Journal of Microbiological Methods 77 (2009) 207–213

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

# Representational Difference Analysis (RDA) of bacteriophage genomes

Elke Lammens, Pieter-Jan Ceyssens, Marleen Voet, Kirsten Hertveldt, Rob Lavigne *, Guido Volckaert

*Division of Gene Technology, Katholieke Universiteit Leuven, Kasteelpark Arenberg 21, Leuven, B-3001, Belgium*

ARTICLE INFO

*Article history:*
Received 6 November 2008
Received in revised form 29 January 2009
Accepted 8 February 2009
Available online 20 February 2009

*Keywords:*
RDA strategy
Bacteriophages
Screening method

ABSTRACT

We implemented the Representational Difference Analysis (RDA) screening method to identify genome variations between related bacteriophages without the need for complete genome sequencing. The strategy, optimized on ϕKMV and LKD16 and further evaluated on the newly isolated phage LUZ19, is based on three successive rounds of reciprocal RDA, with an increasing driver/tester molar ratio from 100/1 to 750/1. Using three relevant restriction endonucleases, only 4 to 6 sequences per restriction enzyme are necessary to provide sufficient discriminatory information to reveal the major genome variations between phages.



## 1. Introduction

Representational Difference Analysis (RDA), first described by Lisitsyn et al. (1993), is a powerful technique to facilitate the identification of differences in the nucleic acid composition of two samples (Sagerström et al., 1997). It combines the use of representations with subtractive hybridization of 'driver' and 'tester' samples and kinetic enrichment (Lisitsyn et al., 1993). The technique underwent several technical improvements (Hubank and Schatz, 1994; Strathdee and Johnson, 1995; Baldocchi and Flaherty, 1997) and a simplified protocol was introduced by Felske (2002).

Whereas RDA was first used to identify polymorphisms in human neoplasia, the approach has now been conducted on other genomes such as those of bacteria (Tinsley and Nassif, 1996), eukaryotic cells (Vanhaaren and Ow, 1993) and their viruses by comparing infected to non-infected cells (Chang et al., 1994). However, this hybridization based technique was not yet conducted to identify variations in small genomes like those of bacterial viruses.

Bacteriophage genome sequencing projects reveal distinct subgroups of closely related virulent phages like the T4-like, ϕKZ-like and ϕKMV-like phages (Nolan et al., 2006; Krylov et al., 2007; Ceyssens et al., 2006). These phage genomes contain hypervariable regions, which seem to be directly related to adaptation of the phage to a particular host or environmental niche (Nolan et al., 2006). When studying newly isolated phages, restriction analysis, electron microscopy (EM) and Southern hybridization experiments can pinpoint phages to a specific subgroup. However, detailed information concerning genomic insertions, deletions and/or recent horizontal exchange typically requires whole-genome sequencing.

A bacteriophage specific RDA screening method was developed to rapidly identify major genetic variations among related phages. All relevant parameters of the approach were first analyzed on the known genomes of ϕKMV and LKD16. The optimized technique was evaluated

**Table 1**
Sequences of the adaptor sets used for Representational Difference Analysis.

| Adaptor set | Adaptor[a] | Sequence |
|---|---|---|
| 1 | DHhaI10 | 5′-CGG-TCA-GAG-G-3′ |
| | DHhaI24 | 5′-AGC-ACT-CTC-CAG-CCT-CTG-ACC-GCG-3′ |
| 2 | DNlaIII10 | 5′-CGG-TCA-GAG-G-3′ |
| | DNlaIII24 | 5′-CAC-TCT-CAC-GCC-TCT-GAC-CGC-ATG-3′ |
| 3 | DCsp6I10 | 5′-TAC-GGT-CAG-A-3′ |
| | DCsp6I24 | 5′-TGA-GCA-CTC-TCC-AGC-CTC-TGA-CCG-3′ |
| 4 | T1HhaI10 | 5′-CTC-CCT-CGG-A-3′ |
| | T1HhaI24 | 5′-GCA-ACT-GTG-CTA-TCC-GAG-GGA-GCG-3′ |
| 5 | T1NlaIII10 | 5′-TCC-CTC-GGA-T-3′ |
| | T1NlaIII24 | 5′-CAA-CTG-TGC-TAT-CCG-AGG-GAC-ATG-3′ |
| 6 | T1Csp6I10 | 5′-TAC-TCC-CTC-G-3′ |
| | T1Csp6I24 | 5′-AGG-CAA-CTG-TGC-TAT-CCG-AGG-GAG-3′ |
| 7 | T2Hha10 | 5′-CCT-GGT-AGA-T-3′ |
| | T2HhaI24 | 5′-CCG-ACG-TCG-ACT-ATC-TAC-CAA-GCG-3′ |
| 8 | T2NlaIII10 | 5′-TTG-GTA-GAT-A-3′ |
| | T2NlaIII24 | 5′-CGA-CGT-CGA-CTA-TCT-ACC-AAC-ATG-3′ |
| 9 | T2Csp6I10 | 5′-TAC-TTG-GTA-G-3′ |
| | T2Csp6I24 | 5′-GAC-CGA-CGT-CGA-CTA-TCT-ACC-AAG-3′ |

[a] Adaptor sets consist of a short (10-mer) and long (24-mer) adaptor. The names of the driver adaptors start with D, tester adaptors with T1 (first and third round) and T2 (second round).

* Corresponding author. Fax: +32 16 32 19 65.
*E-mail address:* rob.lavigne@biw.kuleuven.be (R. Lavigne).

0167-7012/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2009.02.006

**Table 2**
*In silico* analysis of LKD16 and ϕKMV genomes cut with AluI, Csp6I, HaeIII, HhaI, HpaII, NlaIII and TaqI.

| Organism | | LKD16 | | | ϕKMV | | |
|---|---|---|---|---|---|---|---|
| Properties | | Average (bp) | Range (bp) | Fragments lost[a] (%) | Average (bp) | Range (bp) | Fragments lost[a] (%) |
| Restriction enzymes | AluI | 216 | 5–1287 | 8 | 241.5 | 5–1605 | 6 |
| | Csp6I | 220 | 6–1244 | 8 | 210 | 4–1403 | 9 |
| | HaeIII | 97 | 4–598 | 30 | 93 | 4–617 | 30 |
| | HhaI | 111 | 2–746 | 23 | 124 | 2–730 | 18 |
| | HpaII | 129 | 5–761 | 19 | 122 | 5–1029 | 18 |
| | NlaIII | 208 | 7–1066 | 8 | 204 | 4–984 | 9 |
| | TaqI | 162 | 5–1280 | 13 | 170 | 4–1311 | 11 |

[a] Fragments smaller than 100 bp, lost due to purification steps in the procedure.

by the identification of the differences between ϕKMV and the newly isolated ϕKMV-like phage LUZ19.

## 2. Materials and methods

### 2.1. Phages, bacterial strains, vectors and growth media

Lytic *Podoviridae* ϕKMV and LKD16 have previously been described (Lavigne et al., 2003; Ceyssens et al., 2006) and are classified as members of the ϕKMV-like viruses within the *Autographivirinae* subfamily (Lavigne et al., 2008).

All RDA cloning steps were performed in the *Escherichia coli* strain DH5α. The bacterial strains were grown in standard Luria–Bertani medium. Plasmid pUC18, digested with SmaI, was used as cloning vector.

### 2.2. Bacteriophage isolation, purification and characterization

LUZ19 was isolated by a standard enrichment method using *P. aeruginosa* PAO1 as host bacterium (Carlson, 2005). The phages were amplified by infecting exponentially growing *P. aeruginosa* PAO1 cultures at an MOI of 0.1, followed by incubation at 37 °C until lysis was visible. Phages were precipitated in 8% PEG8000 and subsequently purified by two successive rounds of CsCl gradient centrifugation. The genomic DNA was isolated as described elsewhere (Sambrook and Russell, 2001).

### 2.3. Representational Difference Analysis

The RDA analyses were generally based on the procedure of Felske (2002). Genomic DNA (1 μg) of both driver and tester strains was digested with either Csp6I, HhaI or NlaIII and purified (QIAquick PCR Purification Kit, QIAGEN, Düsseldorf, Germany). Driver representations were ligated to a set of driver-specific (D) adaptors (0.5 nmol adaptor/μg digested DNA, adaptor sets 1, 2 and 3 in Table 1). The driver samples were prepared by subsequent amplification using a standard PCR (30 s 95 °C followed by 40 cycles of 10 s 95 °C, 30 s 52 °C, 2 min 72 °C) with GoTaq DNA polymerase (Promega, Madison, Wisconsin, USA). These driver mixtures, further referred to as 'drivers', were purified and digested with the same enzyme used in the first step. The tester mixtures were obtained by ligation of the purified digestions to tester-specific (T1) adaptor sets (0.5 nmol adaptor/μg DNA, adaptor sets 4, 5 and 6 in Table 1) followed by purification, and are further referred to as 'testers-1'.

The drivers and testers-1 samples were hybridized (16 h, 67 °C) in a 100/1 molecular ratio in the presence of 5× EE buffer (Straus and Ausubel, 1990; 50 mM EPPS (4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid), 5 mM EDTA, pH 8) and 0.25 M sodium chloride, followed by amplification with the long tester adaptors. The purified PCR products were digested and ligated to a new set of tester-specific T2 adaptors (Table 1: adaptor sets 7, 8 and 9). The samples, referred to as 'testers-2', were used for a second round of RDA with 500/1 driver/tester ratios. Subsequently, T2 adaptors were removed from the differential products and T1 adaptors were linked, creating 'testers-3'. These were further subjected to a third round of RDA with a 750/1 molecular ratio (Supplementary Fig. 1).

The final differential products of each round and each restriction enzyme were ligated in the SmaI site of pUC18 and *E. coli* DH5α were transformed with these constructs. Extracted plasmid DNA was sequenced using the forward M13 primer. The resulting sequences of all restriction enzymes were pooled and examined with ContigExpress (Vector NTI Advance 10, Invitrogen, Carlsbad, California, USA) and were compared to those available in the GenBank database, using Blastn (Altschul et al., 1990).

### 2.4. Bacteriophage genome sequencing and annotation

Due to previous studies of ϕKMV and LKD16, a large set of 253 primers specific for ϕKMV-like phages was available. DNA sequencing reactions performed with these ϕKMV primers and genomic LUZ19 DNA as template resulted in an 83% success rate. Remaining gaps and uncertainties were determined by 27 LUZ19-specific primers, completing the entire genome sequence with an average fourfold redundancy. Despite availability of compatible primers, a total of 400 sequencing runs were necessary to assemble the genome sequence. Potential open reading frames (ORFs), promoters and conserved sequences were identified as described elsewhere (Ceyssens et al.,

**Fig. 1.** Optimization of the RDA parameters. *A and B. Successive rounds with equal driver/tester ratio.* Representational Difference Analysis was performed with HhaI and NlaIII representations of LKD16 (tester) and ϕKMV divided per 50 base pairs, as a percentage, are indicated in grey. The percentages of the differential sequences per round and per 50 base pairs (pooled for all restriction enzymes) are visualized with colored column bars (part A; round 1: purple, round 2: red, round 3: green). The cumulative percentage of the differential sequences per 50 base pairs is shown as a colored line, analogous to the color of the successive rounds (part B). A schematic overview of LKD16's open reading frames was provided on the *X* axis. The RDA output of the third round reveals several differential regions, such as the early region up to 5000 bp, between 9600 and 12,000 bp (containing ORFs 16 to 19) and the tail fiber region up to ORF 43 (between 36,000 and 40,000 bp). The regions between 20,200 and 22,500 bp (ORFs 26 to 30) and between 26,000 and 26,700 (ORFs 33 and 34) were also represented, although those sequences are only slightly differential (less than 20% differences per 50 base pairs). *C. Successive rounds with increasing driver/tester ratio.* RDA was performed with HhaI and NlaIII representations of LKD16 (tester) and ϕKMV (driver). In the first round, a ratio driver/tester of 100/1 was tested. The second round, starting with the amplified differential products of the first round, was performed with a driver/tester ratio of 500/1. The output of the latter was further subjected to a third round using a 750/1 driver/tester ratio. The cumulative percentage of the differential sequences per 50 base pairs is shown as a colored line, analogous to the colors of the successive rounds in A and B. The RDA output reveals all major differential regions, with a high amplification of the early region (up to 5000 bp) and the tail fibers. The low differential regions containing ORFs 26 to 30 and 33 and 34 were also represented in the first and second round, but almost not in the third round (5% of the output sequences). *D. Appropriate enzymes to create representations.* Representational Difference Analysis was performed with Csp6I, HhaI and NlaIII representations of LKD16 (tester) and ϕKMV (driver), with increasing driver/tester ratio. The cumulative output of the third round is demonstrated with a green line. Per enzyme, the location of differential sequences after the third round was analyzed. The cumulative graphs are indicated in yellow (Csp6I), turquoise (HhaI) and blue (NlaIII). Csp6I differential products tended to be mostly fragments of the region coding for ORFs 16 till 17.1 (50%) and of the early differential region (45%). HhaI differential sequences were mainly clustered in the early region (84%), while NlaIII fragments were divided between the early (53%) and the late region (33%).