

Article

Simple method for assessing diversity and dynamics of microbial community: comparison of dairy phages from industrial and spontaneous fermentation

Agnieszka Olejnik-Schmidt ¹, Bernadeta Pietrzak ¹, Iwona Kawacka ¹, Klaudia Malak ¹, Weronika Wawrzyniak ¹ and Marcin Schmidt ^{1,*}

¹ Department of Food Biotechnology and Microbiology, Poznan University of Life Sciences, Wojska Polskiego 48, 60-627 Poznań, Poland; agnieszka.olejnik-schmidt@up.poznan.pl (A.O.-S.), bernadeta.pietrzak@up.poznan.pl (B.P.), iwona.kawacka@up.poznan.pl (I.K.), 49702@uppoz.onmicrosoft.com (K.M.), weronika.wawrzyniak@up.poznan.pl (W.W.), marcin.schmidt@up.poznan.pl (M.S.)

* Correspondence: marcin.schmidt@up.poznan.pl; Tel.: +48 61 8466024

Featured Application: Assessing diversity and dynamics of low-complexity microbial communities e.g., in fermentation processes.

Abstract: Background: The dairy industry heavily relies on fermentation processes driven in high proportion by *Lactococcus lactis*. The fermentation process can be perturbed or even stopped by bacteriophage activity leading to complete loss of fermentation batch or decreased quality product. Monitoring of the phage diversity and dynamics in the process allows to implement protective measures (e.g. starter rotation) in order to maintain unperturbed production.; Methods: Universal primers were used to amplify sequences of the 936, c2, and P335 *Lactococcus* phage types. The amplicons were sequenced with Sanger method and obtained degenerate sequences were analyzed using simple bioinformatic pipeline in R environment.; Results: The most prevalent phage type is 936, followed by P335, whereas c2 type is less frequent.; Conclusions: Curd cheeses prepared on non-pasteurized milk based on native milk microbiota had higher diversity of phages distinct of those found in dairy plants. Sanger sequencing of heterogeneous amplicons generated on metagenome DNA can be used to assess low-complexity microbiota diversity.

Keywords: phage; dairy; *Lactococcus*; diversity; community

1. Introduction

The dairy industry heavily relies on fermentation processes driven in high proportion by *Lactococcus lactis*. The fermentation process can be perturbed or even stopped by bacteriophage activity leading to huge economic losses [1–3]. Among bacteriophages infecting *L. lactis* [4], those belonging to species c2, 936, or P335 are most commonly encountered in dairy plants [5,6]. Analysis of diversity and dynamics of *lactococcal* bacteriophages is crucial for preventing the spread of infections in dairy environments.

Analysis of microbial community composition and dynamics especially in case of unculturable bacteria or viruses, where lack of knowledge on growth requirements or undetermined permissive host is a limiting factor, always posed a challenge. Today in the era of Next Generation Sequencing (NGS) the problem seems to be solved [7]. However, the NGS is still an expensive method. Its ability to produce a massive output of data in a relative short time can sometimes be a disadvantage. The relatively low price per sample in case of the NGS can be achieved only by filling the whole throughput (lane, chamber or chip of the sequencer) with samples. In case of low complexity samples or targeted approaches to use the whole throughput the extensive barcoding is necessary. It is often necessary to collect a large number of samples to cover the whole throughput of a chip to run the sequencing. Here, we compared diversity of *Lactococcus* phages from dairy

industrial and spontaneous fermentation processes using simple and easy method allowing to study diversity and dynamics of microbial community with well-established and affordable techniques and simple bioinformatics pipeline.

2. Materials and Methods

Cheese samples and DNA isolation. A series of curd cheese samples derived from Polish cheese production facilities (n = 25) made from cow's milk and spontaneous non-pasteurized milk fermentations from in-house cheese production (n = 14) made from cow's, goat's or sheep's milk were collected in 2017-2018. Metagenomic DNA was isolated from the samples of cheeses using Genomic Mini AX Food Kit (A&A Biotechnology, Poland) according to the kit procedure, and used as template for PCR reaction.

PCR amplification and sequencing. A previously described multiplex polymerase chain reaction (PCR) system [8] with latter modification [5] was employed to screen for the phages present in the cheeses. Where multiple of the three major groups of *Lactococcus* phages, i.e., the 936, c2 or P335 groups, were identified in the same sample the separate PCR reactions were performed with single type-specific primers pair for the phage group. The amplicons were purified with Clean-Up Kit or Gel-Out Kit (A&A Biotechnology, Poland) in case of multiple amplicons obtained and submitted to Genomed S.A. (Poland) for Sanger sequencing.

Sequence analysis. The sequencing trace files (.ab1) were inspected for sequencing quality in Sequence Scanner Software 2 (Applied Biosystems). The sequences were submitted to BLAST service analysis [9] to recover the best matched sequences set from the NCBI database. The sequences set obtained from BLAST results were used to build a local database and reference matrix composed of sequence distance measures. The measures were calculated in local alignment of the ambiguous sequence separately to all sequences from the BLAST result sequences set. A principal component analysis (PCA) plot further simplified the interpretation of the community diversity. The bioinformatics analyses were performed in R [10] using DECIPHER [11] and PCATools [12] libraries.

3. Results

3.1. Phage survey

The curd cheese samples derived from Polish dairy facilities (n = 25) and spontaneous milk fermentations from in-house cheese production (n = 14) from different locations were tested for the presence of *Lactococcus* phages with multiplex PCR [8]. All of the dairy plant derived cheeses were tested positive for both P335 and 936 phages, and only 16% of them contained additional c2 phages (Fig. 1, lanes 1-12). All cheeses derived from spontaneous non-pasteurized milk fermentations contained the 936 phage type, 93% contained the P335 phage type, and 14% contained the c2 phage type (Fig. 1, lanes 13-26). However, the multiplex PCR reactions made with metagenomic DNA from cheeses prepared with spontaneous fermentations resulted in generation of multiple non-specific amplicons. Sequence analysis of selected non-specific amplicons (Supplementary Materials: File_S1.fas) indicated their significant homology to *Lactococcus* phages. Non-specific size amplicons generated with primer pair specific to the c2 phage type purified and sequenced (n = 8) had sequence of highest homology to: *Lactococcus* phage CHPC967 (*Siphoviridae*; 88.01% sequence identity), *Lactococcus* phage CHPC959 (*Siphoviridae*; 98.80% sequence identity), *Lactococcus* phage vB_LacS_15 (*Siphoviridae*; 98.28% and 97.62% sequence identity), *Siphoviridae* sp. ctGas1 (*Siphoviridae*; 90.99%, 90.75%, and two of 91.13% sequence identity); with primer pair specific to the P335 phage type (n = 2): *Lactococcus* phage Tuc2009 (*Siphoviridae*; 84.53% sequence identity), *Lactococcus* phage bIL286 (*Siphoviridae*; 82.71% sequence identity); nucleotide BLAST search performed on August 11th, 2021. This indicates possible existence of non-dairy plant environment-adopted phages.

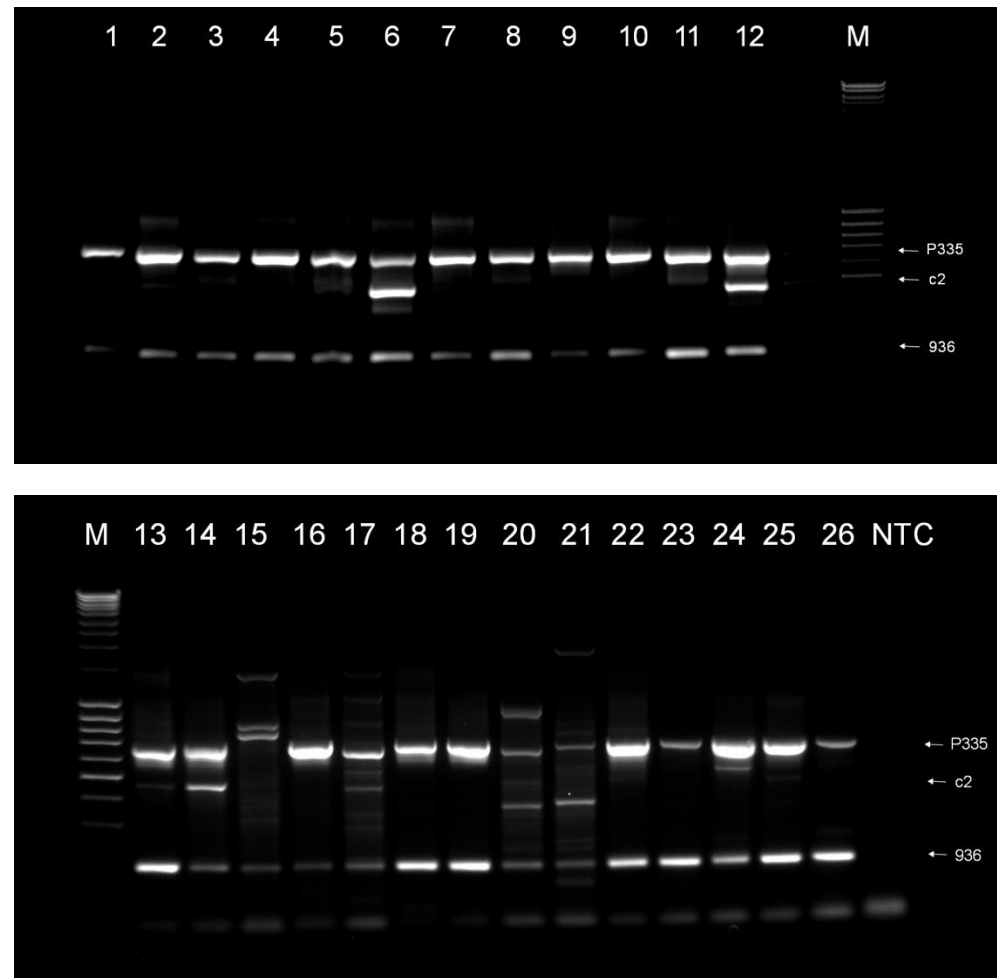


Figure 1. Exemplary multiplex PCR results for detection of *lactococcal* phages in curd cheeses from dairy plant mass production (lanes 1-12) and spontaneous milk fermentation from in-house chees production (lanes 13-26). Arrows indicate positions of specific amplicons for P335, c2, and 936 *Lactococcus* phage types (M – molecular weight marker, NTC – non-template control).

3.2. Phage diversity analysis between samples

To compare a diversity of *Lactococcus* phages between samples we performed PCR amplifications in separate PCR reactions with single type-specific primers pair for each of the phage type. In case of several samples of metagenomic DNA obtained from spontaneous fermentations multiple amplicons were generated with use of single pair of type-specific primers. Such amplicons were purified from gel slabs cut-out from electrophoretic separation gel prior to submission to sequencing. Inspection of sequencing trace files indicated ambiguous nucleotide positions (data not shown) that resulted from sequencing of heterogenic amplicons obtained on metagenomic DNA. We performed multiple alignment of the obtained sequences with DECIPHER library [11] with substitution matrix accounting for ambiguous nucleotides for all IUPAC nucleic acid codes based upon match and mismatch parameters (match = 5, mismatch = -4) and calculated a distance matrix for each of sequences set with Jukes-Cantor substitution model used for distance correction [13]. Each element of the distance matrix corresponds to the dissimilarity between two sequences in the sequence set.

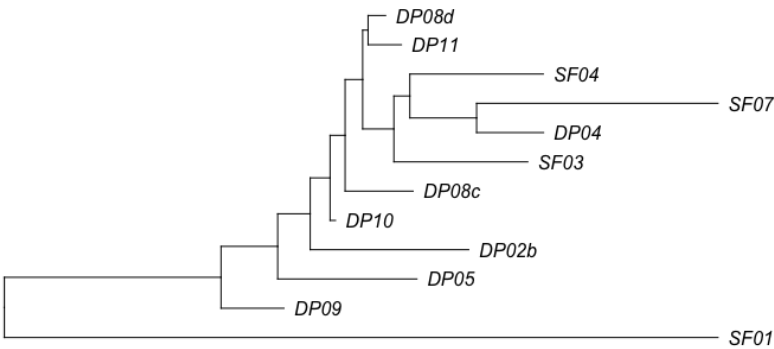


Figure 2. Phylogenetic tree obtained for sequences of amplicons generated with *Lactococcus* 936 phage type primers with metagenomic DNA from curd cheeses from dairy plants (DP--) and spontaneous in-house fermentations (SF--).

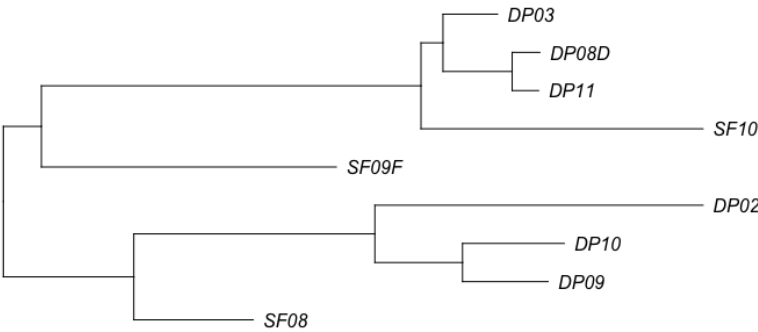


Figure 3. Phylogenetic tree obtained for sequences of amplicons generated with *Lactococcus* c2 phage type primers with metagenomic DNA from curd cheeses from dairy plants (DP--) and spontaneous in-house fermentations (SF--).

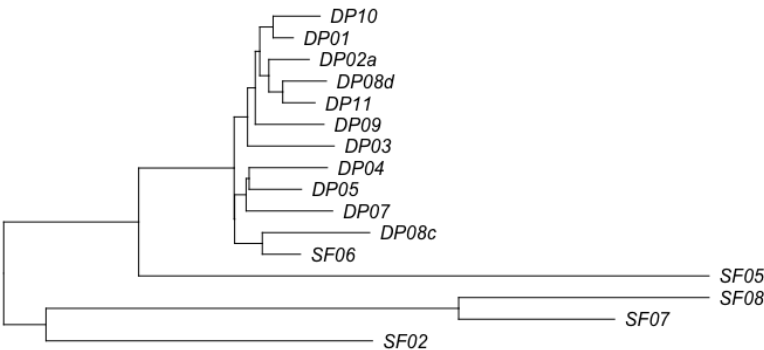


Figure 4. Phylogenetic tree obtained for sequences of amplicons generated with *Lactococcus* P335 phage type primers with metagenomic DNA from curd cheeses from dairy plants (DP--) and spontaneous in-house fermentations (SF--).

Comparison of the phylogenetic trees generated for each of the *Lactococcus* phage types (Figs. 2-4) indicated that the P335 type phages from dairy plant cheeses cluster together whereas the phages from spontaneous fermentations are more divergent.

3.3. Global phage diversity analysis

The BLAST search of the obtained phage sequences resulted with hit list of *Lactococcus* phage sequences of high sequence identity. However, the degenerate nucleotide codes are treated by BLAST algorithm as mismatches in nucleotide alignment [9]. The hit lists for all of the sequence searches were combined and duplicate records removed to build a local database. The local database was used to calculate reference matrix of sequence distance measures with DECIPHER library [11] with substitution matrix accounting for ambiguous nucleotides for all IUPAC nucleic acid codes as described previously. This way we created a sample-composition independent measures for global phage diversity analysis. The global distance measures were used for principal component analysis (PCA) to visualize differences and similarities in phage type composition in analyzed samples. Sequences derived from BLAST search hits used for distance measures cluster tightly together (Fig. 5), metagenome derived sequences form distant, more dispersed clusters separate for dairy plant and spontaneous fermentation cheeses with single outliers DP02(c2), SF01(936), and SF06(P335).

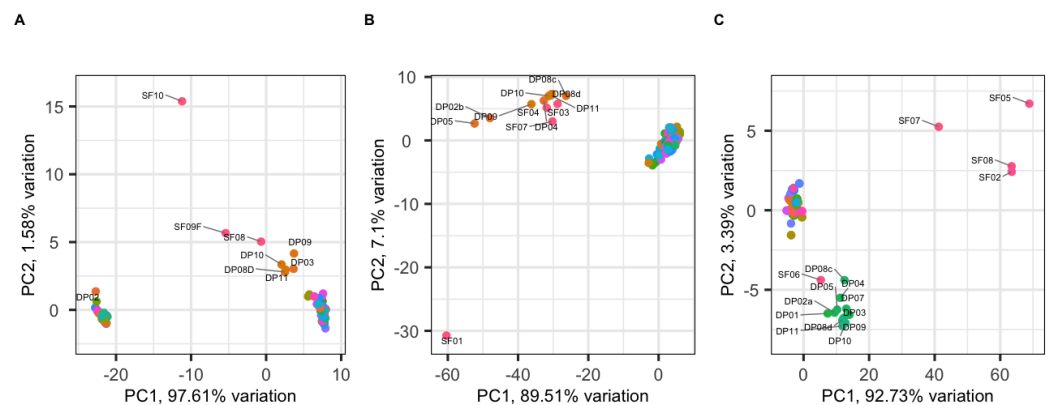


Figure 5. Principal component analysis biplots for sample composition-independent diversity comparison based on sequence distance measures as obtained for *Lactococcus* phage: c2 (A), 936 (B), and P335 (C) types. Labeled points represent phage sequences amplified on metagenomic DNA from dairy plants (DP--) or spontaneous fermentation (SF--) produced curd cheeses; unlabeled points represent sequences derived from BLAST search hits.

4. Discussion

Our study indicated higher prevalence of the *Lactococcus* phage types detected in curd cheeses from Polish dairy plants than previous reports [6,14] showing that the problem of phage persistence in dairy plant environments persists. We also surveyed in-house prepared cheeses from non-pasteurized milk based on spontaneous fermentation proceeded by native milk microbiota. Based on our findings we can assume that the native milk microbiota contains much higher diversity of phages and *L. lactis* strains (as sources of prophages). Especially in case of c2 and P335 phage type we obtained non-specific size amplicons that were of phage origin. The P335 nonspecific amplicons might be derived from prophages of P335-type or phage DNA remnants in chromosomes of *L. lactis*. The higher diversity of spontaneous fermentation cheeses can be explained with sensitivity of some phages to thermal treatments [15].

Here, we described also a simple and easy way to study diversity and dynamics of low-complexity microbial community with well-established and affordable techniques and simple bioinformatics pipeline. Previously reported methods employed laborious and susceptible strains requiring methods involving phage genomic DNA isolation and sequencing [15], low resolution sequence analysis with RFLP [6,16] or MLST [17]. Our method involves PCR amplification of metagenomic DNA using universal primers hybridizing to conserved regions spanning variable sequences (e.g. in 16S rDNA, ITS, etc.). Obtained pool of amplicons is sequenced with Sanger method resulting in read-out

sequence with ambiguous bases at several position. Care must be taken to submit for Sanger sequencing amplicons of the same length to prevent shifts of homologous sequence regions or take into account only selected parts of the homologous regions proximal to sequencing primer. In case of large differences in amplicon length electrophoretic separation and from-gel purification can solve the problem. This sequence corresponds to consensus sequence as it would be generated by bioinformatic sequence alignment of several homologous sequences. The sequences from different samples can be aligned and distance between the calculated with DECIPHER library [11] with substitution matrix accounting for ambiguous nucleotides for all IUPAC nucleic acid codes. Such analysis allows to compare samples composition or dynamics (e.g., time series) in form of phylogenetic tree or PCA plots. To generate the PCA plots in a global scale (sample composition-independent context) a local database of homologous sequences is needed. The local database can be easily prepared by submitting a sequence to BLAST search and downloading the BLAST hits. The BLAST algorithm accepts the IUPAC ambiguity codes however treats them as mismatches in the alignment. Therefore, a lot of information is lost. The sequences set build on the BLAST hits is used to build a local database and reference matrix composed of sequence distance measures. The measures are calculated in local alignment of the ambiguous sequence separately to all sequences from the BLAST recovered sequences set. A phylogenetic tree is generated showing most similar samples. A principal component analysis plot further simplifies the interpretation of samples diversity or dynamics.

The metagenome DNA from dairy plant and spontaneous fermentation in-house prepared curd cheeses analyzed for diversity within major types of *Lactococcus* phages revealed distinct composition of phages detected in cheeses from dairy plant and in-house productions, and higher diversity of phages from spontaneous fermentation on non-pasteurized milk as compared to starter culture-based productions.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title, Video S1: title.

Author Contributions: Conceptualization, visualization, supervision and project administration, M.S.; methodology, formal analysis, data curation, resources, funding acquisition, and writing—original draft preparation, M.S. and A.O.-S.; laboratory work, B.P., I.K., K.M. and W.W.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data reported is available in Supplementary Materials.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marcó, M.B.; Moineau, S.; Quiberoni, A. Bacteriophages and dairy fermentations. *Bacteriophage* **2012**, *2*, 149–158, doi:10.4161/bact.21868.
2. Kleppen, H.P.; Bang, T.; Nes, I.F.; Holo, H. Bacteriophages in milk fermentations: Diversity fluctuations of normal and failed fermentations. *Int. Dairy J.* **2011**, *21*, 592–600, doi:10.1016/j.idairyj.2011.02.010.
3. Schmidt, M.T.; Olejnik-Schmidt, A.K.; Zareba, A.; Pezacki, M.; Wojewoda, I.; Grajek, W. Induction of loci mutation during *Lactococcus lactis* spontaneous conversion to bacteriophage-insensitive phenotype. *Food Biotechnol.* **2010**, *24*, 332–348, doi:10.1080/08905436.2010.524470.
4. Mahony, J.; van Sinderen, D. Current taxonomy of phages infecting lactic acid bacteria. *Front. Microbiol.* **2014**, *5*, doi:10.3389/fmicb.2014.00007.
5. Deveau, H.; Labrie, S.J.; Chopin, M.C.; Moineau, S. Biodiversity and classification of lactococcal phages. *Appl. Environ.*

-
- Microbiol.* **2006**, 72, 4338–4346, doi:10.1128/AEM.02517-05.
6. Oliveira, J.; Mahony, J.; Hanemaaijer, L.; Kouwen, T.R.H.M.; van Sinderen, D. Biodiversity of bacteriophages infecting *Lactococcus lactis* starter cultures. *J. Dairy Sci.* **2018**, 101, 96–105, doi:10.3168/jds.2017-13403.
 7. Cattonaro, F.; Spadotto, A.; Radovic, S.; Marroni, F. Do you cov me? Effect of coverage reduction on metagenome shotgun sequencing studies. *F1000Research* **2020**, 7, doi:10.12688/f1000research.16804.1.
 8. Labrie, S.; Moineau, S. Multiplex PCR for detection and identification of lactococcal bacteriophages. *Appl. Environ. Microbiol.* **2000**, 66, 987–994, doi:10.1128/AEM.66.3.987-994.2000.
 9. Eric, S.D.; Nicholas, T.K.D.D.; Theophilus, K.A. Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA). *J. Bioinforma. Seq. Anal.* **2014**, 6, 1–6, doi:10.5897/ijbc2013.0086.
 10. Ihaka, R.; Gentleman, R. R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* **1996**, 5, 299–314, doi:10.1080/10618600.1996.10474713.
 11. Wright, E.S. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J.* **2016**, 8, 352–359, doi:10.32614/rj-2016-025.
 12. PCAtools GitHub - kevinblighe/PCAtools: PCAtools: everything Principal Components Analysis Available online: <https://github.com/kevinblighe/PCAtools?s=03>.
 13. Som, A. Theoretical foundation to estimate the relative efficiencies of the Jukes-Cantor+gamma model and the Jukes-Cantor model in obtaining the correct phylogenetic tree. *Gene* **2006**, 385, 103–110, doi:10.1016/j.gene.2006.03.027.
 14. Szczepańska, A.K.; Hejnowicz, M.S.; Kołakowski, P.; Bardowski, J. Biodiversity of *Lactococcus lactis* bacteriophages in Polish dairy environment. *Acta Biochim. Pol.* **2007**, 54, 151–158, doi:10.18388/abp.2007_3281.
 15. Mahony, J.; Moscarelli, A.; Kelleher, P.; Lugli, G.A.; Ventura, M.; Settanni, L.; van Sinderen, D. Phage biodiversity in artisanal cheese wheys reflects the complexity of the fermentation process. *Viruses* **2017**, 9, doi:10.3390/v9030045.
 16. Lavelle, K.; Martinez, I.; Neve, H.; Lugli, G.A.; Franz, C.M.A.P.; Ventura, M.; Dal Bello, F.; van Sinderen, D.; Mahony, J. Biodiversity of *Streptococcus thermophilus* phages in global dairy fermentations. *Viruses* **2018**, 10, doi:10.3390/v10100577.
 17. Moisan, M.; Moineau, S. Multilocus sequence typing scheme for the characterization of 936- like phages infecting *Lactococcus lactis*. *Appl. Environ. Microbiol.* **2012**, 78, 4646–4653, doi:10.1128/AEM.00931-12.