

# The mobilome of *Lactobacillus crispatus* M247 includes two novel genetic elements: Tn7088 coding for a putative bacteriocin and the siphovirus prophage $\Phi$ M247

Lorenzo Colombini<sup>1,\*</sup>, Francesco Santoro<sup>1</sup>, Mariana Tirziu<sup>1</sup>, Elisa Lazzeri<sup>1</sup>, Lorenzo Morelli<sup>2</sup>, Gianni Pozzi<sup>1</sup> and Francesco Iannelli<sup>1,\*</sup>

## Abstract

*Lactobacillus crispatus* is a member of the vaginal and gastrointestinal human microbiota. Here we determined the complete genome sequence of the probiotic strain M247 combining Nanopore and Illumina technologies. The M247 genome is organized in one circular chromosome of 2336109 bp, with a GC content of 37.04% and 2303 ORFs, of which 1962 could be annotated. Analysis of the M247 mobilome, which accounts for 14% of the whole genome, revealed the presence of: (i) Tn7088, a novel 14105 bp long integrative and mobilizable element (IME) containing 16 ORFs; (ii)  $\Phi$ M247, a novel 42510 bp long siphovirus prophage containing 52 ORFs; (iii) three clustered regularly interspaced short palindromic repeats (CRISPRs); and (iv) 226 insertion sequences (ISs) belonging to 14 different families. Tn7088 has a modular organization including a mobilization module encoding FtsK homologous proteins and a relaxase, an integration/excision module coding for an integrase and an excisionase, and an adaptation module coding for a class I bacteriocin and homologous to the listeriolysin S (*lts*) locus of *Listeria monocytogenes*. Genome-wide homology search analysis showed the presence of Tn7088-like elements in 12 out of 23 *L. crispatus* complete public genomes. Mobilization and integration/excision modules are essentially conserved, while the adaptation module is variable since it is the target site for the integration of different ISs. Prophage  $\Phi$ M247 contains genes for phage structural proteins, DNA replication and packaging, lysogenic and lytic cycles.  $\Phi$ M247-like prophages are present in seven *L. crispatus* complete genomes, with sequence variability mainly due to the integration of ISs. PCR and sequencing showed that the Tn7088 IME excises from the M247 chromosome producing a circular form at a concentration of  $4.32 \times 10^{-5}$  copies per chromosome, and reconstitution of the Tn7088 chromosomal target site occurred at  $6.65 \times 10^{-4}$  copies per chromosome. The  $\Phi$ M247 prophage produces an excised form and a reconstituted target site at a level of  $3.90 \times 10^{-5}$  and  $2.48 \times 10^{-5}$  copies per chromosome, respectively. This study identified two novel genetic elements in *L. crispatus*. Tn7088 represents the first example of an IME carrying a biosynthetic gene cluster for a class I bacteriocin in *L. crispatus*.

## DATA SUMMARY

The complete genome sequence of *Lactobacillus crispatus* M247 is available under GenBank accession no. CP088015, whereas Nanopore and Illumina sequencing reads are available under Sequence Read Archive (SRA) accession no. SRR17479173 and SRR17479172, respectively.

Received 25 July 2023; Accepted 20 November 2023; Published 12 December 2023

**Author affiliations:** <sup>1</sup>Laboratory of Molecular Microbiology and Biotechnology, Department of Medical Biotechnologies, University of Siena, 53100 Siena, Italy; <sup>2</sup>Università Cattolica del Sacro Cuore, Department of Food Science and Technologies for a Sustainable Agri-food Supply Chain (DiSTAS), University of Piacenza, 53100 Piacenza, Italy.

**\*Correspondence:** Lorenzo Colombini, [lorenzo.colombini2@unisi.it](mailto:lorenzo.colombini2@unisi.it); Francesco Iannelli, [francesco.iannelli@unisi.it](mailto:francesco.iannelli@unisi.it)

**Keywords:** bacteriocins; complete genome; CRISPRs; insertion sequences (ISs); integrative and mobilization elements (IMEs); *Lactobacillus crispatus*; M247; mobile genetic element (MGEs); mobilome; nanopore sequencing; probiotic; prophage; siphovirus; Tn7088;  $\Phi$ M247.

**Abbreviations:** *attB*, bacteria attachment site; *attL*, left attachment site; *attP*, phage attachment site; *attR*, right attachment site; *attTn*, transposon attachment site; bp, base pairs; CRISPR, clustered regularly interspaced short palindromic repeat; DR, direct repeat; EPS, exopolysaccharide; ICE, integrative and conjugative element; IME, integrative and mobilizable element; IS, insertion sequence; Mb, megabase; MGE, mobile genetic element; MRS, DeMan-Rogosa-Sharpe; nts, nucleotides; ORFs, open reading frames; qPCR, quantitative PCR; SSC, saline-sodium citrate; TOMM, thiazole/oxazole-modified microcins.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary figure and four supplementary tables are available with the online version of this article.

001150 © 2023 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

**Impact Statement**

Despite a huge number of bacterial genomic sequences being available in public databases, only a few are complete genomes. Furthermore, the bacterial mobilome, which can constitute up to 25% of the whole genome, remains poorly annotated and characterized. In this work, we determined the complete genome sequence of the probiotic strain *Lactobacillus crispatus* M247 and analysed its mobilome, revealing the presence of two novel genetic elements: the integrative and mobilizable element Tn7088 carrying an adaptation module homologous to the listeriolysin S locus of *Listeria monocytogenes*, and the siphovirus prophage  $\Phi$ M247. We also demonstrated that both elements are able to produce excised forms and reconstitute the target sites in *L. crispatus* chromosomes. These findings suggest that both elements have the potential to horizontally transfer among different *L. crispatus* strains and possibly to other bacterial species of the vaginal and gastrointestinal human microbiota. These data will contribute to the understanding of the evolution of this important microorganism.

**INTRODUCTION**

*Lactobacillus crispatus* is a member of the vaginal and gastrointestinal human microbiota. In healthy women, it is the most commonly isolated species among the vaginal lactobacilli and its presence correlates with a reduced risk of preterm delivery, sexually transmitted infections and bacterial vaginosis [1, 2]. The ability of *L. crispatus* to colonize the vagina following oral or local administration has also been reported [3, 4]. Exopolysaccharide (EPS) production has been associated with beneficial effects of *L. crispatus* against vaginal pathogens and with its persistence in the gut environment [5, 6]. Furthermore, the production of bacteriocins, ribosomally synthesized peptides with antibiotic properties, may enhance bacterial fitness [7, 8]. Despite a large number of *L. crispatus* genomic sequences being available in public databases, to date only 23 are complete genomes. Furthermore, the *L. crispatus* mobilome, defined as the entire set of mobile genetic elements (MGEs), remains poorly annotated and characterized. Genomic variability among *L. crispatus* strains correlates with the site of isolation and is due to allelic variation in the EPS biosynthesis locus, in metabolic genes or to the presence of MGEs, including prophages and clustered regularly interspaced short palindromic repeats (CRISPRs) [9–11]. Lysogeny has been frequently observed in *L. crispatus* strains isolated from the vagina of healthy women [12], and some of these bacteriophages could be induced and had a lytic activity [13], but their nucleotide sequence has not been characterized. A more recent comparative genomic study identified prophage sequences in 90 out of 105 *L. crispatus* strains isolated from various sources, with complete prophage more likely in human strains [10]. In this work, we determined the complete genome sequence of *L. crispatus* strain M247, a newborn faecal isolate largely studied for its probiotic activity and characterized by a strong aggregation phenotype and adherence to intestinal mucus [14–19]. It has been reported that M247 has beneficial effects on intestinal inflammatory disorders [20–22], helps to counteract vaginal dysbiosis [23] and probably contributes to papilloma virus clearance [24]. The M247 genome sequence was obtained combining long Nanopore reads and Illumina reads and mobilome analysis revealed the presence of two novel MGEs: the integrative and mobilizable element (IME) Tn7088 coding for a putative bacteriocin and the siphovirus prophage  $\Phi$ M247.

**METHODS****Bacterial strains and growth conditions**

*L. crispatus* strain M247 was isolated from faeces of human newborns [15]. Bacteria were grown in DeMan-Rogosa-Sharpe (MRS) broth (Oxoid) or in MRS supplemented with 1.5% agar (BD Difco) in the presence of 5% CO<sub>2</sub> at 37°C.

**Genomic DNA purification and quantification**

Bacterial cells were grown at 37°C in 250 ml of MRS broth until reaching middle exponential phase (OD<sub>590</sub> of 1.9), and then harvested by centrifugation at 5000 g for 30 min at 4°C. High-molecular-weight genomic DNA was purified using a raffinose-based method [25, 26]. Briefly, a cell pellet was dry vortex-mixed for 2–3 min and incubated for 1 h at 37°C in Protoplasting Buffer [20% raffinose, 50 mM Tris/HCl (pH 8.0), 5 mM EDTA] containing 4 mg ml<sup>-1</sup> lysozyme. Protoplasts were centrifuged at 5000 g for 5 min, resuspended in 15 ml of deionized H<sub>2</sub>O containing 100 µg ml<sup>-1</sup> proteinase K (Merck) and 0.5% SDS, and incubated for 30 min at 37°C to obtain osmotic lysis. Then, 0.55 M NaCl was added and the mixture was incubated for 10 min at room temperature. High-molecular-weight DNA was purified three times with 1 volume of chloroform-isoamyl alcohol (24:1, v/v), precipitated in 0.6 volumes of ice-cold isopropanol and spooled on a glass rod. DNA was resuspended in 10-fold diluted saline-sodium citrate (SSC) 1× buffer, then adjusted to 1× SSC and maintained at 4°C. The DNA solution was homogenized using a rotator mixer. DNA was quantified with a Qubit 2.0 Fluorometer (Invitrogen, Life Technologies) by using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific) and results were confirmed via a spectrophotometer (Implen). DNA integrity and size were assayed by horizontal gel electrophoresis using 0.6% Seakem LE (Lonza) agarose in 0.5× Tris borate EDTA running buffer.

## Illumina sequencing

Illumina sequencing was performed at MicrobesNG (University of Birmingham, UK) using a Nextera library preparation kit (Illumina) followed by sequencing on a NovaSeq 6000 device (Illumina) (2×250 bp paired-end sequencing). Illumina reads were trimmed using Trimmomatic v0.30 (<https://github.com/usadellab/Trimmomatic>) and analysed with FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Illumina read properties are reported in Table S1, available in the online version of this article.

## Nanopore sequencing

Nanopore sequencing was carried out essentially as already described [27, 28]. Briefly, sequencing reactions were carried out in 1.5 ml LoBind tubes (Sarstedt) using wide bore (Ø1.2 mm) tips for DNA manipulation to reduce physical shearing. Genomic DNA was size selected with 0.5 volumes of AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions. Two micrograms of size-selected DNA was employed for library construction using the SQK-LSK 108 kit (Oxford Nanopore Technologies). Library preparation was performed following the manufacturer's protocol with modifications: (i) incubation on rotator mix for 15 min; and (ii) the Library Loading Beads were not added. Finally, 1 µg of DNA library was loaded onto a R9.4 flow cell (FLO-MIN106) (Oxford Nanopore Technologies). A 21 h sequencing run was performed on a GridION device (Oxford Nanopore Technologies). Real-time base calling was performed with Guppy v3.2.6 (Oxford Nanopore Technologies), filtering out reads with a quality cut-off <Q7. Base called reads were analysed with NanoPlot v1.18.2 (<https://github.com/wdecoster/NanoPlot>). Nanopore read properties are reported in Table S1.

## Genome assembly and annotation

The M247 Nanopore reads were filtered to obtain 95× coverage taking 2.3 Mb as the genome size estimate using FilTlong v0.2.0 software (<https://github.com/rrwick/FilTlong>) with parameter `--target_bases` and assembled using Flye v2.7.1 (<https://github.com/fenderglass/Flye>). The resulting circular contig was polished with Medaka v0.7.1 (<https://github.com/Nanoporetech/medaka>) using all the Nanopore reads, followed by two polishing rounds with Pilon v1.22 (<https://github.com/broadinstitute/pilon>) using the Illumina reads. Assembly completeness was assessed with Bandage v0.8.1 (<https://github.com/rrwick/Bandage>), whereas assembly quality was evaluated with Ideel (<https://github.com/mw55309/ideel>) and CheckM v1.1.3 (<https://github.com/ECogenomics/CheckM>). Bwa v0.7.17 (<https://github.com/lh3/bwa>) and minimap2 v2.13 (<https://github.com/lh3/minimap2>) were used to align Illumina and Nanopore reads to the assembled genome, respectively. Aligned reads were visually inspected with Tablet v1.17.08.17 (<https://github.com/cropgeeks/tablet>) and used to further verify the assembled structure. The M247 genome was automatically annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.1 [29]. Default parameters were used for all software unless otherwise specified.

## Mobilome analysis

Mobilome analysis was performed as already described [30]: ICEfinder (<https://bioinfo-mml.sjtu.edu.cn/ICEfinder/ICEfinder.html>) was used to investigate the presence of bacterial integrative and conjugative elements (ICEs) and IMEs in the M247 genome, ISSaga ([http://issaga.biotoul.fr/issaga\\_index.php](http://issaga.biotoul.fr/issaga_index.php)) for insertion sequences (ISs), and PHASTER (<http://phaster.ca>), Virfam (<http://biodev.cea.fr/virfam/>) and Viridic (<http://rhea.icbm.uni-oldenburg.de/VIRIDIC/>) for prophages. The presence of CRISPRs was evaluated with CRISPRCasFinder (<https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>). Analysis of antibiotic resistance genes analysis was performed using RGI (Resistance Gene Identifier) (v3.2.1) (<https://card.mcmaster.ca/analyze/rgi>), based on CARD (Comprehensive Antibiotic Resistance Database), with parameter `'-loose_criteria=no'`. DNA sequence analysis was performed with Artemis/ACT v17.0.1 (<http://sanger-pathogens.github.io/Artemis/>). Manual annotation of MGEs was carried out by BLAST homology searches of the databases available at the National Center for Biotechnology Information (NCBI) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>), and the Pfam protein family database (available under the InterPro consortium, <https://www.ebi.ac.uk/interpro/search/sequence/>). A transposon name was assigned by the Tn Registry website curators (<https://transposon.lstmed.ac.uk/tn-registry>). The *L. crispatus* complete genomes were downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/1815/>) and the GenBank accession numbers are reported in Table S2.

## Mitomycin C induction and phage preparation

Mitomycin C induction and phage preparation were obtained essentially as previously reported [31]. Briefly, bacterial cells were grown in 600 ml of MRS broth until early exponential phase, and the culture was then split into three aliquots of which two were treated with 200 and 400 ng ml<sup>-1</sup> of mitomycin C. After 21 h of incubation at 37°C, EDTA was added at a final concentration of 10 mM and the samples were centrifuged twice at 5000 g for 40 min at 4°C in 50 ml tubes to eliminate bacterial cells and cellular debris. The recovered supernatants were transferred into six-polyallomer centrifuge tubes and ultracentrifuged at 20000 g for 2 h at 10°C in an Optima L-90K ultracentrifuge with the SW 32 Ti rotor (Beckman Coulter). The phage pellets were resuspended in 350 µl of TM buffer (50 mM Tris/HCl, 10 mM MgSO<sub>4</sub>) and the phage preparations were stored at 4°C.

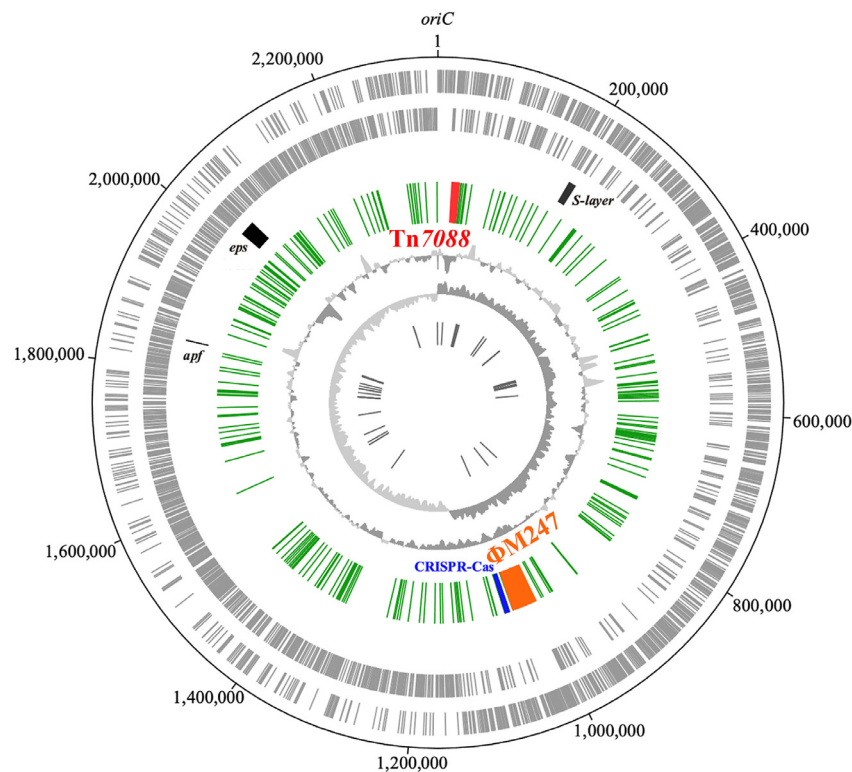
## PCR and sequencing

PCR and direct PCR sequencing were carried out as previously described [32–35]. Oligonucleotide primers and their properties are reported in Table S3. The circular forms of Tn7088 and  $\Phi$ M247 were detected using divergent primers directed at the ends of the elements, while reconstitution of the target sites using primers directed at the chromosomal junction fragments as reported [36, 37]. The following templates were used: (i) the purified high-molecular-weight DNA and (ii) the phage preparations from mitomycin C-treated and untreated M247 cultures. Briefly, quantitative PCR (qPCR) was carried out with the KAPA SYBR FAST qPCR kit Master Mix Universal (2 $\times$ ) (Merck) on a LightCycler 1.5 apparatus (Roche Diagnostics). The real-time PCR mixture contained, in a final volume of 20  $\mu$ l, 1 $\times$  KAPA SYBR FAST qPCR mix, 5 pmol of each primer, and 20 ng of bacterial genomic DNA or 2  $\mu$ l of phage preparation as starting template. The thermal profile was an initial 3 min denaturation step at 95 $^{\circ}$ C followed by 40 cycles of repeated denaturation (0 s at 95 $^{\circ}$ C), annealing (20 s at 62 $^{\circ}$ C) and polymerization (30 s at 72 $^{\circ}$ C). The temperature transition rate was 20 $^{\circ}$ C s $^{-1}$  in the denaturation and annealing steps and 5 $^{\circ}$ C s $^{-1}$  in the polymerization step. Primer pair IF1487/IF1488 amplifying a 350 bp fragment, and primer pair IF1513/IF1514 amplifying a 426 bp fragment were used for quantification of circular forms of Tn7088 and  $\Phi$ M247, respectively; while IF1349/IF1350 amplifying a 227 bp fragment and IF1511/IF1512 amplifying a 133 bp fragment were used for free locus quantification of Tn7088 and  $\Phi$ M247, respectively; a 292 bp fragment of the chromosomal *gyrB* gene, obtained with primers IF1352/IF1353, was used to standardize results. A standard curve for the *gyrB* gene of *L. crispatus* M247 was built by plotting the threshold cycle against the number of chromosome copies using serial dilutions of chromosomal DNA with known concentration. Melting curve analysis was performed to differentiate the amplified products from primer dimers. A t-test was used to assess the statistical significance of differences in quantification values of excised forms.

## RESULTS

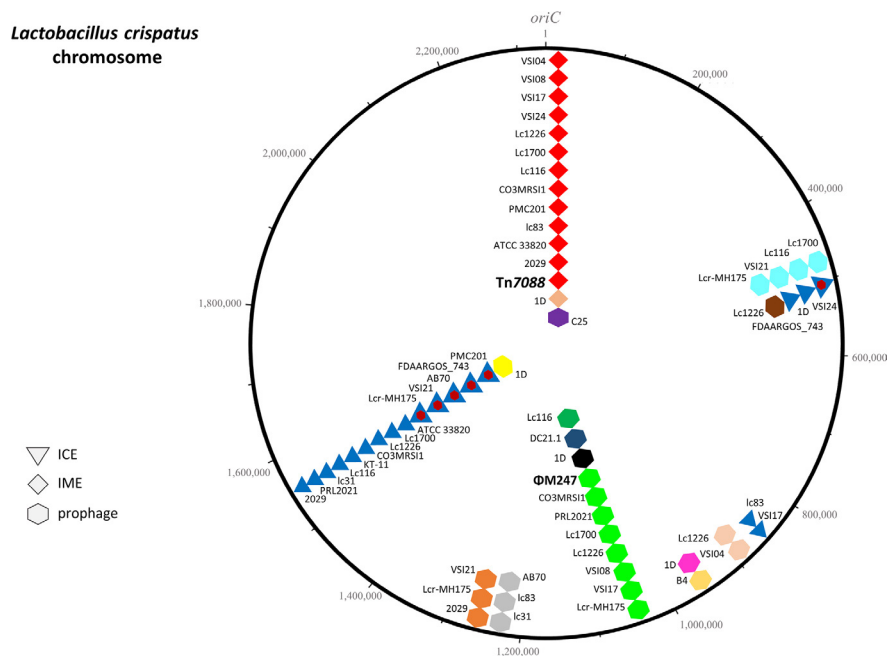
### The M247 genome

Sequence analysis showed that the M247 genome is organized in one circular chromosome 2336109 bp in length, with an average GC content of 37.04% (Fig. 1). The genome contains 2303 ORFs which are equally distributed on both strands (1161



**Fig. 1.** Circular representation of the *L. crispatus* M247 genome. Circle 1 (outer) and circle 2 show the predicted coding regions located on the plus and minus strands, respectively. S-layer, aggregation promoting factor (*apf*) and EPS biosynthesis (*eps*) loci are reported as black blocks. The third circle represents the mobilome including the 14105 bp Tn7088 IME (red block), the 42510 bp  $\Phi$ M247 prophage (orange block), the CRISPR-Cas locus (blue block) and the 226 ISs (green ticks). The fourth and the fifth circles show GC content and GC skew, respectively. The innermost circle indicates RNA genes. The image was created using Artemis DNA-Plotter (v17.0.1).

on sense and 1142 on antisense). An annotation with prediction of a biological function was possible for 1962 ORFs. rRNA genes are grouped into four rRNA operons, 28 out of the 65 tRNA genes are not adjacent to rRNA operons, and three structural RNAs are also present: (i) tRNA-like/mRNA-like RNA, (ii) signal recognition particle RNA and (iii) ribonuclease P RNA. The M247 genome contains a single copy of the *apf* gene [18], spanning nucleotides 1837153–1837824, while the already described *S-layer* locus (GenBank no. AY941197) spans nucleotides 197946–209531, and includes the paralogous genes *slpA* and *slpB* transcribed in opposite directions and spaced by a 4624 bp fragment. A 79695 bp EPS biosynthesis gene cluster, at nucleotides 1956291–2035986, contains 57 ORFs, 52 of which have the same direction of transcription and are potentially involved in EPS biosynthesis. Five conserved genes predicted to encode a transcriptional regulator, a polymerization and chain length determination protein, a tyrosine-protein kinase, a protein-tyrosine phosphatase, and the priming glycosyltransferase are present at the 5' end of the locus as reported in other *L. crispatus* strains [9]. Downstream of this conserved region, nine genes encoding (i) five glycosyltransferases, (ii) a UDP-*N*-acetylglucosamine, (iii) an acyltransferase, (iv) a polysaccharide polymerase and (v) a flippase are probably involved in the synthesis of EPS repeating units, polymerization and export. The remaining genes, as in other lactobacilli, probably contribute to EPS biosynthesis, for example through the generation of activated sugar precursors and the chemical decoration of the EPS [38]. A search carried out on the 23 publicly available *L. crispatus* complete genomes showed that the *L. crispatus* mobilome includes, in addition to plasmids, ISs and CRISPRs, prophages, ICEs and IMEs which are currently not annotated (Fig. 2). Analysis indicates the presence of 10 chromosomal sites for the integration of these elements. The M247 mobilome accounts for 14% (328388 bp) of the whole genome and includes two novel genetic elements, IME Tn7088 (Fig. 3), and the siphovirus prophage  $\Phi$ M247 (Fig. 4), three CRISPRs and 226 ISs belonging to 14 different families (Table S4). No plasmids were detected. Three CRISPR loci were found in the M247 mobilome: (i) locus 1 contains five 23 bp direct repeats (DRs) interspersed by four spacers; (ii) locus 2 contains six 36 bp DRs interspersed by five spacers, and is paired to the CRISPR-associated (Cas) *cas9*, *cas1*, *cas2* and *csn2* genes constituting a type II-A CRISPR-Cas system [39] (Fig. 1); and (iii) locus 3 contains two 25 bp DRs interspersed by one spacer. Despite the M247 genome being rich in ISs, which make up 11.16% (260880 bp) of its length, only 10 ORFs are disrupted by ISs, of which two are located on Tn7088 and one on  $\Phi$ M247. The transposase gene of 26 ISs is truncated or presents a frameshift mutation (Table S4).



**Fig. 2.** Integration sites of MGEs in the *L. crispatus* chromosome. Analysis of the 23 *L. crispatus* complete genomes available in the NCBI database (last accessed July 2023) revealed that MGEs integrate at 10 chromosomal sites. The *L. crispatus* mobilome consists of one ICE, two IMEs, 13 different prophages, and a composite element constituted by a prophage inserted in an ICE. Bacterial chromosome is represented by a circle, and the origin of replication (*oriC*) is indicated; nucleotides reported on the map refer to the M247 strain used as a reference. Triangles, diamonds and hexagons on the circle indicate the insertion sites of ICEs, IMEs and prophages, respectively. Homologous elements are depicted with symbols of the same colour. IME Tn7088 and prophage  $\Phi$ M247 found in strain M247 are indicated with their names, while non-annotated elements are indicated with the names of the strain harbouring them. GenBank accession numbers of the genome sequences are reported in Table S2.



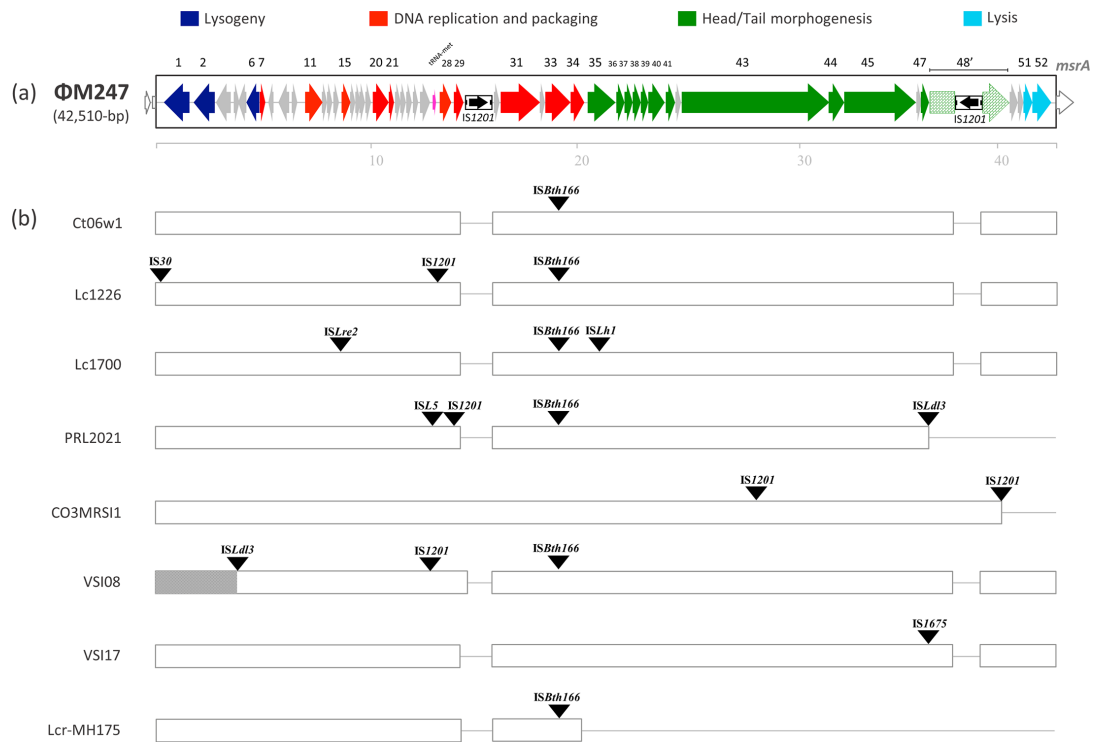
**Fig. 3.** (a) Structure of *L. crispatus* IME Tn7088 and (b) schematic comparison of Tn7088 with other *L. crispatus* Tn7088-like elements. (a) Tn7088 is a 14105 bp long IME, and contains 16 ORFs and two ISs. ORFs and their direction of transcription are represented by arrows, and annotated ORFs are indicated. The element shows a modular organization characterized by three modules: mobilization, integration/excision and adaptation (indicated by solid bars). ISs are reported as thinner, black, boxed arrows. ISLhe5 disrupts *orf14*, while IS1201 disrupts *orf16*. Chromosomal genes flanking Tn7088 are represented by white arrows. The scale is in kilobases. (b) IME Tn7088 is compared with 12 *L. crispatus* Tn7088-like elements present in complete genomes available in public databases. Disrupted ORFs are reported as pattern filled arrows. Non-annotated elements are indicated with the names of the strains harbouring them. Size varies from 11678 bp for the 2029 element to 22175 bp for the CO3MRSI1 element. Genome sequence GenBank accession numbers are reported in Table S2.

### The integrative and mobilizable element Tn7088

Tn7088 is 14105 bp long, spans nucleotides 21914–36018 of the M247 genome and has an average GC content of 30.97%. Tn7088 contains 16 ORFs, of which 15 have the same direction of transcription, two have a GTG alternative start codon and two are disrupted by different ISs (Fig. 3). Manual homology-based annotation attributed a putative function to 14 out of 16 ORFs (Table 1). Tn7088 has a typical IME modular organization [40] composed of (i) a mobilization module where *orf3* and *orf4* code for FtsK homologous proteins and *orf5* for a relaxase; (ii) an integration/excision module constituted by *orf7* and *orf8*, coding for a putative excisionase and an integrase, respectively; and (iii) an adaptation module (*orf9* to *orf16*) homologous to the listeriolysin S (*lIs*) locus of *Listeria monocytogenes* [41, 42] (Fig. 3a, Table 1). The 44 aa Orf9 is homologous to LlsA, the pro-peptide listeriolysin S family of thiazole/oxazole-modified microcins (TOMM) class I bacteriocin, and contains a 13 aa long serine/threonine/cysteine-rich motif. Orf10–Orf11 are homologous to LlsG–LlsH that constitute an ATP-binding cassette transporter, whereas Orf12 is homologous to the hypothetical protein LlsX. Orf13 is homologous to the LlsB dehydrogenase, while Orf14 and Orf15 are homologous to the LlsY and LlsD cyclodehydratases, respectively. Finally, Orf16 is homologous to the metalloprotease LlsP. The NCBI database of 95347 complete microbial genomes (accessed in July 2023) was interrogated using Tn7088 DNA sequence as a query. In 12 out of 23 *L. crispatus* complete genomes, Tn7088-like elements, ranging in length from 11678 bp for strain 2029 to 22175 bp for strain CO3MRSI1, were present. Sequence comparison showed that the mobilization and integration/excision modules are essentially conserved, while variability in the adaptation module is introduced by the integration of distinct ISs disrupting different ORFs (Fig. 3b). It is of note that only the Tn7088-like elements carried by strains 2029 and VSI24 contain an intact adaptation module. In the CO3MRSI1 element a copy of ISLd3 disrupts *orf3* in the mobilization module, whereas the adaptation module carries a DNA inversion spanning the 3' end of *orf12* to *orf16*.

### Prophage ΦM247

The 42510 bp ΦM247 prophage spans nucleotides 1001143–1043652 of the M247 genome, carries 52 ORFs, has a GC content of 35.18% and was classified as a siphovirus phage. Manual homology-based annotation with functional prediction of the



**Fig. 4.** (a) Structure of *L. crispatus* prophage  $\Phi$ M247 and (b) schematic comparison of  $\Phi$ M247 with other  $\Phi$ M247-like prophages. (a)  $\Phi$ M247 is a 42510 bp long prophage, and contains 52 ORFs and two copies of the IS1201 one of which disrupts *orf48* (pattern filled arrow). ORFs and their direction of transcription are represented by arrows, while annotated ORFs are indicated only by their numbers. ISs are reported as thinner, black, boxed arrows. Chromosomal genes flanking the  $\Phi$ M247 insertion site are represented by white arrows. The scale is in kilobases. (b) The 42510 bp  $\Phi$ M247 prophage is compared with the siphovirus prophage Ct06w1 (GenBank BK036340) and seven *L. crispatus*  $\Phi$ M247-like prophages present in complete genomes. For a better alignment of the sequences, the elements were devoid of additional integrated IS elements indicated by solid triangles, while the deletions are represented by lines. Size varies from 37638 bp for the CO3MRSI1 element to 44067 bp for the Lc1700 element. Non-annotated elements are indicated with the names of the strains harbouring them. The element of strain VSI08 contains a 3391 bp inversion at the 5' end depicted as a grey box. Genome sequence GenBank accession numbers are reported in Table S2.

hypothetical gene product was possible only for 27 out of 52 ORFs, including genes for phage structural proteins (*orf35* to *orf48*), DNA replication (*orf7*, *orf11*, *orf15*, *orf20*, *orf21*, *orf28*) and packaging (*orf29* to *orf34*), lysogeny (*orf1*, *orf2*, *orf6*) and lytic cycle-related proteins (*orf51*, *orf52*) (Table 2, Fig. 4a). Furthermore,  $\Phi$ M247 contains a methionyl-tRNA gene which might contribute to modulate the tRNA pools of the bacterium, improving the translation efficiency of viral genes [43, 44]. Two copies of IS1201 are also present, one of which disrupts *orf48*, predicted to encode the tail protein. Homology searches of the NCBI databases using nucleotide BLAST (accessed in July 2023) showed that  $\Phi$ M247 is highly homologous (99.98% identity) to the ct06w1 siphovirus prophage characterized during the metagenomic sequencing of a human vaginal fornix sample [45]. In addition,  $\Phi$ M247-like prophages are present only in seven *L. crispatus* complete genomes, namely Lc1226, PRL2021, Lc1700, CO3MRSI1, VSI08, VSI17 and Lcr-MH175, out of the 95347 complete microbial genomes of the NCBI database (accessed in July 2023). Variability in  $\Phi$ M247-like prophage sequences is due to the integration of distinct ISs in different positions except for strains CO3MRSI1, PRL2021 and Lcr-MH175 where a 3' end deletion of 3.4, 4.6 and 22.14 kb, respectively, was also present (Fig. 4b).

#### Quantification of Tn7088 and $\Phi$ M247 excised forms and reconstituted *attB* sites

PCR and sequencing analysis carried out on M247 genomic DNA showed that the Tn7088 IME is able to excise from the bacterial chromosome producing a circular form where the left and right ends are joined by a 90 bp sequence (*attTn*). Excision of the element reconstitutes the chromosomal 79 bp attachment site (*attB*). Upon integration into the M247 chromosome, Tn7088 is flanked by *attL* and *attR*, identical to *attTn* and *attB*, respectively. *attL-attTn* contain 11 nt insertions and 12 nt changes compared to *attR-attB* (Fig. S1). *attR-attB* contain the last 11 nt of the threonine-tRNA gene (LQF73\_00105). In liquid culture of M247, the circular form of Tn7088 was present at a concentration of  $4.32 \times 10^{-5} \pm 2.17 \times 10^{-7}$  copies per chromosome, whereas the reconstituted *attB* site was at  $6.65 \times 10^{-4} \pm 1.32 \times 10^{-6}$  copies per chromosome. Also, prophage  $\Phi$ M247 produces an excised form, where the left and right ends are joined by *attP*, restoring the *attB* insertion site. *attP* is 138 bp long and is identical to *attR*, while *attB* is 139 bp long and is identical to *attL*. *attR-attP* differ from *attL-attB* by 15 nt changes and 1 nt nucleotide deletion. *attL-attB* of

**Table 1.** Annotated ORFs of IME Tn7088

ORF (aa)*	Annotation and comments (reference)	Homologous protein		Pfam domain (aa)‡ [E value]	
		Protein ID/origin [E value]†	No. of amino acids identical/total		No. of amino acids similar/total
<i>orf1</i> (254)	Transcriptional regulator, putative [63]			HTH_3 (6–66) [1.0e–12]	
<i>orf3</i> (183)	Cell division protein FtsK, putative [64]				
<i>orf4</i> (264)	Cell division protein FtsK [64]			FtsK_SpoIIIE (2–109) [1.1e–05]	
<i>orf5</i> (273)	Relaxase [65]			Rep_trans (133–273) [1.1e–19]	
<i>orf7</i> (59)	Excisionase, putative				
<i>orf8</i> (409)	Tyrosine-type DNA integrase [66]			Phage_integrase (183–397) [9.8e–14]	
<i>orf9</i> (44)	Listeriolysin S family TOMM class I bacteriocin [54]	WP_071661762.1/ <i>L. monocytogenes</i> [3e–04]	12/26 (46%)	14/26 (53%)	
<i>orf10</i> (284)	ABC transporter: ATP-binding protein [54], similar to <i>L. monocytogenes</i> LlsG	WP_003724649.1/ <i>L. monocytogenes</i> [5e–69]	111/284 (39%)	182/284 (64%)	ABC_tran (21–150) [7.6e–22]
<i>orf11</i> (251)	ABC transporter: permease [54], similar to <i>L. monocytogenes</i> LlsH	WP_003730942.1/ <i>L. monocytogenes</i> [2e–70]	96/238 (40%)	160/238 (67%)	ABC2_membrane (5–212) [3.2e–12]
<i>orf13</i> (292)	SagB dehydrogenase family [54], similar to <i>L. monocytogenes</i> LlsB	WP_003730944.1/ <i>L. monocytogenes</i> [6e–95]	135/291 (46%)	192/291 (65%)	Nitroreductase (105–285) [5.8e–13]
<i>orf14'</i> (314)	Listeriolysin S biosynthesis cyclodehydratase [42], similar to <i>L. monocytogenes</i> LlsY, disrupted by IS <i>Lhe5</i>	WP_010958846.1/ <i>L. monocytogenes</i> [3e–72]	128/306 (42%)	192/306 (62%)	
<i>tmp</i> (285)	IS <i>Lhe5</i> transposase [67]	WP_012211839/ <i>L. helveticus</i> [0.0]	240/285 (84%)	259/285 (90%)	
<i>orf15</i> (438)	YcaO-like cyclodehydratase family [54], similar to <i>L. monocytogenes</i> LlsD	WP_003740559.1/ <i>L. monocytogenes</i> [3e–172]	240/440 (55%)	312/440 (70%)	YcaO (70–407) [1.5e–28]
<i>orf16'</i> (195)	CPBP intramembrane metalloprotease family [42], similar to <i>L. monocytogenes</i> , LlsP disrupted by IS1201	WP_226989712.1/ <i>L. monocytogenes</i> [7e–31]	56/147 (38%)	92/147 (62%)	CPBP (5–170) [0.045]
<i>tmp</i> (408)	IS1201 transposase [68]	P35880/ <i>L. helveticus</i> [0.0]	333/368 (90%)	352/368 (95%)	

\*The number of amino acids of the predicted protein is shown in parentheses, and disrupted ORFs are indicated with an apostrophe.

†Determined by compositional matrix adjustment.

‡Numbers in parentheses indicate the part of the predicted protein with homology to the Pfam domain.

ΦM247 includes 96 nt at the 5' end of the peptide-methionine (S)-S-oxide reductase encoding gene *msrA* (LQF73\_05265). The excised form of the phage genome was present at a concentration of  $3.90 \times 10^{-5} \pm 9.43 \times 10^{-6}$  copies per chromosome, whereas the reconstituted *attB* site was at  $2.48 \times 10^{-5} \pm 7.18 \times 10^{-7}$  copies per chromosome.

### ΦM247 excised forms are not enriched upon mitomycin C exposure

Liquid cultures of M247 were treated with mitomycin C at a final concentration of 200 and 400 ng ml<sup>-1</sup>, and culture supernatants were recovered and concentrated by sequential centrifugation and ultracentrifugation steps without filtering to minimize particle breakage. ΦM247 excised forms were detected and quantified with qPCR on concentrated supernatants. The number of excised forms per millilitre was essentially the same ( $P=0.09$ ) in phage preparations obtained from the mitomycin C-treated cultures (mean value  $2.14 \times 10^2 \pm 1.07 \times 10^2$ ) and the untreated control culture ( $7.66 \times 10^2 \pm 6.17 \times 10^2$ ), suggesting that, under our experimental conditions, mitomycin C does not induce excision of ΦM247.

## DISCUSSION

In this study, complete genome sequence analysis allowed us to define the M247 mobilome which includes: (i) the novel IME Tn7088, (ii) the novel ΦM247 prophage, (iii) three CRISPRs and (iv) 226 ISs. More than 10% of the M247 genome length is constituted by ISs, which only interrupt 10 predicted coding sequences, suggesting that there is a selective pressure against gene disruption. Interestingly, three out of 10 interrupted genes belong to ΦM247 and Tn7088, which are probably less subject to selective constraints. The presence of many ISs in the M247 genome possibly confers genome plasticity, favouring chromosomal rearrangements and duplications mediated by sequence homology [46]. ΦM247 was capable of excision from the bacterial chromosome, but could not be induced by mitomycin C. Some active prophages are not responsive to genotoxic drugs such as



**Table 2.** Annotated ORFs of  $\Phi$ M247

ORF (aa)*	Annotation and comments (reference)	Virfam homologous protein (identity) [E value or HHsearch probability]†	Homologous protein ID/origin No. of amino acids identical/total [E value]‡	Pfam domain (aa)§ [E value]
<i>orf1</i> (407)	Tyrosine-type DNA integrase [66]		DAW29718.1/ <i>Siphoviridae</i> sp. isolate ct06w1 407/407 (100%) [0.0]	Phage_integrase (180–373) [1.8e–25] Phage_int_SAM_5 (32–164) [1.8e–05]
<i>orf2</i> (334)	Abortive infection protein [69]		DAW29696.1/ <i>Siphoviridae</i> sp. isolate ct06w1 334/334 (100%) [0.0]	Abi_2 (31–239) [1.5e–36]
<i>orf6</i> (208)	Repressor protein CI [70]		DAW29717.1/ <i>Siphoviridae</i> sp. isolate ct06w 208/208 (100%) [9e–27]	Peptidase_S24 (86–202) [3.1e–26]
<i>orf7</i> (71)	Helix-turn-helix XRE-family-like protein [63]		DAW29695.1/ <i>Siphoviridae</i> sp. isolate ct06w 74/74 (100%) [9e–56]	
<i>orf11</i> (285)	DNA polymerase B		DAW29710.1/ <i>Siphoviridae</i> sp. isolate ct06w1 285/285 (100%) [1e–158]	HTH_36 (24–74) [4.4e–05]
<i>orf15</i> (145)	HNH endonuclease [71]		DAW29694.1/ <i>Siphoviridae</i> sp. isolate ct06w1 145/145 (100%) [3e–111]	HNH_3 (66–112) [7.5e–10]
<i>orf20</i> (248)	Phage antirepressor KilAC domain-containing protein [72]		DAW29715.1/ <i>Siphoviridae</i> sp. isolate ct06w1 247/248 (99%) [0.0]	AntA (17–85) [5.7e–19] ANT (124–236) [8.9e–32]
<i>orf21</i> (74)	Restriction alleviation protein, putative		DAW29742.1/ <i>Siphoviridae</i> sp. isolate ct06w1 73/73 (100%) [2e–55]	
<i>orf27</i>	tRNA-Met			
<i>orf28</i> (176)	HNH endonuclease [73]		DAW29714.1/ <i>Siphoviridae</i> sp. isolate ct06w1 176/176 (100%) [2e–138]	HNH (88–131) [8.1e–08]
<i>orf29</i> (156)	Phage terminase, small subunit [74]		WP_060463559.1/ <i>L.crispatus</i> 154/156 (99%) [1e–55]	Terminase_4 (29–140) [4.3e–16]
<i>tmp</i> (392)	IS1201 transposase [68]		P35880/ <i>L. helveticus</i> 333/368 (90%) [0.0]	
<i>orf31</i> (624)	Phage terminase, large subunit [74]	Phi adh phage TermL (60%) [0]	DAW29733.1/ <i>Siphoviridae</i> sp. isolate ct06w1 624/624 (100%) [0.0]	Terminase_1 (100–587) [1.5e–50]
<i>orf33</i> (392)	Phage portal protein [75]	Phi adh phage Portal (54%) [100%]	WP_060464314.1/ <i>L.crispatus</i> 392/392 (100%) [3e–66]	Phage_portal (47–354) [4.9e–39]
<i>orf34</i> (228)	ATP-dependent Clp protease [76]		DAW29711.1/ <i>Siphoviridae</i> sp. isolate ct06w1 228/228 (100%) [7e–174]	CLP_protease (32–175) [2.0e–33]
<i>orf35</i> (452)	Phage major capsid protein	DT1 phage MCP (36%) [100%]	DAW29706.1/ <i>Siphoviridae</i> sp. isolate ct06w1 452/452 (100%) [0.0]	Phage_capsid (132–421) [2.8e–17]
<i>orf36</i> (129)	Phage head-tail adaptor	Phi adh phage Ad1 (42%) [100%]	DAW29705.1/ <i>Siphoviridae</i> sp. isolate ct06w1 129/129 (100%) [2e–98]	
<i>orf37</i> (121)	Phage head closure knob	Phi adh phage Hc1 (34%) [100%]	DAW29704.1/ <i>Siphoviridae</i> sp. isolate ct06w1 121/121 (100%) [8e–94]	Phage_H_join (12–110) [0.00065]
<i>orf38</i> (132)	Phage type I neck protein	DT1 phage Ne1 (36%) [100%]	DAW29703.1/ <i>Siphoviridae</i> sp. isolate ct06w1 132/132 (100%) [4e–101]	HK97-gp10_like (6–98) [0.00044]
<i>orf39</i> (124)	Phage tail completion protein	Phi adh phage Tc1 (24%) [100%]	DAW29702.1/ <i>Siphoviridae</i> sp. isolate ct06w1 124/124 (100%) [4e–94]	
<i>orf40</i> (258)	Phage major tail protein [77]	DT1 phage MTP (30%) [3e–20]	DAW29701.1/ <i>Siphoviridae</i> sp. isolate ct06w1 258/258 (100%) [0.0]	Phage_TTP_1 (8–211) [1.7e–42]
<i>orf41</i> (137)	Phage tail assembly chaperone [77]		DAW29701.1/ <i>Siphoviridae</i> sp. isolate ct06w1 137/137 (100%) [4e–103]	Phage_TAC_3 (7–122) [1.5e–11]
<i>orf43</i> (2339)	Phage tail tape measure protein [77]		DAW29700.1/ <i>Siphoviridae</i> sp. isolate ct06w1 2338/2339 (99%) [0.0]	PhageMin_Tail (315–532) [1.1e–38]
<i>orf44</i> (253)	Phage distal tail protein [77]		DAW29699.1/ <i>Siphoviridae</i> sp. isolate ct06w1 252/253 (99%) [0.0]	
<i>orf45</i> (1135)	Phage tail protein [77]		DAW29698.1/ <i>Siphoviridae</i> sp. isolate ct06w1 1135/1135 (100%) [0.0]	Prophage_tail (73–441) (744–824) [7.2e–13]
<i>orf47</i> (127)	BppU family phage baseplate upper protein		WP_005719141.1/ <i>Lactobacillus crispatus</i> 127/127 (100%)	BppU_N (1–120) [5.6e–09]
<i>orf48'</i> (795)	Phage tail protein [78], disrupted by IS1201		DAW29725.1/ <i>Siphoviridae</i> sp. isolate ct06w1 795/795 (100%) [0.0]	Bppu_N (3–164) [3.7e–08] Lipase_GDSL (575–776) [8.4e–12]

Continued

Table 2. Continued

ORF (aa)*	Annotation and comments (reference)	Virfam homologous protein (identity) [E value or HHsearch probability]†	Homologous protein ID/origin No. of amino acids identical/total [E value]‡	Pfam domain (aa)§ [E value]
<i>tnp</i> (392)	IS1201 transposase [68]		P35880/ <i>L. helveticus</i> 333/368 (90%) [0.0]	
<i>orf51</i> (142)	Phage holin LLH family [79]		DAW29722.1/ <i>Siphoviridae</i> sp. isolate ct06w1 142/142 (100%) [8e-107]	Phage_holin_6_1 (3–110) [1.3e-05]
<i>orf52</i> (294)	Cpl1 lysin [80]		DAW29716.1/ <i>Siphoviridae</i> sp. isolate ct06w1 294/294 (100%) [0.0]	Glyco_hydro_25 (9–196) [1.8e-21]

\*The number of amino acids of the predicted protein is shown in parentheses, and disrupted ORFs are indicated with an apostrophe.

†Virfam performs PSI-blast on the ACLAME database, and the HHsearch probability is provided when PSI-blast search does not detect homology.

‡Determined by compositional matrix adjustment.

§Numbers in parentheses indicate the part of the predicted protein with homology to the Pfam domain.

mitomycin C, but are induced by other triggering factors [47–49]. Further investigations will clarify if also  $\Phi$ M247 is responsive to triggering factors other than mitomycin C. Genomic analysis showed that prophages homologous to  $\Phi$ M247 are present in seven other complete genomes belonging only to *L. crispatus*, suggesting host specificity of  $\Phi$ M247 [50]. Tn7088 contains three functional modules: (i) mobilization, (ii) integration/excision and (iii) adaptation. In the mobilization module *orf3*, *orf4* and *orf5* encode FtsK homologous proteins and a relaxase, respectively, which probably constitute the relaxosome [51]. The integration/excision module contains *orf7* and *orf8* coding for a putative excisionase and for a site-specific integrase that belongs to the family of tyrosine recombinases, as reported for many IMEs [52]. IMEs integrate into the bacterial genome and are capable of intracellular transposition to a new genomic site or intercellular transposition to a new host using *in trans* the conjugation machinery of a helper ICE or conjugative plasmid [40, 53]. The presence of circular intermediates of Tn7088 was detected by PCR analysis, proving that the integration/excision module is functional and the element has the potential to be horizontally transferred. Biosynthetic gene clusters for the synthesis of class I bacteriocins are widely distributed in both Gram-positive and Gram-negative bacteria [54]. These bacteriocin gene clusters were reported in *L. crispatus*, although the genomic location was not indicated [55, 56]. Recently, it has been demonstrated that listeriolysin S does not contribute to *L. monocytogenes* virulence and tissue damage, but instead targets exclusively prokaryotic cells, playing a role in modulation of the host microbiota [57–59] and acting via a cell-to-cell contact mechanism of action [60]. We found that a bacteriocin gene cluster homologous to that of the *L. monocytogenes* listeriolysin S constitutes the adaptation module of the IME Tn7088 and Tn7088-like elements, and possibly mediates an interaction with other bacteria of the microbiota. To the best of our knowledge, Tn7088 is the first example of a mobile element containing a predicted biosynthetic gene cluster for a class I bacteriocin in *L. crispatus*. The Tn7088 adaptation module contains an *orf* coding for a bacteriocin pro-peptide carrying a serine/threonine/cysteine-rich motif which probably acts as target site in post-translational modifications [61] and *orfs* coding for all enzymes required for post-translational modifications of the bacteriocin pro-peptide [62]. However, the disruption of various ORFs of Tn7088 and Tn7088-like elements by integration of ISs leaves an open question regarding whether the adaptation module does or does not encode for a functional bacteriocin. Further investigations are needed to elucidate the contribution of Tn7088 to the niche-adaptive traits of the M247 host.

#### Funding information

This work was supported in part by the Italian Ministry of Education, University and Research (MIUR-Italy) under grant number 20177J5Y3P (call 'Progetti di Ricerca di Rilevante Interesse Nazionale – Bando 2017'), and in part by EU funding within the MUR PNRR Extended Partnership Initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT, 'One Health Basic and Translational Research Actions addressing Unmet Needs on Emerging Infectious Diseases'). L.C. holds a junior lecturer position funded by the PNRR Extended Partnership initiative on Emerging Infectious Diseases.

#### Acknowledgements

We are grateful to Nagaia Ciacci for helpful discussions. Illumina genome sequencing was provided by MicrobesNG (<https://microbesng.com>).

#### Author contributions

This study was conceptualized by L.C., F.S., F.I. and G.P. The M247 strain was supplied by L.M. L.C. performed sequencing and computational analysis with contributions from F.S. Formal genomic data analysis was undertaken by L.C., M.T., E.L. and F.S. The original draft was written by L.C. and F.I. while L.C., F.S., M.T., E.L., L.M., F.I., and G.P. revised the final manuscript. G.P. supervised the study.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### References

- Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S. *Lactobacillus* species as biomarkers and agents that can promote various aspects of vaginal health. *Front Physiol* 2015;6:81.
- Walter J. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol* 2008;74:4985–4996.

3. Antonio MAD, Meyn LA, Murray PJ, Busse B, Hillier SL. Vaginal colonization by probiotic *Lactobacillus crispatus* CTV-05 is decreased by sexual activity and endogenous lactobacilli. *J Infect Dis* 2009;199:1506–1513.
4. De Leo V, Lazzeri E, Governini L, Cuppone AM, Colombini L, et al. Vaginal colonization of women after oral administration of *Lactobacillus crispatus* strain NTCVAG04 from the human microbiota. *Minerva Obstet Gynecol* 2023;75:432–439.
5. Castro-Bravo N, Wells JM, Margolles A, Ruas-Madiedo P. Interactions of surface exopolysaccharides from *Bifidobacterium* and *Lactobacillus* within the intestinal environment. *Front Microbiol* 2018;9:2426.
6. Donnarumma G, Molinaro A, Cimini D, De Castro C, Valli V, et al. *Lactobacillus crispatus* L1: high cell density cultivation and exopolysaccharide structure characterization to highlight potentially beneficial effects against vaginal pathogens. *BMC Microbiol* 2014;14:14.
7. Heilbronner S, Krismer B, Brötz-Oesterhelt H, Peschel A. The microbiome-shaping roles of bacteriocins. *Nat Rev Microbiol* 2021;19:726–739.
8. Stoyancheva G, Marzotto M, Dellaglio F, Torriani S. Bacteriocin production and gene sequencing analysis from vaginal *Lactobacillus* strains. *Arch Microbiol* 2014;196:645–653.
9. Ojala T, Kankainen M, Castro J, Cerca N, Edelman S, et al. Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. *BMC Genomics* 2014;15:1070.
10. Pan M, Hidalgo-Cantabrana C, Barrangou R. Host and body site-specific adaptation of *Lactobacillus crispatus* genomes. *NAR Genom Bioinform* 2020;2:lqaa001.
11. van der Veer C, Hertzberger RY, Bruisten SM, Tytgat HLP, Swanenburg J, et al. Comparative genomics of human *Lactobacillus crispatus* isolates reveals genes for glycosylation and glycogen degradation: implications for in vivo dominance of the vaginal microbiota. *Microbiome* 2019;7:49.
12. Damelin LH, Paximadis M, Mavri-Damelin D, Birkhead M, Lewis DA, et al. Identification of predominant culturable vaginal *Lactobacillus* species and associated bacteriophages from women with and without vaginal discharge syndrome in South Africa. *J Med Microbiol* 2011;60:180–183.
13. Kiliç AO, Pavlova SI, Alpays S, Kiliç SS, Tao L. Comparative study of vaginal *Lactobacillus* phages isolated from women in the United States and Turkey: prevalence, morphology, host range, and DNA homology. *Clin Diagn Lab Immunol* 2001;8:31–39.
14. Cesena C, Morelli L, Alander M, Siljander T, Tuomola E, et al. *Lactobacillus crispatus* and its nonaggregating mutant in human colonization trials. *J Dairy Sci* 2001;84:1001–1010.
15. Hynönen U, Palva A. *Lactobacillus* surface layer proteins: structure, function and applications. *Appl Microbiol Biotechnol* 2013;97:5225–5243.
16. Kirjavainen PV, Ouwehand AC, Isolauri E, Salminen SJ. The ability of probiotic bacteria to bind to human intestinal mucus. *FEMS Microbiol Lett* 1998;167:185–189.
17. Marcotte H, Ferrari S, Cesena C, Hammarström L, Morelli L, et al. The aggregation-promoting factor of *Lactobacillus crispatus* M247 and its genetic locus. *J Appl Microbiol* 2004;97:749–756.
18. Siciliano RA, Cacace G, Mazzeo MF, Morelli L, Elli M, et al. Proteomic investigation of the aggregation phenomenon in *Lactobacillus crispatus*. *Biochim Biophys Acta* 2008;1784:335–342.
19. Bertuccioli A, Cardinali M, Zonzini G, Cazzaniga M, Di Pierro F. *Lactobacillus crispatus* M247: characteristics of a precision probiotic instrument for gynecological and urinary well-being. *Microbiol Res* 2022;13:963–971.
20. Castagliuolo I, Galeazzi F, Ferrari S, Elli M, Brun P, et al. Beneficial effect of auto-aggregating *Lactobacillus crispatus* on experimentally induced colitis in mice. *FEMS Immunol Med Microbiol* 2005;43:197–204.
21. Voltan S, Castagliuolo I, Elli M, Longo S, Brun P, et al. Aggregating phenotype in *Lactobacillus crispatus* determines intestinal colonization and TLR2 and TLR4 modulation in murine colonic mucosa. *Clin Vaccine Immunol* 2007;14:1138–1148.
22. Voltan S, Martines D, Elli M, Brun P, Longo S, et al. *Lactobacillus crispatus* M247-derived H2O2 acts as a signal transducing molecule activating peroxisome proliferator activated receptor-gamma in the intestinal mucosa. *Gastroenterology* 2008;135:1216–1227.
23. Pierro FD, Bertuccioli A, Cattivelli D, Soldi S, Elli M. *Lactobacillus crispatus* M247: A Possible Tool to Counteract CST IV. 2018, pp. 169–172.
24. Di Pierro F, Criscuolo AA, Dei Giudici A, Senatori R, Sesti F, et al. Oral administration of *Lactobacillus crispatus* M247 to papillomavirus-infected women: results of a preliminary, uncontrolled, open trial. *Minerva Obstet Gynecol* 2021;73:621–631.
25. Pinzauti D, Iannelli F, Pozzi G, Santoro F. DNA isolation methods for Nanopore sequencing of the *Streptococcus mitis* genome. *Microb Genom* 2022;8:000764.
26. Teodori L, Colombini L, Cuppone AM, Lazzeri E, Pinzauti D, et al. Complete genome sequence of *Lactobacillus crispatus* type strain ATCC 33820. *Microbiol Resour Announc* 2021;10:e0063421.
27. Cuppone AM, Colombini L, Fox V, Pinzauti D, Santoro F, et al. Complete genome sequence of *Streptococcus pneumoniae* strain Rx1, a Hex mismatch repair-deficient standard transformation recipient. *Microbiol Resour Announc* 2021;10:e0079921.
28. Pinzauti D, De Giorgi S, Fox V, Lazzeri E, Messina G, et al. Complete genome sequences of *Mycobacterium chimaera* strains 850 and 852, isolated from heater-cooler unit water. *Microbiol Resour Announc* 2022;11:e0102121.
29. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 2016;44:6614–6624.
30. Colombini L, Cuppone AM, Tirziu M, Lazzeri E, Pozzi G, et al. The mobilome-enriched genome of the competence-deficient *Streptococcus pneumoniae* BM6001, the original host of integrative conjugative element Tn5253, is phylogenetically distinct from historical pneumococcal genomes. *Microorganisms* 2023;11:1646.
31. Santoro F, Pastore G, Fox V, Petit M-A, Iannelli F, et al. *Streptococcus pyogenes*  $\Phi$ 1207.3 is a temperate bacteriophage carrying the macrolide resistance gene pair *mef(A)*-*msr(D)* and capable of lysogenizing different streptococci. *Microbiol Spectr* 2023;11:e0421122.
32. Santoro F, Oggioni MR, Pozzi G, Iannelli F. Nucleotide sequence and functional analysis of the *tet(M)*-carrying conjugative transposon Tn5251 of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* 2010;308:150–158.
33. Iannelli F, Santagati M, Santoro F, Oggioni MR, Stefani S, et al. Nucleotide sequence of conjugative prophage  $\Phi$ 1207.3 (formerly Tn1207.3) carrying the *mef(A)*/*msr(D)* genes for efflux resistance to macrolides in *Streptococcus pyogenes*. *Front Microbiol* 2014;5:687.
34. Iannelli F, Santoro F, Oggioni MR, Pozzi G. Nucleotide sequence analysis of integrative conjugative element Tn5253 of *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* 2014;58:1235–1239.
35. Iannelli F, Giunti L, Pozzi G. Direct sequencing of long polymerase chain reaction fragments. *Mol Biotechnol* 1998;10:183–185.
36. Santoro F, Romeo A, Pozzi G, Iannelli F. Excision and circularization of integrative conjugative element Tn5253 of *Streptococcus pneumoniae*. *Front Microbiol* 2018;9:1779.
37. Santoro F, Fox V, Romeo A, Lazzeri E, Pozzi G, et al. Chromosomal integration of Tn5253 occurs downstream of a conserved 11-bp sequence of the *rbgA* gene in *Streptococcus pneumoniae* and in all the other known hosts of this integrative conjugative element (ICE). *Mob DNA* 2021;12:25.
38. Deo D, Davray D, Kulkarni R. A diverse repertoire of exopolysaccharide biosynthesis gene clusters in *Lactobacillus* revealed by comparative analysis in 106 sequenced genomes. *Microorganisms* 2019;7:444.

39. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 2011;9:467–477.
40. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev* 2014;38:720–760.
41. Cotter PD, Draper LA, Lawton EM, Daly KM, Groeger DS, et al. Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I *Listeria monocytogenes*. *PLoS Pathog* 2008;4:e1000144.
42. Lee S. Bacteriocins of *Listeria monocytogenes* and their potential as a virulence factor. *Toxins* 2020;12:103.
43. Limor-Waisberg K, Carmi A, Scherz A, Pilpel Y, Furman I. Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies. *Nucleic Acids Res* 2011;39:6016–6028.
44. Alamos P, Tello M, Bustamante P, Gutiérrez F, Shmaryahu A, et al. Functionality of tRNAs encoded in a mobile genetic element from an acidophilic bacterium. *RNA Biol* 2018;15:518–527.
45. Tisza MJ, Buck CB. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc Natl Acad Sci U S A* 2021;118:e2023202118.
46. Ring N, Abrahams JS, Jain M, Olsen H, Preston A, et al. Resolving the complex *Bordetella pertussis* genome using barcoded nanopore sequencing. *Microb Genom* 2018;4:e000234.
47. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol* 2015;197:410–419.
48. Cornuault JK, Moncaut E, Loux V, Mathieu A, Sokol H, et al. The enemy from within: a prophage of *Roseburia intestinalis* systematically turns lytic in the mouse gut, driving bacterial adaptation by CRISPR spacer acquisition. *ISME J* 2020;14:771–787.
49. Henrot C, Petit M-A. Signals triggering prophage induction in the gut microbiota. *Mol Microbiol* 2022;118:494–502.
50. Koskella B, Meaden S. Understanding bacteriophage specificity in natural microbial communities. *Viruses* 2013;5:806–823.
51. Shoemaker NB, Wang GR, Salyers AA. Multiple gene products and sequences required for excision of the mobilizable integrated *Bacteroides* element NBU1. *J Bacteriol* 2000;182:928–936.
52. Guédon G, Libante V, Coluzzi C, Payot S, Leblond-Bourget N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes* 2017;8:337.
53. Burrus V, Pavlovic G, Decaris B, Guédon G. The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* 2002;48:77–97.
54. Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, et al. Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc Natl Acad Sci USA* 2008;105:5879–5884.
55. Collins FWJ, O'Connor PM, O'Sullivan O, Gómez-Sala B, Rea MC, et al. Bacteriocin gene-trait matching across the complete *Lactobacillus* pan-genome. *Sci Rep* 2017;7:3481.
56. Fontana F, Alessandri G, Lugli GA, Mancabelli L, Longhi G, et al. Probiogenomics analysis of 97 *Lactobacillus crispatus* strains as a tool for the identification of promising next-generation probiotics. *Microorganisms* 2020;9:73.
57. Mohammadzadeh R, Azadegan A, Kalani BS. Listeriolysin S may inhibit the anti-listerial properties of *Lactobacillus plantarum*. *Microb Pathog* 2019;137:103744.
58. Quereda JJ, Dussurget O, Nahori M-A, Ghozlane A, Volant S, et al. Bacteriocin from epidemic *Listeria* strains alters the host intestinal microbiota to favor infection. *Proc Natl Acad Sci U S A* 2016;113:5706–5711.
59. Quereda JJ, Meza-Torres J, Cossart P, Pizarro-Cerdá J. Listeriolysin S: a bacteriocin from epidemic *Listeria monocytogenes* strains that targets the gut microbiota. *Gut Microbes* 2017;8:384–391.
60. Meza-Torres J, Lelek M, Quereda JJ, Sachse M, Manina G, et al. Listeriolysin S: a bacteriocin from *Listeria monocytogenes* that induces membrane permeabilization in a contact-dependent manner. *Proc Natl Acad Sci U S A* 2021;118:e2108155118.
61. McAuliffe O, Ross RP, Hill C. Lantibiotics: structure, biosynthesis and mode of action. *FEMS Microbiol Rev* 2001;25:285–308.
62. Melby JO, Li X, Mitchell DA. Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* 2014;53:413–422.
63. Brennan RG, Matthews BW. The helix-turn-helix DNA binding motif. *J Biol Chem* 1989;264:1903–1906.
64. Begg KJ, Dewar SJ, Donachie WD. A new *Escherichia coli* cell division gene, *ftsK*. *J Bacteriol* 1995;177:6211–6222.
65. Balson DF, Shaw WV. Nucleotide sequence of the *rep* gene of staphylococcal plasmid pCW7. *Plasmid* 1990;24:74–80.
66. Kwon HJ, Tirumalai R, Landy A, Ellenberger T. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 1997;276:126–131.
67. Callanan M, Kaleta P, O'Callaghan J, O'Sullivan O, Jordan K, et al. Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. *J Bacteriol* 2008;190:727–735.
68. de Los Reyes-Gavilán CG, Limsowtin GK, Tailliez P, Séchaud L, Accolas JP. A *Lactobacillus helveticus*-specific DNA probe detects restriction fragment length polymorphisms in this species. *Appl Environ Microbiol* 1992;58:3429–3432.
69. Garvey P, Fitzgerald GF, Hill C. Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Appl Environ Microbiol* 1995;61:4321–4328.
70. Paetzl M, Dalbey RE, Strynadka NCJ. Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature* 1998;396:186–190.
71. Krishna SS, Majumdar I, Grishin NV. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res* 2003;31:532–550.
72. Sandt CH, Hopper JE, Hill CW. Activation of prophage eib genes for immunoglobulin-binding proteins by genes from the IbrAB genetic island of *Escherichia coli* ECOR-9. *J Bacteriol* 2002;184:3640–3648.
73. Edgell DR. Selfish DNA: homing endonucleases find a home. *Curr Biol* 2009;19:R115–7.
74. Schouler C, Ehrlich SD, Chopin MC. Sequence and organization of the lactococcal prolate-headed bIL67 phage genome. *Microbiology* 1994;140 (Pt 11):3061–3069.
75. Moore SD, Prevelige PEJ. DNA packaging: a new class of molecular motors. *Curr Biol* 2002;12:R96–8.
76. Wang J, Hartling JA, Flanagan JM. The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell* 1997;91:447–456.
77. Pell LG, Cumby N, Clark TE, Tuite A, Battaile KP, et al. A conserved spiral structure for highly diverged phage tail assembly chaperones. *J Mol Biol* 2013;425:2436–2449.
78. Upton C, Buckley JT. A new family of lipolytic enzymes? *Trends Biochem Sci* 1995;20:178–179.
79. Gindreau E, Lonvaud-Funel A. Molecular analysis of the region encoding the lytic system from *Oenococcus oeni* temperate bacteriophage phi 10MC. *FEMS Microbiol Lett* 1999;171:231–238.
80. Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, et al. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA* 1995;92:7090–7094.