Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium Streptococcus thermophilus

Peer-reviewed author version

McDonnell, Brian; Mahony, Jennifer; Neve, Horst; Hanemaaijer, Laurens; NOBEN, Jean-Paul; Kouwen, Thijs & van Sinderen, Douwe (2016) Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium Streptococcus thermophilus. In: APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 82(17), p. 5153-5165.

DOI: 10.1128/AEM.00835-16
Handle: http://hdl.handle.net/1942/22500

Identification and analysis of a novel group of bacteriophages infecting the lactic acid bacterium *Streptococcus thermophilus*

Brian McDonnell[a], Jennifer Mahony[a], Horst Neve[b], Laurens Hanemaaijer[c], Jean-Paul Noben[d], Thijs Kouwen[c], and *Douwe van Sinderen[a, e].

[a] School of Microbiology, University College Cork, Western Road, Cork, Ireland.

[b] Department of Microbiology and Biotechnology, Max Rubner Institut, Kiel, Germany.

[c] DSM Biotechnology Centre, Delft, The Netherlands.

[d] Biomedical Research Institute, Hasselt University, B-3590 Diepenbeek, Belgium.

[e] APC Microbiome Institute, University College Cork, Western Road, Cork, Ireland.

Running title: Novel phage group infecting *Streptococcus thermophilus*.

Key words: *Lactococcus lactis*, dairy, phage, recombination, antireceptor, receptor binding protein, adsorption, proteome.

*Corresponding author: d.vansinderen@ucc.ie

Abstract

We present the complete genome sequences of four members of a novel group of phages infecting *Streptococcus thermophilus*, designated here as the 987 group. Members of this phage group appear to have resulted from genetic exchange events, as evidenced by their 'hybrid' genomic architecture, exhibiting DNA sequence relatedness to the morphogenesis module of certain P335 group *Lactococcus lactis* phages and to the replication modules of *S. thermophilus* phages. All four identified members of the 987 phage group were shown to elicit adsorption affinity to both their cognate *S. thermophilus* hosts as well as to a particular *L. lactis* starter strain. The receptor binding protein of one of these phages (as a representative of this novel group) was defined using an adsorption inhibition assay. The emergence of a novel phage group infecting *S. thermophilus* phage highlights the continuous need for phage monitoring and development of new phage control measures.

Importance:

Phage predation of *S. thermophilus* is an important issue for the dairy industry, where viral contamination can lead to fermentation inefficiency or complete fermentation failure. Genome information (and phage-host interaction studies) of *S. thermophilus* phages, particularly those emerging in the marketplace, are an important part of limiting the detrimental impact of these viruses in the dairy environment.

Introduction

Streptococcus thermophilus is a globally employed dairy bacterium used in the production of a variety of cheeses and yoghurt. Having been safely consumed by humans for millennia, this bacterium is now a mainstay of the dairy industry due its favourable acidification and texturing properties (1, 2). Despite advances in the available knowledge regarding dairy phage containment (3, 4), and *S. thermophilus* phage genetics and biology (5, 6), contamination of dairy production lines by *S. thermophilus*-infecting (bacterio)phages remains a persistent problem (for a review, see 7).

Classification of phages of *S. thermophilus* (reviewed by J. Mahony & D. van Sinderen [8]) has long been based on (i) morphology, i.e. as *Siphoviridae*, corresponding to group B as defined by D. E. Bradley (9), and (ii) a combination of the mode of DNA packaging (i.e. *cos*- or *pac*-site containing) and major structural protein content (10). A variable genomic region thought to be (at least in part) responsible for host determination (VR2 region; 11) can also be used to categorize the majority of isolated *S. thermophilus* phages (12). More recently, however, a morphologically distinct and genetically divergent *S. thermophilus* phage named 5093, containing neither *cos*/*pac*-defining structural elements, nor a confirmed antireceptor-encoding gene, was described (13), prompting the creation of a third *S. thermophilus* phage group (henceforth termed the '5093 group'). The genomic content of phage 5093 (containing several genes of non-dairy streptococcal phage origin) highlights the genetic plasticity of *S. thermophilus* phages, thus explaining the appearance of such diverse phage lineages.

A total of thirteen complete genome sequences of *S. thermophilus*-infecting phages have been published to date; with a large degree of conservation observed within the defined groupings. Phage groups have been defined as follows: (i) *cos* site-containing, with members

Sfi19 and Sfi21 (lytic and temperate; 14), DT1 (lytic; 15), 7201 (lytic; 16), and Abc2 (lytic; 6); (ii) *pac* site-containing phages O1205 (temperate; 17), Sfi11 (lytic; 18), 2972 (lytic; 19), 858 (lytic; 20), ALQ13.2 (lytic; 6), and TP-J34 & TP-778L (temperate; 5); and (iii) the 5093 group archetype 5093 (lytic; 13).

Whole genome sequencing of *S. thermophilus*-infecting phages has enabled their genome-wide, nucleotide-level comparison and elucidation of their putative mechanisms of evolution. It was postulated (18) that the main modes of *S. thermophilus* phage evolution are represented by the rearrangement (or recombination) of discrete genomic modules, as well as by insertions, deletions and point mutations – of which the latter is likely to function as a means to evade active CRISPR systems of their hosts (20). Consistent monitoring of phage populations in dairy plants in this manner is necessary to ensure that adequate knowledge-based rotational schemes are in place so as to avoid fermentation inconsistencies, or even complete failure. This must initially include host sensitivity profiling and phage typing studies, yet may be extended to whole phage genome sequencing in the case of newly emerging groups and/or persistent or highly virulent phages.

Here, we present the complete genome sequences of four novel phages capable of infecting *S. thermophilus* ST64987, an industrial dairy starter strain. The 987 group phages were categorized as novel based on their recalcitrance to typing using a previously designed multiplex PCR protocol, their distinct morphology, and finally their genetic content which differed from previously described groups of *S. thermophilus* phages. Comparative genomic analysis was performed on all four phages. The structural protein complement of one representative phage of this group was confirmed by mass spectrometry. The phages were further characterized by microscopic analysis and adsorption analyses, and the antireceptor of one phage was defined (as a representative) using an adsorption inhibition assay.

Materials & methods

Bacteriophage isolation, propagation, enumeration & storage

Bacterial strains were routinely grown from single colonies or reconstituted skimmed milk (RSM) stocks overnight at 30 °C (*L. lactis*) or 42 °C (*S. thermophilus*) in M17 Broth (Oxoid, Hampshire, U.K.) containing 0.5 % glucose (Sigma-Aldrich, St. Louis, MO, U.S.A; GM17; *L. lactis*) or lactose (Sigma-Aldrich; LM17; *S. thermophilus*). Phage enumeration was performed based on standard spot or plaque assay methods (21) in which LM17 broth was supplemented with 0.25 % glycine (Oxoid), 10 mM $CaCl_2$ (Oxoid) and either 10 g/L (solid agar base) or 4 g/L (semi-solid overlay) technical agar (Merck, Darmstadt, Germany). Industrially derived cheese whey samples from dairy plants producing fermented milk products (such as cheeses and yoghurts) were obtained (stored at -20 °C) and analysed for the presence of phages against *S. thermophilus* using the spot and plaque assay methods mentioned above. These samples now form part of the DSM phage collection (Delft, The Netherlands). Single plaque isolates were then propagated following the method of S. Moineau, *et al*. (22) in LM17 at 42 °C. The lysed culture was filtered (0.45 μm; Sarstedt, Nümbrecht, Germany) and stored at 4 °C for use in subsequent assays. Single plaque isolation and propagation was performed at least twice to ensure the purity of phage preparations.

Bacteriophage purification & DNA preparation

Individual phages were propagated in a 2 L volume before concentration by PEG_{8000} (Sigma-Aldrich) precipitation and purification using a discontinuous cesium chloride (Sigma-

Aldrich) block gradient as described by Sambrook & Fritsch (23), using a Beckman 50 Ti rotor (Beckman Coulter, Brea, CA, U.S.A). Phage DNA was prepared using a method adapted from S. Moineau, *et al*. (22) and Sambrook & Fritsch (23). Briefly, 20 µl proteinase K (20 mg/ml; Fisher Scientific, Waltham, MA, U.S.A.) was added to 500 µl of CsCl purified phage and the mixture heated at 56 °C for 20 minutes. Sodium dodecyl sulphate solution (SDS; Sigma Aldrich) was then added to a final concentration of 1.5 % before heating at 65 °C for 30 min. Potassium acetate was added to a final concentration of 1 M and the mixture placed on ice for 30 min. Centrifugation at 13,200 x *g* for 10 min was followed by two phenol:chloroform:isoamyl alcohol (25:24:1; Sigma Aldrich) extractions and the addition of 0.1 volume of 3 M sodium acetate (pH 4.8; Lancaster Synthesis, Ward Hill, MA, U.S.A.) and 2.5 volumes of ice cold 96 % ethanol. Precipitated phage DNA was pelleted at 21,000 x *g* for 15 min and resuspended in 50 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA [Sigma-Aldrich]; pH 7.5). Phage DNA was visualised on 1 % agarose (Sigma-aldrich) gels stained with Midori Green Advance DNA stain (Nippon Genetics Europe GmbH, Dueren, Germany) using the method of Sambrook & Fritsch (23).

DNA sequencing & in silico analysis

Approximately 20 µg phage DNA was extracted and verified by nanodrop (Nanodrop 2000, Thermo Scientific) quantification. Confirmatory molecular ID tests were also conducted on the DNA extract prior to shipment to the contract sequencing facility (Macrogen Inc., Geumcheon-gu, Seoul, South Korea). At least 100-fold sequencing coverage was obtained using pyrosequencing technology on a 454 FLX instrument. The individual sequence files generated by the 454 FLX instrument were assembled with GSassembler (454 Lifesciences, Branford, CT, U.S.A.) to generate a consensus sequence. Quality improvement of the genome

sequence involved Sanger sequencing (Eurofins MWG, Ebersberg, Germany) of at least three PCR products across each entire genome to ensure correct assembly, double stranding and the resolution of any remaining base-conflicts occurring within homopolymer tracts. Genomes were annotated using a Heuristic approach (Genemark; [24]) and manually using the Basic Local Alignment Search Tool (NCBI; [25]). Conserved protein domains (where relevant) were detected using Pfam (26), HHpred (27) and/or CDD (28). Complete genomes were visualised using Artemis (29). Phylogenetic trees were generated using the FigTree tool (http://tree.bio.ed.ac.uk/software/figtree/).

Electron microscopic analysis

Cesium chloride phage samples were dialysed (as above) and subjected to further purification by ultracentrifugation (and dialysis) according to the method of M. Briggiler Marco, *et al*. (30), using a Beckman VTi 65.2 rotor (Beckman Coulter). Dialysis was performed twice for 24 hr and 45 min, respectively, against 2 L phage buffer (0.05 M Tris-Hcl [pH 7.5], 0.1 M NaCl, 8 mM $MgSO_4$; 30). Electron microscopy was performed as previously described by E. Casey and colleagues (31).

Structural protein identification

Phage protein extraction (including methanol-chloroform extraction), SDS-PAGE visualisation and preparation of phage structural protein samples were performed as described by E. Casey, *et al*. (31). Electrospray ionization-tandem mass spectrometry (ESI-MS/MS) was performed as previously described (32, 33). Coverage levels of at least two

unique peptides for each structural protein, or 5 % of the total protein length, were used as cut-off values when identifying gene products as components of the viral particle (31).

Adsorption assays

Quantification of phage adsorption to bacterial strains was determined using a method adapted from P. Garvey, *et al*. (34). 10 ml LM17 or GM17 broth was inoculated (2-4 %, strain dependent) with the appropriate *S. thermophilus* (either ST64987 as sample strain or ST67368 as adsorption control) or *L. lactis* (LL64981 as sample strain) strain from a fresh overnight culture and grown at 42 °C or 30 °C, respectively, until the OD_{600nm} reached a value between 0.5 and 0.54. 700 µl of the growing culture was transferred to a microcentrifuge tube and centrifuged at 5000 x *g* for 10 min to pellet the cells. The supernatant was removed and the cells were resuspended in 700 µl of ¼ strength Ringers solution (Merck). An equal volume of the appropriate phage lysate (diluted to an approximate titer of 10^{5-6} pfu/ml) was added to the tube or to 700 µl 1/4 strength Ringer's solution (Merck), which served as a negative control. The mixture was incubated at 30 °C or 42 °C for 12 min, centrifuged at 15,000 x *g* for 3 min to remove bacterial cells before 200 µl of the residual phage-containing supernatant was removed for enumeration as described above. Calculation of adsorption levels (as a percentage of total number of phages present) was performed as follows: ([Control phage titre – Free phage titre in supernatant] / Control phage titre) × 100.

Antireceptor purification & adsorption inhibition assays

177 The protein product of *ORF19*$_{9871}$ (predicted to encode the phage antireceptor, termed here
178 the receptor binding protein or RBP$_{9871}$) was purified using a previously described method
179 (35). Briefly, the *ORF19*$_{9871}$ gene was amplified using Phusion polymerase (New England
180 Biolabs, Ipswich, MA, U.S.A.) and employing primers that incorporate a sequence encoding
181 an N-terminal His$_6$-purification tag and appropriate restriction enzyme sites (namely
182 RBP$_{9871}$F, 5'-
183 AGCAGCCCATGGCACACCATCACCATCACCATTCTTCTGGTGAACATAAGATAAT
184 TTTAAGT-3' and RBP$_{9871}$R, 5'-AGCAGCTCTAGATTAATATATACTTGGATATGA-3'),
185 and cloned behind the Nisin-inducible promoter of plasmid pNZ8048 (36). The ligation
186 mixture was dialysed against sterile distilled (sd) H_2O for 10 min and introduced into
187 electrocompetent *L. lactis* NZ9000 cells (36). Plasmid DNA was then extracted using a
188 GeneJet Plasmid Miniprep Kit (Thermo Scientific) and subjected to Sanger sequencing (as
189 above) to verify the integrity of the DNA sequence. For target protein induction, NZ9000
190 strains containing the required plasmid were grown to an OD_{600nm} of 0.2 prior to the addition
191 of nisin (10 ng/ml) using Nisaplin (Danisco, Copenhagen, Denmark). Growth was continued
192 for 3.5 hr prior to cell lysis and sonication as per B. Collins, *et al*. (35), with the following
193 modifications: the concentration of $CaCl_2$ (Sigma-Aldrich) in the lysis buffer (10 mM Tris,
194 300 mM NaCl, 10 mM $CaCl_2$, 25 mg/ml lysozyme [Sigma-Aldrich]; pH8) was increased to
195 50 mM, and a further 200 µl 1 M $CaCl_2$ was added to the lysed cells prior to sonication
196 (Soniprep 150; MSE, London, U.K.) cycles. Sonicated cells were then centrifuged and target
197 protein purification was performed using a Ni-nitrilotriacetic acid agarose (Qiagen, Hilden,
198 Germany) column (Bio-Rad, Hercules, CA, U.S.A.), using varying concentrations of
199 imidazole buffer (10 mM Tris-HCl, 50 mM $CaCl_2$, 300 mM NaCl, 50-200 mM imidazole; pH
200 7.5) according to the manufacturer's instructions. Protein fractions were visualised by
201 separation on a 12.5 % SDS-PAGE gel at 160 V for 90 minutes. Fractions containing bands

of the correct size with minimal contamination were dialysed against 100 ml protein buffer (as above) three times for 40 minutes each to remove remaining imidazole. Dialysed fractions were stored at 4 °C for use in subsequent adsorption inhibition assays.

Adsorption inhibition assays were performed as described by B. Collins, *et al*. (35), with the following modification: both the antireceptor incubation and phage adsorption temperatures were increased to 42 °C. Adsorption to wild-type and antireceptor-incubated cells was calculated as described above. Adsorption inhibition, expressed as a percentage of phage adsorption to wild type (WT) cells, was calculated as follows: ([% adsorption on WT – % adsorption on preincubated cells] / % adsorption on WT) x 100.

Nucleotide sequence accession numbers

Whole genome sequence data for phages 9871, 9872, 9873 and 9874 are available in the GenBank database under the following accession numbers: KU678389 (9871), KU678390 (9872), KU678391 (9873) and KU678392 (9874).

Results & Discussion

Isolation of phages

The bacterial strains and phage isolates (Table 1, an exception being NZ9000), which formed part of a larger industrial starter strain and phage sample collection, were initially subjected to a phage-host survey to determine the host ranges of isolated phages. This was followed (in the case of the phages) by further characterization and genome sequencing of a representative selection, the results of which will be published elsewhere. The phage isolates characterized as part of the current study, named 9871, 9872, 9873 and 9874 (together referred to here as 9871-4 or the 987 phage group), originated from distinct dairy fermentation samples from a range of geographical locations and time points, specifically: Portugal in 2008 (isolate 9871); Slovakia in 2008 (isolate 9872); U.K. in 2009 (isolate 9873); and Australia in 2010 (isolate 9874). Of 90 industrial strains tested, phages 9871-4 were found to infect just a single strain (named ST64987) with subsequent experiments revealing the ability of low level infection of a second strain (ST47795) by 9872, 9873 and 9874 only (data not shown). This observed narrow host range is typical of *S. thermophilus* phages (12, 37). The four phages were shown to reach a high titer during standard propagations (approximately 10^9 pfu/ml), and DNA could readily be extracted from both crude lysate and CsCl-purified preparations. However, a standard *cos*/*pac* phage typing PCR (38) on either lysate or DNA preparations repeatedly failed to yield a product (data not shown). For this reason, these phages were identified as phage isolates that potentially belong to a novel group, and they were therefore subjected to genome sequencing.

239 *Genome analysis*

240

241 *General characteristics*

242 The salient genome characteristics of phages 9871-4 are outlined in Table 2, with a detailed
243 list of top BLAST identities provided for phage 9871 (as a representative of the group, due to
244 overall conservation of the four genome structures) in Supplementary Table S1. Genome
245 sizes ranged from 32.6 – 33.1 kilo base pairs (kbp) - making these genomes the shortest thus
246 far described for *S. thermophilus* phages. Initial analysis of the DNA sequences revealed a
247 high level of nucleotide identity (greater than 90 % across approximately a third of the length
248 of their genomes) with phage ul36 (39), and also to Tuc2009, TP901-1 and the archetype
249 P335, which are related phages, all belonging to P335 subgroup II (40, 41). In contrast, the
250 rightward end of each of the 9871-4 phage genomes appears to bear more similarity to *S.*
251 *thermophilus* phage replication modules (42; Fig. 1). Each module of this apparent 'hybrid'
252 987 phage group is discussed below.

253

254 *Structural modules & structural protein determination*

255 The structural gene module in the four 987 group members in each phage, spanning the
256 region starting at the putative portal protein-encoding gene to the predicted serine
257 acetyltransferase-encoding gene, are remarkably conserved at the deduced amino acid level
258 (Fig. 1). For this reason, structural proteins present in purified phage particles prepared from
259 a cell lysate were determined by mass spectrometry for phage 9871 as a representative of the
260 group (Fig. 2), and this phage alone will be discussed further (unless otherwise indicated).
261 The deduced products of *ORF*s *4-8*$_{9871}$ were all confirmed as structural proteins, and are

262 presumed to be involved in phage head morphogenesis (based on their positions in the
263 genome, as well as amino acid identities to known phage head proteins; Fig. 2). The proteins
264 encoded by *ORF9*$_{9871}$, *ORF10*$_{9871}$ and *ORF11*$_{9871}$ were not detected during mass
265 spectrometric analysis, possibly due to their low abundance in the 9871 particle. *ORF9*$_{9871}$
266 and *ORF10*$_{9871}$ appear to encode a so-called head-tail connector or adapter (43, 44), based on
267 conserved phage head-tail connector domains (specifically, those present in proteins GP15
268 and GP16 of the well characterised *Bacillus subtilis* phage SPP1) being detected using CDD
269 (*ORF9*$_{9871}$) and HHPred (*ORF10*$_{9871}$), respectively (Fig. 1).

270 The defined tail morphogenesis gene cluster in phage 9871 commences with
271 *ORF11*$_{9871}$, which is predicted to encode a putative tail component (Fig. 1). *ORF12*$_{9871}$ was
272 not confirmed as a structural protein-encoding gene, however, a homologue of this gene
273 product present in lactococcal phage TP901-1 has recently been annotated as the tail
274 terminator protein (45), with apparently unchecked tail extension observed in mutant phages
275 containing a stop codon in this gene. *ORF13*$_{9871}$ specifies the presumed major tail protein,
276 with *ORF14*$_{9871}$ and *ORF15*$_{9871}$ encoding putative tail assembly chaperone proteins (46, 47)
277 of *ORF16*$_{9871}$ (predicted to encode the tail tape measure protein). Indeed, *ORF14*$_{9871}$ shows
278 evidence of a 'slippery sequence' (5'-AAAAAAA-3'), a feature present in some genes
279 involved in tail assembly which leads to an alternative frame translation (46) and production
280 of an essential tail chaperone in bacteriophage λ (48). The product of *ORF16*$_{9871}$ (TMP$_{9871}$)
281 was not confirmed as a structural protein, suggesting that it is present in low amounts in the
282 phage particle.

283 *ORF17*$_{9871}$, encoding the putative distal tail protein (the product of which was
284 confirmed as a structural protein, Fig. 2), which is homologous to its functional equivalent in
285 the lactococcal phage TP901-1, the latter forming the core of the phage tail tip (49). The
286 putative tail-associated lysin (TAL) is encoded by *ORF18*$_{9871}$ (confirmed as a structural

protein, Fig. 2) and shares significant amino acid similarity (particularly at the N-terminus) with the corresponding genes in phages Tuc2009, TP901-1 (in which it was defined as the tail fiber), ul36 and P335 (39, 50-53). The location of the endopeptidase-encompassing domain (M23 family), including the catalytic His residue (residing within amino acid sequence ATGVHLHF, being the equivalent of VTGPHLHF in Tuc2009 and TP901-1; [54, 55]), in this protein appears to be conserved in phage 9871, based on CDD (28) search results. Previously, it has been reported that the TAL of *L. lactis* phage Tuc2009 undergoes autocleavage at a specific GGSSG*GG amino acid sequence, where * indicates the cleavage site (54, 55). In TAL_{9871}, this site appears to be replaced by AASGGGG, with underlined residues indicating amino acid substitutions relative to the site in $TAL_{Tuc2009}$.

The final structural protein of phage 9871 as determined by mass spectrometry is the product of $ORF19_{9871}$, which encodes the putative receptor binding protein, henceforth referred to as RBP_{9871}. While the 'tripods' (as defined by D. Veesler, *et al.* (49)) of *L. lactis* phages that are closely related to the 987 group phages are encoded by at least two genes (Tuc2009 (35, 56) and TP901-1 (57)), the baseplate in phage 9871 appears to be encoded by a single gene (for reasons outlined below), perhaps akin to the arrangement in several P335 phages of *Lactococcus lactis* including BK5-T, LC3, BM13 and Q33, which belong to the P335 subgroups I, III and IV, respectively (41). Several *S. thermophilus* phages (including DT1) apparently share this arrangement – with a single antireceptor gene containing at least one variable region, (one of) which (termed 'VR2') was shown to be correlated to host specificity (11). Here, the N-terminal end of RBP_{9871} shares a high level of amino acid identity (approximately 85 %) with the N-terminal portion of the upper baseplate protein (BppU) of TP901-1, Tuc2009, P335 and ORF322 of ul36, and then appears to be extended (relative to BppU) at the C-terminal end. This composite arrangement is visualized in Figure 3. A parallel beta helix domain at the C-terminal end of RBP_{9871} (identified using Pfam) is a

member of clan CL0268, members of which include glycosyl hydrolases, pectate lyases, pectin esterases and *Salmonella* phage P22-like tail-spikes. Similarly, using a CDD search, pectate lyase domains were found to be present toward the C-terminal end of the protein which bear similarity to glycosyl hydrolase family 28 - members of which hydrolyse glycosidic bonds in the heteropolysaccharide pectin (58). Taken together, these findings suggest that RBP_{9871} has a carbohydrate binding function - leading us to hypothesize that this protein incorporates the receptor binding activities of the BppU and BppL proteins of TP901-1, where BppL is known to be responsible for host interaction and specificity (59). Importantly, the other three members of the 987 group each encompass an $ORF19_{9871}$ homologue, which exhibit near complete nucleotide identity to each other, being consistent with the extremely narrow host range of these phages.

$ORF20_{9871}$ (highly conserved in the 987 group phages; Fig. 1) was not detected during mass spectrometry and its function is currently unknown, though it appears (using a BLAST search) to be related to a family of serine acetyltransferases. A search using the CDD database confirms the presence of the serine acetyltransferase domain as well as a sugar O-acetyltransferase domain of the NeuD family, identified as a sialic acid O-acetyltransferase in group B streptococci (60). O-acetylation has been shown to be present at precise locations of the sialic acid component of the capsular polysaccharide of group B *Streptococcus* (61). The presumed sugar interaction of the predicted O-acetyltransferase enzyme may thus be significant in the context of the outer cell layer encountered by the phage during host adsorption, where the product of $ORF20_{9871}$ may perhaps play an accessory role in host recognition, similar to that exhibited by BppA in Tuc2009 (35, 62) - particularly considering its proposed position in the tail morphogenesis module of the 987 group phage genomes.

Lysis & lysogeny modules

Approximately half of the currently sequenced *S. thermophilus* phages possess two distinct holin-encoding genes, being largely conserved, with the exception of phage 2972 (19). The 987 group phages also appear to possess two distinct holin-encoding genes, one gene product being closely related to holins found in *L. lactis* phages, and one to those found in *S. thermophilus* phages (Fig. 1; Table S1). The lysin-encoding gene of 9871 ($ORF23_{9871}$) is located immediately downstream of the holin-encoding genes, and appears to be interrupted by a putative group I intron, a feature previously described in other phages of *S. thermophilus* (63). This is indicated by the presence of a predicted endonuclease-encoding open reading frame, known to be a feature of certain group I introns (64), as well as the presence of a 14 bp consensus sequence (surrounding the predicted intron splice site) correlated with intron possession (63) in all four phages, with varying nucleotide identity.

The predicted lysogeny modules present in the 9871-4 phages appear in each case to have been subjected to genetic decay and therefore redundant, as based on the smaller size of the region relative to proven lysogenic phages (5) and the absence of certain genes (most notably, in this case, an integrase-encoding gene) typically associated with these modules in genuine temperate phages (65). These regions are commonly known as lysogeny 'replacement' modules, which are a feature of lytic *S. thermophilus* phages (14, 17).

Replication modules

The gene products encoded by the individual replication modules present in phages 9871-4 (downstream of the lysogeny replacement modules) are largely conserved at the amino acid level (Fig. 1) and appear to belong to the '7201-like' grouping (66), which has previously

been identified in phages 7201 (42), Abc2 (6) and, more recently, 5093 (13). Despite this general conservation, however, the replication module of phage 9874 (more so than those of phages 9871-3) is characterized by deletions, insertions and point mutations – a common feature of this region in *S. thermophilus* phage genomes (18). Various genes encoding proteins of apparent non-streptococcal phage origin are positioned downstream of the replication module (detailed in the legend of Fig. 1 and in Table S1), including a RecT recombinase-encoding gene. Interestingly, these genes are often associated with exonuclease-encoding genes which together form so-called 'recombination modules' (67). Indeed, this is the case for the 987 group phages, with the exonuclease-encoding gene being located immediately downstream of the recombinase-encoding genes (Fig. 1). Phages 9871, 9872 and 9873 are also predicted to encode a cytosine-5 methyltransferase (*ORF46*$_{9871}$). In general, phage-encoded methyltransferases are thought to be an anti-defensive response to the DNA-targeting activity of restriction-modification systems in bacterial hosts, but potentially also function in other viral and cellular processes (for a review, see 68).

The proposed 'terminal' ORFs are defined here as those ORFs preceding the small subunit of the terminase in the genomes of the 987 group phages (*ORF50*$_{9871/9872}$, *ORF49*$_{9873}$ or *ORF48*$_{9874}$; Fig. 1). These ORFs are defined as 'terminal' based on homologues being present upstream of the defined *cos*-site in several *cos*-containing phages of *S. thermophilus* such as DT1 (15), Sfi19 (14), 7201 (16), Sfi21 (14) and Abc2 (6); the *cos*-site, in turn, being located upstream of the small subunit of the terminase. The protein products of the terminal ORFs in the 987 group phages appear to be conserved in 9871-3, with that of 9874 being divergent. A Pfam search using these proteins in phages 9871-3 indicate that they belong to the DUF1492 family, which was recently found to be one of several major groups of 'late transcriptional regulators' (ltr) in phages of Gram-positive bacteria (69). Similarly, transcriptional regulation appears to be the primary function of the product of *ORF48*$_{9874}$,

384 which shows approximately 50 % amino acid identity with ArpU family transcriptional
385 regulators of various streptococcal species (69).

386

387 *Morphological characteristics*

388 The morphology of the *Siphoviridae* family has been well documented (group B as defined
389 by D. E. Bradley [9]), and siphophages infecting *L. lactis* and *S. thermophilus* exhibit the
390 expected morphology, with icosahedral heads and non-contractile tails. Phages infecting *S.*
391 *thermophilus* in general possess longer tails than their lactococcal counterparts (with an
392 exception being the 949 group of lactococcal phages [40, 70]), their long tails being
393 consistent with their long TMP-encoding genes (71). Upon electron microscopic analysis, it
394 was found that phages 9871-4 exhibit icosahedral heads and relatively short tails (see Fig. 4E
395 for exact dimensions). The distal tail-associated baseplate, which generally functions in the
396 attachment of the phage to the bacterial cell (discussed above), is clearly visible (Fig. 4A -
397 D). The presence and observed features of a base plate are consistent with those previously
398 observed in P335 species *L. lactis* phages such as Tuc2009 and TP901-1 (57, 71, 72), and
399 indeed with the observed similarity between the tail structural gene products and those
400 encoded by the 987 group phages. The measured head diameter is similar among each of the
401 987 group phages, yet slightly larger than those previously reported for Tuc2009 and TP901-
402 1 - similarly, the tail lengths of the phages in the 987 group are within the same range relative
403 to each other, but slightly shorter than previously reported for those P335 phages (72-74). In
404 keeping with this distinction between phages infecting *S. thermophilus* and *L. lactis* P335
405 group phages, the observed tail lengths of the 987 group phages are lower than previously
406 reported for *S. thermophilus* phages (71).

407

Adsorption & adsorption inhibition assays

Consistent with the genetic composition of the structural module (and, in particular, the antireceptor-encoding genes) of the four 987 group phages, it appears that these phages are able to adsorb to certain *L. lactis* strains as well as the primary *S. thermophilus* host. Ten *L. lactis* strains, which are routinely combined with ST64987 in industrial fermentations, were initially tested for phage adsorption (as described above), of which one (LL64981) appeared to adsorb all four 987 group phages at a level of approximately 50 % (Fig. 5). For this assay, a negative control phage (3681, a *cos*-containing phage infecting industrial *S. thermophilus* strain ST67368) was used to illustrate the specific adsorption affinity of strains ST64987 & LL64981 for the 987 group phages. Following the adsorption assay, a DNA transduction experiment was performed using this phage/strain combination, as well as a positive control combination, according to a previously reported method (75) - however, no confirmed *L. lactis* transductants were obtained (data not shown). Adsorption to both strains of *L. lactis* and *S. thermophilus* suggests that a common cell-surface molecule is recognized by these hybrid phages (discussed further below), complemented by the observed genetic similarity of the tail tip regions of the 987 group phages to those regions in phages infecting *L. lactis*. Indeed, a phage infecting *L. lactis* LL64981 (termed 98103) was also shown to exhibit adsorption affinity to both its host and to *S. thermophilus* ST64987 (Fig. 5).

In order to further investigate the interaction between the 987 group phages and their host(s), we performed a competitive phage adsorption inhibition assay using the presumed host-recognition protein RBP_{9871}. This protein product is proposed to represent the antireceptor of this phage group based on the position of the encoding gene in all four phage genomes (Fig. 1), its confirmed presence as a structural protein in the viral particle (Fig. 2A and C), and for reasons discussed in detail above. RBP_{9871} was overexpressed and purified (Fig. 6A, Lane 2), and then used in adsorption inhibition assays (as described in Materials

433 and Methods). Figure 6 clearly shows that RBP_{9871}, when incubated with wild-type *S.*
434 *thermophilus* and *L. lactis* cells, inhibits adsorption of phage 9871 to both strains in a dose-
435 dependent manner (Fig. 6B and C). Maximal (an average of approximately 80 %) adsorption
436 inhibition was achieved using a concentration of 69.7 pM in both cases, a concentration
437 comparable to that observed by B. Collins, *et al*. (35) using a lactococcal phage RBP and host
438 combination. Any observed difference in the potency of the respective RBPs may possibly be
439 accounted for by an increased (or decreased) amount of available binding sites present on the
440 cell surface, particularly considering the differences in host genera (for *S. thermophilus*
441 ST64987) and subspecies (for *L. lactis* subsp. *lactis* LL64981).

442 These adsorption inhibition data have a number of implications. Firstly, it may be
443 postulated that the cell surface target used by the 987 group phages is carbohydrate in nature,
444 considering the putative carbohydrate-binding function of the antireceptor protein (discussed
445 above), as well as the homology of the tail tip regions to phages of *L. lactis*, which target cell
446 surface carbohydrate moieties (76). These data also suggest that the phage target expressed
447 on the cell surface of *S. thermophilus* ST64987 and on *L. lactis* LL64981 is at least similar in
448 nature. Considering the observed similarity between some *S. thermophilus* and *L. lactis* genes
449 encoding exopolysaccharide (EPS) biosynthetic elements (1, 77), combined with the
450 observed heterogeneity of the EPS clusters of *S. thermophilus* (1, 77, 78), this is conceivable.
451 Furthermore, considering the genetic divergence between the antireceptors of the 987 group
452 phages and those of previously sequenced phages such as DT1 (11), and phage 5093 (13), it
453 is possible that alternative cell surface targets are recognized by these phages during the
454 initial phage-host interaction.

455

456 *Evolutionary aspects*

457 Phages 9871-4 represent a group of *S. thermophilus*-infecting phages which is distinct from
458 documented *cos*-containing and *pac*-containing phage groups, as well as the more recently
459 discovered 5093 group. While phage 5093 appears to have acquired several genes from non-
460 dairy streptococcal phages such as those infecting *Streptococcus pneumoniae*, *Streptococcus*
461 *gordonii* and *Streptococcus pyogenes* (13), a genetic crossover previously observed in
462 prophages of *S. pyogenes* (79), the 987 group phages appear to have been the result of a
463 genomic recombination event between a (temperate) P335 phage of *L. lactis*, and an
464 unknown *S. thermophilus* phage. Figure 7 shows the genetic distance between the currently
465 known (thirteen) fully sequenced *S. thermophilus* phages, ten sequenced *L. lactis* phages of
466 the P335 group and the four 987 group phages. Using this (unrooted) visualization, it appears
467 that the 987 group phages are derivatives of the P335 subgroup of which Tuc2009, TP901-1,
468 P335 and ul36 are members, and are more closely related to this group than to the other
469 known phages that infect *S. thermophilus*. Considering the high level of nucleotide identity to
470 these phages across the (generally) more conserved structural regions, this is not surprising.
471 Alignments of the structural modules (Fig. 7B; comprising the TerS-encoding gene to the
472 holin-encoding gene) and the replication modules (Fig. 7C; comprising the lysin-encoding
473 gene to the terminal ORF-encoding gene) of the 987 group with the relevant comparators, i.e.
474 phages which also harbor group II/7201-like replication modules (42, 66), show a clear
475 difference in clustering, indicating the diverse lineage of these respective modules in the 987
476 group phages.

477 The genetic/structural similarity of these phages to those of *L. lactis* - with the
478 retention of the ability to infect *S. thermophilus* - may be considered a form of adaptive
479 mosaicism, a known evolutionary strategy common to phages infecting a wide range of
480 bacteria (80, 81). Due to the close association of *L. lactis* and *S. thermophilus* both in raw
481 milk and in the dairy processing environment, gene transfer between (the phages of) these

482 species has been the subject of speculation – with perhaps the most striking phage example of
483 this phenomenon being observed in the case of phage BK5-T, a temperate phage of *L. lactis*
484 H2L (82), which shares significant sequence similarity with *S. thermophilus* phage Sfi21
485 (83). Further examples of this phenomenon include the genomes of phage 1358 (infecting *L.*
486 *lactis* SMQ-388), with homology to phages infecting *Listeria monocytogenes* (84); phage
487 Q54 (infecting *L. lactis* SMQ-562) which appears to be a hybrid of the 936 and c2
488 lactococcal phage species (85), and phage 1706 (infecting *L. lactis* SMQ-450), proposed to be
489 derived from a number of prophages of other *Firmicutes* (86). The mechanisms by which
490 such horizontal gene transfer events between phages occur have also been proposed. S.
491 Moineau, *et al*. (22) and E. Durmaz & T. R. Klaenhammer (87) have shown that lytic phages
492 can evolve by acquiring segments of DNA from the host chromosome (including, potentially,
493 remnant prophage) sometimes in response to pressure from abortive infection (Abi) phage
494 resistance systems (88). More recently it has been shown that transduction in *L. lactis* is
495 possible using *S. thermophilus* phages (75), clearly demonstrating that phage-mediated
496 horizontal DNA transfer between these two species is possible. Furthermore, considering the
497 rapid nature of phage infection (reviewed by A. Quiberoni, *et al*. (7)) and, in turn (by
498 necessity), the acquisition of phage resistance, a common genetic lineage in phages of *S.*
499 *thermophilus* and *L. lactis* phages may be reflected in the numerous phage resistance
500 mechanisms of their hosts. Indeed, this has been shown by X. Sun and colleagues (89), who
501 demonstrated that the superinfection exclusion (sie) phage resistance protein Ltp confers
502 phage resistance to both *S. thermophilus* and *L. lactis* hosts, against their respective attacking
503 phages (89). Such multi-genus protection is indicative of co-evolution of both phage and
504 host, possibly accelerated by their continuous mutual exposure in the dairy environment.

505 The impact of genetic mosaicism in phages on the marketplace is illustrated by the
506 987 group phages above, which retain infective ability in *S. thermophilus*, despite having

many genetic and morphological characteristics of phages of *L. lactis*. This is an important consideration in the dairy industry, an environment in which lactic acid bacteria are in close proximity on a regular basis, and which may well present further examples of genetic mosaicism as an evolutionary strategy in dairy phages.

Conclusions

Here we report the complete genome sequences of four novel phages infecting the dairy bacterium *S. thermophilus*. Comparative genomic analysis revealed high level of nucleotide homology to the replication modules of *S. thermophilus* phages and the structural modules of *L. lactis* phages, suggesting a relatively recent horizontal gene transfer or recombination event. These genome sequences represent a significant divergence compared to the previously published thirteen *S. thermophilus* phage genomes, being highly mosaic in nature, and are the first members of this phage group to be sequenced. The structural protein complement of one of these phages (as a representative of the group) was determined and found to be similar to previously characterised phages of *L. lactis*. Morphological similarity to phages of *L. lactis* was also observed using electron microscopic analysis, in which short tails and claw-like baseplates were observed in all four members of the group.

Adsorption studies revealed the ability of members of this group of phages to adsorb to both their native *S. thermophilus* hosts as well as an *L. lactis* strain with which it is routinely combined in dairy fermentations, suggesting that certain cell surface molecules are shared between the genera. This finding also hints at the event by which the hybrid genomes of these phages may have begun to be replicated, possibly being facilitated by mutually expressed cell surface proteins in combination with a favourable *S. thermophilus* phage co-infection or prophage-mediated evolutionary event. The phage gene product responsible for

this adsorption was defined by the use of purified protein to inhibit phage adsorption to both strains, providing a more detailed analysis of the initial phage-host interaction.

Genetic mosaicism is a common trait of bacteriophages and, in the context of dairy fermentations, may represent a new challenge to phage control methods, which usually consist of traditional bacteriophage insensitive mutant (BIM) generation and rotational schemes. In light of the rapidly increasing genetic diversity being observed in phages of *S. thermophilus*, continual monitoring of phage populations in dairy productions will be necessary to ensure that BIM generation methods and knowledge-based rotational systems can be used effectively to ameliorate phage spoilage of industrial fermentations.

Funding Information

The authors gratefully acknowledge the financial support of DSM Food Specialties. JM is in receipt of a Technology Innovation Development Award (TIDA) (Ref. No. 14/TIDA/2287) funded by Science Foundation Ireland (SFI). DvS is supported by a Principal Investigator award (Ref. No. 13/IA/1953) through Science Foundation Ireland (SFI).

Acknowledgements

The authors gratefully acknowledge the technical assistance of Eoghan Casey and Erik Royackers, and Emiel Ver-Loren-van-Themaat for useful discussion.

552 **Figure legends**

553 **Figure 1.**

554 Comparative analysis of the genetic organization and content of phages 9871-4 with
555 archetypes TP901-1 (P335 *L. lactis* phage species) and 7201 (*cos*-containing *S. thermophilus*
556 phage). Predicted ORFs (indicated by arrows) and gene products (putative function indicated
557 by colour coding) are aligned with adjacent genomes according to % amino acid identity
558 (indicated by shaded boxes). Gene products considered to be notable are marked in black,
559 with accompanying legend.

560 **Figure 2.**

561 Structural proteome analysis of phage 9871. (A) Deduced structural proteins (and
562 corresponding ORF number) as identified by ESI-MS/MS (threshold: two unique peptides or
563 5 % ORF coverage). No. of peptides = number of distinct polypeptide strings identified
564 during the analysis. Amino acids = the total number of amino acids identified in each protein.
565 Coverage = the number of amino acids identified expressed as a percentage of the number of
566 amino acids in the entire protein. (B) SDS-PAGE gel (12 %) showing the structural protein
567 profile of phage 9871. Lane 1: Broad range protein ladder (New England Biolabs); Lane 2:
568 phage 9871 protein extraction. (C) ORF schematic of phage 9871 highlighting confirmed
569 structural protein-encoding genes (bold outline).

570 **Figure 3.**

571 Schematic representation of ORFs predicted to encode the tail proteins of *S. thermophilus*
572 phage 7201 (*cos*-containing), 2972 (*pac*-containing) 9871 (987 group), Tuc2009 and phiLC3
573 (*L. lactis* phage P335 group). Predicted ORFs (indicated by arrows) and gene products

574 (putative function indicated by colour coding) are aligned with adjacent genomes according
575 to % amino acid identity (indicated by shaded boxes).

576 **Figure 4.**

577 Uranyl acetate stained transmission electron micrograph of phages 9871 (A) 9872 (B), 9873
578 (C) and 9874 (D), and discerned head and tail measurements (E).

579 **Figure 5.**

580 Adsorption analysis of phages 9871-4 on primary host *S. thermophilus* ST64987 and *L. lactis*
581 LL64981 at adsorption temperature (T_A) = 42 °C. Phage 3681 (a *cos*-containing lytic phage
582 of *S. thermophilus*) is included as an adsorption-negative control for ST64987 and LL64981,
583 and was found separately to adsorb optimally (>90%) to its primary host (*S. thermophilus*
584 ST67368) at T_A = 42 °C. Phage 98103 (P335 species *L. lactis* phage infecting LL64981) was
585 also shown to exhibit adsorption affinity to both strains. Comparable adsorption data for all
586 strains were generated at T_A = 30 °C.

587 **Figure 6.**

588 Phage 9871 adsorption inhibition analysis using varying concentrations of purified RBP_{9871}
589 on strains *S. thermophilus* ST64987 and *L. lactis* LL64981 by blocking assay. (A) SDS-
590 PAGE gel (12 %) showing purified antireceptor of phage 9871. Lane 1: Blue prestained
591 protein standard, Broad range (New England Biolabs), Lane 2: purified 9871 antireceptor.
592 (B) inhibition (%) of 9871 adsorption on ST64987. (C) inhibition (%) of 9871 adsorption on
593 LL64981.

594 **Figure 7.**

595 Unrooted phylogenetic tree showing the genetic relatedness between the 987 group phages,
596 *cos*- and *pac*-containing *S. thermophilus*-infecting phages, as well as *L. lactis* infecting
597 phages of the P335 group (colour coding indicating the respective groupings is indicated in
598 the accompanying legend). (A) whole genome nucleotide comparison; (B) structural module
599 comparison, with those *S. thermophilus* phages also harbouring a group II/7201-like
600 replication module; (C) replication module comparison, with those *S. thermophilus* phages
601 also harbouring a group II/7201-like replication module.

602

603 **References**

604 1. **Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R.** 2011. Specialized
605 adaptation of a lactic acid bacterium to the milk environment: the comparative
606 genomics of *Streptococcus thermophilus* LMD-9. Microb Cell Fact **10 Suppl 1:**S22.
607 2. **Ott A, Hugi A, Baumgartner M, Chaintreau A.** 2000. Sensory investigation of
608 yogurt flavor perception: mutual influence of volatiles and acidity. J Agric Food
609 Chem **48:**441-450.
610 3. **Guglielmotti DM, Mercanti DJ, Reinheimer JA, Quiberoni AdL.** 2011. Review:
611 efficiency of physical and chemical treatments on the inactivation of dairy
612 bacteriophages. Front microbiol **2.** doi:10.3389/fmicb.2011.00282.
613 4. **Capra ML, Neve H, Sorati PC, Atamer Z, Hinrichs J, Heller KJ, Quiberoni A.**
614 2013. Extreme thermal resistance of phages isolated from dairy samples: Updating
615 traditional phage detection methodologies. Int Dairy J **30:**59-63.
616 5. **Ali Y, Koberg S, Hessner S, Sun X, Rabe B, Back A, Neve H, Heller KJ.** 2014.
617 Temperate *Streptococcus thermophilus* phages expressing superinfection exclusion
618 proteins of the Ltp type. Front Microbiol **5:**98. doi:10.3389/fmicb.2014.00098.
619 6. **Guglielmotti DM, Deveau H, Binetti AG, Reinheimer JA, Moineau S, Quiberoni**
620 **A.** 2009. Genome analysis of two virulent *Streptococcus thermophilus* phages isolated
621 in Argentina. Int J Food Microbiol **136:**101-109.
622 7. **Quiberoni A, Moineau S, Rousseau GM, Reinheimer J, Ackermann H-W.** 2010.
623 *Streptococcus thermophilus* bacteriophages. Int Dairy J **20:**657-664.
624 8. **Mahony J, van Sinderen, D.** 2014. Current taxonomy of phages infecting lactic acid
625 bacteria. Front Microbiol **5:**7. doi:10.3389/fmicb.2014.00007
626 9. **Bradley DE.** 1967. Ultrastructure of bacteriophage and bacteriocins. Bacteriol Rev
627 **31:**230-314.

628 10. **Le Marrec C, van Sinderen D, Walsh L, Stanley E, Vlegels E, Moineau S, Heinze**
629 **P, Fitzgerald G, Fayard B.** 1997. Two groups of bacteriophages infecting
630 *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging
631 and genetic determinants for major structural proteins. Appl Environ Microbiol
632 **63:**3246-3253.
633 11. **Duplessis M, Moineau S.** 2001. Identification of a genetic determinant responsible
634 for host specificity in *Streptococcus thermophilus* bacteriophages. Mol Microbiol
635 **41:**325-336.
636 12. **Binetti AG, Del Rio B, Martin MC, Alvarez MA.** 2005. Detection and
637 characterization of *Streptococcus thermophilus* bacteriophages by use of the
638 antireceptor gene sequence. Appl Environ Microbiol **71:**6096-6103.
639 13. **Mills S, Griffin C, O'Sullivan O, Coffey A, McAuliffe O, Meijer W, Serrano L,**
640 **Ross R.** 2011. A new phage on the 'Mozzarella'block: Bacteriophage 5093 shares a
641 low level of homology with other *Streptococcus thermophilus* phages. Int Dairy J
642 **21:**963-969.
643 14. **Lucchini S, Desiere F, Brussow H.** 1999. The genetic relationship between virulent
644 and temperate *Streptococcus thermophilus* bacteriophages: whole genome comparison
645 of *cos*-site phages Sfi19 and Sfi21. Virology **260:**232-243.
646 15. **Tremblay DM, Moineau S.** 1999. Complete genomic sequence of the lytic
647 bacteriophage DT1 of *Streptococcus thermophilus*. Virology **255:**63-76.
648 16. **Proux C, van Sinderen D, Suarez J, Garcia P, Ladero V, Fitzgerald GF, Desiere**
649 **F, Brussow H.** 2002. The dilemma of phage taxonomy illustrated by comparative
650 genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J Bacteriol **184:**6026-
651 6036.

17. **Stanley E, Fitzgerald GF, Le Marrec C, Fayard B, van Sinderen D.** 1997. Sequence analysis and characterization of phi O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. Microbiology **143 (Pt 11):**3417-3429.

18. **Lucchini S, Desiere F, Brussow H.** 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J Virol **73:**8647-8656.

19. **Levesque C, Duplessis M, Labonte J, Labrie S, Fremaux C, Tremblay D, Moineau S.** 2005. Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. Appl Environ Microbiol **71:**4057-4068.

20. **Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol **190:**1390-1400.

21. **Lillehaug D.** 1997. An improved plaque assay for poor plaque-producing temperate lactococcal bacteriophages. J Appl Microbiol **83:**85-90.

22. **Moineau S, Pandian S, Klaenhammer TR.** 1994. Evolution of a Lytic Bacteriophage via DNA Acquisition from the *Lactococcus lactis* Chromosome. Appl Environ Microbiol **60:**1832-1841.

23. **Sambrook J, Fritsch E. F., Maniatis, T.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

24. **Besemer J, Borodovsky M.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res **27:**3911-3920.

25. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403-410.

677 26. **Sonnhammer EL, Eddy SR, Durbin R.** 1997. Pfam: a comprehensive database of
678 protein domain families based on seed alignments. Proteins **28:**405-420.
679 27. **Soding J, Biegert A, Lupas AN.** 2005. The HHpred interactive server for protein
680 homology detection and structure prediction. Nucleic Acids Res **33:**W244-248.
681 28. **Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY,**
682 **Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song**
683 **JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH.** 2015.
684 CDD: NCBI's conserved domain database. Nucleic Acids Res **43:**D222-226.
685 29. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A,**
686 **Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics
687 **16:**944-945.
688 30. **Briggiler Marco M, Garneau JE, Tremblay D, Quiberoni A, Moineau S.** 2012.
689 Characterization of two virulent phages of *Lactobacillus plantarum*. Appl Environ
690 Microbiol **78:**8719-8734.
691 31. **Casey E, Mahony J, O'Connell-Motherway M, Bottacini F, Cornelissen A, Neve**
692 **H, Heller KJ, Noben JP, Dal Bello F, van Sinderen D.** 2014. Molecular
693 characterization of three *Lactobacillus delbrueckii* subsp. *bulgaricus* phages. Appl
694 Environ Microbiol **80:**5623-5635.
695 32. **Ceyssens PJ, Mesyanzhinov V, Sykilinda N, Briers Y, Roucourt B, Lavigne R,**
696 **Robben J, Domashin A, Miroshnikov K, Volckaert G, Hertveldt K.** 2008. The
697 genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage
698 resembling M6. J Bacteriol **190:**1429-1435.
699 33. **Vanheel A, Daniels R, Plaisance S, Baeten K, Hendriks JJ, Leprince P, Dumont**
700 **D, Robben J, Brone B, Stinissen P, Noben JP, Hellings N.** 2012. Identification of
701 protein networks involved in the disease course of experimental autoimmune

encephalomyelitis, an animal model of multiple sclerosis. PLoS One **7:**e35544. doi:10.1371/journal.pone.0035544.

34. **Garvey P, Hill C, Fitzgerald GF.** 1996. The Lactococcal Plasmid pNP40 Encodes a Third Bacteriophage Resistance Mechanism, One Which Affects Phage DNA Penetration. Appl Environ Microbiol **62:**676-679.

35. **Collins B, Bebeacua C, Mahony J, Blangy S, Douillard FP, Veesler D, Cambillau C, van Sinderen D.** 2013. Structure and functional analysis of the host recognition device of lactococcal phage Tuc2009. J Virol **87:**8429-8440.

36. **Kuipers OP, de Ruyter PG, Kleerebezem M, de Vos WM.** 1998. Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol **64:**15-21.

37. **Zinno P, Janzen T, Bennedsen M, Ercolini D, Mauriello G.** 2010. Characterization of *Streptococcus thermophilus* lytic bacteriophages from mozzarella cheese plants. Int J Food Microbiol **138:**137-144.

38. **Quiberoni A, Tremblay D, Ackermann HW, Moineau S, Reinheimer JA.** 2006. Diversity of *Streptococcus thermophilus* phages in a large-production cheese factory in Argentina. J Dairy Sci **89:**3791-3799.

39. **Labrie S, Moineau S.** 2002. Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. Virology **296:**308-320.

40. **Samson JE, Moineau S.** 2010. Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. Appl Environ Microbiol **76:**6843-6852.

41. **Mahony J, Martel B, Tremblay DM, Neve H, Heller KJ, Moineau S, van Sinderen D.** 2013. Identification of a new P335 subgroup through molecular analysis of lactococcal phages Q33 and BM13. Appl Environ Microbiol **79:**4401-4409.

42. **Stanley E, Walsh L, van der Zwet A, Fitzgerald GF, van Sinderen D.** 2000. Identification of four loci isolated from two *Streptococcus thermophilus* phage genomes responsible for mediating bacteriophage resistance. FEMS Microbiol Lett **182:**271-277.

43. **Lurz R, Orlova EV, Gunther D, Dube P, Droge A, Weise F, van Heel M, Tavares P.** 2001. Structural organisation of the head-to-tail interface of a bacterial virus. J Mol Biol **310:**1027-1037.

44. **Bebeacua C, Lai L, Vegge CS, Brondsted L, van Heel M, Veesler D, Cambillau C.** 2013. Visualizing a complete *Siphoviridae* member by single-particle electron microscopy: the structure of lactococcal phage TP901-1. J Virol **87:**1061-1068.

45. **Stockdale SR, Collins B, Spinelli S, Douillard FP, Mahony J, Cambillau C, van Sinderen D.** 2015. Structure and Assembly of TP901-1 Virion Unveiled by Mutagenesis. PloS one **10:**e0131676. doi:10.1371/journal.pone.0131676.

46. **Xu J, Hendrix RW, Duda RL.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol Cell **16:**11-21.

47. **Siponen M, Sciara G, Villion M, Spinelli S, Lichiere J, Cambillau C, Moineau S, Campanacci V.** 2009. Crystal structure of ORF12 from *Lactococcus lactis* phage p2 identifies a tape measure protein chaperone. J Bacteriol **191:**728-734.

48. **Xu J, Hendrix RW, Duda RL.** 2013. A balanced ratio of proteins from gene G and frameshift-extended gene GT is required for phage lambda tail assembly. J Mol Biol **425:**3476-3487.

49. **Veesler D, Spinelli S, Mahony J, Lichiere J, Blangy S, Bricogne G, Legrand P, Ortiz-Lombardia M, Campanacci V, van Sinderen D, Cambillau C.** 2012. Structure of the phage TP901-1 1.8 MDa baseplate suggests an alternative host adhesion mechanism. Proc Natl Acad Sci U S A **109:**8954-8958.

50. **Seegers JF, Mc Grath S, O'Connell-Motherway M, Arendt EK, van de Guchte M, Creaven M, Fitzgerald GF, van Sinderen D.** 2004. Molecular and transcriptional analysis of the temperate lactococcal bacteriophage Tuc2009. Virology **329:**40-52.

51. **Brondsted L, Ostergaard 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.

52. **Labrie SJ, Josephsen J, Neve H, Vogensen FK, Moineau S.** 2008. Morphology, genome sequence, and structural proteome of type phage P335 from *Lactococcus lactis*. Appl Environ Microbiol **74:**4636-4644.

53. **Bebeacua C, Bron P, Lai L, Vegge CS, Brondsted L, Spinelli S, Campanacci V, Veesler D, van Heel M, Cambillau C.** 2010. Structure and molecular assignment of lactococcal phage TP901-1 baseplate. J Biol Chem **285:**39079-39086.

54. **Kenny JG, McGrath S, Fitzgerald GF, van Sinderen D.** 2004. Bacteriophage Tuc2009 encodes a tail-associated cell wall-degrading activity. J Bacteriol **186:**3480-3491.

55. **Stockdale SR, Mahony J, Courtin P, Chapot-Chartier MP, van Pijkeren JP, Britton RA, Neve H, Heller KJ, Aideh B, Vogensen FK, van Sinderen D.** 2013. The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for infection specialization. J Biol Chem **288:**5581-5590.

56. **Sciara G, Blangy S, Siponen M, Mc Grath S, van Sinderen D, Tegoni M, Cambillau C, Campanacci V.** 2008. A topological model of the baseplate of lactococcal phage Tuc2009. J Biol Chem **283:**2716-2723.

775 57. **Vegge CS, Brondsted L, Neve H, Mc Grath S, van Sinderen D, Vogensen FK.**
776 2005. Structural characterization and assembly of the distal tail structure of the
777 temperate lactococcal bacteriophage TP901-1. J Bacteriol **187:**4187-4197.
778 58. **Markovic O, Janecek S.** 2001. Pectin degrading glycoside hydrolases of family 28:
779 sequence-structural features, specificities and evolution. Protein Eng **14:**615-631.
780 59. **Vegge CS, Vogensen FK, Mc Grath S, Neve H, van Sinderen D, Brondsted L.**
781 2006. Identification of the lower baseplate protein as the antireceptor of the temperate
782 lactococcal bacteriophages TP901-1 and Tuc2009. J Bacteriol **188:**55-63.
783 60. **Lewis AL, Hensler ME, Varki A, Nizet V.** 2006. The group B streptococcal sialic
784 acid O-acetyltransferase is encoded by neuD, a conserved component of bacterial
785 sialic acid biosynthetic gene clusters. J Biol Chem **281:**11186-11192.
786 61. **Lewis AL, Nizet V, Varki A.** 2004. Discovery and characterization of sialic acid O-
787 acetylation in group B *Streptococcus*. Proc Natl Acad Sci U S A **101:**11123-11128.
788 62. **Legrand P, Collins B, Blangy S, Murphy J, Spinelli S, Gutierrez C, Richet N,**
789 **Kellenberger C, Desmyter A, Mahony J, van Sinderen D, Cambillau C.** 2016.
790 The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host
791 Recognition Involves Two Different Carbohydrate Binding Modules. mBio **7**(1),
792 e01781-15. doi:10.1128/mBio.01781-15.
793 63. **Foley S, Bruttin A, Brussow H.** 2000. Widespread distribution of a group I intron
794 and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus*
795 bacteriophages. J Virol **74:**611-618.
796 64. **Shub DA, Goodrich-Blair H, Eddy SR.** 1994. Amino acid sequence motif of group I
797 intron endonucleases is conserved in open reading frames of group II introns. Trends
798 Biochem Sci **19:**402-404.

65. **Lucchini S, Desiere F, Brussow H.** 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content gram-positive bacteria. Virology **263:**427-435.

66. **Brussow H, Desiere F.** 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol Microbiol **39:**213-222.

67. **Datta S, Costantino N, Zhou X, Court DL.** 2008. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. Proc Natl Acad Sci U S A **105:**1626-1631.

68. **Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D.** 2013. Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. Appl Environ Microbiol **79:**7547-7555.

69. **Quiles-Puchalt N, Tormo-Mas MA, Campoy S, Toledo-Arana A, Monedero V, Lasa I, Novick RP, Christie GE, Penades JR.** 2013. A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. Nucleic Acids Res **41:**7260-7275.

70. **Mahony J, Randazzo W, Neve H, Settanni L, van Sinderen D.** 2015. Lactococcal 949 group phages recognize a carbohydrate receptor on the host cell surface. Appl Environ Microbiol **81:**3299-3305.

71. **Pedersen M, Ostergaard S, Bresciani J, Vogensen FK.** 2000. Mutational analysis of two structural genes of the temperate lactococcal bacteriophage TP901-1 involved in tail length determination and baseplate assembly. Virology **276:**315-328.

72. **Mc Grath S, Neve H, Seegers JF, Eijlander R, Vegge CS, Brondsted L, Heller KJ, Fitzgerald GF, Vogensen FK, van Sinderen D.** 2006. Anatomy of a lactococcal phage tail. J Bacteriol **188:**3972-3982.

73. **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.

74. **Johnsen MG, Neve H, Vogensen FK, Hammer K.** 1995. Virion positions and relationships of lactococcal temperate bacteriophage TP901-1 proteins. Virology **212:**595-606.

75. **Ammann A, Neve H, Geis A, Heller KJ.** 2008. Plasmid transfer via transduction from *Streptococcus thermophilus* to *Lactococcus lactis*. J Bacteriol **190:**3083-3087.

76. **Ainsworth S, Sadovskaya I, Vinogradov E, Courtin P, Guerardel Y, Mahony J, Grard T, Cambillau C, Chapot-Chartier MP, van Sinderen D.** 2014. Differences in lactococcal cell wall polysaccharide structure are major determining factors in bacteriophage sensitivity. mBio **5**(3)**:**e00880-00814. doi:10.1128/mBio.00880-14.

77. **Bourgoin F, Pluvinet A, Gintz B, Decaris B, Guedon G.** 1999. Are horizontal transfers involved in the evolution of the *Streptococcus thermophilus* exopolysaccharide synthesis loci? Gene **233:**151-161.

78. **Pluvinet A, Charron-Bourgoin F, Morel C, Decaris B.** 2004. Polymorphism of *eps* loci in *Streptococcus thermophilus*: sequence replacement by putative horizontal transfer in *S. thermophilus* IP6757. Int Dairy J **14:**627-634.

79. **Desiere F, McShan WM, van Sinderen D, Ferretti JJ, Brussow H.** 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic Streptococci: evolutionary implications for prophage-host interactions. Virology **288:**325-341.

80. **Hatfull GF, Hendrix RW.** 2011. Bacteriophages and their genomes. Curr Opin Virol **1:**298-303.

81. **Casjens SR, Thuman-Commike PA.** 2011. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology **411:**393-415.

82. **Boyce JD, Davidson BE, Hillier AJ.** 1995. Sequence analysis of the *Lactococcus lactis* temperate bacteriophage BK5-T and demonstration that the phage DNA has cohesive ends. Appl Environ Microbiol **61:**4089-4098.

83. **Desiere F, Mahanivong C, Hillier AJ, Chandry PS, Davidson BE, Brussow H.** 2001. Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology **283:**240-252.

84. **Dupuis ME, Moineau S.** 2010. Genome organization and characterization of the virulent lactococcal phage 1358 and its similarities to *Listeria* phages. Appl Environ Microbiol **76:**1623-1632.

85. **Fortier LC, Bransi A, Moineau S.** 2006. Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of *Lactococcus lactis*. J Bacteriol **188:**6101-6114.

86. **Garneau JE, Tremblay DM, Moineau S.** 2008. Characterization of 1706, a virulent phage from *Lactococcus lactis* with similarities to prophages from other *Firmicutes*. Virology **373:**298-309.

87. **Durmaz E, Klaenhammer TR.** 2000. Genetic analysis of chromosomal regions of *Lactococcus lactis* acquired by recombinant lytic phages. Appl Environ Microbiol **66:**895-903.

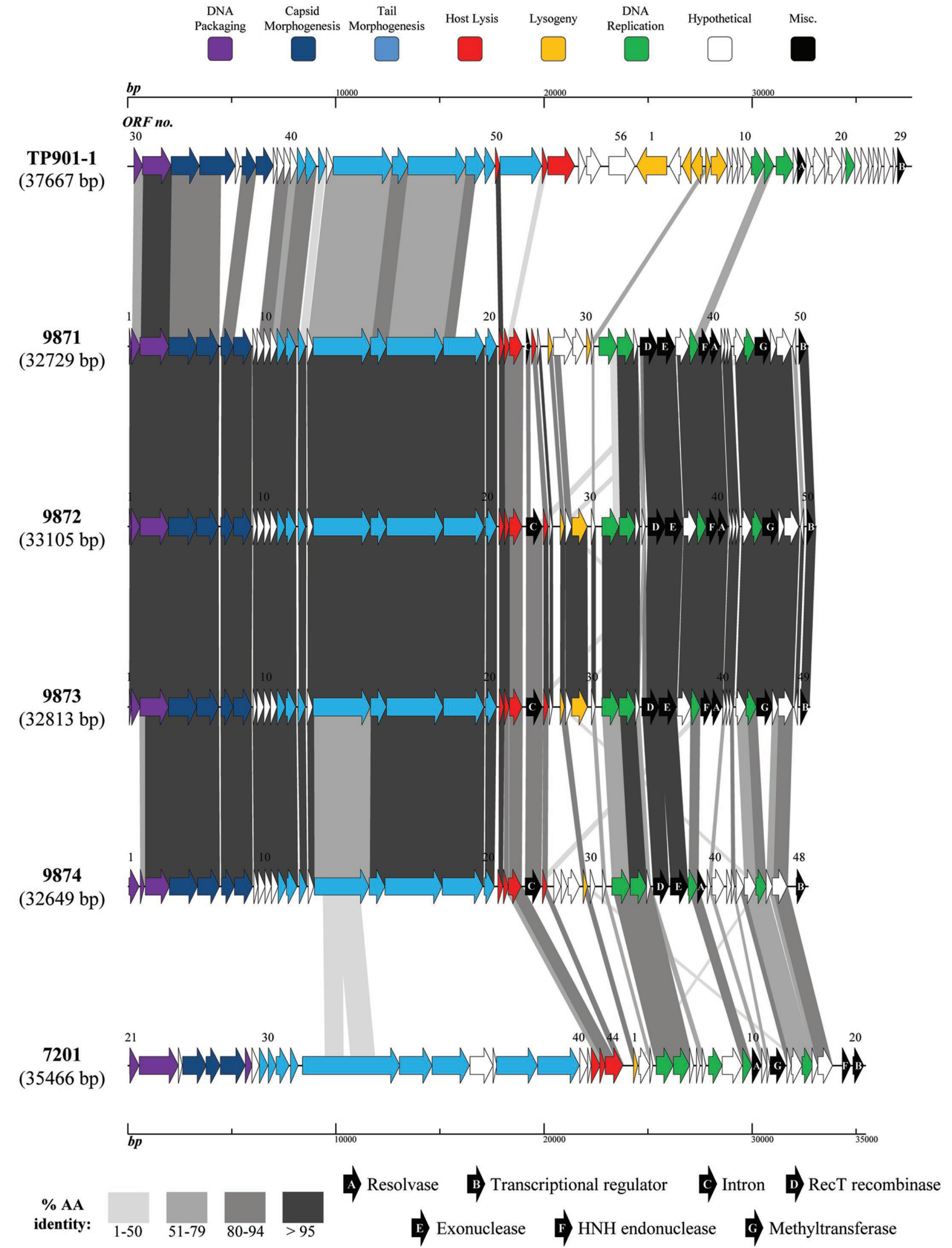
88. **Labrie SJ, Moineau S.** 2007. Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. J Bacteriol **189:**1482-1487.

872 89. **Sun X, Gohler A, Heller KJ, Neve H.** 2006. The ltp gene of temperate *Streptococcus*
873 *thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus*
874 *thermophilus* and *Lactococcus lactis*. Virology **350:**146-157.

875

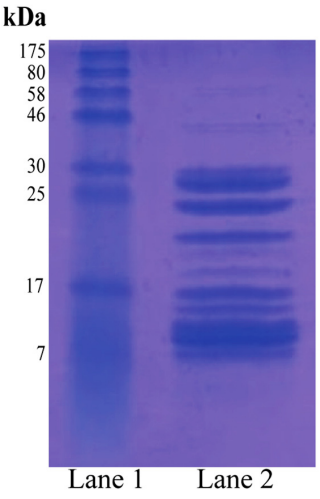
876

877

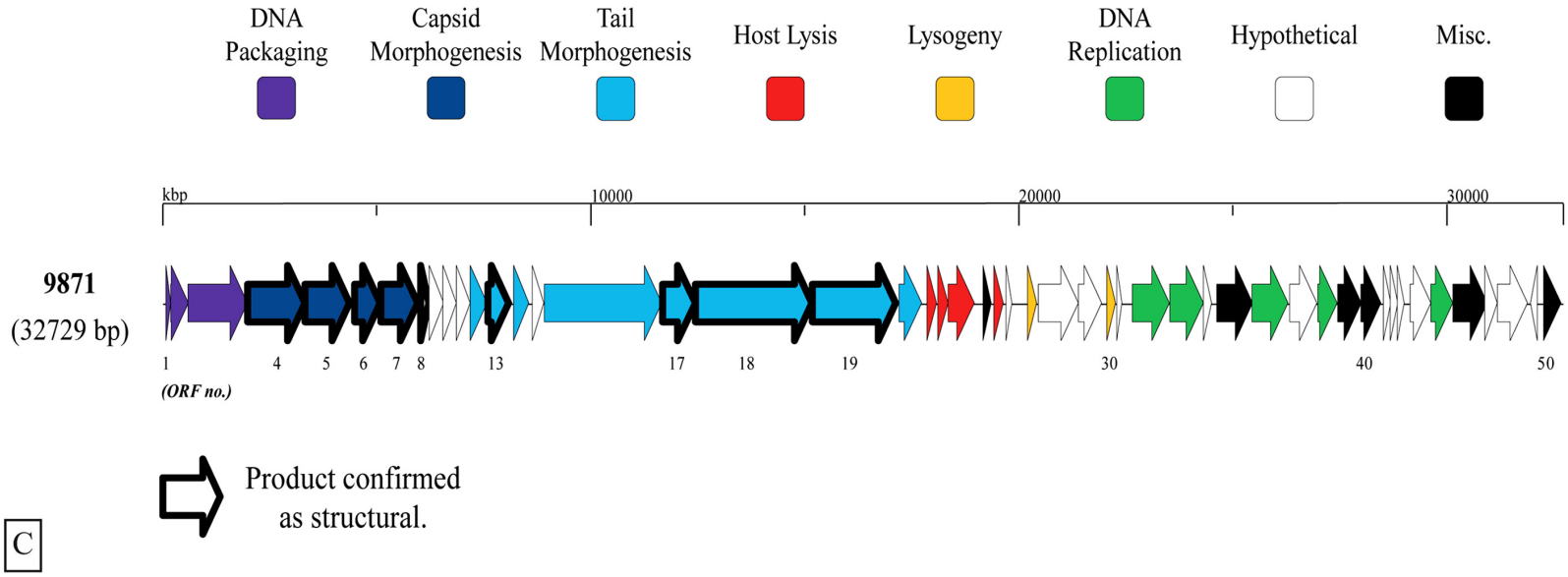

DNA Packaging
Capsid Morphogenesis
Tail Morphogenesis
Host Lysis
Lysogeny
DNA Replication
Hypothetical
Misc.
bp
10000
20000
30000
ORF no.
TP901-1
(37667 bp)
9871
(32729 bp)
9872
(33105 bp)
9873
(32813 bp)
9874
(32649 bp)
7201
(35466 bp)
35000
% AA identity:
1-50
51-79
80-94
> 95
A Resolvase
B Transcriptional regulator
C Intron
D RecT recombinase
E Exonuclease
F HNH endonuclease
G Methyltransferase

Phage	ORF no.	Putative function	No. of peptides	Amino acids	Coverage (%)
9871	4	Portal protein	4	53	11.9
	5	Minor capsid protein	4	41	11.8
	6	Scaffolding protein	3	42	21.1
	7	Major capsid protein	8	94	32.8
	8	Hypothetical protein	2	28	44.4
	13	Major tail protein	2	19	11.5
	17	Dit	2	27	10.7
	18	Tal	6	73	8.0
	19	Antireceptor	5	71	10.9

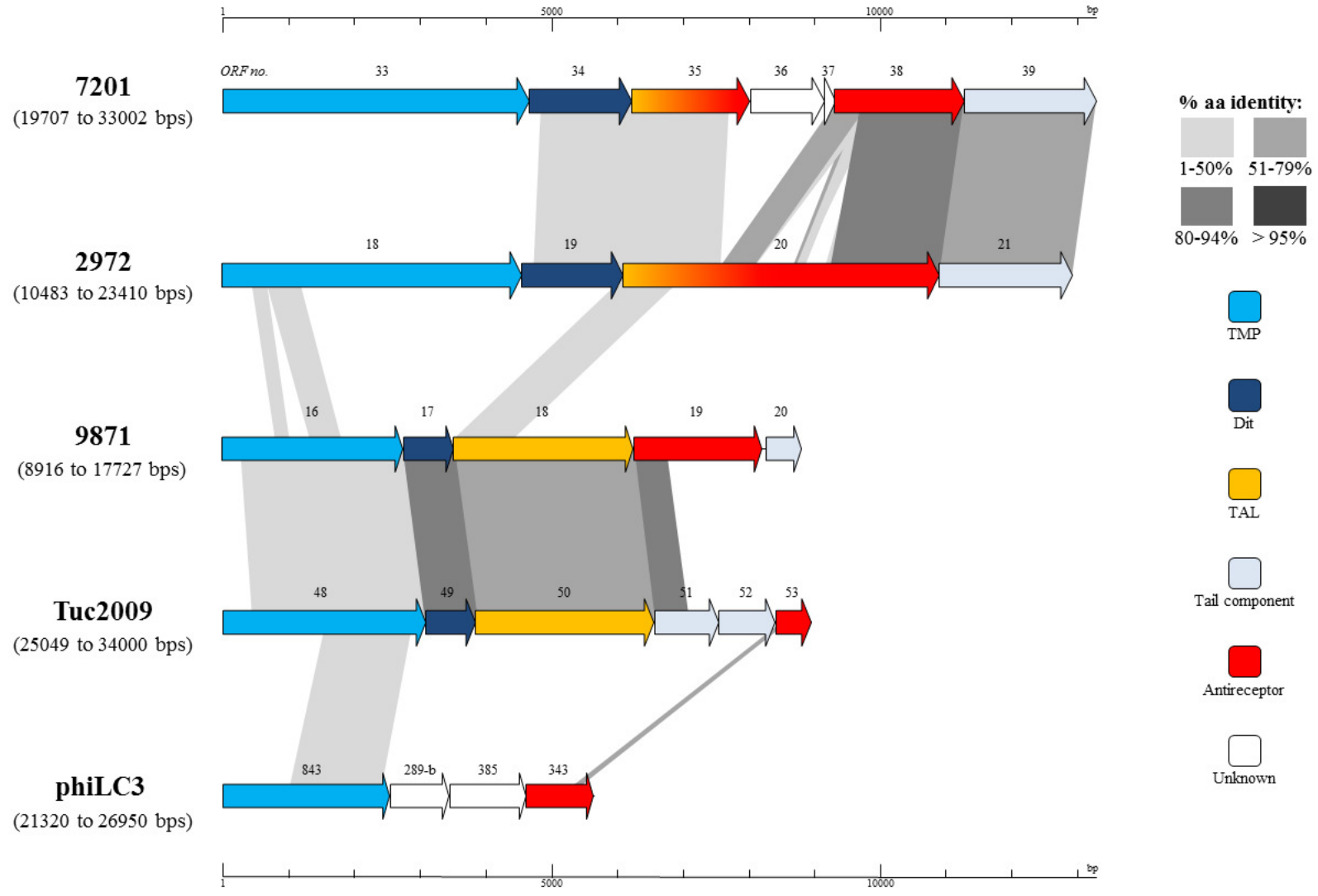
A

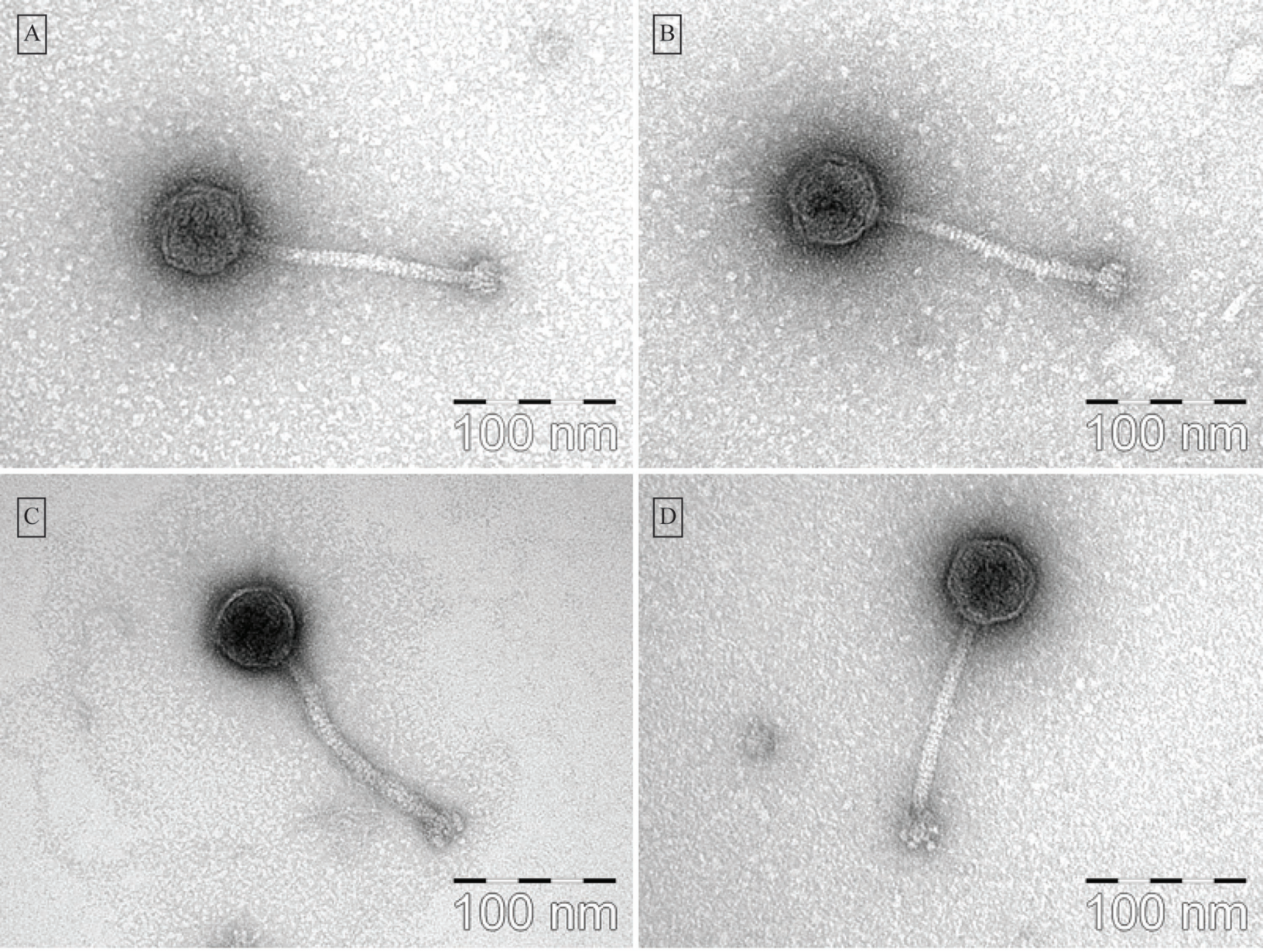


B



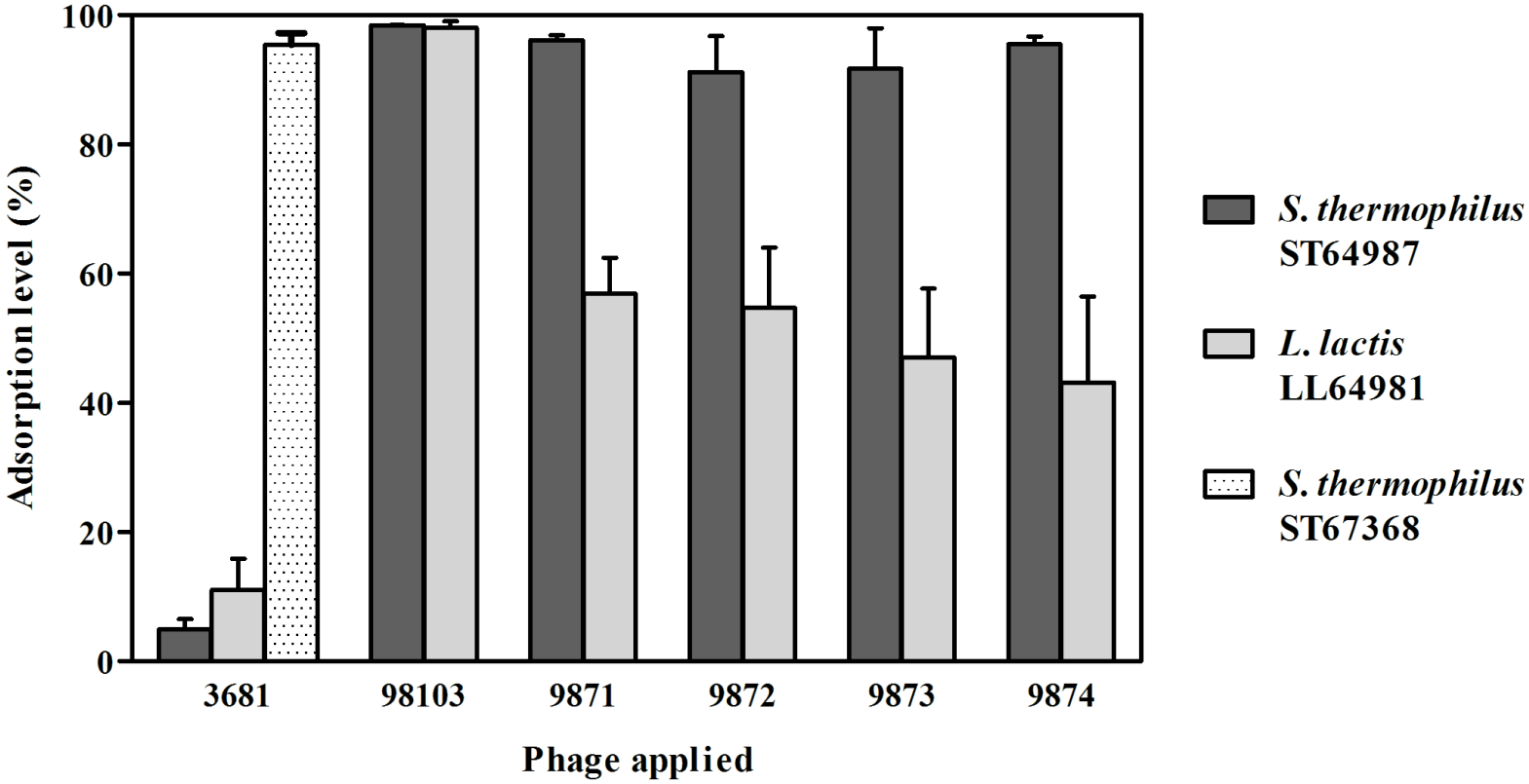
C

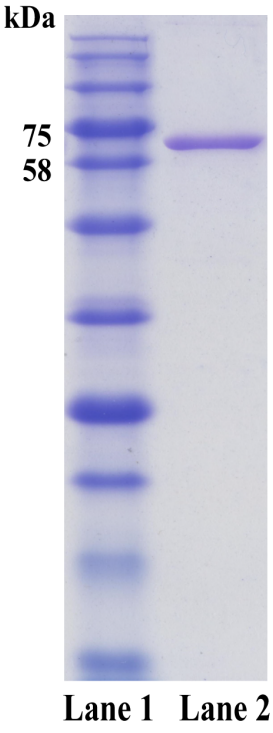

1
5000
10000
bp
ORF no.
7201
(19707 to 33002 bps)
33
34
35
36
37
38
39
2972
(10483 to 23410 bps)
18
19
20
21
9871
(8916 to 17727 bps)
16
17
18
19
20
Tuc2009
(25049 to 34000 bps)
48
49
50
51
52
53
phiLC3
(21320 to 26950 bps)
843
289-b
385
343
% aa identity:
1-50%
51-79%
80-94%
> 95%
TMP
Dit
TAL
Tail component
Antireceptor
Unknown



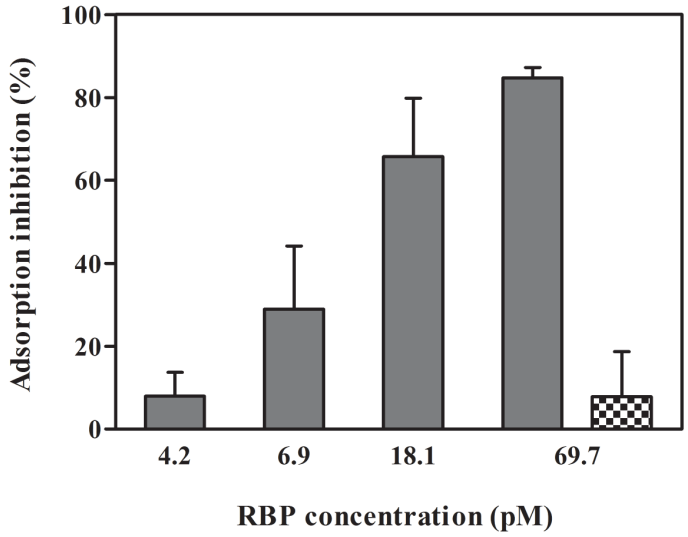
E

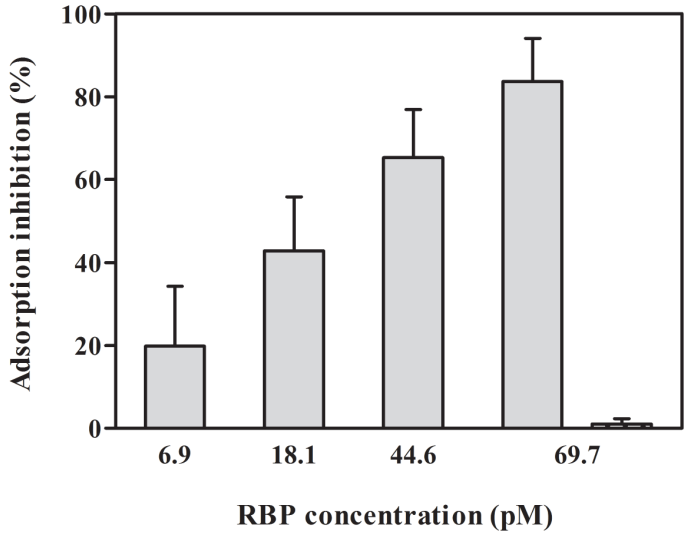
Characteristic	9871	9872	9873	9874
Head diameter (nm)	58.6 ± 1.7 (n=16)	58.4 ± 1.1 (n=11)	56.8 ± 1.8 (n=13)	58.3 ± 1.5 (n=14)
Tail length (nm)	139.9 ± 1.2 (n=13)	137.8 ± 2.3 (n=13)	135.1 ± 2.9 (n=12)	132.2 ± 2.7 (n=14)


100
80
60
40
20
0
Adsorption level (%)
3681
98103
9871
9872
9873
9874
Phage applied
S. thermophilus ST64987
L. lactis LL64981
S. thermophilus ST67368


kDa
75
58
Lane 1
Lane 2

A


Adsorption inhibition (%)
100
80
60
40
20
0
4.2
6.9
18.1
69.7
RBP concentration (pM)
Host / Phage combination:
ST64987 / 9871
ST67368 / 3681


Adsorption inhibition (%)
100
80
60
40
20
0
6.9
18.1
44.6
69.7
RBP concentration (pM)
Host / Phage combination:
LL64981 / 9871
LL64981 / 98103

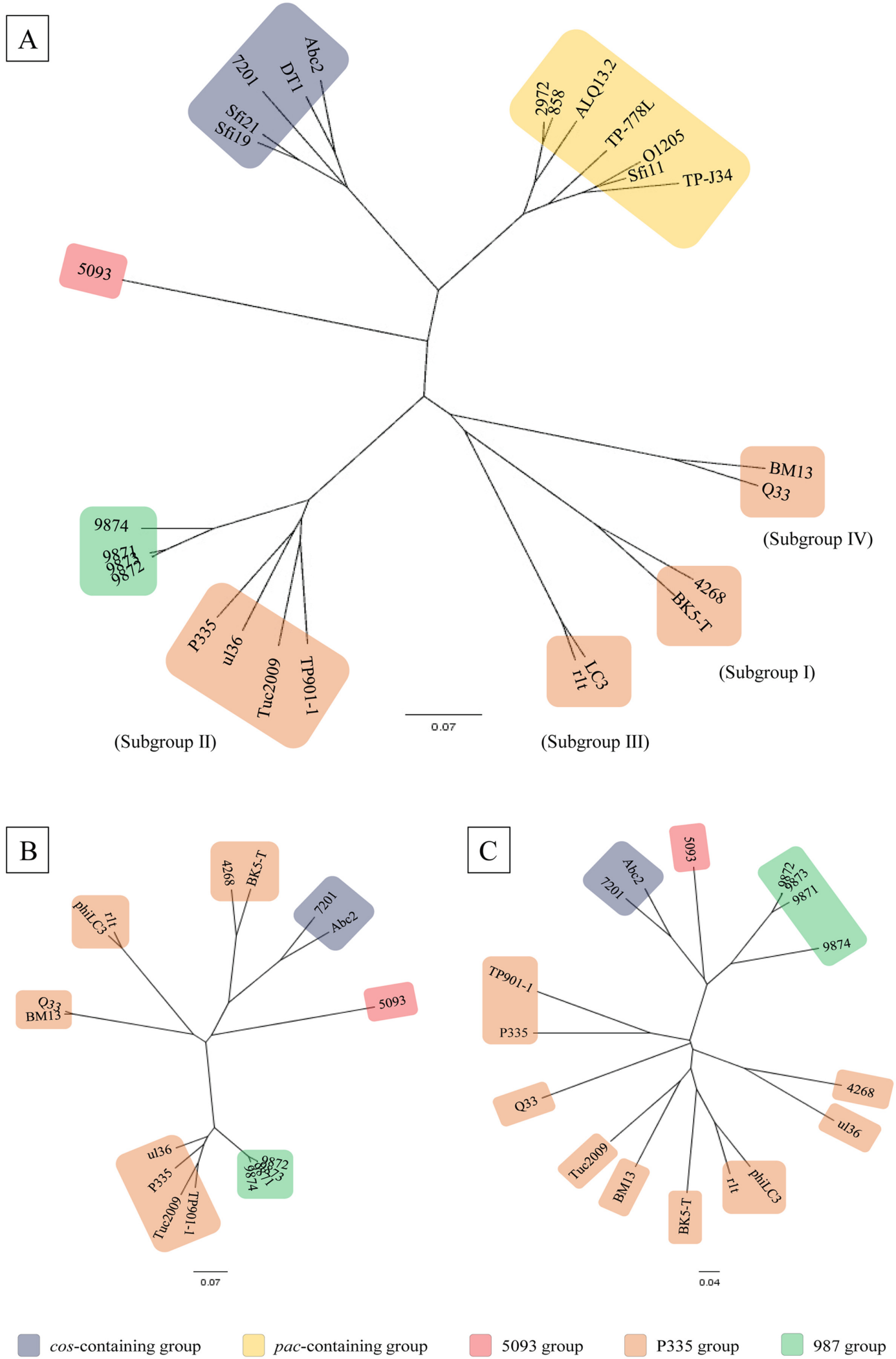

A
Abc2
DT1
7201
Sfi21
Sfi19
2972
858
ALQ13.2
TP-778L
O1205
Sfi11
TP-J34
5093
BM13
Q33
(Subgroup IV)
9874
9871
9873
9872
4268
BK5-T
P335
ul36
Tuc2009
TP901-1
LC3
r1t
(Subgroup I)
0.07
(Subgroup II)
(Subgroup III)
B
4268
BK5-T
7201
Abc2
phiLC3
r1t
Q33
BM13
5093
ul36
P335
Tuc2009
TP901-1
9872
9873
9871
9874
0.07
C
Abc2
7201
5093
9872
9873
9871
9874
TP901-1
P335
Q33
4268
ul36
Tuc2009
BM13
BK5-T
r1t
phiLC3
0.04
cos-containing group
pac-containing group
5093 group
P335 group
987 group

Table 1: Bacterial strains and bacteriophages used in this study.

Bacteria/Phages	**Description**	**Origin**
Bacterial strains		
ST64987	*Streptococcus thermophilus* host for phages 9871-4	DSM, The Netherlands
LL64981	*Lactococcus lactis* subsp. *lactis* host for phage 98103	"
ST67368	*S. thermophilus* host for phage 3681, as adsorption control	"
NZ9000	Transformation host	Kuipers *et al*., 1998
Phages		
9871	Lytic phage of *S. thermophilus* ST64987	DSM, The Netherlands
9872	"	"
9873	"	"
9874	"	"
3681	Lytic phage of *S. thermophilus* ST67368, as adsorption control	"
98103	Lytic phage of *L. lactis* LL64981, as adsorption control	"

Table 2: General characteristics of the genomes of phages 9871-4.

Characteristic	**9871**	**9872**	**9873**	**9874**
Length (bp)	32,729	33,105	32,813	32,649
Predicted ORFs (no.)	50	50	49	48
Coding (%)	92.2	91.4	91.7	89.0
GC content (%)	37.06	36.84	36.9	36.62