



DPS – A rapid method for genome sequencing of DNA-containing bacteriophages directly from a single plaque



Witold Kot^{a,*}, Finn K. Vogensen^b, Søren J. Sørensen^a, Lars H. Hansen^{a,c}

^a Department of Biology, Faculty of Science, University of Copenhagen, Universitetsparken 15, DK-2100 København Ø, Denmark

^b Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

^c Department of Environmental Science, Aarhus University, Frederiksborgvej 399, Roskilde, Denmark

A B S T R A C T

Article history:

Received 16 July 2013

Received in revised form 25 October 2013

Accepted 29 October 2013

Available online 14 November 2013

Keywords:

Phages

High-throughput sequencing

Full genome sequencing

Single plaque sequencing

Bacteriophages (phages) coexist with bacteria in all environments and influence microbial diversity, evolution and industrial production processes. As a result of this major impact of phages on microbes, tools that allow rapid characterization of phages are needed. Today, one of the most powerful methods for characterization of phages is determination of the whole genome using high throughput sequencing approaches. Here a direct plaque sequencing (DPS) is described, which is a rapid method that allows easy full genome sequencing of DNA-containing phages using the Nextera XT™ kit. A combination of host–DNA removal followed by purification and concentration of the viral DNA, allowed the construction of Illumina-compatible sequencing libraries using the Nextera™ XT technology directly from single phage plaques without any whole genome amplification step. This method was tested on three *Caudovirales* phages; ϕ 29 *Podoviridae*, P113g *Siphoviridae* and T4 *Myoviridae*, which are representative of >96% of all known phages, and were sequenced using the Illumina MiSeq platform. Successful de novo assembly of the viral genomes was possible.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Bacteriophages (phages) are the most abundant biological systems on the planet (Chibani-Chennoufi et al., 2004), and can be found and isolated from different environments including water, soil, air and the intestinal tract (Bergh et al., 1989; Vos et al., 2009; Reyes et al., 2010; Verreault et al., 2011). It has been reported that phages are major players in horizontal gene transfer and influence the ecology of microbial populations (Suttle, 2007; Ogunseitan, 2008). Phages and phage-derived enzymes have multiple applications including phage therapy, molecular biology, and food safety (Silander and Saarela, 2008; Wright et al., 2009; Kuchment, 2011). They are also major facilitators of dairy fermentation failures, contributing to large economic losses in this industry (Moineau et al., 2002).

With the introduction of the high-throughput DNA sequencing (HTS), phage research has entered the new age (Mann, 2005). There is a significant increase of complete phage genome in sequence databases, e.g. GenBank. For many years, sequencing of a phage derived from a culturable phage–host system was a laborious process. In order to obtain sufficient amounts of DNA that is required

for preparation of the sequencing library, multi-step protocols were needed: (1) production of a high phage-titer lysate, (2) removal of bacterial cells and debris, (3) precipitation of phage particles with polyethylene glycol (PEG), (4) cesium chloride purification of the phage solution, and (5) phenol–chloroform DNA isolation that could be used in a standard high-throughput sequencing workflow (Sambrook and Russell, 2001).

Recently, DePew et al. (2013) described a novel method for sequencing viral genomes derived from single isolated plaques using the Sequence Independent Single Primer Amplification (SISPA). This method significantly reduced time compared with the traditional method of Sambrook and Russell (2001), by omitting the time-consuming lysate production and CsCl centrifugation steps. However, with the new transposon-based sequencing library preparation kits such as the Nextera XT (Illumina, CA, USA), which is designed to produce Illumina compatible sequencing libraries from as little as ≤ 1 ng of DNA, making it now possible to significantly speed up the whole sequencing process. In the transposon-based method used by the Nextera XT kit, the DNA is fragmented and tagged in a single reaction, thereby decreasing the amount of required input DNA.

More rapid and inexpensive sequencing of culturable phages derived from a single plaque would advance phage genomics and shed a new light on the evolutionary mechanisms of phages in various host–phage systems. Rapid phage sequencing can also be useful to monitor phages in the industrial fermentation environment.

* Corresponding author at: Department of Biology, Faculty of Science, University of Copenhagen, Universitetsparken 15, DK-2100 København Ø, Denmark.
Tel.: +45 51 82 70 30.

E-mail address: nws772@ku.dk (W. Kot).

In this study, direct plaque sequencing (DPS) is described as a rapid method for single-plaque sequencing of DNA-containing phages. DPS provides nearly a tenfold reduction of the sample preparation time compared with the method described by DePew et al. (2013). DPS with plaques was examined from three phages belonging to the *Podoviridae*, *Siphoviridae* and *Myoviridae* families, respectively. Phages belonging to these three families constitute more than 96% of all known phages (Ackermann, 2011).

2. Materials and methods

2.1. Phages, bacterial strains and media

Phages and bacterial strains used in this study are listed in Table 1. Phages T4 and ϕ 29 and their hosts were obtained from DSMZ (Braunschweig, Germany). Phage P113g was obtained from the phage collection at the Max Rubner-Institut (Kiel, Germany). *Lactococcus lactis* IL1403 was obtained from INRA (Jouy en Josas, France). *Escherichia coli* and *Bacillus subtilis* were grown in tryptic soya broth (TSB) or on tryptic soya agar (TSA) (Oxoid, Basingstoke, UK) and incubated overnight at 37 °C. *L. lactis* was grown in M17 medium (Oxoid, Basingstoke, UK) supplemented with 0.5% (wt/vol) of glucose (GM17) and incubated overnight at 28 °C. All media was supplemented with 10 mM CaCl₂ for phage propagation, and plaques were produced using the double agar method (Kropinski et al., 2009) with 0.8% agarose in the top layer.

2.2. Plaque harvesting and processing

Well-separated single plaques were cut out from the agar plate with a trimmed 1000 μ l-pipette tip (diameter ~1–2 mm), and placed in a 100 μ l of 1 \times DNase I buffer (Thermo Scientific, Waltham, USA). The mixture with the agarose plug was briefly vortexed and incubated for 30 min at 40 °C. The solution was then filtered using an ultrafiltration spin column with 0.45 μ m cutoff (Millipore, Billerica, USA). Subsequently 1 U of DNase I (Thermo Scientific, Waltham, USA) was added followed by 30 min incubation at 37 °C. DNase I was inactivated by addition of 10 μ l of 50 mM EDTA (Thermo Scientific, Waltham, USA), leaving a dilute but host DNA reduced phage solution.

2.3. Phage DNA extraction

The phage solution was treated with 5 μ l (approx. 3 U) of PCR-grade Proteinase K, (Thermo Scientific, Waltham, USA). After which, 10% SDS solution was added to a final concentration of 1%, and the solution was incubated for 30 min at 55 °C. Proteinase K was inactivated by 10 min incubation at 70 °C. The DNA was purified and concentrated using a DNA Clean & Concentrator™-5 kit (Zymo Research, Irvine, USA) according to the manufacturer's protocol and eluted with 6 μ l of elution buffer. The schematic workflow for the entire phage DNA extraction is shown in Fig. 1.

2.4. DNA sequencing

DNA sequencing libraries were prepared using the Nextera® XT DNA kit (Illumina, San Diego, USA) according to the manufacturer's protocol with the following modifications: input DNA was not adjusted to 0.2 ng/ μ l, instead 5 μ l of purified DNA was used directly regardless of concentration. To compensate for the <1 ng amount of input DNA, two extra cycles during the PCR amplification step were added. Finally, the library normalization and pooling process was based solely on the concentration of the final libraries measured using Qubit fluorometer (Life Technologies, Carlsbad, USA). Individually tagged libraries were sequenced as a part of a flowcell (0.8%, 0.1% and 0.7% of the total sequencing run yield, for

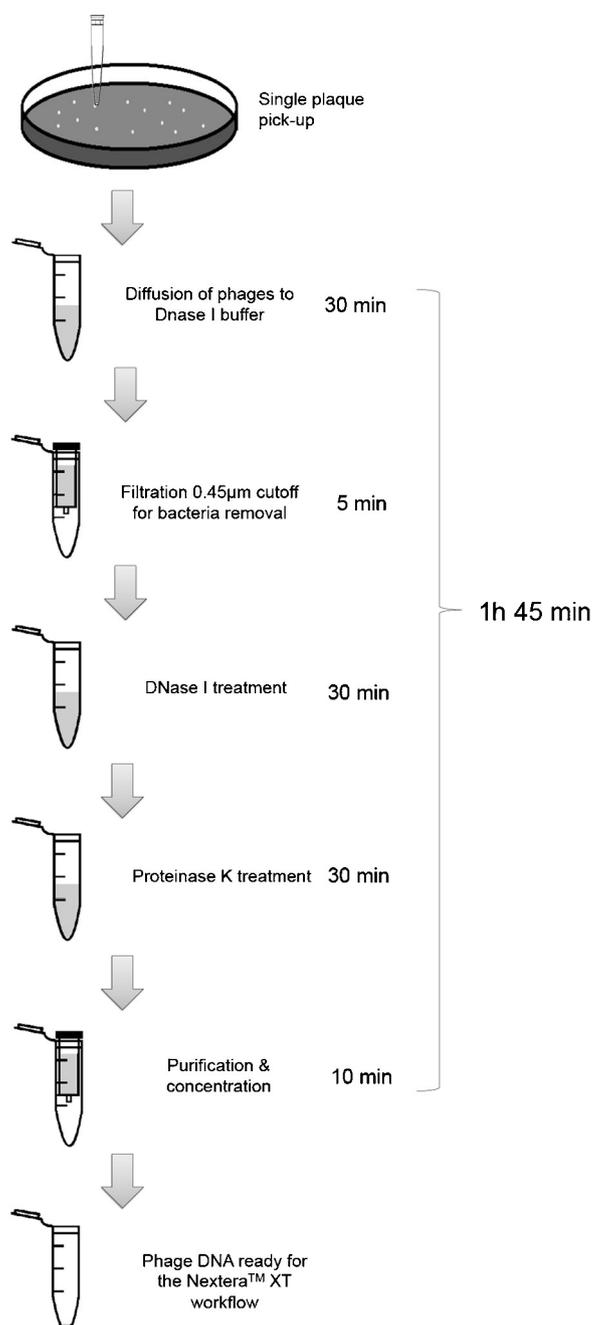


Fig. 1. The schematic workflow of DNA isolation used in this study.

phages ϕ 29, P113g and T4, respectively) as 2 \times 250 base paired-end reads using the Illumina MiSeq platform (Illumina, San Diego, USA).

2.5. Sequencing analysis and assembly

Reads were trimmed, analyzed and assembled using CLC Genomic Workbench 6.0.4 (CLC bio, Aarhus, Denmark). Trimming was performed using the “trim sequences” tool. Trimming settings were set to low quality limit of 0.03, with no ambiguous nucleotides allowed, and removal of 20 nucleotides from the 3' end, which usually have a lower quality. Additionally, Nextera® adapter sequences were removed using the adapter trimming option and the Nextera® transposase sequences (mismatch cost = 2, gap cost = 3 minimum score of 10, minimum score at end = 4). After trimming, reads were assembled de novo using “De Novo Assembly” tool with the default

Table 1
Bacterial and phage strains used in this study.

Phage	Host strain	Family	Genome size (kb)	Reference genome accession nr	Reference
φ29 (DSM 5546)	<i>Bacillus subtilis</i> 110NA (DSM 5547)	Podoviridae	19	EU771092	Anderson and Mosharrafa (1968)
P113g	<i>Lactococcus lactis</i> IL1403	Siphoviridae	30	KC182548	Mahony et al. (2013)
T4 (DSM 4505)	<i>Escherichia coli</i> B (DSM 613)	Myoviridae	168	HM137666	Miller et al. (2003)

settings and mapping, automatic bubble size and 64-word size. The scaffolding was performed with auto-detect paired distances.

Furthermore, trimmed reads were mapped to the corresponding reference genome (accession numbers: EU771092, KC182548, or HM137666) using “Map Reads to Reference” tool with the default settings. Obtained read mapping was analyzed using the “Probabilistic Variant Detection” tool.

3. Results

3.1. Sequencing result

The sequencing run yielded a total of 421,338 paired-ended reads for the three different phage libraries. After trimming and adaptor removal, the number of total reads dropped to 404,576 with the average read length of 220 bp. Table 2 contains the specific statistics for each sample.

3.2. Mapping to reference genome

Reads were mapped to the respective reference phage genome as indicated in Table 1. In case of phage P113g, 96.95% of the trimmed reads were mapped to the reference genome. For phages φ29 and T4, this number was smaller, 82.79% and 16.63%, respectively. The average coverage can be found in Table 2. All three genomes were sequenced with a minimum depth of coverage of 10% of the average coverage, except the proximate regions of 5' and 3' ends in phages φ29 and P113g (Fig. 2A and B, respectively). Two regions with no coverage were identified. The first region was 8 bp at the start of the φ29 phage genome, and the second region was an 853 bp fragment of the T4 phage within the *rII* genes region (Fig. 2C). However, detailed analysis of the reads spanning this region revealed that the *rIIA* and *rIIB* genes were fused as is frequently seen in the laboratory strains of T4 phage (Silhavy, 2000). P113g was fully covered, although the depth of coverage in the *cos*-site region was very low (only six reads).

3.3. De novo assembly

In order to evaluate DPS for sequencing of novel phages, de novo assembly using trimmed reads was carried out using CLC Genomic Workbench 6.0.4. The longest contig in each of the assemblies represented nearly the full genome of the corresponding phages, again except of the proximate terminal region of φ29 phage where 79 bp and 62 bp were absent, respectively. The entire genomes of phages P113g and T4 were assembled in to one contig, respectively. The minimum fraction of the reads to obtain a near full genome by de novo assemble was 1% for φ29 phage (1967 reads), 18% for phage P113g (4517 reads) and 27% for phage T4 (48,660 reads), respectively. Differences between the reference genomes and sequenced genomes were identified. For phage φ29, there was one insertion of a triplet CGA resulting in the insertion of an additional arginine residue, and one single nucleotide variation (SNV) resulting in an amino acid change. In P113g, there was an 18 nt deletion in one of the CDSs and one SNV resulting in amino acid change. In phage T4, there were 12 SNVs with eight of them resulting in amino acid changes as well as 1 nt deletion in a non-coding region.

4. Discussion

An alternative, pre-amplification-free and more rapid method is described in comparison with the recently published method by DePew et al. (2013). It was important to develop this simple, rapid and robust method that can be used for genome sequencing of diverse type of phages, despite the use of the Nextera XT system, which limits DPS to dsDNA containing phages. DPS was tested with phages from three different taxonomic groups (*Podoviridae*, *Siphoviridae* or *Myoviridae* family) and infecting different bacterial hosts.

One of the challenges in sequencing of phages genomes from a single plaque is a low amount of starting DNA. To overcome the small abundance of DNA in the sample, several approaches were proposed including whole genome amplification using φ29 polymerase, linker adapter shotgun library (LASL) or SISPA (Allen et al., 2011; Marine et al., 2011; DePew et al., 2013). These methods are less cost- and time-efficient since they require separate pre-amplification, DNA-fragmentation, and library construction steps, whereas DPS does this all in one step without any pre-amplification. These methods can also introduce amplification biases, however that seems to be of a bigger concern in case of metagenomic samples rather than single-strain samples. With the Nextera technology, biases have also been reported, specifically in connection to GC% content of regions (Marine et al., 2011). Furthermore, the system currently supports a maximum of 96-sample multiplexing, which is lower than e.g. SISPA, which can support up to 1544 bar-coded primer sequences (Hamady et al., 2008).

To increase the ratio of phage derived DNA to the host DNA, we decided to use the filtration process, similarly to DePew et al. This approach proved to be very efficient, especially in phages φ29 and P113g, where more than 80% of the total reads could be mapped into their respective reference genomes. In the case of phage T4, only 16.63% of the reads had a T4 genome origin. That might be an effect of ineffective separation of host cells during the filtration step or inhibition of DNase I by e.g. salt carry over from the media. It should be noted that host DNA contamination did not fully block the assembly process, however more reads had to be used to obtain the full genome sequence of the T4 phage. The coverage of the T4 genome had approx. tenfold higher coverage than contigs that showed similarity to the host genomic DNA (data not showed). Overall, the phage-DNA enrichment step allowed us to omit time and cost inefficient steps like precipitation of phages, ultracentrifugation or whole genome amplification (WGA).

The workflow used by DPS is very quick and can be carried out in most microbiology labs. The time from plaque harvesting to a ready-to-use DNA eluent is less than two hours. The whole process is performed in a convenient, single tube setup and it includes only column-based filtration and purification steps (Fig. 1).

The removal of contaminating host DNA permits a more resourceful use of the capacity of the sequencing platform. The double indexing system offered by the Nextera XT workflow makes it possible to sequence up to 96 samples in one MiSeq run. This significantly decreases the sequencing cost. The number of samples could be increased by using custom-made index primers, thus using more efficiently the MiSeq platform to sequence hundreds of single-plaque phages in one run.

Table 2
Summary of the sequencing results for each individual sample.

Phage	Total reads	Reads after trimming	Average read length (after trimming)	Average coverage	Reads aligned to reference genome (%)
φ29	209,532	198,353	224	1912.4	82.79
P113g	26,680	25,289	227	184.4	96.95
T4	185,126	180,934	216	40.6	16.63

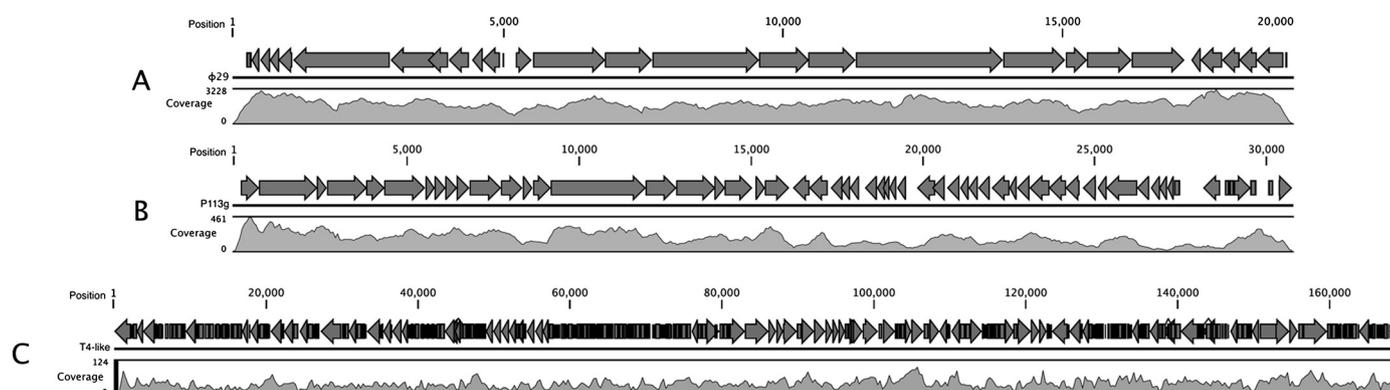


Fig. 2. Depth of coverage plots for the mapping assembly of phages φ29 (A), P113g (B), and T4 (C), where x-axis represents the reference genome and y-axis represents the coverage. Each sample has a different scale for depth of coverage indicated on the left side of the plot. Above the reference genomes, nucleotide position is indicated. Predicted ORF are indicated with arrows. The region of fused *rII* genes in the T4 genome is indicated with a black bar on the coverage plot.

The region proximal to the genomic termini of the φ29 and P113g phages had much lower coverage than the rest of the genome. This is caused by several factors including the Nextera XT transposome technology, which has an expected drop off in sequencing coverage of about 50 bp from each distal end or single-stranded *cos*-site overhangs as seen in phage P113g. In case of phage φ29, absence of the sequence of the termini is likely due to ineffective removal of the terminal protein by the proteinase K treatment. Generally, determination of the sequence of the genomic termini should be performed as an additional step using the remaining 1 μl from the total yield of 6 μl of purified DNA. For example, Sanger sequencing of a PCR product that spans the ligated *cos*-site region could be used. This approach is limited to already known phages where PCR amplification of the region of interest can be performed prior to Sanger sequencing. For novel phages, it might be necessary to isolate larger amount of DNA or use the WGA technique. Determination of the genomic termini appeared not to be necessary in phages that employ headful-packaging system, e.g. T4-like phages where depth of coverage remains on a similar level throughout the whole genome.

In conclusion, DPS was tested and optimized for diverse phage–host system, however it can be further optimized for a specific phage–host pair by changing the cutoff value of the filter and/or adding additional enzymatic treatments. In addition, the application of this method enables sequencing of low-frequency phage mutants directly from a plate or sequencing of phages in which it is particularly difficult to obtain a high-titer lysates. DPS has potential to influence whole genome sequencing of phages by significant reducing the turnaround time and cost.

Acknowledgments

This work was supported by the MetaPhageLAB project financed by The Danish Research Council for Technology and Production (Project: 11-106991, Doc#: 2105568).

We would like to acknowledge The Danish National High-Throughput DNA-Sequencing Centre for sequencing the samples.

We would like to acknowledge Arshnee Moodley for her invaluable help with proof reading this manuscript.

References

- Ackermann, H.W., 2011. Bacteriophage taxonomy. *Microbiol. Aust.* 32, 90–94.
- Allen, L.Z., Ishoey, T., Novotny, M.A., McLean, J.S., Lasken, R.S., Williamson, S.J., 2011. Single virus genomics: a new tool for virus discovery. *PLoS ONE* 6, e17722.
- Anderson, D.L., Mosharraf, E.T., 1968. Physical and biological properties of phage Phi29 deoxyribonucleic acid. *J. Virol.* 2, 1185–1190.
- Bergh, O., Børshheim, K.Y., Bratbak, G., Heldal, M., 1989. High abundance of viruses found in aquatic environments. *Nature* 340, 467–468.
- Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L., Brüßow, H., 2004. Phage–host interaction: an ecological perspective. *J. Bacteriol.* 186, 3677–3686.
- DePew, J., Zhou, B., McCarrison, J.M., Wentworth, D.E., Purushe, J., Koroleva, G., Fouts, D.E., 2013. Sequencing viral genomes from a single isolated plaque. *Virol. J.* 10, 181.
- Hamady, M.M., Walker, J.J.J., Harris, J.K.J., Gold, N.J.N., Knight, R.R., 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 5, 235–237.
- Kropinski, A.M., Mazzocco, A., Waddell, T., Lingohr, E., Johnson, R., 2009. Bacteriophages. Humana Press, Totowa, NJ.
- Kuchent, A., 2011. *The Forgotten Cure: The Past and the Future of Phage Therapy*. The Forgotten Cure. Springer New York, New York, NY.
- Mahony, J., Kot, W., Murphy, J., Ainsworth, S., Neve, H., Hansen, L.H., Heller, K.J., Sorensen, S.J., Hammer, K., Cambillau, C., Vogensen, F.K., van Sinderen, D., 2013. Investigation of the relationship between Lactococcal host cell wall polysaccharide genotype and 936 phage receptor binding protein phylogeny. *Appl. Environ. Microbiol.* 79, 4385–4392.
- Mann, N.H., 2005. The third age of phage. *PLoS Biol.* 3, 753–755.
- Marine, R., Polson, S.W., Ravel, J., Hatfull, G., Russell, D., Sullivan, M., Syed, F., Dumas, M., Wommack, K.E., 2011. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl. Environ. Microbiol.* 77, 8071–8079.
- Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 67, 86–156.
- Moineau, S., Tremblay, D., Labrie, S., 2002. Phages of lactic acid bacteria: from genomics to industrial applications. *ASM News* 68, 388–393.
- Ogunseitan, O.A., 2008. Genetic transduction in freshwater ecosystems. *Freshwater Biol.* 53, 1228–1239.
- Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., Gordon, J.I., 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338.
- Sambrook, J., Russell, D.W., 2001. *Molecular Cloning: A Laboratory Manual*. 3rd edition. Cold Spring Harbor Press, ISBN: 0879695773.
- Silander, K., Saarela, J., 2008. Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. In: *Genomics Protocols, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 1–18.
- Silhavy, T.J., 2000. Gene fusions – guest commentary. *J. Bacteriol.* 182, 5935–5938.

- Suttle, C.A., 2007. Marine viruses – major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812.
- Verreault, D., Gendron, L., Rousseau, G.M., Veillette, M., Masse, D., Lindsley, W.G., Moineau, S., Duchaine, C., 2011. Detection of airborne Lactococcal bacteriophages in cheese manufacturing plants. *Appl. Environ. Microbiol.* 77, 491–497.
- Vos, M., Birkett, P.J., Birch, E., Griffiths, R.I., Buckling, A., 2009. Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* 325, 833.
- Wright, A., Hawkins, C.H., Ånggård, E.E., Harper, D.R., 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clin. Otolaryngol.* 34, 349–357.