerase (Fermentas, Germany). Each PCR probe was amplified from 100 pg of template DNA and the reaction followed this thermal profile: initial heating to 95 °C for 5 min followed by 36 cycles of 95 °C for 1 min and 72 °C for 1 min.

The quality of PCR amplicons was examined by electrophoresis on 1.5% agarose gel with GeneRuler 100 bp DNA Ladder (Fermentas, Germany). The quality of PCR amplicons was examined by electrophoresis on 1.5% agarose gel with GeneRuler 100 bp DNA Ladder (Fermentas, Germany) as a marker. If non-specific DNA fragments were present, the specific PCR amplicon was cut out from the gel and isolated. Diluted DNAs from specific PCR products were subsequently used as the templates for the second-round PCR amplification following the same conditions as mentioned above.

After verification, PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Germany) and stored lyophilized at −70 °C until printing.

PCR amplification and purification

BFK20 probes were amplified from purified phage genomic DNA. The control probes were prepared by PCR using genomic DNA from B. flavum CCM 251 or pDsRed and pBSdotA plasmid DNA as the templates. pBSdotA contains 483 bp dotA fragment originated from C. burnetti (GenBank accession no. NC_002971) (gift from Dr. J. Melnicakova, Institute of Virology, Bratislava, Slovakia). PCR was conducted on the UNO-II thermocycler (Biometra, Germany) in total volume of 50 μl using 2.5 U of Pfu DNA polymerase (Fermentas, Germany). Each PCR probe was amplified from 100 pg of template DNA and the reaction followed this thermal profile: initial heating to 95 °C for 5 min followed by 36 cycles of 95 °C for 1 min and 72 °C for 1 min.

The quality of PCR amplicons was examined by electrophoresis on 1.5% agarose gel with GeneRuler 100 bp DNA Ladder (Fermentas, Germany) as a marker. If non-specific DNA fragments were present, the specific PCR amplicon was cut out from the gel and isolated. Diluted DNAs from specific PCR products were subsequently used as the templates for the second-round PCR amplification following the same conditions as mentioned above.

After verification, PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Germany) and stored lyophilized at −70 °C until printing.

Microarray printing and processing

The lyophilized PCR probes were dissolved in 50% DMSO to a final concentration of 200 ng/μl and transferred into sterile 384-well microplate. The plate was subsequently attached into PixSys 5500 microarrayer (Cartesian Technologies, USA) equipped with one ChipMaker CMP4 Micro Spotting Pin (Telechem International, USA).

Probes were deposited onto the GAPS II Coated Slides (Corning, USA) in four replicates per grid and two grids per slide, which allowed for two separate experiments at one time. DNA microarray printing was performed at constant temperature (22 °C) and humidity (47%). DNA microarrays were dried overnight at room temperature. Afterwards, the slides were crosslinked by UV irradiation at 250 mJ (UV Stratalinker 1800, Stratagene, USA) and stored in the dark.

The spotting result was verified using CheckIt Chips Kit (Telechem International, USA).

Hybridization and washing

Hybridization protocol followed the standard hybridization procedure for DNA microarrays (Hegde et al., 2000). Microarray slides were incubated in preheated (42 °C) prehybridization buffer (5× SSC, 0.1% SDS and 1% BSA) for 1.5 h with initial 20 min vigorous shaking. After incubation, slides were rinsed 4× with ddH2O and subsequently immersed into boiling ddH2O for 2 min. DNA probes were subsequently fixed by dipping slide into prechilled 100% isopropanol for 5 s. Slides were dried with compressed air and placed into a thermostat at 42 °C until hybridization.

Hybridization was carried out in a Hybridization Chamber (Corning, USA). Differentially labeled cDNAs were combined together and dried at 85 °C. Pellets were completely dissolved in 12 μl of the hybridization buffer (50% formamide, 6× SSC, 5× Denhardt’s solution, 0.5% SDS and 50 mM KH2PO4). Samples were subsequently denatured at 95 °C for 5 min, immediately chilled on ice and spun in a microcentrifuge. Afterwards, they were incubated at 60 °C until applied onto the DNA microarray. The hybridization was performed overnight for 16 h at 42 °C. Then the slides were washed using an initial wash in 2× SSC and 0.1% SDS preheated to 50 °C for 5 min, followed by 5 min wash in 2× SSC and finished by 1 min wash in 0.5× SSC both at room temperature. The slides were immediately dried with compressed air.

Image capture and data analysis

Fluorescence images were recorded by scanning the slides with a GeneTAC UC4 microarray scanner (Genomic Solutions, USA) at 10 μm resolution. Fluorescence signals from both channels were subsequently quantified using Spotfinder (Saeed et al., 2003). The mean signal intensity was determined for each spot, while local background surrounding the spot area was subtracted from each intensity value. The signals were taken account only if they had been 3 times over the average background. Exported tab-delimited files were trimmed in MS Excel to integrate mean intensity values from replicate spots and to add the annotation.

Within-array normalization of obtained spot mean intensities was done manually, separately for each channel using linear regression. The regression curve was constructed from the set of controls presented on the array. Between-array normalization was performed by comparison of reverse channels’ individual intensities and scaling the loop of experiments according the obtained factors. Normalized data for each experiment in the loop were converted back to tab-delimited files.

The normalized and filtered expression files were finally analyzed and statistics was performed by using MultiExperiment Viewer (TM4 package, TIGR, USA).

SDS-PAGE

The phage BFK20 structural proteins were separated on a 12% SDS–PAGE gel using the Emperor Penguin Dual Gel...
apparatus (Owl, USA) according to Laemmli (1970). Gel was stained with silver for detection.

**Real-time RT-PCR**

We selected eight genes and for each of them five different times according to minor loop sampling for validation using real-time RT-PCR approach (Fig. 2). Target genes represented one proposed early gene, one regulatory gene, three proposed late genes (Bukovska et al., 2006) and three controls (16 S rDNA, atpA, DsRed) (Supplementary Table).

The first strand cDNAs were prepared following the same protocol as for microarray experiment, where 5 μg total RNA and 25 ng DsRed spike were primed with 100 ng random hexamers in the absence of labeled deoxynucleotide. RNA strands were removed by RNase H treatment.

PCR reactions were performed on the 7900HT Fast Real-Time PCR System (Applied Biosystems, USA) in a total volume of 20 μl. Each PCR reaction was performed in 3 replicates. Each well in 96-well PCR plate contained 10 pmol of forward and reverse primer, template cDNA in the desired replicates. Each well in 96-well PCR plate contained 10 pmol of forward and reverse primer, template cDNA in the desired replicates. Time PCR System (Applied Biosystems, USA) in a total volume of 20 μl. Each PCR reaction was performed in 3 replicates. Each well in 96-well PCR plate contained 10 pmol of forward and reverse primer, template cDNA in the desired replcates. Each well in 96-well PCR plate contained 10 pmol of forward and reverse primer, template cDNA in the desired replicates. Each well in 96-well PCR plate contained 10 pmol of forward and reverse primer, template cDNA in the desired replicates. Each well in 96-well PCR plate contained 10 pmol of forward and reverse primer, template cDNA in the desired replicates.

PCR amplification followed this thermal profile: 95 °C for 10 min followed by 40 cycles of 95 °C for 20 s, 56 °C for 45 s and 72 °C for 45 s and was finished by melting curve analysis. Raw data were obtained by using SDS software (Applied Biosystems, USA) and the results were analyzed in MS Excel. Expression rate of the selected genes was calculated using comparative CT method ($\Delta\Delta CT$).

**Southern blot analysis**

Sample preparation and Southern blot hybridization was performed according to Halgasova et al. (2005). Total DNAs from 0, 10, 20, 40 and 60 min after phage infection digested by restriction endonuclease KpnI were separated in 0.9% (w/v) TAE agarose gel and subsequently transferred on Hybond N membrane (Amersham Biosciences, UK). The DNA fragments were hybridized with random-prime digoxigenin (DIG)-labeled DNA probe (Roche, Germany) derived from ORF21 coding for tail fiber protein of BFK20 DNA. Hybridization signals were detected by DIG Detection Kit using NBT/BCIP (Roche, Germany).

**Acknowledgments**

We would like to thank Marek Basler (Institute of Microbiology ASCR, Prague) for assistance in DNA microarrays’ printing and Dr. Hana Drahovska (Faculty of Natural Sciences CU, Bratislava) for help with the quantitative PCR experiment. We are grateful to Dr. Imrich Barak (Institute of Molecular Biology SAS, Bratislava), Dr. Vladimir Kery (Systems Biology PNLL, WA, USA) and Prof. Jan P. Kraus (UCHSC, CO, USA) for critical reading of manuscript and helpful comments. This work was supported by VEGA grant 2/5068/25 from the Slovak Academy of Sciences and in part by grant APVT-21-052602 from Science and Technology Assistance Agency.

**Appendix A. Supplementary data**


**References**


