

# Characterization of the Genome of the Dairy *Lactobacillus helveticus* Bacteriophage $\Phi$ AQ113

Miriam Zago,<sup>a</sup> Erika Scaltriti,<sup>b</sup> Lia Rossetti,<sup>a</sup> Alessandro Guffanti,<sup>c</sup> Angelarita Armiento,<sup>b</sup> Maria Emanuela Fornasari,<sup>a</sup> Stefano Grolli,<sup>b</sup> Domenico Carminati,<sup>a</sup> Elena Brini,<sup>c</sup> Paolo Pavan,<sup>c</sup> Armando Felsani,<sup>c</sup> Annalisa D'Urzo,<sup>d</sup> Anna Moles,<sup>c</sup> Jean-Baptiste Claude,<sup>e</sup> Rita Grandori,<sup>d</sup> Roberto Ramoni,<sup>b</sup> Giorgio Giraffa<sup>a,f</sup>

Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per le Produzioni Foraggere e Lattiero-Casearie (CRA-FLC), Lodi, Italy<sup>a</sup>; Department of Veterinary Sciences, University of Parma, Parma, Italy<sup>b</sup>; Genomnia srl, Milan, Italy<sup>c</sup>; Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy<sup>d</sup>; Genostar, Montbonnot, France<sup>e</sup>; Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Unità di Ricerca per la Maiscoltura (CRA-MAC), Bergamo, Italy<sup>f</sup>

The complete genomic sequence of the dairy *Lactobacillus helveticus* bacteriophage  $\Phi$ AQ113 was determined. Phage  $\Phi$ AQ113 is a *Myoviridae* bacteriophage with an isometric capsid and a contractile tail. The final assembled consensus sequence revealed a linear, circularly permuted, double-stranded DNA genome with a size of 36,566 bp and a G+C content of 37%. Fifty-six open reading frames (ORFs) were predicted, and a putative function was assigned to approximately 90% of them. The  $\Phi$ AQ113 genome shows functionally related genes clustered together in a genome structure composed of modules for DNA replication/regulation, DNA packaging, head and tail morphogenesis, cell lysis, and lysogeny. The identification of genes involved in the establishment of lysogeny indicates that it may have originated as a temperate phage, even if it was isolated from natural cheese whey starters as a virulent phage, because it is able to propagate in a sensitive host strain. Additionally, we discovered that the  $\Phi$ AQ113 phage genome is closely related to *Lactobacillus gasserii* phage KC5a and *Lactobacillus johnsonii* phage Lj771 genomes. The phylogenetic similarities between *L. helveticus* phage  $\Phi$ AQ113 and two phages that belong to gut species confirm a possible common ancestral origin and support the increasing consideration of *L. helveticus* as a health-promoting organism.

Lactic acid bacteria (LAB) are important for the food industry because they are used as starter or adjunct cultures for the production of fermented foods. One of the main problems encountered in food fermentation is the ubiquitous presence of virulent bacteriophages that can alter the quality of fermented products or delay the manufacturing processes. In the dairy industry, vast quantities of milk are transformed daily to produce fermented dairy products (1–5). Phage infection represents a significant risk to this industry; therefore, phage populations must be kept under control and at a low level on a day-to-day basis (1). Considering the economic impact of phage infections on the dairy industry, the morphology, physiology, and genetics of LAB bacteriophages have been extensively studied. Most of these studies have concerned phages that infect *Lactococcus lactis* and *Streptococcus thermophilus*, but the increasing use of LAB, especially *Lactobacillus* spp., as health-promoting agents has led to a growing interest in *Lactobacillus* phages (2–4). As a matter of fact, phage infections are even more worrying when probiotic bacteria are the target. The manufacturing of certain types of probiotics involves the propagation of single strains as starters, and these strains are often slow growing and, therefore, particularly vulnerable to phage (5). In recent years, probiotic lactobacilli have become increasingly important in fermented foods and nutraceuticals; thus, we need to increase our knowledge of the phylogenesis, genomics, and ecology of the phages that infect lactobacilli (6–9). Despite the fact that some phage genomes of *Lactobacillus* spp. have already been sequenced, most of the sequences available in public databases belong to phages that infect *Lactococcus* spp. or *Streptococcus* spp. The genomes available for phages of *Lactobacillus* spp. include those of *Lactobacillus plantarum*  $\Phi$ g1e (accession number X98106),  $\Phi$ LP65 (AY682195), and  $\Phi$ JL1 (AY236756); *Lactobacillus gasserii*  $\Phi$ adh (AJ131519) and  $\Phi$ KC5a (DQ320509); *Lactobacillus delbrueckii* subsp. *lactis*  $\Phi$ LL-H (EF455602) and  $\Phi$ mv4;

*Lactobacillus casei*  $\Phi$ A2 (AJ251789) and  $\Phi$ AT3 (AY605066); and *Lactobacillus rhamnosus*  $\Phi$ Lc-Nu (AY236756). Surprisingly, no sequences of *Lactobacillus helveticus* bacteriophages are available, although this species has a recognized role in the starter cultures used for the production of Italian, French, and Swiss cheeses and is increasingly viewed as an emerging probiotic (10–12). Early studies on *L. helveticus* phages isolated from Emmental starters and cheese whey in French factories were carried out by Sozzi and Maret (13) and Séchaud et al. (14). More recently, Zago et al. demonstrated the presence of *L. helveticus* phages in natural whey starters for Grana Padano cheese (15, 16). This study had the following objectives: (i) to determine and analyze the complete genome sequence of the *L. helveticus* phage  $\Phi$ AQ113; (ii) to evaluate phage morphology, to identify the structural genes, and to analyze the structural proteins; and (iii) to explore the genomic organization of the phage and to compare it with that of related phages. In this paper, we report the analysis of the complete genome of an *L. helveticus* bacteriophage and one of the first-ever small genomes to be fully sequenced and assembled *ab initio* with the SOLiD short-read next-generation sequencing (NGS) technology.

Received 26 February 2013 Accepted 25 May 2013

Published ahead of print 31 May 2013

Address correspondence to Miriam Zago, miriam.zago@entecra.it.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.00620-13>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AEM.00620-13

## MATERIALS AND METHODS

**Bacteria and bacteriophage culture conditions.** *Lactobacillus helveticus* phage  $\Phi$ AQ113 and its host strain Lh1405 were isolated from a starter culture for Grana Padano cheese (15). Bacterial cells were maintained as a frozen stock at  $-80^{\circ}\text{C}$  in the presence of 15% glycerol as cryoprotective agent. MRS (Merck, Germany) broth, supplemented with 10 mM  $\text{CaCl}_2$ , was routinely used to grow bacteria and to propagate the phages at  $42^{\circ}\text{C}$ .  $\Phi$ AQ113 lysates and stocks were prepared as described by Zago et al. (16).

**Electron microscopy (EM).** Phage  $\Phi$ AQ113 particles were concentrated from 500 ml of a phage lysate by polyethylene glycol (PEG) precipitation and resuspended in 500  $\mu\text{l}$  of TMN solution (10 mM Tris-HCl, pH 7.7, 10 mM  $\text{MgSO}_4$ , 5 M NaCl) to obtain a titer of  $10^{12}$  CFU/ml. To avoid further loss of phage particles, no additional purification steps were performed. Samples of phage suspensions were then negatively stained with 2% uranyl acetate and examined in a JEOL JEE 1200 EXII (JEOL, London, United Kingdom) transmission electron microscope at an accelerating voltage of 80 kV.

**Phage DNA preparation.** Phages were propagated in their host in 2 liters of MRS-Ca broth. After lysis, the phage particles were filtered and concentrated by subsequent centrifugations (10,000  $\times$  g, 1 h,  $4^{\circ}\text{C}$ ; 30,000  $\times$  g, 30 min,  $4^{\circ}\text{C}$ ). The final suspension of approximately 2 ml was treated for 30 min with 1  $\mu\text{g/ml}$  DNase I and 1  $\mu\text{g/ml}$  RNase (Sigma-Aldrich, Milan, Italy). After incubation, the phage suspension was centrifuged (8,000  $\times$  g, 1 h,  $4^{\circ}\text{C}$ ). The phage pellet was then resuspended in 500  $\mu\text{l}$  of 20 mM EDTA containing 50  $\mu\text{g/ml}$  proteinase K and 0.5% (wt/vol) SDS. After incubation at  $56^{\circ}\text{C}$  for 1 h, DNA was then purified according to the protocol of Brown et al. (17).

**Restriction enzyme analysis.** Phage DNA was cleaved with the restriction enzymes AluI, AvaII, BamHI, HaeIII, HhaI, MspI, PstI, PvuI, and SmaI (New England BioLabs, Hertfordshire, United Kingdom) according to the manufacturer's instructions. The digested DNA fragments were separated on a 1.5% (wt/vol) agarose gel and visualized with GelRed solution (Biotium, Hayward, CA, USA) staining. A 1-kb Plus DNA ladder and a  $\lambda$ -HindIII DNA ladder (Invitrogen, Milan, Italy) were used as DNA molecular weight markers.

**Phage DNA NGS library construction and sequencing.** For the sequence determination, one fragment library (50:50 bp) and two mate-paired sequencing libraries with different insert sizes (MP\_1500 library, 1.5 kb; MP\_650 library, 650 bp) were created from the phage genomic DNA according to the manufacturer's instructions using the SOLiD fragment library construction kit and the SOLiD long mate-paired (LMP) library construction kit (Life Technologies, Monza, Italy), respectively.

**Mate-paired library construction.** Two aliquots of 12  $\mu\text{g}$  each of DNA were fragmented by sonication using the Covaris S2 system (Covaris, Woburn, MA, USA) to an average size of 650 bp or 1,500 bp. The fragments were then end repaired and size selected by gel electrophoresis. The size-selected genomic DNA fragments were ligated to long mate-paired (LMP) SOLiD cleaved amplified polymorphic (CAP) adaptors and circularized with internal adaptors. Due to the phosphorylation state of the adaptors, the procedure results in a nick on each strand when the DNA is circularized. These nicks were translated by nick translation into the genomic DNA region. The DNA was cut with T7 exonuclease and S1 nuclease on the strand opposite from the nick, and the DNA mate pair was released. The final product contained two genomic DNA tags derived from the ends of the initial DNA insert flanked by the P1 and P2 adaptors and was amplified by PCR for 14 cycles. The two 50-bp mate-paired libraries were checked for size distribution using the Agilent 1000 kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Waldbronn, Germany).

**Fragment library construction.** Six micrograms of DNA was sheared using the Covaris S2 system (Covaris). The sheared DNA was end repaired, adaptors P1 and P2 were ligated to the end-repaired DNA, and the DNA was size selected on a gel. The fragment library was obtained by amplifying the nick-translated DNA for 3 cycles. The fragment library dimension was checked using the Agilent 1000 kit on an Agilent 2100 Bioanalyzer.

**Sequencing.** Template beads were prepared from both the fragment and mate-paired libraries according to the manufacturer's instructions using the ePCR kit v.2 and the Bead Enrichment kit (Life Technologies) for SOLiD 3.5. The quality of the libraries was checked using a Workflow Analysis kit (Life Technologies) according to the manufacturer's instructions. A sufficient number of template P2 beads was deposited onto a slide according to the manufacturer's instructions using the Bead Deposition kit (Life Technologies). One quadrant of a slide was run for each library using the SOLiD mate-paired library sequencing kit with Master Mix 50 and SOLiD fragment library sequencing kit with Master Mix 50 (Life Technologies) using cyclic ligation sequencing on a SOLiD 3 Plus system. In this system, five rounds of primers (primers A, B, C, D, and E) are used to sequence template by ligation of dibase-labeled probes. In particular, each base of the sequencing template is interrogated in two independent ligation reactions by two different primers producing sequences in a numerical format (Color Space) instead of the standard nucleotide notation normally associated with pyrosequencing or Sanger sequencing. This ensures an optimal sequencing quality and the possibility of distinguishing errors from variants before the actual sequence mapping.

**Bioinformatic analysis.** Both the fragment and mate-paired libraries were separately assembled in Color Space with the SOLiD *de novo* accessory tools 2.0; the resulting contigs for each F3 and R3 were separately assembled using the Cap3 assembly program (<http://seq.cs.iastate.edu>) (18). The assemblies were compared pairwise using the EMBOSS Needle program, and chimeric contigs were eliminated from further analysis. An NCBI BLASTX (translated nucleotide query against protein database) search of the remaining contigs was performed against a subset of the NCBI nonredundant protein sequence database to identify the most similar phages; the same analysis performed with NCBI BLASTN was used to check the nucleotide-nucleotide similarity with other phage genomes. The EMBOSS Restrict program was used to produce an *in silico* restriction map and to compare it with the ones obtained in the laboratory. The GeneMarkS program (<http://exon.gatech.edu/genemarks.cgi>) (19), complemented with BLASTP searches against the NCBI nr database, was used to create a preliminary list of open reading frames (ORFs) in the assembled genome. As a second, complementary approach for comprehensive ORF prediction and functional annotation, we used the Metabolic Pathway Builder 3.10 software. ORFs were predicted *ab initio* with a Markov model, confirmed, and, if necessary, corrected by homology searches (BLAST). Each ORF was functionally annotated according to results from the BLAST (Swiss-Prot and nr databases) and HMM (PFAM) searches.

We performed a functional analysis by comparing the predicted ORFs against a nonredundant proteome collection of 4,731 proteins that corresponded to 96 selected phage genomes retrieved from the NCBI genome database. The search was performed with the NCBI BLASTP software (Blosum45 substitution matrix; gap costs: open, 15; extend, 2). The alignment program used was Clustalw2 (<http://www.ebi.ac.uk>) with Blosum45 matrix, a gap cost opening penalty of 15, and an extending penalty of 2. The phylogenetic tree was obtained using the Jukes-Cantor algorithm for evaluating genetic distance and neighbor joining for clustering, resampling by bootstrap 100 times, and creating a consensus tree. A Mauve genome alignment of  $\Phi$ AQ113 against the two more similar phages *L. gasseri*  $\Phi$ KC5a and *Lactobacillus johnsonii*  $\Phi$ Lj771 was performed (20).

**Sanger sequencing.** The phage AQ113 DNA sequence was validated in selected regions by Sanger sequencing using an ABI Prism 310 automated DNA sequencer (Life Technologies) as previously described (21). The primer pairs used for the DNA sequencing are listed in Table 1. The obtained sequences were compared and grouped into clusters according to the sequence distance between all pairs. The clusters were aligned as pairs and then collectively as sequence groups to produce the overall alignment. Sequence alignments were performed with the Sequence Navigator software using the ClustalW algorithm (Life Technologies).

**Analysis of structural proteins.** We analyzed purified phage particles for structural protein composition using 12% SDS-PAGE according to the method of Laemmli (22). Concentrated phage particles were prepared

TABLE 1 Primers used in this study for DNA sequencing

| Primer       | Sequence (5'→3')             | Position in the sequence | Amplicon size (bp) |
|--------------|------------------------------|--------------------------|--------------------|
| HEAD9 for    | GACAGAAGTTATTAACATGCTGAT     | 4850–4874                | 1,044              |
| HEAD9 rev    | AGTAGTCTTACCAGAGAAGTCAA      | 5894–5871                |                    |
| SHEATH15 for | GGCAATGGTAAGCCAGTATTAACC     | 7499–7522                | 812                |
| SHEATH15 rev | CTTCATAGTTATAAGCCACGCCACT    | 8311–8288                |                    |
| SHEATH16 for | GGCAACAACCTTTAGAACAGGTT      | 8910–8931                | 468                |
| SHEATH16 rev | CTGAATACCATCAAAAGGTTGAA      | 9378–9356                |                    |
| CDS16        | GCTGGTAAGCAAATTATCCAACCTCGAT | 9232–9258                | 875                |
| CDS19        | GGATTGCTAAAAGCGTCAACAACTCTA  | 10107–10081              |                    |
| CDS41        | GCACAGTCACTTCTTATAGACAAATT   | 26894–26919              | 837                |
| CDS43        | GTCAATTGGCGTATTGAAACACCCTT   | 27731–27707              |                    |
| CDS51        | GCTAATAATAATAATGACCAAGCCAC   | 30922–30948              | 911                |
| CDS52        | GTCAATATTTAAAGCCACCGCTTCTAT  | 31833–31808              |                    |
| CDS54        | GGAGACTACGAATATGTTAGAAAAATA  | 32966–32992              | 711                |
| CDS56        | CCACGCAACGCAATTAATGAAAT      | 33677–33655              |                    |
| CDS57 for    | GATTCAAATGGACAAGTAGATATTGCA  | 33888–33915              | 541                |
| CDS57 rev    | GGGATGCTTAGTTCTTCAAATTA      | 34429–34405              |                    |

from a 2-liter culture of infected host bacteria, as described for the electron microscopy visualization. The phages were lyophilized and resuspended in 30  $\mu$ l (final titer,  $10^{12}$  CFU/ml) of  $1\times$  phosphate buffer solution (137 mM NaCl, 2.7 mM KCl, 10 mM sodium phosphate dibasic, 2 mM potassium phosphate monobasic, pH 7.4). Fifteen microliters of phage solution was mixed with the same volume of Laemmli sample loading buffer that had been prepared with a slightly modified composition (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue, 0.125 M Tris HCl, pH 6.8) and boiled for 5 min. After electrophoresis, protein bands were visualized by Coomassie blue staining. Upon band excision, the proteins were destained, reduced, alkylated, and digested by trypsin. The peptides were extracted by the addition of 2 volumes of acetonitrile and 5% formic acid and were then lyophilized and desalted by ZipTip (Millipore, Billerica, MA), according to the manufacturer's instructions.

The peptides were analyzed on a hybrid quadrupole time of flight (Q-TOF) mass spectrometer (Qstar Elite; ABSciex, Framingham, MA) equipped with a nano-electrospray ionization sample source. Metal-coated borosilicate capillaries (Proxeon, Odense, Denmark) with medium-length emitter tips of 1- $\mu$ m internal diameter were used to infuse the samples. The instrument was calibrated by the standard Renin substrate solution (ABSciex, Framingham, MA) on the molecular ion  $MH^{2+}$  (879.97 Da) and its fragment  $MH^+$  (110.07 Da).

Mass spectrometry (MS) spectra were acquired in the 400-to-2,000  $m/z$  range with an accumulation time of 1.0 s, an ion spray voltage of 1,200 V, and a declustering potential of 60 V. The proteins were identified using Bioanalyst (ABSciex, Framingham, MA) software to interrogate a database containing the translated open reading frames of the entire bacteriophage  $\Phi$ AQ113 genome with a sequence coverage of at least 36% for peptide mass fingerprinting and at least 5 peptides matched by tandem MS (MS/MS) data for each protein.

**Nucleotide sequence accession number.** The nucleotide sequence of the phage  $\Phi$ AQ113 genome has been deposited in the EMBL database under accession no. [HE956704](https://www.ebi.ac.uk/EMBL/nuccore/HE956704).

## RESULTS AND DISCUSSION

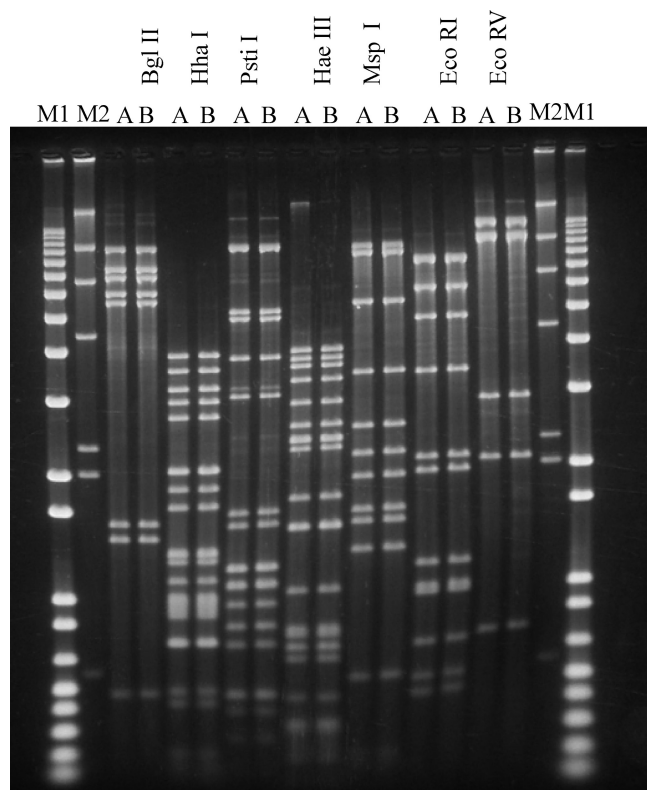
**$\Phi$ AQ113 genome characterization and proteome analysis.** *Lactobacillus helveticus* phage  $\Phi$ AQ113 was isolated from a natural whey starter culture used for the production of Grana Padano, an Italian hard-cooked cheese (15). Then, its titer was determined ( $10^{12}$  CFU/ml) and it was visualized by electron microscopy (EM). Although phage  $\Phi$ AQ113 belongs to the *Myoviridae* family and has a short tail, the contractile sheath was not visualized by EM. Additionally, we observed that this phage has an isometric capsid

(diameter,  $55.2 \pm 2.23$  nm) with a contractile tail (length,  $147.25 \pm 3.47$  nm; diameter,  $21.7 \pm 2.45$  nm) (Fig. 1). The final assembled consensus sequence revealed a linear, circularly permuted, double-stranded DNA genome with a size of 36,566 bp and a G+C content of 37%. The sequence assembly was validated by a comparative restriction analysis (data not shown). No evidence could be found for the presence of cohesive, protruding ends (*cos*) in the  $\Phi$ AQ113 genome, and ligation of its DNA did not alter the restriction patterns (as shown in Fig. 2). Therefore,  $\Phi$ AQ113 likely utilizes the *pac* mechanism of DNA packaging. The size of this phage genome is similar to that of the virulent *Siphoviridae* phages *L. plantarum*  $\Phi$ phiJL-1 (36,677 bp) and *L. rhamnosus*  $\Phi$ Lc-Nu (36,466 bp) and the temperate *Myoviridae* phage *L. gasseri*  $\Phi$ KC5a (38,239 bp) (23). The G+C content is similar to that (35.3%) of *L. gasseri* phage  $\Phi$ adh (2).

Fifty-six open reading frames (ORFs) were predicted both *ab initio* and by homology, and a putative function was assigned to approximately 90% of these ORFs (see Table S1 in the supplemental material). Five ORFs showed no homology to existing sequences, and at least five others had very low homology matches (<40%). Only 15 ORFs showed homology to annotated phage proteins; the majority of the ORFs showed similarity to sequences from bacterial genomes. As many of these ORFs are typical phage genes, it may be inferred that the matches in the bacterial genomes represented prophages or remnants of prophages. This finding might indicate that  $\Phi$ AQ113 itself has a prophage origin or a close



FIG 1 Transmission electron micrograph of phage  $\Phi$ AQ113.

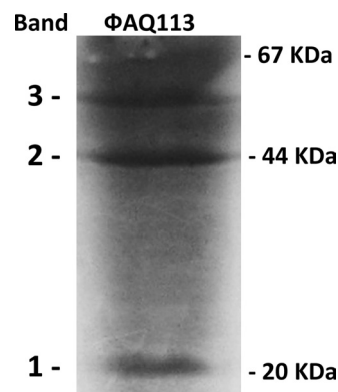


**FIG 2** Restriction analysis of  $\Phi$ AQ113 DNA. The phage DNA was digested with BglII, HhaI, PstI, HaeIII, MspI, EcoRI, and EcoRV. Lanes M1 and M2, 1-kb Plus and  $\lambda$ -HindIII DNA ladders; A and B indicate that the digests were unheated and heated prior to electrophoresis, respectively.

evolutionary relationship to temperate phages. In fact, this new phage genome is modularly organized, containing DNA packaging, head and tail morphogenesis, cell lysis, and DNA replication/regulation modules, as well as elements typically associated with lysogeny controls, such as a repressor, an antirepressor, and a Cro-like protein.

The bacteriophage structural proteins were visualized by SDS-PAGE. The high background of the staining is most likely due to interactions between the Coomassie brilliant blue stain and the PEG that was used for the purification of the phage. Three major bands were visible (Fig. 3). The smallest band was identified in the bacteriophage genome database as ORF16 (Bioanalyst score 121 and 63.7% sequence coverage for the peptide fingerprint data), a putative tail sheath protein (calculated molecular mass, 17,579 Da). The intermediate band was identified as ORF9 (Bioanalyst score 228 and 36% sequence coverage for the peptide fingerprint data), a putative major head protein (calculated molecular mass, 40,293 Da). The largest band was identified as ORF15 (Bioanalyst score 516 and 48% sequence coverage for the peptide fingerprint data), a putative tail sheath protein (calculated molecular mass, 52,698 Da). Therefore, on the basis of their putative functions, it is not surprising that these polypeptides might represent the major structural proteins of the bacteriophage.

**Genome analysis.** The *L. helveticus* bacteriophage genome was sequenced using the SOLiD system. The sequencing was performed at a very high coverage and with two different approaches (fragment and mate pair libraries), the latter at two different insert



**FIG 3** Analysis of the phage  $\Phi$ AQ113 structural proteins. After purifications of  $\Phi$ AQ113 by PEG precipitation, the purified phage proteins were separated by SDS-PAGE. The structural proteins from the main bands were identified by MS and MS/MS (see the text for details). Band 1, sheath tail protein; band 2, major head protein; band 3, sheath tail protein.

sizes (650 bp and 1,500 bp). The high sequence coverage was driven by a physical requirement of the sequencing: in one quad, not less than 50 M beads has to be deposited to reach high-quality data for a single sample. Moreover, we chose to analyze these three different types of libraries in order to identify the best strategy to solve the problem of sequencing a very heterogeneous small genome in a single shotgun step. The 50-bp fragment sequence library (FRAG 50) resulted in 69,493,353 reads; the mate pair library with a 650-bp insert (MP\_650) resulted in a total of more than 162 million reads, divided almost equally between the forward (F3) and reverse (R3) 50-nucleotide (50-nt) pairs; and the mate pair library with a 1,500-bp insert (MP\_1500) resulted in a number of sequences similar to that of the MP\_650 library. In total, the genome phage sequencing yielded 395,665,425 sequence reads of 50 nt each, for a grand total of 19,783,271,250 nt of sequence. Preliminary sequence assembly performed in the Sequencing By Ligation format (Color Space) resulted in three different series of contigs: 28 contigs from FRAG 50 with an  $N_{50}$  value of 2,081 nt, 10 contigs from MP\_650 with an  $N_{50}$  value of 32,438 nt, and 9 contigs derived from the Color Space assembly of MP\_1500 with an  $N_{50}$  value of 30,033 nt. Thus, sequencing with mate pairs at an insert length of 1,500 bp was identified as the best strategy to sequence a new phage genome. Further bioinformatic analysis of these contigs produced a single assembly of 34,928 nt, with extended similarities at the protein level with the *L. gasseri* phage Kc5a and in good agreement with the results of the restriction analysis. Further restriction analysis (Fig. 2) and comparison with the restriction profiles developed *in silico* prompted direct Sanger sequencing of the contig end from divergent primers (data not shown). The sequencing revealed a circular genome with an additional 1.3-kb segment. The genome was then reassembled from the start with the addition of the Sanger sequencing results, and we obtained a final contig of 36,307 nucleotides. An *in silico* restriction analysis of the new assembly, as well as a quantitative comparison with the previous restriction pattern and the one obtained in the laboratory, confirmed definitive progress toward the final genome sequence of 36,566 bp. An *ab initio* prediction performed with two differential methodologies and the functional annotation by similarity resulted in a total of 56 ORFs that can be divided into the six functional modules detailed below based on

their putative predicted functions (see Table S1 in the supplemental material): (i) DNA packaging, (ii) head morphogenesis, (iii) tail morphogenesis, (iv) lysis, (v) lysogeny, and (vi) DNA replication/regulation.

**(i) Genes encoding proteins involved in phage DNA packaging.** The products of the genes ORF 1, ORF 3, and ORF 4 most likely represent the small and large subunits of the phage terminase, found in *Lactobacillus vaginalis* strain ATCC 49540 and *L. gasseri* phage  $\Phi$ KC5a with a similarity that ranged from 66 to 83%. These results suggest that these three proteins are likely to be involved in phage DNA packaging. The protein encoded by ORF 5 shows a similarity of 71% with the portal protein of *L. gasseri* strain ATCC 33323. The portal protein is the second component of the bacteriophage DNA packaging. The gene product of ORF 2 shares homology (46.4%) with the HNH endonuclease of *Bacteroides* sp. strain 3\_1\_33FAA. These endonucleases can be found as free-standing ORFs between genes or contained within introns. However, it is likely that the HNH homing endonuclease is involved in DNA packaging because of the proximity of ORF 2 to the terminase genes, as has been observed for *L. plantarum* phage  $\Phi$ JL-1 (24).

**(ii) Genes involved in head morphogenesis.** The gene products of ORF 6 and ORF 8 show an identity of 54% with *L. johnsonii* prophage  $\Phi$ Lj771 and of 58.8% with the *L. gasseri* strain, respectively. ORF 9 is similar to the phage major capsid protein of *L. gasseri* phage  $\Phi$ KC5a (71% identity). The presence of this protein was confirmed by the analysis of the structural proteins (Fig. 3), and the 41-kDa protein appears to be the major head protein according to the amino acid sequence. The predicted protein encoded by ORF 10 shares a sequence similarity of 47.3% with the head-tail connector protein of the bacteriophage SPPI, suggesting that this ORF product may be involved in phage assembly.

**(iii) Genes involved in tail morphogenesis.** ORFs 15, 16, 19, 21, and 26 are proposed to encode components of the  $\Phi$ AQ113 tail. They exhibit similarity to tail components of *L. gasseri* strains and phage  $\Phi$ KC5a. In particular, the ORF 15 and 16 proteins are similar to phage sheath proteins that are involved in the creation of a channel for viral genome delivery by driving the tail tube through the cell outer membrane (25). The ORF 15 and 16 proteins were also observed by SDS-PAGE (Fig. 3) with molecular masses of 17.5 and 52.7 kDa, respectively. ORF 19 is predicted to encode a tape measure protein, which determines the length of the phage tail during morphogenesis (26). The sequence of ORF 19 shows a degree of similarity of 60% with an analogous protein from *L. gasseri* strain 224-1. The ORF 21 protein was identified as the minor tail protein, and ORF 26 appears to be a baseplate protein with a high degree of similarity (62.7%) to *L. gasseri* strain 224-1. Kondou et al. (27) showed that the baseplate protein is essential for phage Mu assembly and the generation of viable phages.

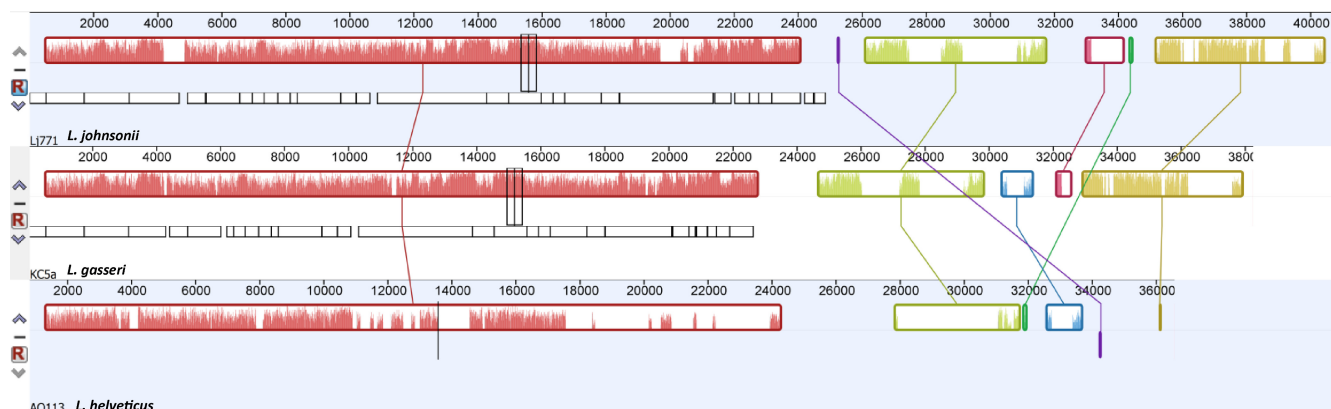
**(iv) Genes encoding host cell lysis proteins.** The protein encoded by ORF 35 displays similarity to a gene belonging to the lysis module. In particular, this ORF showed a high degree of identity (88%) with the endolysin (Mur-LH) from the *L. helveticus* temperate phage  $\Phi$ -0303 (28). This gene has been detected in other *L. helveticus* phages with 99% homology (16). Deutsch et al. (28) found a truncated integrase gene downstream from the *lys* gene that encodes Mur-LH; this genomic organization is similar to that of other LAB phages (29, 30). However, no integrase genes were found in the sequence of *L. helveticus* phage  $\Phi$ AQ113. ORF 22

showed homology to a Lys-M-like protein of *L. gasseri* phage  $\Phi$ KC5a (48% of identity), but ORF 22 is not adjacent to the previous endolysin gene.

**(v) Genes involved in the lysogeny module.** ORFs 40, 43, and 44 exhibit similarity to genes involved in the establishment of the lysogeny in other *Lactobacillus* phages. In particular, ORF 40 is implicated in the transcription of prophage gene expression (82% similarity), and ORF 43 codes for a Cro-like repressor protein, given its similarity with the Cro repressor of *Lactobacillus* phage Sal2. Cro repressors are antagonistic to *cI* repressors and therefore function to block or terminate lysogeny. ORF 44 appears to be an antirepressor by virtue of its similarity with the prophage antirepressor protein of *Lactobacillus jensenii* 269-3 (80.7% identity). This result may suggest an antirepression-mediated induction similar to that observed for a large family of lambdoid prophages found in *Salmonella* genomes. The repressors of these phages are not cleaved upon induction; rather, they are inactivated by the binding of small antirepressor proteins. Therefore, the formation of the complex causes the repressor to dissociate from DNA (31). ORF 49 encodes a prophage protein similar to that of *Lactobacillus crispatus* strain CTV-05 (74.8% identity). The product of ORF 47 shows similarity to the HNH endonuclease of the *Lactobacillus rhamnosus* strain Lc-Nu (42.8%). The HNH motif was originally identified in the subfamily of the HNH homing endonucleases that initiate the process of the insertion of mobile genetic elements into specific sites (32); therefore, the position of this ORF near the lysogeny module suggests that it may be involved in the excision/integration process of temperate phages, as has been shown for *Leuconostoc mesenteroides*  $\Phi$ 1-A4 (33). Overall, the data suggested that the virulent phage  $\Phi$ AQ113 may have originated from a temperate phage.

**(vi) Genes involved in DNA replication and regulation.** Beginning with ORF 51, the remaining genes are related to the replication and regulation of the transcription of phage DNA. ORFs 52 and 53 display an identity of 45% with DNA replication proteins from *L. crispatus* strain CTV-05 and *L. johnsonii* prophage Lj771, respectively, suggesting that the two proteins are involved in DNA synthesis. The predicted proteins of ORFs 54 and 63 show homology to the putative transcriptional regulators of the *L. gasseri* phage  $\Phi$ KC5a (73%) and the *L. gasseri* strain JV-V03 (38%), respectively. ORF 51 contains a conserved domain of the ERF (essential recombination function) superfamily of single-strand DNA annealing proteins that mediate the circularization of linear DNA following infection of the host cell (34). Additionally, ORF 56 displays a 55% identity to a single-stranded DNA binding protein of the temperate phage  $\phi$ NIH1.1. ORF 60 is also in the replication/regulation module of *L. helveticus* phage, and it encodes a protein that belongs to the restriction nuclease superfamily and has significant identity (83.6%) to *L. gasseri* strain JV-V03.

A successful bacteriophage infection requires the regulation of gene expression, DNA replication, formation of the phage capsid, and the release of the new phage particles from the infected host (35). In most bacteriophages, the genes encoding related biological functions are clustered together in functional groups or modules, and they are turned on and off in coordination (36). In the case of  $\Phi$ AQ113, the 56 ORFs are divided into six functional modules that also include lysogenic functions. Lysogeny is a peculiar trait in *L. helveticus* and is considered an important source of virulent phages in dairy factories (14, 37). Despite the fact that numerous strategies of phage control have been developed and are



**FIG 4** Comparison of the gene order and orientation of the genomes of *L. helveticus*  $\Phi$ AQ113, *L. gasseri*  $\Phi$ KC5a, and *L. johnsonii*  $\Phi$ Lj771 using Mauve genome alignment software. The figure was generated by the Mauve rearrangement viewer. Each of the three genomes is displayed horizontally with homologous locally colinear blocks outlined and connected with lines. The height of the similarity profile inside each block corresponds to the average level of conservation in that region of the genome sequence.

now used in the dairy industry, the risk of phage infections is still significantly elevated, because virulent bacteriophages are present and cannot be completely eliminated from food fermentation environments. Phage  $\Phi$ AQ113 was isolated from natural cheese whey starters as a virulent phage because it is able to propagate in a sensitive host strain (15). Nevertheless, the identification of genes involved in the establishment of lysogeny (ORFs 40, 43, and 44) indicates that it may have originated as a temperate phage. Several virulent *Lactobacillus* phages have been postulated to have a prophage origin, such as *L. delbrueckii* phage  $\Phi$ LL-H (38) and *L. casei* phage  $\Phi$ S1 (39). This study confirmed the assumption that temperate phages are one possible source of virulent phages, although it is difficult to estimate the extent of their contribution to the outbreaks of new virulent phages observed in industrial processes.

**Phylogenetic analysis.** A phylogenetic analysis was performed with the ORFs predicted to encode the major head protein (ORF 9) and a putative transcriptional regulator (ORF 54). The major head protein was one of the structural genes used for phylogenetic studies because the packaging-head gene cluster has a structural gene order that is generally very well conserved (40). ORF 54 was selected using the approach of clustering predicted proteins by sequence similarity. Beginning with all 56 predicted ORFs, we searched for highly similar sequences against a selection of 96 phage genomes by Tblastn (search translated nucleotide target genomes using protein queries). We finally selected the two ORFs with the highest similarities that can be found in the highest number of phages in the database. Because viruses are known to be greatly variable, we used a Blosum45 substitution matrix (gap costs: open, 15; extend, 2) that is recommended for highly divergent alignments and contains distances calculated to take into account a greater recombination ratio.

In a few cases, the resulting blast hits were split between two neighboring high-scoring segment pairs (HSPs), indicating that the ORFs had relatively large insertions in that genome. For these cases, the entire genome sequence was retrieved, translated again, and replaced with the source hits. We thus obtained a new target proteome that was subjected to a new alignment. We also searched the same 56 predicted ORFs against a nonredundant proteome collection of 4,731 proteins belonging to the same 96 phages re-

trieved from NCBI. The search was performed by BLASTP (Blosum45 substitution matrix; gap costs: open, 15; extend, 2).

All of the dendrograms confirmed the similarity of *L. helveticus*  $\Phi$ AQ113 to other *Lactobacillus* thermophilic phages such as *L. gasseri*  $\Phi$ KC5a and *L. johnsonii*  $\Phi$ Lj771 (see Fig. S1 in the supplemental material), but *Lactobacillus delbrueckii* phages were not even present in the dendrogram, highlighting their lack of similarity to *L. helveticus*  $\Phi$ AQ113. Mauve genome alignment revealed many clusters of genes or colinear blocks that are in the same order in both  $\Phi$ AQ113, *L. gasseri*  $\Phi$ KC5a, and *L. johnsonii*  $\Phi$ Lj771 (Fig. 4). The gene order of the different modules was largely colinear, although *L. gasseri*  $\Phi$ KC5a and *L. johnsonii*  $\Phi$ Lj771 were 2 and 4 kb longer than  $\Phi$ AQ113, respectively. As expected, the three phages revealed a high level of sequence similarity in the packaging and head and tail morphogenesis modules; however, a greater level of diversity was observed for the lysis, lysogeny, and replication/regulation modules. In particular, the modular arrangement of the genes responsible for packaging and head and tail morphogenesis was similar to that of the two bacteriophages of *L. johnsonii* and *L. gasseri*. This finding suggests that these viruses and the *L. helveticus* phage  $\Phi$ AQ113 have a common ancestry, as shown by Slattery et al. (41) concerning *L. helveticus* bacteria. It is worth noting that *L. helveticus* strain DPC4571, isolated from cheddar cheese, shows a remarkable similarity in gene content with intestinal lactobacilli such as *L. acidophilus* NCFM, supporting the concept that the sequence differences between the dairy and the gut species may be caused by a relatively small but highly specific gene set (42). Phylogenetic similarities between *L. helveticus* phage  $\Phi$ AQ113 and two phages that belong to gut species confirm a possible common ancestral origin and support the increasing consideration of *L. helveticus* as a health-promoting organism (43, 44).

In conclusion, *L. helveticus* phage  $\Phi$ AQ113 possesses a mosaic genome with predicted proteins that have homologies to proteins from a wide range of bacteria. The majority of these are intestinal bacteria; this assumption emphasizes the probiotic characteristics of *L. helveticus* species. Furthermore, the fact that many of the  $\Phi$ AQ113 genes matched sequences from bacterial genomes highlights the high propensity to genetic exchange and recombination between bacteriophages and bacteria.  $\Phi$ AQ113 shows also a lysogenic module that underlines its possible origin from a temperate

phage. This study may provide new insights into phage genetics and phage-host interactions in *L. helveticus* and between *Lactobacillus* phages.

## ACKNOWLEDGMENTS

This work was supported by grants from the ASER/COL-MIA project (D.M. MIPAF 16101/7301/08) and Genomina research and development funds (for NGS and bioinformatic analysis). The strain used in this study belongs to the ASER/COL-MIA collection.

We thank Maria Luisa Callegari from the University of Piacenza for performing the electron microscopy.

## REFERENCES

- Moineau S, Levesque C. 2005. Control of bacteriophages in industrial fermentations, p 285–296. In Kutter E, Sulakvelidze A (ed), *Bacteriophages: biology and applications*. CRC Press, Boca Raton, FL.
- Altermann E, Klein JR, Henrich B. 1999. Primary structure and features of the genome of the *Lactobacillus gasserii* temperate bacteriophage  $\Phi$ adh. *Gene* 236:333–346.
- Brussow H. 2001. Phages of dairy bacteria. *Annu. Rev. Microbiol.* 55:283–303.
- Brussow H, Suarez JE. 2006. *Lactobacillus* phages, p 653–666. In Calendar R, Abedon ST (ed), *The bacteriophages*, 2nd ed. Oxford University Press, New York, NY.
- Mozzi F, Raya RR, Vignolo GM. 2010. Biotechnology of lactic acid bacteria—novel applications. Wiley-Blackwell, Blackwell Publishing, Ames, IA.
- Brussow H, Hendrix RW. 2002. Phage genomics: small is beautiful. *Cell* 108:13–16.
- Desiere F, Lucchini S, Brüssow H. 1998. Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. *Virology* 241:345–356.
- Desiere F, Mahanivong C, Hillier AJ, Chandry PS, Davidson BE, Brüssow H. 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology* 283:240–252.
- Hendrix RW. 2003. Bacteriophage genomics. *Curr. Opin. Microbiol.* 6:506–511.
- Lazzi C, Rossetti L, Zago M, Neviani E, Giraffa G. 2004. Evaluation of bacterial communities belonging to natural whey starters for Grana Padano cheese by length-heterogeneity-PCR. *J. Appl. Microbiol.* 96:481–490.
- Nakamura Y, Yamamoto N, Sakai K, Okubo A, Yamazaki S, Takano T. 1995. Purification and characterization of angiotensin I-converting enzyme inhibitors from sour milk. *J. Dairy Sci.* 78:777–783.
- Rossetti L, Fornasari ME, Gatti M, Lazzi C, Neviani E, Giraffa G. 2008. Grana Padano cheese whey starters: microbial composition and strain distribution. *Int. J. Food Microbiol.* 127:168–171.
- Sozzi T, Maret R. 1975. Isolation and characteristics of *Streptococcus thermophilus* and *Lactobacillus helveticus* phages from Emmental starters. *Lait* 55:269–288.
- Séchaud L, Rousseau M, Fayard B, Callegari ML, Quéneé P, Accolas JP. 1992. Comparative study of 35 bacteriophages of *Lactobacillus helveticus*: morphology and host range. *Appl. Environ. Microbiol.* 58:1011–1018.
- Zago M, Comaschi L, Neviani E, Carminati D. 2005. Investigation on the presence of bacteriophages in natural whey starters used for the production of Italian long-ripened cheeses. *Milchwissenschaft* 60:171–174.
- Zago M, Rossetti L, Reinheimer J, Carminati D, Giraffa G. 2008. Detection and identification of *Lactobacillus helveticus* bacteriophages by PCR. *J. Dairy Res.* 75:196–201.
- Brown JCS, Ward LJH, Davey GP. 1994. Rapid isolation and purification of lactococcal bacteriophage DNA without the use of caesium chloride gradients. *Lett. Appl. Microbiol.* 18:292–293.
- Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9:868–877.
- Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29:2607–2618.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.
- Suarez V, Zago M, Giraffa G, Reinheimer J, Quiberoni A. 2009. Evidence for the presence of restriction/modification systems in *Lactobacillus delbrueckii*. *J. Dairy Res.* 76:433–440.
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.
- Villion M, Moineau S. 2009. Bacteriophages of *Lactobacillus*. *Front. Biosci.* 14:1661–1683.
- Lu Z, Altermann E, Breidt F, Predki P, Fleming HP, Klaenhammer T. 2005. Sequence analysis of the *Lactobacillus plantarum* bacteriophage  $\Phi$  JL-1. *Gene* 348:45–54.
- Aksyuk AA, Leiman PG, Kurochkina LP, Shneider MM, Kostyuchenko VA, Mesyanzhinov VV, Rossmann MG. 2009. The tail sheath structure of bacteriophage T4: a molecular machine for infecting bacteria. *EMBO J.* 28:821–829.
- Katsura I. 1987. Determination of bacteriophage lambda tail length by a protein ruler. *Nature* 327:73–75.
- Kondou Y, Kitazawa D, Takeda S, Tsuchiya Y, Yamashita E, Mizuguchi M, Kawano K, Tsukihara T. 2005. Structure of the central hub of bacteriophage Mu baseplate determined by X-ray crystallography of gp44. *J. Mol. Biol.* 352:976–985.
- Deutsch SM, Guezenec S, Piot M, Foster S, Lortal S. 2004. Mur-LH, the broad-spectrum endolysin of *Lactobacillus helveticus* temperate bacteriophage  $\Phi$ -0303. *Appl. Environ. Microbiol.* 70:96–103.
- Arendt EK, Daly C, Fitzgerald GF, van de Guchte M. 1994. Molecular characterization of lactococcal bacteriophage Tuc2009 and identification and analysis of genes encoding lysin, a putative holin, and two structural proteins. *Appl. Environ. Microbiol.* 60:1875–1883.
- van Sinderen D, Karsens H, Kok J, Terpstra P, Ruiters MH, Venema G, Nauta A. 1996. Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage r1t. *Mol. Microbiol.* 19:1343–1355.
- Lemire S, Figueroa-Bossi N, Bossi L. 2011. Bacteriophage cross-talk: coordination of prophage induction by trans-acting antirepressors. *PLoS Genet.* 7:e1002149. doi:10.1371/journal.pgen.1002149.
- Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* 29:3757–3774.
- Lu Z, Altermann E, Breidt F, Kozyavkin S. 2010. Sequence analysis of *Leuconostoc mesenteroides*  $\Phi$ 1-A4 isolated from an industrial vegetable fermentation. *Appl. Environ. Microbiol.* 76:1955–1966.
- Poteete AR. 1982. Location and sequence of the *erf* gene of phage P22. *Virology* 119:422–429.
- Brøndsted L, Østergaard S, Pedersen M, Hammer K, Vogensen FK. 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. *Virology* 283:93–109.
- Ptashne M. 2004. A genetic switch: phage lambda revisited, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Carminati D, Mazzucotelli L, Giraffa G. 1997. Incidence of inducible bacteriophage in *Lactobacillus helveticus* strains isolated from natural whey starter cultures. *J. Dairy Sci.* 80:1505–1511.
- Mikkonen M, Dupont L, Alatosava T, Ritzenthaler P. 1996. Defective site-specific integration elements are present in the genome of virulent bacteriophage LL-H of *Lactobacillus delbrueckii*. *Appl. Environ. Microbiol.* 62:1847–1851.
- Shimizu-Kadota M, Sakurai T, Tsuchida N. 1983. Prophage origin of a virulent phage appearing on fermentations of *Lactobacillus casei* S-1. *Appl. Environ. Microbiol.* 45:669–674.
- Brussow H, Desiere F. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insight from dairy phages. *Mol. Microbiol.* 39:213–222.
- Slattery L, O'Callaghan J, Fitzgerald GF, Beresford T, Ross RP. 2010. Invited review: *Lactobacillus helveticus*—a thermophilic dairy starter related to gut bacteria. *J. Dairy Sci.* 93:4435–4454.
- Callanan MP, Kaleta P, O'Callaghan J, O'Sullivan O, Jordan K, McAuliffe O, Sangrador-Vegas A, Slattery L, Fitzgerald GF, Beresford T, Ross RP. 2008. Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. *J. Bacteriol.* 190:727–735.
- Taverniti V, Guglielmetti S. 2011. The immunomodulatory properties of probiotic microorganisms beyond their viability (ghost probiotics: proposal of paraprobiotic concept). *Genes Nutr.* 6:261–274.
- Vinderola G, Matar C, Perdígón G. 2007. Milk fermentation products of *Lactobacillus helveticus* R389 activate calcineurin as a signal to promote gut mucosal immunity. *BMC Immunol.* 8:19. doi:10.1186/1471-2172-8-19.