Biology and Genome Sequence of *Streptococcus mutans* Phage M102AD

Allan L. Delisle,^{a*} Ming Guo,^{a*} Natalia I. Chalmers,^{a*} Gerard J. Barcak,^b Geneviève M. Rousseau,^c and Sylvain Moineau^c

Department of Biomedical Sciences, School of Dentistry, University of Maryland, Baltimore, Maryland, USA^a; Department of Biochemistry and Molecular Biology, School of Medicine, University of Maryland, Baltimore, Maryland, USA^b; and Département de Biochimie, de Microbiologie et de Bio-informatique, Faculté des Sciences et de Génie, Groupe de Recherche en Écologie Buccale, Faculté de Médecine Dentaire, Université Laval, Félix d'Herelle Reference Center for Bacterial Viruses, Québec City, QC, Canada^c

M102AD is the new designation for a *Streptococcus mutans* phage described in 1993 as phage M102. This change was necessitated by the genome analysis of another *S. mutans* phage named M102, which revealed differences from the genome sequence reported here. Additional host range analyses confirmed that *S. mutans* phage M102AD infects only a few serotype c strains. Phage M102AD adsorbed very slowly to its host, and it cannot adsorb to serotype e and f strains of *S. mutans*. M102AD adsorption was blocked by c-specific antiserum. Phage M102AD also adsorbed equally well to heat-treated and trypsin-treated cells, suggesting carbohydrate receptors. Saliva and polysaccharide production did not inhibit plaque formation. The genome of this siphophage consisted of a linear, double-stranded, 30,664-bp DNA molecule, with a GC content of 39.6%. Analysis of the genome extremities indicated the presence of a 3'-overhang *cos* site that was 11 nucleotides long. Bioinformatic analyses identified 40 open reading frames, all in the same orientation. No lysogeny-related genes were found, indicating that phage M102AD is strictly virulent. No obvious virulence factor gene candidates were found. Twelve proteins were identified in the virion structure by mass spectrometry. Comparative genomic analysis revealed a close relationship between *S. mutans* phages M102AD and M102 as well as with *Streptococcus thermophilus* phages. This study also highlights the importance of conducting research with biological materials obtained from recognized microbial collections.

The group of oral bacteria generally associated with caries formation in mammals is referred to as the “mutans” group of streptococci. These Gram-positive lactic acid-producing organisms consist of at least nine serotypes (a to h and k) (10, 32, 40) distributed among seven phenotypically related species (49): *Streptococcus criceti* (serotype a), *Streptococcus rattii* (serotype b), *Streptococcus mutans* (serotypes c, e, f, and k), *Streptococcus sobrinus* (serotypes d and g), *Streptococcus downei* (serotype h), *Streptococcus macacae* (serotype c), and *Streptococcus ferus* (serotype c). It is generally accepted that *S. mutans* is primarily responsible for the initiation of human enamel caries since it has been identified as the presumed etiologic agent(s) in 70 to 100% of many patient populations studied (26). *S. mutans* typically produces a highly adhesive dextran that enables it to colonize tooth surfaces (18, 39).

Bacterial viruses (bacteriophages or phages) play an important role in the evolution of human complex communities, including the oral cavity (21, 34). Even though relatively little is known about *S. mutans* phages (48), they may be important ecological factors in determining the establishment, maintenance, and pathogenicity of this cariogen in the oral cavity. Phages of *S. mutans* may also be of considerable interest for the control and prevention of oral infections (18). In addition, some of their encoded gene products such as peptidoglycan hydrolases (endolysins) may have clinical utility as diagnostic and therapeutic agents to treat or prevent dental caries (17). They have the potential to lyse and kill this species of cariogenic bacteria *in vivo*, possibly in a targeted manner, leaving intact noncariogenic organisms believed to be important for maintaining good oral health.

All serotypes of *S. mutans* produce cell wall rhamnose-glucose polysaccharides (RGPs) consisting of a backbone structure of α 1,2- and α 1,3-linked rhamnosyl units with glucose side chains linked to alternate molecules on the cell wall (35, 41). Each sero-

type RGP has unique linkages of its glucose side chains (35). It has been shown recently that the glucose side chain of RGP is indispensable for adsorption and infection of *S. mutans* phage M102 (41).

It has been recently suggested that *S. mutans* is often attacked by M102-like bacteriophages (48), and the complete genomic sequence of the reference phage M102 was determined. This phage was isolated and reported by Tiraby's group (Université Paul Sabatier, France) in 1988 (5). It contains 31,147 bp with a GC content of 39.2% and cohesive ends (3' overhangs of 11 nucleotides [nt]) and carries 41 open reading frames (ORFs) (47). This phage, which belongs to the *Siphoviridae* family, has a tail of 269 nm in length and an icosahedral capsid with a diameter of 63 nm. This reference phage is also conserved at the Félix d'Herelle Reference Center for Bacterial Viruses.

Another stock of phage M102 was received in the late 1980s at the University of Maryland from the same French group (5), and

Received 29 November 2011 Accepted 18 January 2012

Published ahead of print 27 January 2012

Address correspondence to Sylvain Moineau, Sylvain.Moineau@bcm.ulaval.ca.

* Present address: Allan L. Delisle, Carolina Preserve, Cary, North Carolina, USA; Ming Guo, Department of Neurology, School of Medicine, University of Maryland, Baltimore, Maryland, USA; Natalia I. Chalmers, Department of Pediatric Dentistry, University of Maryland, Baltimore, Maryland, USA.

This publication is dedicated to the memory of Gerard J. Barcak, who died 22 September 2007, after a valiant year-long battle with cancer. He was a driving force behind this project and will be sorely missed by his many friends and colleagues.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/AEM.07726-11

this phage was characterized to a limited extent a few years later (11). This phage was believed to be identical to the reference phage M102 described above by van der Ploeg (47), who obtained it from the same original source. However, genome sequencing of the American stock revealed some distinct differences from the reference M102 phage. Because of these differences, we have renamed this phage M102AD (vB_SmuS_M102AD).

Here, we present the complete genomic sequence of the serotype c-specific *S. mutans* phage M102AD as well as its structural proteome. We also report the adsorption features of this strictly virulent phage in comparison with those of two other *S. mutans* phages, namely, e10 (infecting serotype e strains) and f1 (infecting serotype f strains). Finally, its evolutionary relatedness to other phages of Gram-positive bacteria is discussed.

MATERIALS AND METHODS

Phages and bacterial hosts. Phages M102AD, e10, and f1 (11) were grown on *S. mutans* host strains P42S (serotype c), LM7S (serotype e), and OMZ 175S (serotype f), respectively. These host bacteria are all spontaneous streptomycin-resistant mutants of the original host strains (11). Strains of *S. mutans* were routinely grown in TSYPS broth (30 g/liter Trypticase soy broth [TSB; BBL], 5 g/liter yeast extract, 5 g/liter K_2HPO_4 , and 0.5 g/liter streptomycin sulfate) at 34°C under static conditions, in air, or on plates containing TSYPS solidified with 15 g/liter agar. Top agar overlays for phage assays contained 7.5 g/liter agar or 3.5 g/liter agarose. For adsorption rate determinations, *S. mutans* cultures were grown in TSYPS pretreated with 50 mg/ml of dextranase (Sigma D8144; 400 to 800 units/mg) and 50 mg/ml invertase (Sigma I4504; grade VII; >300 units/mg) to remove trace amounts of sucrose and glucans (30). After the enzymes were added, the medium was incubated for 1 h at 50°C before autoclaving. Both phage M102AD and its host are stored at the Felix d'Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca).

Phage amplification and purification. Phage were grown by infecting log-phase TSYPS broth cultures of the appropriate hosts at a multiplicity of infection (MOI) of 0.1 and incubating them for 16 h at 34°C. Phage DNA was extracted and purified using a standard proteinase K-cetyltrimethylammonium bromide (CTAB) procedure (19), modified as previously described (12), and used as sequencing templates (see below). For genomic and proteomic studies, phage were concentrated by polyethylene glycol precipitation, followed by two rounds of ultracentrifugation through a CsCl gradient (9).

Host range analysis. The host range of phage M102AD was tested by spreading log-phase bacterial cultures on TSB agar with BHI top agar overlay and, after solidification, spotting the overlays with undiluted and 10-fold serial dilutions of phage. The TSAYE broth, top agar, and agar plates were also used (31). Plates were then incubated overnight at 37°C. Phage sensitivity was indicated by the presence of discrete plaques in areas spotted with serially diluted phage suspensions.

Phage adsorption studies. Cells grown overnight in enzyme-treated TSYPS medium (see above) were harvested by centrifugation for 20 min at $8,000 \times g$ (Sorvall SS34 rotor) and resuspension of the pellets in STMC buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 2 mM $CaCl_2$, pH 7.8). This wash was repeated, and the cells were diluted into sterile STMC buffer to an optical density at 600 nm (OD_{600}) of 1.0. We added 0.1 ml of phage suspension (10^6 phage/ml) to 0.9 ml of the washed cell suspension (approximately 5×10^9 cells/ml) at various time intervals during incubation at 37°C, and samples were removed and centrifuged (5 min at 8,000 rpm in a Sorvall SS34 rotor). The number of free unadsorbed phage in the supernatants was then determined by plating 0.1 ml of serial 10-fold dilutions with 0.3 ml of the appropriate host cells (grown overnight in TSYPS broth at 37°C) in 3 ml of molten top agar. Controls consisted of the equivalent amounts of phage added to sterile buffer, which were incubated and centrifuged and the titers of which were estimated to determine the amount of phage added to each experimental mixture. Adsorption

rates were calculated from the slope of log plots of unadsorbed phage at each time interval.

The ability of phage to adsorb to heat-treated cells was determined with log-phase cells grown in enzyme-treated medium (30) after washing and resuspending them in STMC buffer to an OD_{600} of 1.0, as above. Heat-treated cells were prepared by placing cell suspensions in an 85°C water bath for 30 min. After cooling to room temperature, 0.1 ml of phage suspension was added to 0.9 ml of cells and adsorption was allowed to proceed for 20 and 40 min at 37°C, after which the mixtures were centrifuged at low speed and the numbers of free unadsorbed phage remaining in the supernatants were determined by diluting and plating with the appropriate host. Controls consisted of phage incubated with unheated cell suspensions.

The effect of saliva on phage adsorption was tested by preincubating 0.9 ml of washed cells with 0.1 ml of freshly collected, stimulated saliva (pooled from three lab personnel) and incubating the mixture for 20 min at 37°C. Phage was then added and allowed to adsorb to the treated cells for 20 or 40 min. The mixtures were centrifuged, and the number of unadsorbed phage remaining in the supernatant was determined as described above. Controls were prepared by adding the same amount of phage to buffer alone and to buffer with 0.1 ml of saliva.

Phage neutralization assays. To determine whether *S. mutans* serotype-specific antisera could block phage adsorption, 20 μ l of antiserum was spotted onto top agar overlays prepared with sufficient phage to produce semiconfluent lysis (ca. 5,000 plaques/plate). After incubation, the absence of plaques in the antiserum spots indicated that phage adsorption had not occurred. Blocking of phage adsorption was also quantitatively assessed by pretreating 0.9 ml of STMC-washed host cells with 10 μ l of serotype c-, e-, or f-specific antiserum for 30 min at 37°C, adding phage, allowing 30 min for phage adsorption, and centrifuging the mixtures at low speed to remove cells and adsorbed phage. The numbers of free unadsorbed phage in each supernatant were determined by titration on plates. The six *S. mutans* rabbit antisera used were specific for serotype a, b, c, d/g, e, or f and were obtained from and prepared by the National Institute of Dental Research (NIH, Bethesda, MD) as part of the 1980 National Caries Program. Each serum was preadsorbed with cells of all five of the nonhomologous serotypes to remove cross-reacting antibodies. Protein concentrations of the resulting sera ranged from 9 to 13 mg/ml. The sera were diluted 1:2 with nonimmune rabbit serum and preserved with 0.1% sodium azide.

Effect of sugars on phage adsorption. Efficiency-of-plaquing (EOP) assays were used to determine if the extracellular glucan polysaccharides that *S. mutans* synthesizes from sucrose reduce or block phage adsorption. Host cells were pregrown overnight in TSY broth containing either 2.5 g/liter glucose or 5.0 g/liter sucrose. The resulting cultures were used as inocula to seed both TSY-glucose and TSY-sucrose top agars containing known dilutions of phage. The mixtures were then overlaid onto TSY-glucose or TSY-sucrose bottom agar-containing plates. The EOP was calculated as the ratio of phage titer under experimental conditions to the titer using glucose-containing medium.

Phage morphology. Electron micrographs were obtained as described elsewhere (14). To prepare the grids, 10 μ l of 2% uranyl acetate was deposited in a sterile petri dish and a copper Formvar carbon-coated grid (200 mesh; Pelco International) was deposited upside down on the uranyl acetate drop. The phage preparation (10 μ l) was added 30 s after the grid and mixed with the uranyl acetate by pipetting up and down. The grid was removed after 60 s and put on a blotting paper facing up. The grids were dried for at least 5 min prior to observation (14). Phages were observed at 80 kV with a JEOL 1230 transmission electron microscope.

Genome sequencing. Initial sequences of the M102AD genome were obtained by sequencing random Sau3AI partial digest fragments cloned (using conventional methods) into the vector pIBI (International Biotechnologies), using standard pUC sequencing primers. The initial sequences were extended in both directions to their 5' ends by primer walking. The Biopolymer Laboratory (University of Maryland, Baltimore,

MD) carried out the sequencing and synthesized all sequencing and cloning primers. The *cos* site was determined by sequencing ligated phage M102AD DNA preparations using converging PCR primers (synthesized by Eurogentec) at the Pavillon Charles-Eugene Marchand on the campus of the Université Laval.

Identification of genes encoding structural proteins. The protein concentrations of CsCl-purified phage preparations were measured in a Beckman DU530 spectrophotometer using the Bradford method (7). Samples (7.2 μg) were mixed with 2 $\times$ sample loading buffer and boiled for 5 min before being loaded onto PAGE gels, composed of a 4% stacking gel and a 12% separating gel (22). After migration, the gel was stained with Coomassie blue and protein bands were extracted from the gel, digested with trypsin, and identified by nano-high-pressure liquid chromatography (HPLC)–tandem mass spectrometry (MS/MS) at the proteomics platform of the Génome Québec Innovation Centre at McGill University.

RESULTS AND DISCUSSION

Phage morphology. Electron microscopy confirmed that phage M102AD belongs to the *Siphoviridae* family with B1 morphology (Fig. 1). The diameter of its capsid was estimated at 67 ± 0.4 nm, while the length of its noncontractile tail was estimated at 283 ± 7 nm with a width of 8.3 ± 0.1 nm. These dimensions are in agreement with those observed for phage M102 (47).

Phage	Host strain	Adsorption rate (ml/min)
M102AD	P42S	1.5×10^{-10}
e10	LM7S	1.7×10^{-11}
f1	OMZ 175S	1.4×10^{-10}

Phages M102AD and f1 adsorbed almost exclusively to their respective host serotypes, explaining why they cannot form plaques on strains of alternative serotypes (Fig. 2). Although phage e10 adsorbed to serotype c strain P42, this phage could not form plaques on P42 or on serotype f strain OMZ175. Thus, each of the three phages is serotype specific.

Then, we tested whether *S. mutans* serotype-specific antisera could block phage adsorption. Indeed, anti-c, anti-e, and anti-f antisera prevented plaque formation by the three phages on their cognate hosts (data not shown). Antisera to other mutans-like species of oral streptococci (serotypes a, b, d/g, and h) did not prevent plaque formation. The carbohydrate surface antigens that define *S. mutans* serotypes are likely involved in phage adsorption. In contrast, protein components of the cell wall do not seem to be involved in the adsorption process since heat- and trypsin-treated cells allowed phage to adsorb to the same extent as did untreated

Applied and Environmental Microbiology

TABLE 2 Effect of sucrose-derived surface polysaccharides on the EOPs of *S. mutans* phages

Phage	Host	Sugar in liquid medium	Sugar in plating medium	EOP
M102AD	P42S	Glucose	Glucose	1.0
M102AD	P42S	Glucose	Sucrose	0.6
M102AD	P42S	Sucrose	Sucrose	0.6
M102AD	P42S	Sucrose	Glucose	0.3
e10	LM7S	Glucose	Glucose	1.0
e10	LM7S	Glucose	Sucrose	0.7
e10	LM7S	Sucrose	Sucrose	0.2
e10	LM7S	Sucrose	Glucose	0.2
f1	OMZ 175S	Glucose	Glucose	1.0
f1	OMZ 175S	Glucose	Sucrose	0.3
f1	OMZ 175S	Sucrose	Sucrose	0.8
f1	OMZ 175S	Sucrose	Glucose	0.7

cells (data not shown). Finally, the lack of inhibition of phage adsorption by saliva indicates that salivary components are not likely to interfere with phage infection and replication in the oral cavity.

Finally, as the data shown in Table 2 indicate, the unique sucrose-derived glucan polymers produced by *S. mutans* do not significantly interfere with phage adsorption or replication. These data support the view that the production of polysaccharides does not confer a significant phage resistance phenotype (8, 13, 45).

Complete nucleotide sequence. The linear genome of phage M102AD contains 30,664 bp with an average G+C content of 39.6% and has 3'-overhanging *cos* sites (5'-CCGCGTGAATA-3') of 11 nucleotides. The genome of phage M102AD is slightly smaller than that of phage M102 (31,147 bp) while the *cos* sites are identical (47). To our knowledge, M102 and M102AD are the only *S. mutans* phage genomes available. All of the 40 ORFs of phage M102AD were identified on the same strand, and they started with ATG, TTG, or GTG initiation codons. Of the 40 ORFs, 23 could be assigned a putative function. Moreover, every ORF was preceded by a region with similarities to the Shine-Dalgarno sequence, AG AAAGGAGG, which is complementary to the 10 nucleotides of the 3' end of the 16S rRNA of *S. mutans*. Two to 11 nucleotides separate the Shine-Dalgarno sequence from the initiation codon (Table 3). No tRNA was found in the genome of phage M102AD. Two significant intergenic regions were observed between *orf37* and *orf38* (385 bp) as well as between *orf20* and *orf21* (290 nt).

Comparison with the M102 phage genome. The M102 and M102AD genomes share 90.8% identity at the nucleotide level and thus are, as expected, highly related. Nonetheless, phage M102AD has one small extra ORF (*orf40*) not present in phage M102 while M102 also has three small supplementary ORFs (*orf22*, *orf30*, and *orf41*). For these four cases, it appears that two genes (*orf22* and *orf30*) were acquired in the genome of phage M102 (or deleted in M102AD) while the other two ORFs (*orf40*/M102AD and *orf41*/M102) are phage specific, as they are located at the same genomic location. Interestingly, only four deduced ORFs share 100% identity (ORF18/holin, ORF19/endolysin, ORF20/endolysin, and ORF25/unknown function) at the amino acid level (Table 3). Overall, 32 ORFs (out of 40) from phage M102AD shared over 90% identity with *S. mutans* phage M102. Only ORF40 had no homology with a known protein, and thus, its origin is unknown (Table 3).

On the other hand, *orf34* of the M102AD genome (Table 3; Fig. 3) matches with a portion of the M102 genome that was not previously identified as an open reading frame (47). Upon further analyses, a new ORF (named *orf35b*) was added to the phage M102 genome, bringing its total to 42 ORFs compared to the 40 ORFs of phage M102AD. The position of *orf35b* on the genome of M102 is between bp 28827 and 29099, and it would encode a protein of 90 amino acids (aa). ORF34 of phage M102AD has 88% identity (79/90 amino acids) with ORF35b of phage M102, but its function is unknown.

Overall, the degree of divergence between two phages (M102 and M102AD) derived from the same laboratory is striking. The relatively recent diversification of the phage M102AD genome was due mainly to point mutations as well as gene deletion and additions (4, 20). At this time, it is unclear why and when these events occurred due to the lack of precise documentation on the handling and serial amplification of M102AD since the late 1980s.

Comparative analysis with *Streptococcus thermophilus* phages.

Twenty ORFs of phage M102AD also share identity with ORFs of some *Streptococcus thermophilus* phages, namely, those of the *cos*-type group (44). A genomic relationship between *S. mutans* phage M102 and *S. thermophilus* phage DT1 (46) was previously observed (47), although here we found phage M102AD to be closer to *S. thermophilus* phage 7201. Interestingly, ORF33 has a 73% amino acid match with ORF9 of 7201, which appears to be a single-stranded binding protein (38).

It has been demonstrated elsewhere that *S. thermophilus* phages can be separated into at least two different clusters based on the mode of packaging and the number of major structural proteins (24). *S. thermophilus cos*-type phages, such as phages 7201 and DT1, have cohesive genomic extremities and two major structural proteins. Interestingly, the genomic sequence of *S. mutans* phage M102AD has a 3'-overhanging *cos* site, and the analysis of its structural proteins clearly identified two major structural proteins, namely, ORF6 (major capsid protein) and ORF11 (major tail protein) (Fig. 4).

Moreover, 9 out of 10 structural proteins of phage M102AD identified by nano-HPLC-MS/MS were similar to structural proteins found in *cos*-type *S. thermophilus* phages such as phage DT1 (46). Many genes coding for these conserved ORF proteins are part of the core genome of *S. thermophilus* phages (36). Almost all the detected phage M102AD proteins had approximately the expected size, except for *orf12*, coding for the tape measure protein (TMP). The TMP, which is responsible for determining tail length, was found in two separate bands (Fig. 4). These bands corresponded to the N-terminal and C-terminal parts of this protein. The tape measure protein, which is very similar to that of the *S. thermophilus* phage MD2 (16), has two domains, namely, a CHAP (cysteine, histidine-dependent aminohydrolase/peptidase) domain and an SLT (soluble lytic transglycosylase) domain. They may be involved, after the phage adsorption process, in locally degrading the peptidoglycan to allow DNA entry into the cell. Finally, the putative receptor binding protein (RBP/ORF14) of phage M102AD has similarities at the amino acid level with two ORFs of phage 7201 (ORF35 and ORF38) and may reflect the modular organization of RBPs (25, 42).

At this time, it is unclear why *S. mutans* phages are closely related to *S. thermophilus* phages, at least from a structural perspective. The latest phylogenetic analyses of streptococci still classify *S. thermophilus*, which is not a natural inhabitant of the hu-

TABLE 3 Features of phage M102AD ORFs and the putative functions of their products

ORF	Start position	Stop position	Size (aa)	Molecular mass (kDa)	pI	SD sequence ^a (AGAAAGGAGG)	Putative function and motif ^a	Best match (% amino acid identity) ^c	E value ^d	Accession no.	Size ^b (aa)
1	170	538	122	14.3	9.5	<u>AGAAAGTTT</u> Tataa ATG	Endonuclease (HNHc)	ORF1 M102 (118/122, 97%)	2E-64	CA077351	122
2	560	898	112	13.1	5.5	<u>AGAGAATTGG</u> cg ATG	Small subunit terminase	ORF2 M102 (112/112, 100%)	7E-59	CA077352	112
3	885	2759	624	71.4	5.1	<u>GAAAGAGGCT</u> Gatatca ATG	Large subunit terminase	ORF3 M102 (609/624, 98%)	0E+0	CA077353	630
4	2961	3905	314	35.6	5.6	<u>AAAAAGGAG</u> Aaaat ATG	Portal protein	ORF4 M102 (307/314, 98%)	3E-179	CA077354	314
5	3902	4798	298	32.4	4.2	<u>TGAAAAGAGGT</u> gattactta TTG	Clp protease	ORF5 M102 (288/298, 97%)	3E-156	CA077355	298
6	4818	5954	378	41.0	5.0	<u>AGAAAAGAGG</u> Gaaaa ATG	Major capsid protein	ORF6 M102 (364/378, 96%)	0E+0	CA077356	378
7	5997	6308	103	11.6	4.4	<u>AGAAAAGAGG</u> Ggatagag GTG	DNA packaging protein	ORF7 M102 (101/103, 98%)	3E-51	CA077357	103
8	6305	6649	114	13.4	9.3	<u>TGAAAAGTGA</u> Gaatgaagc ATG	Head-tail joining protein	ORF8 M102 (112/114, 98%)	5E-58	CA077358	114
9	6642	7034	130	15.0	5.6	<u>TGTAAGGAA</u> Accgaaaac ATG	Tail component protein	ORF9 M102 (125/130, 96%)	1E-67	CA077359	130
10	7018	7374	118	13.4	4.4	<u>AGAAAAGGAT</u> Attgaaga ATG	Tail component protein	ORF10 M102 (112/118, 95%)	2E-57	CA077360	118
11	7392	8075	227	24.3	5.4	<u>AGAAAAGAGG</u> Aaataatca ATG	Major tail protein	ORF11 M102 (221/227, 97%)	8E-127	CA077361	227
12	8249	13450	1,733	188.7	9.7	<u>AGAAAGTCT</u> Gataat ATG	TMP ^e /CHAP/SLT	ORF12 M102 (1,693/1,733, 98%)	0E+0	CA077362	1,733
13	13450	14976	508	57.5	6.5	<u>ATTAAGGGT</u> Attatttg ATG	Tail component protein	ORF13 M102 (496/508, 98%)	0E+0	CA077363	508
14	14973	17129	718	81.0	5.0	<u>AGAGAACGGG</u> catatca ATG	Receptor-binding protein	ORF14 M102 (698/718, 97%)	0E+0	CA077364	718
15	17130	18887	585	64.7	5.0	<u>ATGTAGGGGG</u> ggccttag ATG	Minor structural protein	ORF15 M102 (565/585, 97%)	0E+0	CA077365	585
16	18908	19411	167	18.9	5.3	<u>CGAAAGGGGG</u> Tgagaat ATG	—	ORF16 M102 (166/167, 99%)	6E-90	CA077366	167
17	19428	19700	90	10.6	8.9	<u>AGATAGGGG</u> Attatgat TTG	—	ORF17 M102 (89/90, 99%)	8E-43	CA077367	90
18	19697	20164	155	16.8	6.6	<u>CGAAAAGGTA</u> Aagaaaa ATG	Holin	ORF18 M102 (155/155, 100%)	2E-81	CA077368	155
19	20180	21001	273	30.0	4.9	<u>AAAAAAGAGG</u> Tatttt ATG	Endolysin/Glyco 25/Acm	ORF19 M102 (273/273, 100%)	6E-159	CA077369	273
20	21001	21516	171	18.9	5.9	<u>ACAAAGGAA</u> Aata ATG	Endolysin/CHAP	ORF20 M102 (171/171, 100%)	5E-96	CA077370	171
21	21806	22660	284	32.3	9.6	<u>TAAATGCA</u> Ggagaa ATG	DNA replication protein	ORF21 M102 (265/272, 97%)	1E-154	CA077371	272
22	22809	23567	252	29.2	9.0	<u>AAATAGAGGG</u> Taagaa ATG	DnaC, replication protein/IstB	ORF23 M102 (231/252, 92%)	4E-134	CA077373	252
23	23567	23770	67	8.1	10.4	<u>AGACGAGAG</u> Gataaa ttttaATG	—	ORF24 M102 (61/67, 91%)	3E-26	CA077374	67
24	23767	23919	50	6.1	9.2	<u>ATTGAAGAGG</u> Tgttaataa ATG	—	ORF25 M102 (43/50, 86%)	4E-17	CA077375	56
25	23930	24178	82	9.3	4.2	<u>TGAGAGGTGA</u> aga ATG	—	ORF26 M102 (82/82, 100%)	3E-39	CA077376	82
26	24188	24934	248	28.1	7.7	<u>GAAATAGGAG</u> Tatgaa ATG	SSAP ^d /RecT	ORF27 M102 (230/248, 93%)	1E-132	CA077377	248
27	24948	25796	282	32.6	7.6	<u>AGAAAGTGT</u> Gattgc ATG	DUF1351	ORF28 M102 (234/282, 83%)	3E-131	CA077378	309
28	25878	26174	98	11.4	9.9	<u>AGAAAAGGTT</u> Aagcta ATG	—	ORF29 M102 (92/98, 94%)	7E-45	CA077379	98
29	26295	26573	92	10.6	9.6	<u>AGGAAAGAGG</u> Taaaacc ATG	—	ORF31 M102 (78/92, 85%)	1E-36	CA077381	93
30	26687	26914	75	9.0	9.8	<u>AGGAAGGAA</u> Aaaaa ATG	—	ORF32 M102 (68/75, 91%)	2E-32	CA077382	75
31	26917	27315	132	15.6	9.9	<u>GGAAAGATGG</u> gtaaa ATG	—	ORF33 M102 (106/131, 81%)	3E-58	CA077383	131
32	27319	27798	159	18.4	5.0	<u>AGAAAGTTT</u> Ggataaaa GTG	—	ORF34 M102 (154/159, 97%)	5E-84	CA077384	159
33	27795	28190	131	15.1	5.5	<u>AAAGAGGAGT</u> tgaa ttttATG	Single-stranded binding protein	ORF35 M102 (124/131, 95%)	2E-65	CA077385	131
34	28209	28481	90	11.2	8.8	<u>AAAAAGGTGG</u> Tgagaaa ATG	—	ORF35b M102 (79/90, 88%)	1E-48		90
35	28456	28890	144	16.8	9.0	<u>GGAGACGAGT</u> gttt ATG	Holliday junction resolvase (RusA)	ORF36 M102 (138/144, 96%)	4E-66	CA077386	145
36	28887	29291	134	16.2	6.0	<u>ATTAAGGAG</u> Atattagaaa ATG	DUF1642	ORF37 M102 (109/134, 81%)	7E-61	CA077387	134
37	29291	29488	65	7.3	9.5	<u>GCTCAGGAGG</u> tcgtata ATG	—	ORF38 M102 (23/33, 70%)	7E-5	CA077388	33
38	29873	30217	114	13.2	5.2	<u>AAAGAGGAG</u> Caatc ATG	—	ORF39 M102 (106/114, 93%)	6E-52	CA077389	114
39	30226	30417	63	7.4	4.0	<u>ATAAGCGAGC</u> ga TTG	—	ORF40 M102 (57/63, 90%)	3E-25	CA077390	63
40	30494	30640	48	5.9	10.2	<u>ATAACGCTA</u> Attaa ATG	—				

^a —, unknown function.^b Size of the aligned protein.^c TMP, tape measure protein.^d SSAP, single-strand annealing protein.^e Percentage calculated on the smallest protein length.^f Corresponding to the best significant expect (E) value.^g Nucleotides identical to the RBS consensus sequence are underlined. The putative RBS sequence is in uppercase letters. The interspaced nucleotides between the RBS sequence and the start codon are in lowercase letters. The start codon is in boldface.

man oral mucosa, into the salivarius group with *Streptococcus salivarius* and *Streptococcus vestibularis* (33). Clearly, other streptococcal phage genomic sequences and the associated analysis of gene functions are needed to get further insight into their evolution.

Comparison with endolysins of other streptococcal phages.

As indicated above, the lysis module is 100% conserved between M102 (47) and M102AD. The M102AD phage has two putative endolysins encoded by *orf19* and *orf20*. ORF19 con-

tains the Acm (1,4- β -N-acetylmuramidase) domain, which belongs to the glycosyl hydrolase 25 family (27). This 273-aa ORF corresponds to amino acids 148 through 366 of the 443 aa of the *Streptococcus agalactiae* B30 phage endolysin (15). ORF20 has a CHAP domain of 171 aa that is similar to amino acids 3 to 138 of the 250 aa of the *Streptococcus pyogenes* MGAS10394 phage endolysin (6). Usually, the CHAP domain can be found at the N-terminal end of the endolysin and the Acm domain can be found at the C-terminal end. Here, both domains were

FIG 3 Schematic representation of the genomic organization of *S. mutans* phage M102AD and M102 genome maps compared to *S. thermophilus* phage 7201. Each line represents a different phage genome, and each arrow represents a putative protein. Gray shading connects regions conserved with the corresponding percentages in amino acid identity. The percentages were calculated using the smallest proteins. The black ORFs are unique. Arrows with thick lines represent structural proteins observed on SDS-PAGE gels and identified by MS sequencing.

found on two deduced proteins. The CHAP domain likely carries out endopeptidase activity responsible for cell lysis, while the Acm domain carries out glycosidase activity.

Microbial resources. Finally, this study reinforces the need to thoroughly characterize any given phage of interest and to carry

out authentication testing (sequencing, restriction patterns, host range, etc.) to avoid the use of the wrong organism in investigations. Our study also points to the need to deposit key phage representatives in public culture collections and, ideally, in more than one collection around the world to ensure proper maintenance.

FIG 4 Analysis of phage M102AD structural proteins. (Left panel) Coomassie blue staining of a 12% SDS-PAGE gel showing structural proteins from M102AD CsCl-purified virions. The sizes (in kDa) of the different proteins from the broad-range molecular mass standards (M) are indicated on the left. Lines and letters on the right indicate bands excised from the gel and identified by nano-HPLC-MS/MS. (Right panel) Identification of phage structural proteins.

nance and continued access to them for future studies (14). Too often, biological materials are either lost or mishandled due to the retirement of principal investigators. A well-ordered detailed documentation system should also be maintained.

ACKNOWLEDGMENTS

We thank B.-A. Conway for editorial assistance; C. Levesque for suggestions; R. Young for helpful discussions; and B. Burall, F. Williams, and B. Saville for performing the phage adsorption and EOP experiments as well as D. Tremblay for the electron micrograph. We also thank G. Nicholas and M. Lavoie for providing additional strains of *S. mutans* for host range testing.

A.L.D. acknowledges the financial support of the NIDCR, NIH, for support of the cloning and sequencing experiments conducted at UM (grant DE13181). S.M. is grateful to the Canadian Institutes of Health Research (Team Grant-Emerging: Novel Alternatives to Antibiotics) for financial support of the sequencing, protein, and computer analyses conducted at UL.

REFERENCES

- Adams MH. 1959. Bacteriophages. Interscience Publishing, Inc, New York, NY.
- Ajdić D, et al. 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. Proc. Natl. Acad. Sci. U. S. A. 99:14434–14439.
- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
- Arber W. 2003. Elements for a theory of molecular evolution. Gene 317: 3–11.
- Armau E, Bousque JL, Boue D, Tiraby G. 1988. Isolation of lytic bacteriophages for *Streptococcus mutans* and *Streptococcus sobrinus*. J. Dent. Res. 67:121.
- Banks DJ, et al. 2004. Progress toward characterization of the group A *Streptococcus* metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. J. Infect. Dis. 190:727–738.
- Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–254.
- Broadbent JR, McMahon DJ, Welker DL, Oberg CJ, Moineau S. 2003. Biochemistry, genetics, and applications of exopolysaccharide production in *Streptococcus thermophilus*: a review. J. Dairy Sci. 86:407–423.
- Chibani Azaiez SR, Fliss I, Simard RE, Moineau S. 1998. Monoclonal antibodies raised against native major capsid proteins of lactococcal c2-like bacteriophages. Appl. Environ. Microbiol. 64:4255–4259.
- Coykendall AL. 1989. Classification and identification of the viridans streptococci. Clin. Microbiol. Rev. 2:315–328.
- Delisle AL, Rostkowski CA. 1993. Lytic bacteriophages of *Streptococcus mutans*. Curr. Microbiol. 27:163–167.
- Delisle AL, Barcak GJ, Guo M. 2006. Isolation and expression of the lysis genes of *Actinomyces naeslundii* phage Av-1. Appl. Environ. Microbiol. 72:1110–1117.
- Deveau H, van Calsteren M-R, Moineau S. 2002. The effect of exopolysaccharides on phage-host interactions in *Lactococcus lactis*. Appl. Environ. Microbiol. 68:4364–4369.
- Deveau H, Labrie SJ, Chopin M-C, Moineau S. 2006. Biodiversity and classification of lactococcal phages. Appl. Environ. Microbiol. 72:4338–4346.
- Donovan DM, et al. 2006. The cell lysis activity of the endolysin from the *Streptococcus agalactiae* bacteriophage B30 relies on the CHAP endopeptidase domain. Appl. Environ. Microbiol. 72:5108–5112.
- Duplessis M, Lévesque CM, Moineau S. 2006. Characterization of *Streptococcus thermophilus* host range phage mutants. Appl. Environ. Microbiol. 72:3036–3041.
- Fischetti VA. 2010. Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. Int. J. Med. Microbiol. 300:357–362.
- Hitch G, Pratten J, Taylor PW. 2004. Isolation of bacteriophages from the oral cavity. Lett. Appl. Microbiol. 39:215–219.
- Kricker M, Carlson K. 1994. Isolation of T4 phage DNA, p 455–456. In Karam JD, et al (ed), Molecular biology of bacteriophage T4. ASM Press, Washington, DC.
- Labrie SJ, Moineau S. 2007. Abortive infection mechanisms and prophage sequences significantly influence the genetic make-up of emerging lytic lactococcal phages. J. Bacteriol. 189:1482–1487.
- Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8:317–327.
- Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11–16.
- Le Marrec C, et al. 1997. Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. Appl. Environ. Microbiol. 63:3246–3253.
- Lévesque C, et al. 2005. Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. Appl. Environ. Microbiol. 71:4057–4068.
- Loesche WJ. 1986. Role of *Streptococcus mutans* in human dental decay. Microbiol. Rev. 50:353–380.
- López R, García E. 2004. Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. FEMS Microbiol. Rev. 28:553–580.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.
- Lukashin AV, Borodovsky M. 1998. GenMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26:1107–1115.
- Lü-Lü, et al. 1992. Chelating agents inhibit activity and prevent expression of streptococcal glucan-binding lectins. Infect. Immun. 60:3807–3813.
- Morency H, Mota-Meira M, Lapointe G, Lacroix C, Lavoie MC. 2001. Comparison of the activity spectra against pathogens of bacterial strains producing a mutacin or a lantibiotic. Can. J. Microbiol. 47:322–331.
- Nakano K, Nomura R, Nakagawa I, Hamada S, Ooshima T. 2004. Demonstration of *Streptococcus mutans* with a cell wall polysaccharide specific to a new serotype, k, in the human oral cavity. J. Clin. Microbiol. 42:198–202.
- Pombert J-F, Sistek V, Boissinot M, Frenette M. 2009. Evolutionary relationships among salivarius streptococci as inferred from multilocus phylogenies based on 16S rRNA-encoding, recA, secA, and secY gene sequences. BMC Microbiol. 9:232.
- Pride DT, et al. 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. Genome Res. 21:126–136.
- Pritchard DG, Gregory RL, Michalek SM, McGhee JR. 1986. Biochemical aspects of serotype carbohydrate antigens of *Streptococcus mutans*, p 39–49. In Hamada S, et al (ed), Molecular microbiology and immunology of *Streptococcus mutans*. Elsevier, Amsterdam, Netherlands.
- Quiberoni A, Moineau S, Rousseau GM, Reinheimer J, Ackermann H-W. 2010. Review: *Streptococcus thermophilus* bacteriophages. Int. Dairy J. 20:657–664.
- Rousseau GM, Moineau S. 2009. Evolution of *Lactococcus lactis* phages within a cheese factory. Appl. Environ. Microbiol. 75:5336–5344.
- Scaltriti E, et al. 2009. Structure and function of phage p2 ORF34_{p2}, a new type of single stranded DNA binding protein. Mol. Microbiol. 73:1156–1170.
- Schilling KM, Bowen WH. 1992. Glucans synthesized in situ in experimental salivary pellicle function as specific binding sites for *Streptococcus mutans*. Infect. Immun. 60:284–295.
- Schleifer KH, Kilpper-Bälz R, Kraus J, Gehring F. 1984. Relatedness and classification of *Streptococcus mutans* and “mutans-like” streptococci. J. Dent. Res. 63:1047–1050.
- Shibata Y, Yamashita Y, van der Ploeg JR. 2009. The serotype-specific glucose side chain of rhamnose-glucose polysaccharides is essential for adsorption of bacteriophage M102 to *Streptococcus mutans*. FEMS Microbiol. Lett. 294:68–73.
- Spinelli S, et al. 2006. Lactococcal bacteriophage p2 receptor binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat. Struct. Mol. Biol. 13:85–89.
- Staden R. 1996. The Staden sequence analysis package. Mol. Biotechnol. 5:233–241.
- 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.
45. Sutherland IW, Hughes KA, Skillman LC, Tait K. 2004. The interaction of phage and biofilms. FEMS Microbiol. Lett. 232:1–6.
46. Tremblay DM, Moineau S. 1999. Complete genomic sequence of the lytic bacteriophage DT1 of *Streptococcus thermophilus*. Virology 255:63–76.
47. van der Ploeg JR. 2007. Genome sequence of *Streptococcus mutans* bacteriophage M102. FEMS Microbiol. Lett. 275:130–138.
48. van der Ploeg JR. 2009. Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. Microbiology 155:1966–1976.
49. Whiley RA, Beighton D. 1998. Current classification of the oral streptococci. Oral Microbiol. Immunol. 13:195–216.