Bacteriophage genotyping using BOXA repetitive-PCR

Dragica Damnjanovic, Xabier Vázquez-Campos, Daniel L. Winter, Melissa Harvey and Wallace J. Bridge *

Abstract

Background: Repetitive-PCR (rep-PCR) using BOXA1R and BOXA2R as single primers was investigated for its potential to genotype bacteriophage. Previously, this technique has been primarily used for the discrimination of bacterial strains. Reproducible DNA fingerprint patterns for various phage types were generated using either of the two primers.

Results: The similarity index of replicates ranged from 89.4–100% for BOXA2R-PCR, and from 90 to 100% for BOXA1R-PCR. The method of DNA isolation (p = 0.08) and the phage propagation conditions at two different temperatures (p = 0.527) had no significant influence on generated patterns. Rep-PCR amplification products were generated from different templates including purified phage DNA, phage lysates and phage plaques. The use of this method enabled comparisons of phage genetic profiles to establish their similarity to related or unrelated phages and their bacterial hosts.

Conclusion: The findings suggest that repetitive-PCR could be used as a rapid and inexpensive method to preliminary screen phage isolates prior to their selection for more comprehensive studies. The adoption of this rapid, simple and reproducible technique could facilitate preliminary characterisation of a large number of phage isolates and the investigation of genetic relationship between phage genotypes.

Keywords: Bacteriophage, Phage genotyping, Repetitive-PCR

Background

Repetitive DNA sequences constitute a substantial component of both eukaryotic and prokaryotic genomes. In some higher plant species, they can account for up to 90% of the genomic DNA [1], while in humans DNA repeats comprise nearly half of the genome [2]. The presence or absence of certain types of repeats, diversity in their nucleotide sequences, their size, location and copy number per genome characterize various bacterial species, even those with the smallest genomes [3]. Interspersed repeats play a significant role in genomic rearrangements, such as inversions, deletions, duplications and translocations [3]. The proposed functional roles of repetitive sequences involve the regulation of coding sequence expression and the formatting necessary for genome packaging; DNA repair and restructuring; genome replication and transmission to progeny cells; formation of nucleoprotein complexes; and formation of a characteristic genome system organization that allows for evolutionary significant changes without altering coding sequences [4].

Specific families of interspersed DNA elements have been observed in many bacterial and archaeal genomes [3, 5], while bacteriophages are considered to carry few repetitive elements [5]. The BOX family of repetitive DNA elements, consisting of different combinations of three sequence sub-motifs, boxA, boxB, and boxC, was originally identified in Gram-positive Streptococcus pneumoniae [6]. Hybridization studies have shown that only boxA sequences are highly evolutionary conserved. Specific families of interspersed DNA elements have been observed in many bacterial and archaeal genomes [3, 5], while bacteriophages are considered to carry few repetitive elements [5]. The BOX family of repetitive DNA elements, consisting of different combinations of three sequence sub-motifs, boxA, boxB, and boxC, was originally identified in Gram-positive Streptococcus pneumoniae [6]. Hybridization studies have shown that only boxA sequences are highly evolutionary conserved. The outwardly facing repetitive primers BOXA1,
BOXA1R and BOXA2R that are complementary to the consensus sequences of boxA, when used as single primers in the repetitive-PCR, generated complex fingerprint patterns in various Gram-positive and Gram-negative bacterial species [7]. BOXA-based primers have since been used for genotyping diverse bacterial species in various ecological [8], epidemiological [9] and industrial application studies [10].

The optimal performance of dairy starter cultures is challenged by the risk of lytic bacteriophage (phage) infection [11, 12] and indeed, phage may pose a problem to any industry based on bacterial fermentation [13]. In recent years, there has been a renewed interest in phage from the perspective of their beneficial applications, such as phage therapy for treating pathogens [14, 15]. Of great utility for the dairy industry are multiplex PCR systems that can detect and classify the three main lactococcal phage species, 936, P335 and c2, as well as Streptococcus thermophilus and Lactobacillus delbrueckii phages [16, 17]. These rapid and sensitive tests are based on the generation of specific PCR amplification products for each phage species; however, they do not enable the identification of individual phage strains. Methods that can be applied for genetic characterization of phages involve restriction digestion of genomic DNA [18]; multi-locus sequence typing (MLST) [19]; restriction fragment length polymorphism (RFLP) or a denaturing gradient gel electrophoresis (DGGE) of a particular gene [20]; random amplification of polymorphic DNA (RAPD)-PCR [18, 21, 22] and genomic sequencing [23]. However, these methods are not necessarily the most suitable for routine use due to time or cost-associated constraints. MLST and DGGE require a priori genetic information [21]. Restriction digestion of genomic DNA and DNA/DNA hybridizations are considered time-consuming and often require large quantities (μg) of pure DNA [21]. Additionally, the genomes of some phages can be resistant to restriction enzymes, which may be due to a scarcity of cleavage sites [24]; a base modification within the recognition sequence, genome methylation or other anti-restriction mechanisms [18]. This imposes the need to use several restriction enzymes to ensure digestion [18]. While MLST can distinguish phages with the same RFLP pattern [19], it may not be universally applicable for fingerprinting all phage types. For example, it has proved suitable for phylogenetic analysis of the 936-like phages of Lactococcus lactis, but could not be used to generate amplification products from the c2 and P335 phage groups [19]. Whole-genome phage sequencing is associated with substantial cost and technical difficulties [22], and may be impractical for routine use.

RAPD-PCR and rep-PCR are in the category of fast and inexpensive genotyping methods, but RAPD-PCR has been reported to have poor reproducibility. This is due to the use of short and arbitrary primers which target randomly distributed sequences [25] and thus requires substantial optimization of conditions [18].

Our preliminary work on genetic characterization of Lactococcus lactis phages showed that restriction digestion was not successful for all of them rendering comprehensive phage comparison unachievable (unpublished data). This prompted our search for an alternative method for phage differentiation. To this end, repetitive-PCR using BOXA-based primers, which we have previously successfully applied to discriminate bacterial and yeast isolates [26] was explored. The aim of this study was to investigate whether repetitive-PCR typing is applicable to bacteriophage, by determining whether amplified viral DNA could produce specific and stable fingerprints that could be used for phage genotyping.

Results

BOXA-PCR for phage fingerprinting

The potential of repetitive-PCR as a fingerprinting method was initially tested with the BOXA2R primer on a set of Lc. lactis phages; one Pseudomonas phage and one Str. thermophilus phage (see Fig. 1). Lactococcal phages were selected to cover isolates from different sources and comprised representatives of the three main species (936, P335 and c2), while the phages that infect two other bacterial species were included to explore whether the method would be applicable to other phage types. The analysis showed that lactococcal phages of the same species clustered together. For example, phage 38 (host Lc. lactis ssp. diacetylactis FD11) and phage BU (host Lc. lactis ssp. cremoris HP) that both belong to the 936-species formed one group together with the reference 936-type phage sk1. Also, the P335-type phages, 63 and CW that infect Lc. lactis ssp. lactis ML8, and 301 that infects Lc. lactis ssp. lactis Ni301, clustered together. Within the c2-like and 936-like clusters, phages that infected the same host, such as 38 and 54 or 15 and R48 (host Lc. lactis ssp. cremoris Mo9) were more related, while within a P335 cluster, phages to the same host were not the most closely related. Out of the three lactococcal phage groups, Str. thermophilus phage BSN3 appeared most related to P335 phages, while Pseudomonas phage Pf4G was an outlier. Similar phage groupings were also generated using the BOXA1R primer, except that phage BU (HP) consistently generated a single intense fragment of ~ 300 bp, which affected the relationship analysis (results not shown).

Comparison of the BOXA-PCR results with phage sequencing data

To determine whether phage relatedness inferred from BOXA-PCR fingerprinting could be supported by another classification method, the genomes of eight
lactococcal (38(FD11), BU (HP), 301(Ni301), 63(ML8), CW (ML8), 15(Mo9), 923 and R48(Mo9)) and two 
Pseudomonas (Pf4G and Ps15) phages were sequenced (manuscript in preparation), and the sequences compared. The genome of R48 was not included into the final analysis due to difficulties with its assembly. The hierarchical clustering was performed using the Prokaryotic Virus Orthologous Groups (pVOGs, formerly termed Phage Orthologous Groups, POGs) resource [27]. The annotation of pVOGs was made using the predicted proteins and the hierarchical clustering was performed on these annotations. It resulted in clustering of lactococcal phages comparable to that seen with BOXA-PCR, where 936-, P335-, and c2-type phages formed well defined separate clusters. The 
Pseudomonas phages were placed outside these groups and were unrelated (see Fig. 2).

An alternative approach based on the correlation of tetranucleotide signature frequencies, that showed high discriminatory power comparing genome fragments even when alignments are not likely [28], produced the same correlations between the tested phages (see Fig. 1, Additional file 1).

**BOXA-PCR method validation**

The method was investigated and validated on a group of well characterised *Lc. lactis* phages using both BOXA1R and BOXA2R primers (see Table 1, Additional File 2). The reproducibility of the method was explored using phage propagated at two different temperatures (30 °C and 37 °C) with the DNA being isolated by two protocols. The details of the statistical analysis descriptive performances performed using the non-parametric Mann-Whitney test are presented in the Supplementary material (see Additional file 3).

Repetitive-PCR with both BOXA1R and BOXA2R primers resulted in reproducible DNA fingerprint profiles of the tested phages, which also included lambda,
phiX174 and T4 phages. The similarity index of the BOXA1R-PCR replicates ranged from 90.2–100% and of BOXA2R-PCR replicates from 89.4–100%. The difference in performance between these two primers was not statistically significant ($p = 0.279$) (see Table 1, Additional File 3).

Influence of the DNA isolation method on the phage profile
Separate PCR amplifications from the same DNA source generated reproducible fingerprints with each BOXA2R primer (see Fig. 3) and BOXA1R (see Fig. 4). The replicates showed very similar (see Fig. 3, Øsk1 and Fig. 4, ØT4) or indistinguishable (see Fig. 4, ØC6A and Fig. 5, ØLambda) profiles. No statistically significant difference ($p = 0.08$) was observed in the fingerprint profiles of the same phage using the two DNA preparation methods, phenol-chloroform and Qiagen commercial kit (see Table 3, Additional File 3).

Rep-PCR produced stable profiles when the same phage DNA was amplified at different times on separate gels, although band migration on gels may fluctuate. An example, ØX174 amplified with BOXA1R is presented in Figs. 4 and 5. The two gels appeared different due to different concentrations of DNA used in the PCR reactions, 100 ng (Fig. 4) vs 25 ng (Fig. 5). When the migrating bands are compared against the marker it can be seen that the banding patterns are consistent between the two gels.
Influence of the propagation temperature on the phage profile

Minor variations in the fingerprint patterns of the same phage cultivated and amplified at two temperatures, 30 °C and 37 °C, were observed, however the differences were not significant (p = 0.527) (see Table 2, Additional File 3). The variability in band intensity, which was occasionally seen, was not associated with the incubation temperature as there was no clear separation between the 30 °C or 37 °C clusters (see Fig. A in Additional File 4).

Influence of the propagating host on the phage profile

The host on which a phage was amplified slightly altered the phage profile, such as seen in the BOXA2R-PCR profiles of the Øsk1 amplified on two bacterial hosts: Lc. lactis ssp. cremoris LMO230 and MG1363 (see Fig. 6) or the ØP087 amplified on Lc. lactis ssp. lactis strains ML8 and C10 (see Fig. B in Additional File 4). Slightly differing bands were seen in the ~2 Kbp region, where Øsk1(MG1363) displayed a single, pronounced band whereas Øsk1(LMO230) displayed two smaller bands.
instead (96% similarity). The differences in the ØP087 profiles were detected in the ~800–1000 bp range, which resulted in the 90% similarity between amplifications when propagated on the two hosts.

Similarly, one band difference between the profiles of the same phage propagated on two different hosts was also observed with BOXA1R. For example, Øc6A amplified on the Lc. lactis ssp. lactis C6 and C10 differed by the presence of a ~580 bp band, which was visible only for the C6 host (results not shown).

**Application of BOXA-PCR genotyping for screening phage isolates**

Repetitive-PCR using either BOXA1R or BOXA2R repetitive primer was further tested on a number of ecological isolates with distinctive fingerprint profiles being generated for all tested phage (see Figs. 7–8).

The amplification of *Staphylococcus aureus* phages with BOXA1R primer produced the lowest number (2–5) of PCR fragments (see Fig. 7). All four isolates appeared highly related, with there being no difference in profiles for Sa1 and Sa7, and for Sa11 and Sa12, whereas the similarity between these two pairs was 90%. Contrary to the *Staph. aureus* phages, the four *Pseudomonas* phages (Ps6, Ps15, Ps19 and Ps21) produced four unique fingerprint profiles. The Ps6, Ps15, Ps19 phages all exhibited one band of ~1.5 kb in common, whereas all other bands were exclusive to each phage isolate.

The PCR profiles of DNA extracted from the purified particles of the lytic ØPf4G obtained using the two boxA primers were compared to that of the wild-type *Ps. aeruginosa* PAO1, and the mutant of PAO1 created by the removal of the complete filamentous Pf4 prophage genome (PAO1-)[31]. While the band patterns of both bacterial strains were similar with either boxA primer, the phage Pf4G displayed a very different profile to these bacteria and yielded DNA fragments of substantially different size and length (see Figs. 7 and 8). This indicated that the superinfective ØPf4G may have undergone significant DNA rearrangement compared to its prophage/temperate form.

Successful amplification of the purified DNA of an archeal phage, DL4HV, which infects *Halorubrum* sp. and some other halobacteriaceae, was also achieved using the BOXA-PCR method (see Figs. 7 and 8).

**Sequencing of the BOXA-PCR fragments**

Eight PCR products randomly selected from fingerprints of different phage isolates were sequenced with either BOXA1R or BOXA2R primer (see Table 1, Additional File 5) to determine the sequence of the genes contained in the rep-PCR products. This experiment resulted in the identification of encoded proteins that are clearly linked to the phages they originated from, although with varying levels of homology (37–89%). The lower levels of homology may have resulted from technical issues or from alternative binding sites of the BOXA primers to the rep-PCR products, thereby producing overlapping sequences that are difficult to deconvolute.
To further validate the presence of the identified genes in the phage genomes, a series of primers were designed to target the exact coding sequences obtained by the BLAST analyses (see Table 2, Additional File 5). The second round of amplifications of four fragments with gene-specific primers generated PCR products that matched the size of the targeted genes (see Fig. 1, Additional File 5). Sequencing of these products confirmed the genes identified in the BLAST results (see Figs. 1–4, Additional File 6).

The comparisons of results obtained by sequencing selected fragments using both the boxA repetitive primers and primers specifically designed for particular annotated genes are presented in Table 1.

The discriminatory capacity of the method was corroborated against the whole genome sequencing data of selected Lc. lactis and Pseudomonas phages. Phage relationships deduced from each method compared well and resulted in clustering of lactococcal phages according to their species while Pseudomonas phages were clearly identified as a separate group.

Reproducibility of the BOXA-PCR was demonstrated using both BOXA1R and BOXA2R primers on a set of selected known phages amplified on different hosts and/

**Table 1** BOXA-PCR fragment sequencing and confirmatory results

| FRAGMENT NAME (Primer) | Rep-PCR fragment size-gel assessment (bp) | UniProt KB annotation of rep-PCR fragment (Top BLAST result) | Rep-PCR BLAST results | Confirmatory BLAST results using primers designed to amplify the putative CDS* |
|------------------------|-----------------------------------------|-------------------------------------------------------------|-----------------------|--------------------------------------------------------------------------------|
| ØX174-1 (BOXA1R)       | 650                                     | P03631 Replication-associated protein A (Enterobacteria phage phiX174)* | E-value: 1.9e-91       | E-value: 0.0 | Score: 745                                | Score: 1819 |
| ØT4-2 (BOXA2R)         | 700                                     | A0A376YLU8 Putative baseplate structural protein (Escherichia coli) | E-value: 1.7e-13       | Score: 1004 | Score: 98.50%                           | Ident: 98.50% |
| ØLambda-1 (BOXA2R)     | 450                                     | P03689 Replication protein P (Escherichia phage lambda)*       | E-value: 1e-15         | Score: 1133 | E-value: 3.6e-156                      | Ident: 100.00% |
| Øc2-1 (BOXA1R)         | 1200                                    | Q38305 Probable tape measure protein (Lactococcus phage c2)    | E-value: 5.8e-17       | Score: 1529 | E-value: 0.0                           | Ident: 94.70% |

*refer to Tables 1 and 2 (Additional file 5) for details

**Discussion**

This work has demonstrated that the BOXA-PCR method is applicable for phage fingerprinting and is discriminatory and reproducible under various conditions. The discriminatory capacity of the method was corroborated against the whole genome sequencing data of selected Lc. lactis and Pseudomonas phages. Phage relationships deduced from each method compared well and resulted in clustering of lactococcal phages according to their species while Pseudomonas phages were clearly identified as a separate group.

Reproducibility of the BOXA-PCR was demonstrated using both BOXA1R and BOXA2R primers on a set of selected known phages amplified on different hosts and/
or amplified in the same host after cultivation at different temperatures (30°C and 37°C). Consistent fingerprint profiles were obtained from phage DNA purified by two protocols \((p = 0.08)\). The amplification of a phage on different hosts resulted in some small, but obvious differences. These variations likely reflect the small genetic differences between the host strains, such as in the case of Lc. lactis ssp. cremoris MG1363 and LMO230, the plasmid-free and prophage-cured derivatives of their closely related parent strains NCDO712 and C2 [32]. This aspect may represent a limitation of the method and requires further work, including investigating fingerprints of phages amenable to propagation on both genetically similar and dissimilar strains.

BOXA-PCR was tested on a range of phages that infect a variety of bacteria, for example E. coli, Pseudomonas sp., Staphylococcus sp., Lc. lactis; and also the haloarchaea algae Haloarcula. Both the BOXA1R and BOXA2R primers used in the PCR amplification of phage DNA produced distinctive individual fingerprint profiles for all tested isolates. This broad application may be due to the lower GC content of these primers and the associated optimal annealing temperature of 40°C [7]. This decreased stringency may have enabled the amplification of various phage types. It is suggested that potentially other repetitive primers, for example ERIC primers for enterococcal phages, could be more suitable for the genotyping of other bacteriophage types.

The method also appears useful for application to phages that are difficult to characterize by other genetic methods. For example, the DNA of the CR1 phage has previously been shown to be refractory to both restriction endonuclease digestion and to genome sequencing, which has prevented its genomic characterization [33]. Fingerprinting analysis with BOXA1R showed that it was unrelated to the E. coli ØT4 phage (results not shown). Lc. lactis ssp. lactis Ø15(Mo9) and some other lactococcal phages investigated in this study were also resistant to cutting by several restriction enzymes hence the direct determination of genetic relationship between various lactococcal phages by restriction enzyme digestion method was not possible.

Phage template DNA obtained from different sources (such as purified phage DNA, phage lysates or a single picked plaque from an agar plate) generated comparable BOXA-PCR products. These findings suggest the method would have utility and convenience for routine phage typing. However, when unpurified phage preparations are used as templates, such as lysate that has not been pre-treated with DNase or a plaque directly picked from the agar plate, there is a possibility that traces of host DNA may be co-amplified and contaminate phage fingerprint profile. The incorporation into the method of appropriate controls to account for any host DNA contamination could provide a solution to this limitation.

The sequencing of PCR fragments produced with the boxA primers verified the phage origin of the amplified DNA. The majority of the genes identified during the BLAST analyses against the UniProt KB database [34] encoded phage-associated proteins involved in different viral functions, such as the Replication protein A in ØX174 and the Replication protein P in ØLambdA, respectively, or baseplate proteins (ØT4 and Ø301). Phage-originating genes identified from the nucleotide sequences might facilitate future work to identify and locate the repeats and elucidate the potential repetitive primer binding sites on the target genomes.

In terms of speed and simplicity the repetitive-PCR described in this work is comparable to RAPD-PCR. Both methods allow for a whole genome comparison. Rep-PCR is commonly applied to bacteria but has been reported as a complement to RAPD for phage characterization [35]. Although RAPD-PCR method has been used to examine genetic diversity of various bacterial and algal phages [18, 21, 22, 35, 36] it has been associated with poor reproducibility [37] and hence requires extensive optimization [21, 37]. This involves selection of a suitable decamer primer [22, 35] followed by optimization of PCR conditions, such as primer and MgCl₂ concentration, annealing temperature and standardization of template concentrations [22, 38]. Furthermore, to increase the sensitivity of the RAPD assay, the pooling of patterns from at least two amplifications with different random primers has often been applied [35, 38]. RAPD cannot discriminate closely related phages and the amplification of the phage and host DNA has not been possible under the same RAPD-PCR assay conditions [22]. Our findings demonstrated that the use of BOXA-PCR under defined conditions could achieve better discrimination, higher reproducibility with fewer optimisation steps and would be more broadly applicable than RAPD-PCR for use in phage genotyping.

Conclusion

The findings of this study suggest that repetitive-PCR could be a useful high resolution method for bacteriophage genotyping. Repetitive-PCR, preferably using specific primers to BOX sequences: BOXA1R and BOXA2R, is applicable to the genetic identification of an isolated phage and for the investigation of genetic relationships between phage genotypes. The technique could well be used in a phage management program in addition to multiplex PCR for phage species determination where the knowledge of genetic profiles of circulating and emerging phages could be advantageous for the selection of the most appropriate strategies to address
contamination problems. It could also be used to complement other phage genotyping methods, such as RFLP, and would be particularly useful for analysing phage which appear refractory to digestion with restriction endonucleases. Given the abundance of phage in the environment [20], having a rapid, inexpensive method for genetic characterization of both phage and its bacterial hosts could be useful especially when preliminary genotyping large numbers of phage isolates is required for screening purposes.

**Methods**

**Bacteria**

*Lc. lactis* ssp. *lactis* Ni301 and *Lc. lactis* ssp. *cremoris* Mo9 were isolated from dairy products. *Lc. lactis* ssp. *lactis* biovar diacetylactis FD11 was isolated from the mesophilic starter type culture Flora Danica (FD). *Lc. lactis* ssp. *cremoris* HP and UK712, and *Lc. lactis* ssp. *lactis* biovar diacetylactis ML8 were retrieved from our internal culture collection. *Lc. lactis* ssp. *lactis* 112, C6, C10 and IL1407 were kindly provided by Dr. Jasna Rakonjac, Massey University, NZ. *Pseudomonas aeruginosa* PAO1 (wild type) and PAO1- (free of endogenous prophage Pf4) were kindly provided by Dr. Vanessa Huron, The Centre for Marine Bio-Innovation and the School of Biotechnology and Biomolecular Sciences (BABS).

The strains were grown anaerobically for 24–48 h at 30 °C on M17 agar (Oxoid) or M17 broth supplemented with 0.5% lactose (LM17) [39]. Stock cultures were prepared in 9.5% (w/v) autoclaved (121 °C for 15 min) reconstituted skim milk and kept at −80 °C.

**Bacteriophages**

The *Lc. lactis* phages used in the study were isolated from dairy whey samples (see Table 2) with other phage types provided by external sources (see Table 3). Phage preparations provided by external sources were analysed by PCR in the form they were supplied without any further modification. The phage coding system for the phages isolated in this study was based on designating the phage name followed by the name of the bacterial strain on which that phage was propagated (in brackets).

**Phage propagation, amplification, and titering**

The lysates of the isolated phages were prepared from single plaques picked from sensitive host lawns of double layer LM17 agar plates containing 10 mM CaCl₂.

### Table 2 Bacteriophage samples isolated from whey samples

| Phage code | Host Species | Titer (pfu/ml) | Year of isolation | Phage species | Phage plaque description | Source |
|------------|--------------|---------------|-------------------|---------------|-------------------------|--------|
| 15(Mo9)    | *Lactococcus lactis* ssp. *cremoris* Mo9 | 2.00E+08       | 2016              | c2+P335       | 2 mm clear + 1 mm zone  | isolated in this study |
| R48(Mo9)   | *Lc. lactis* ssp *cremoris* Mo9 | 1.90E+08       | 2016              | c2           | 2 mm and 4 mm clear     | isolated in this study |
| 301(Ni301) | *Lc. lactis* ssp *lactis* Ni301 | 2.20E+10       | 2016              | P335         | 1 mm clear              | isolated in this study |
| 54(FD11)   | *Lc. lactis* ssp *lactis* var diacetylactis FD11 | 2.80E+09       | 2016              | 936          | 3 mm clear + 2 mm margin | isolated in this study |
| 38(FD11)   | *Lc. lactis* ssp *lactis* var diacetylactis FD11 | 6.00E+09       | 2016              | 936          | 2 mm clear + 2 mm margin | isolated in this study |
| BU(HP)     | *Lc. lactis* ssp *cremoris* HP | 7.20E+09       | 2016              | 936          | 6 mm clear              | isolated in this study |
| 63(ML8)    | *Lc. lactis* ssp *lactis* var diacetylactis ML8 | 3.70E+09       | 1981              | P335         | 2 mm clear, sharp edge  | BABS, UNSW culture collection |
| CW(ML8)    | *Lc. lactis* ssp *lactis* var diacetylactis ML8 | 1.50E+08       | 1981              | P335         | 1 mm clear              | BABS, UNSW culture collection |
| BSN3 (France) | *Streptococcus thermophilus* BSN3 | 5.30E+08       | 1988              | -            | -                       | BABS, UNSW culture collection |

*15(Mo9) has been identified as a *Lactococcus* virus unclassified c2-like (unclassified Ceduovirus) bacteriophage by DNA sequencing (manuscript in preparation) - Not determined

### Table 3 Bacteriophage samples provided by external sources

| Phage code | Host Species | Preparation | Provider |
|------------|--------------|-------------|----------|
| Pf4        | *Pseudomonas aeruginosa* | purified DNA | A/Prof. Scott Rice, UNSW |
| DL4HV      | *Halorubrum* sp. | purified DNA | Dr Susanne Erdmanne, Max Planck Institute for Marine Microbiology, Germany |
| T4         | *Escherichia coli* | phage lysate | Dr Nicola Petty, The iThree institute, UTS |
| CR1        | *Citrobacter rodentium* | phage lysate | Dr Nicola Petty, The iThree institute, UTS |
| Ps6, Ps15, Ps19, Ps21 | *Pseudomonas* sp. | phage lysate | Dr Lisa Elliott, AusPhage |
| Sa1, Sa7, Sa11, Sa12 | *Staphylococcus aureus* | phage lysate | Dr Lisa Elliott, AusPhage |
| 923, 943   | *Lactococcus lactis* | phage lysate | A/Prof. Jasna Rakonjac, Massey Univ., NZ |
| PRD1       | *E. coli* | plate | Culture collection, BABS, UNSW |
and were purified twice. Following the amplification and filtration with a 0.22 μm sterile Millipore filter, broth lysates with titers equal or higher than 10^8 pfu/ml were stored in 1.5 ml cryovials at −20 °C.

**Phage purification and DNA isolation**
Crude phage lysate (1.5 ml) was treated with 10 μg/ml RNase A and 1 μg/ml DNase I (Sigma) final concentration for 30 min at 37 °C to remove bacterial nucleic acids. Following centrifugation at 22000 X G for 10 min, the phage particles in the supernatant were PEG-salt precipitated and further purified as previously described [40].

**Multiplex PCR for phage typing**
The determination of lactococcal phage type was performed by multiplex PCR as described by Labrie and Moineau (2000) [16] using isolated phage DNA as template.

**Repetitive polymerase chain reaction (rep-PCR)**
Phage genotyping was performed using puReTaq Ready-To-Go Polymerase Chain Reaction (PCR) Beads (GE Healthcare) with either BOXA1R (5′-CTACGGCAAG GCCACGCTGACG-3′) or BOXA2R (5′-ACGTGG TTTGAAGAGATTTTCG-3′) [7] as a single primer.

For each reaction, 50 pmol of the single primer BOXA1R or BOXA2R, 50–100 ng Lc. lactis phage DNA in DNase/RNase - free water at a final volume of 25 μl was added to a tube containing a single PCR bead. The rep-PCR method was also tested using phage lysates (10^5 pfu/ml) and plaques as DNA sources. The negative control reaction contained primer only, with water substituted for the template DNA. PCR amplifications were performed in an automated thermal cycler (Eppendorf) with an optimal cycling program set for each primer (see Table 4). The amplification products (5 μl) were electrophoresed on 15 x 20 cm 1.2% (w/v) molecular grade agarose (Bio-Rad) gels in 1 x TAE (Tris-acetate, EDTA, pH 8.1) at a constant 100 V for 3 h. The amplicons were assessed against the molecular size marker HyperLadder I (M I) or HyperLadder I (M II) (Bioline). After staining with GelRed 3 x staining solution in water (Biotium) for 15 min, the gels were photographed with a camera attached to a tripod mounted on a UV-transilluminator (Ultra Lum).

**Phage genome sequencing and bioinformatic analysis**
The sequencing of phage DNA was performed in the Ramaciotti Center for Genomics, UNSW, Sydney on a MiSeq system (Illumina) (1x MiSeq reagent kit v2, 2 x 150 bp Nano Sequencing Run). The sequencing library was prepared with 10x Nextera XT DNA Library preparation kit (Illumina). Phage genomes were assembled with SPAdes v3.13.1 [42]. Gene-calling was performed with multiPhATE v1.0 [43]. Predicted proteins were annotated against Prokaryotic Virus Orthologous Groups (pVOGs, downloaded Nov. 2019) [27] with hmmscan (HMMER v3.2.1) [44].

Cluster analysis on the presence-absence of pVOGs was performed in R v3.6.1 [45] with hclust on a distance matrix.
matrix calculated with the “binary” method in the stats package.

Correlation based on tetranucleotide composition was calculated with pyani v0.2.9 [46].

**Sequencing of the PCR bands**

Randomly selected individual DNA amplicons (see Table 1, Additional File 5) generated with either primer were separated from a mixture of other PCR products by applying the band stab method [47]. The subsequent re-amplification with the same primer in a new PCR reaction was run on a gel to verify that a single band of the same size was obtained. The size of the amplified fragment was measured against the molecular size marker, and the concentration and purity were determined by Nanodrop. After purification with GenElute PCR Clean-Up kit (Sigma), Sanger sequencing of the selected amplicons was performed in the Ramaciotti Centre for Genomics using either BOXA1R or BOXA2R as a sequencing primer. BLAST runs of the obtained nucleotide sequences against the UniProt KB database were used to deduce the protein sequences and their functional information [34].

**Validation of the phage origin of rep-PCR products**

Primers targeting specific phage genes were designed based on the results from BLAST analyses (see Table 2, Additional File 5). PCR was performed using the puReTaq Ready-To-Go Polymerase Chain Reaction (PCR) Beads (GE Healthcare) and the appropriate purified phage DNA as template. The sizes of PCR products were evaluated by electrophoresis in 1% (w/v) agarose gels against the molecular size marker. Selected PCR products were sequenced in the Ramaciotti Centre for Genomics for final validation using the corresponding forward primer as the sequencing primer. Sequencing results were aligned against the coding sequence of genes identified during BLAST analyses using the MAFFT algorithm in the Benchling platform [48].

**Supplementary information**

Supplementary information accompanies this paper at https://doi.org/10.1186/s12866-020-01770-2.

**Additional file 1** Analysis of the sequenced phage genomes based on tetranucleotide frequencies. This file provides the figure that shows a Pearson correlation of tetranucleotide frequencies including the clustering trees and heatmaps with Pearson correlation values.

**Additional file 2** List of phages used in the reproducibility study. This data outlines the phages used in the study, which strain they were propagated on, and the culture conditions.

**Additional file 3** Statistical analysis of the BOXA-PCR for phage fingerprinting. This file provides the details of the statistical analysis of the reproducibility testing.

**Additional file 4** Influence of the propagating temperature and propagating host on the phage profile. This file provides the.

**Additional file 5** Sequenced phage fragments. This file provides the list of sequenced phage fragments and details relating to their analysis, including the primers used for the amplification of selected phage genes; results from the second round of PCR amplifications and BLAST analyses that used gene-specific primers; and a figure showing PCR amplifications of the selected phage DNA fragments using the forward and reverse primer pair corresponding to the annotated genes in each phage.

**Additional file 6** Sanger sequencing of the BOXA-PCR fragments. This file provides the alignment of the sequencing of the PCR products to the sequences of the corresponding proteins.

**Abbreviations**

Rep-PCR: Repetitive-PCR; MLST: Multilocus sequence typing; RFLP: Restriction fragment length polymorphism; DGGE: Denaturing gradient gel electrophoresis; RAPD-PCR: Random amplification polymorphic DNA; ERIC: Enterobacterial repetitive intergenic consensus; UPGMA: Unweighted pair group method with arithmetic mean; pvOGs: Prokaryotic Virus Orthologous Groups.

**Acknowledgements**

The authors would like to express their gratitude to Prof. Peter White and A/Prof. Scott Rice, School of Biotechnology and Biomolecular Sciences, UNSW, Sydney; Dr. Nicola Petty, The TThree Institute, University of Technology, Sydney; Dr. Lisa Elliott, AusPhage Pty Ltd, Queensland, A/Prof. Jasna Rakonjac, Massey University, NZ and Dr. Susanne Erdmanne, Max Planck Institute for Marine Microbiology, Germany, for their provision of phage samples used in this study. The assistance of Dr. Veljko Prodanovic, UNSW, with statistical analysis is greatly acknowledged.

**Authors’ contributions**

DO planned and designed the experiments, performed the experiments, analysed and interpreted the results, wrote the manuscript; XVC performed the bioinformatic analysis of the sequenced phages; DLW designed the primers listed in Table 3. Additional file 5, and analysed the sequencing data showing in Figs. 1–4 Additional file 6; MH participated in data analysis and reviewed drafts; WJB contributed discussions, reviewed drafts and edited the final manuscript. All authors have read and approved the final manuscript.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Availability of data and materials**

All data generated or analysed during this study are included in the Results section of this published article and its supplementary information files (Additional files 1–6).

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Received** 21 August 2019 **Accepted** 29 March 2020 **Published online** 11 June 2020

**References**

1. Melkonian S, Goyal Y. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genom Proteom Bioinf 2014;12(4):164–171. https://doi.org/10.1016/j.gpb.2014.07.003.
2. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2012;13:36–46. https://doi.org/10.1038/nrg3117.
3. Smirnov GB. Repeats in Bacterial Genome: Evolutionary Considerations. Mol Gen Genet. 2010;283(2):56–65.
4. Shapiro JA, von Stemberg R. Why repetitive DNA is essential to genome function. Biol Rev. 2005;80(03):227–250. https://doi.org/10.1017/S1464793104006657.
5. Treangen TJ, Abraham AL, Tousson M, Rocha EPC. Genesis, effects and fates of repeats in prokaryotic genomes. FEMS Microbiol Rev. 2009;33(5):559–71.
6. Martin B, Humbert O, Camara M, Guerzu E, Walker J, Mitchell T, et al. A highly conserved repeated DNA element located in the chromosome of Streptococcus pneumoniae. Nucleic Acid Res. 1992;20(13):3479–83. https://doi.org/10.1093/nar/20.13.3479.
7. Koeuth T, Versalovic J, Lupski JR. Differential subsequence conservation of unusually large genomes. Appl Environ Microbiol. 1990;56(7):2180–85. https://doi.org/10.1128/AEM.56.7.2180-2185.1990.
8. https://doi.org/10.1128/AEM.00931-12.
9. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 1990;18(24):7213–8.
10. Damjanovic D, Harvey M, Bridge WI. Application of colony BOX-PCR for the differentiation and identification of lactic acid coconi. Food Microbiol. 2019;82:277–86. https://doi.org/10.1016/j.fm.2019.02.011.
11. Marcó MB, Moineau S, Quiberoni A. Bacteriophages and dairy fermentations. BMC Microbiol. 2011;11:188. doi.org/10.1186/1471-2180-11-188.
12. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 1990;18(24):7213–8.
13. Damjanovic D, Harvey M, Bridge WI. Application of colony BOX-PCR for the differentiation and identification of lactic acid coconi. Food Microbiol. 2019;82:277–86. https://doi.org/10.1016/j.fm.2019.02.011.
14. Koc M, Cokmus C, Cihan AC. The genotypic diversity and lipase production of some thermophilic bacilli from different genera. Braz J Microbiol. 2015;46(4):734–41. https://doi.org/10.1590/1517-8382-2015-04942.
15. Marcó MB, Moineau S, Quiberoni A. Bacteriophages and dairy fermentations. Bacteriophage. 2012;2(3):95–8. https://doi.org/10.4161/bact.21868.
16. Mahony J, Moscarelli A, Kelleher P, Lugli GA, Ventura M, Settanni L, van Sinderen D. Phage biodiversity in artisanal cheese wheys reflects the complexity of the fermentation process. Viruses. 2017;9(3):45. https://doi.org/10.3390/v9030045.
17. Vunsche L. Importance of bacteriophages in fermentation processes. Acta Biotechnol 1989;9(5):349–41. https://doi.org/10.1080/01475430.1989.11054351.
18. Jeon J, Park JH, Yong D. Efficacy of bacteriophage treatment against carbapenem-resistant Acinetobacter baumannii in Galleria mellonella larvae and a mouse model of acute pneumonia. BMC Microbiol 2019;19(1):70. https://doi.org/10.1186/s12866-019-1443-5.
19. Yang H, Liang L, Lin S, Jia S. Isolation and characterization of a virulent bacteriophage AB1 of Acinetobacter baumannii. BMC Microbiol 2010;10:131. https://doi.org/10.1186/1471-2180-10-131.
20. Labrie S, Moineau S. Multiplex PCR for detection and identification of lactococcal bacteriophages. Appl Environ Microbiol. 2000;66(3):987–94. https://doi.org/10.1128/AEM.66.3.987-994.2000.
21. de la Rúa B, Biñetí A, Martin MC, Fernández M, Magadan AH, Alvarez MA. Multiplex PCR for the detection and identification of dairy bacteriophages in milk Food Technol. 2007;24:75–81. https://doi.org/10.1016/j.lwt.2006.03.001.
22. Barrangou R, Yoon SS, Breidt F Jr, Fleming HP, Kleinhamper TR. Characterization of six Leucostoc faeticus bacteriophages isolated from an industrial sauerkraut fermentation. Appl Environ Microbiol. 2002;68(11):5452–8. https://doi.org/10.1128/AEM.68.11.5452-5458.2002.
23. Moisan M, Moineau S. Multilocus sequence typing scheme for the characterization of 936-like phages infecting Lactococcus lactis. Appl Environ Microbiol. 2012;78(13):4646–53. https://doi.org/10.1128/AEM.00931-12.
24. Clokie MR, Millard AD, Letarov AV, Heaphy AV, Heaphy S. Phages in nature. Bacteriophage. 2011;1(1):31–45. https://doi.org/10.4161/bact.1.1.14942.
25. Conesa AM, Short S, Suttle CA. The use of degenerate-primed random amplification of polymorphic DNA (DP-RAPD) for strain-typing and inferring the genetic similarity among closely related viruses. J Virol Methods. 2004;112(2):95–102. https://doi.org/10.1016/j.jvmt.2004.01.020.
26. Gutierrez D, Martin-Platero A, Rodriguez A, Martinez-Bueno M, Garcia P, Martinez B. Typing of bacteriophages by randomly amplified polymorphic DNA (RAPD)-PCR to assess genetic diversity. FEMS Microbiol Letters. 2011;322:90–7. https://doi.org/10.1111/j.1574-6968.2011.02342.x.
27. Klumpp J, Fouet DE, Sothamanthan S, Never SN. Bacteriophage identification: new generation sequencing technologies and the changing landscape of phage genomes. Bacteriophage. 2012;2(3):190–9. https://doi.org/10.4161/bact.21211.
28. Prevots F, Mata M, Ritzenthaler P. Taxonomic differentiation of 101 lactococcal bacteriophages and characterization of bacteriophages with unusually large genomes. Appl Environ Microbiol. 1990;56(7):1280–5.
47. Bjouzon AJ, Cooper JE. Band-stab PCR: a simple technique for the purification of individual PCR products. Nucleic Acids Res. 1992;20(17):4675.

48. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT; a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66. https://doi.org/10.1093/nar/gkf436.

**Publisher’s Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.