JOURNAL OF BACTERIOLOGY, Apr. 2003, p. 2362–2368 Vol. 185, No. 7
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.7.2362–2368.2003

Genome Organization and Molecular Analysis of the Temperate Bacteriophage MM1 of *Streptococcus pneumoniae*

Virginia Obregón, José L. García, Ernesto García, Rubens López,* and Pedro García

Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas, CSIC, 28006 Madrid, Spain

Received 16 October 2002/Accepted 7 January 2003

The genome of MM1 (40,248 bp), a temperate bacteriophage from the Spain^{23F}-1 multiresistant epidemic clone of *Streptococcus pneumoniae*, is organized in 53 open reading frames (ORFs) and in at least five functional clusters. Bioinformatic and N-terminal amino acid sequence analyses enabled the assignment of possible functions to 26 ORFs. Analyses comparing the MM1 genome with those of other bacteriophages revealed similarities, mainly with genomes of phages infecting gram-positive bacteria, which suggest recent exchange of genes between species colonizing the same habitat.

Streptococcus pneumoniae (the pneumococcus) is a gram-positive human pathogen that is the leading cause of pneumonia, meningitis, and bloodstream infections in the elderly and that is one of the main pathogens responsible of middle ear infections in children (23). A better knowledge of the molecular biology of the pneumococcus has been achieved through the study of pneumococcal phages (11). The abundance of temperate phages in clinical isolates of the pneumococcus was suggested some years ago (3), and it was recently proposed that up to 75% of the samples analyzed contained phages (26). Pneumococcal phages have been a subject of continuous interest in our laboratory since the isolation of these phages was reported (22, 36). We have recently isolated and partly characterized a temperate phage (MM1) belonging to the *Siphoviridae* family from a clinical isolate of the multiply antibiotic-resistant Spain^{23F}-1 strain (12). This strain best illustrates the rapid spread of drug resistance, because it was originally detected in Spain and then was rapidly disseminated to other parts of the world (24). A study conducted in 38 states of the United States revealed that, of 328 isolates highly resistant to penicillin (MIC $\geq$ 2.0 μg/ml), about 40% belonged to the Spanish/American *S. pneumoniae* 23F clone Spain^{23F}-1 (23).

Despite the observation that most recent clinical isolates of *S. pneumoniae* carry prophages, information on temperate pneumococcal phages is insufficient, and only a limited amount of data about gene expression and the function of the gene products is currently available. This, together with the documented interchanges between phage and host lytic genes that seem to play a role in pneumococcal virulence (19), prompted our interest in studying the genomics of these temperate phages. This approach appears to be a fundamental step in determining the contribution of phage genes to the virulence of this clinically important microorganism. Since DNAs highly similar to that of MM1 have been detected by Southern hybridization in several clinical isolates of pneumococci of different capsular types, a finding which indicates the widespread presence of closely related pneumococcal phages in pathogenic strains (12), this approach will also facilitate a comparative analysis of the genomes of these temperate phages.

Methods. *S. pneumoniae* 949 (24) was grown in C medium (17) supplemented with yeast extract (0.8 mg/ml; Difco laboratories; C+Y medium) at 37°C without shaking, and growth was monitored with a Hach 2100N nephelometer. Phage MM1 was induced from the lysogenic strain 949 (12). At a concentration of 1.2×10^8 CFU/ml, mitomycin C was added to a final concentration of 75 ng/ml, and the culture was incubated in the dark at 37°C until lysis occurred. MM1 purification (9) and preparation of proteinase K-treated MM1 DNA (28) and DNA-protein complexes (9) were carried out as previously described. Amplified or restricted DNA fragments for cloning or sequencing were isolated from 0.7% (wt/vol) agarose gels with the Geneclean II kit (Bio 101). Restriction endonucleases (New England Biolabs, Beverly, Mass.) and T4 DNA ligase (Amersham-Pharmacia Biotech.) were used as recommended by the suppliers. DNA sequencing was carried out by using an ABI Prism 3700 DNA sequencer (Applied Biosystems, Inc.). DNA and protein sequences were analyzed with the Genetics Computer Group software package (version 10.0) (7) or with the programs at the Deambulum (http://www.infobiogen.fr), the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov), or the European Bioinformatics Institute (http://www.ebi.ac.uk) site. Sequence similarity searches were performed by using the EMBL/GenBank and SWISS-PROT databases. To localize putative functional motifs the PROSITE and Pfam databases were also employed (http://hits.isb-sib.ch/cgi-bin/PFSCAN). N-terminal sequence analyses were carried out according to a published procedure (31).

Determination of the complete nucleotide sequence of the MM1 genome. The DNA from mature phage particles appears to contain a covalently bound protein, as reported for other pneumococcal phages (9, 10, 28). Moreover, preliminary assays (restriction analyses, denaturation-renaturation analyses combined with electron microscopy, etc.) suggested that MM1 DNA is circularly permuted, terminally redundant, and packaged via a headful mechanism (data not shown). Initially, we decided to determine the nucleotide sequence of the entire genome of MM1 by using a shotgun sequencing approach, and the overlapping sequences were assembled into several contiguous stretches. The remaining gaps were closed by PCR am-

* Corresponding author. Mailing address: Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas, CSIC, Velázquez 144, 28006 Madrid, Spain. Phone: (34) 91 561 1800. Fax: (34) 91 562 7518. E-mail: ruben@cib.csic.es.

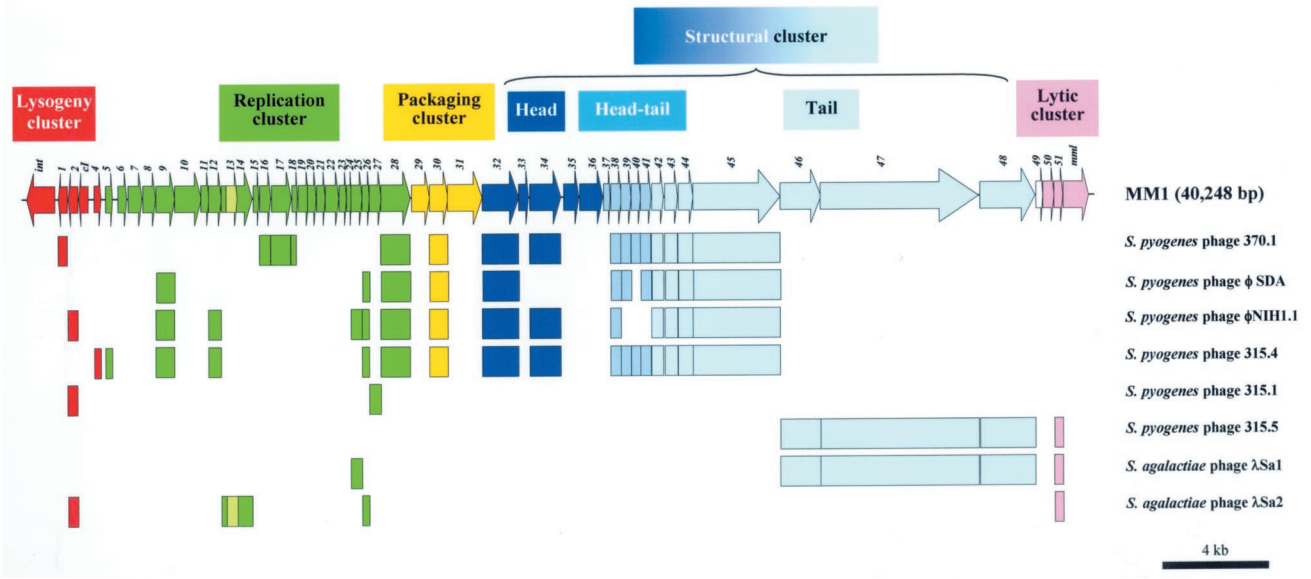

FIG. 1. Schematic representation of the proposed functional organization of MM1 DNA. Genes likely to belong to the same cluster were marked with the same colors. Genes from *S. pyogenes* or *S. agalactiae* phages similar to MM1 ORFs are also shown. In all cases the corresponding gene products showed amino acid identities ranging from 24 to 73% (see Table 1).

plification, with the entire phage genome as the template, and sequencing with specific primers. Our analyses revealed the presence of 53 open reading frames (ORFs) in a unit genome size of 40,248 bp in the prophage state (Fig. 1). The correctness of the sequence assembly was confirmed by comparing the predicted map from this sequence with that experimentally obtained by using restriction enzymes. The average G+C content of MM1 DNA was calculated as 38.4%, which is only slightly lower than the 39.7% reported for the host genome (14, 35).

Analysis and organization of the MM1 genome. The 53 ORFs analyzed potentially code for polypeptides with more than 50 amino acid (aa) residues. Every ORF is preceded by a putative ribosome binding site and begins with either an ATG, GTG, or TTG initiation codon. The most common stop codon used was TAA (33 ORFs). The MM1 genome is apparently organized into five major clusters (schematically represented in Fig. 1) in the prophage state and starts with the *int* gene (Table 1). The five leftmost genes (from *int* to *orf4*) comprise the lysogeny cluster and are organized in a way characteristic of temperate bacteriophages of the *Siphoviridae* family from low-G+C-content gram-positive bacteria (20). *orf2* codes for a protein that is 54% identical to a protein of *Streptococcus agalactiae* phage λSa2. These proteins share the Zn recognition motif (V-X-X-H-E[I]-G-H) characteristic of metalloproteinases (16). This sequence is named the HD domain and defines a new superfamily of metal-dependent phosphohydrolases (Pfam database accession no. PF01966). The *cI* gene codes for a 120-aa protein that exhibits high similarity to several phage-related transcriptional repressors and that represents a λ CI analogue. Most probably, Orf4 represents the Cro-like repressor of MM1.

As in other *pac* site phages, the replication cluster of MM1 follows the lysogeny cluster. In MM1 DNA, there is a very A+T-rich sequence (76% A+T within 231 bp) located between *orf5* and *orf6* that contains many direct and reverse repeats (not shown) and that may correspond to the origin of replication (*ori*) of the phage genome. The *ori* region includes, among others, three copies of a tandem, 15-bp direct repeat (TTTT ACAAATCTGTA) as well as two copies of a 25-bp direct repeat (AATAAATACTAACTAACAACAAGTA). Moreover, this area contains three palindromes capable of forming weak stem-loop structures with free energies ranging between −8.5 and −9.2 kcal/mol. These structures may be involved in the melting of the DNA strands and most likely represent the starting points of DNA replication. The *orf9* and *orf10* gene products are similar to two proteins that are also encoded by adjacent genes in the genome of the *Streptococcus thermophilus* phage 7201 (32). The protein similar to Orf9 is assumed to be a recombinase, whereas Orf10 appears to be a histone-like protein since it contains a Pfam PF00216-like motif that binds DNA. *orf13* and *orf14* correspond to two overlapped genes since they are transcribed from two different frames and encode proteins that are very similar to two proteins encoded by the *S. pneumoniae* transposon Tn*5252*, identified as components of a 5-cytosine-specific DNA methyltransferase (C5-MTase) (PF00145) (29). The presence and significance of genes coding for C5-MTases in MM1 and related phages have been recently discussed (25). These genes as well as that coding for the lytic amidase (*mml*) are examples of high sequence similarities between genes of the host bacterium and those of a pneumococcal temperate phage. Note especially that the G+C contents of *orf13* (41.4%) and *orf14* (44.1%) are clearly different from the average for the MM1 genome, 38.4% (see above). Most of the proteins encoded by the ORFs identified in this cluster strongly resemble proteins of unknown function encoded by phage genomes of the *Siphoviridae* family from low-G+C-content gram-positive bacteria although a ParB-like nuclease domain (PF02195) has been found in Orf28. ParB preferentially cleaves single-stranded DNA. Interestingly, *orf27* is

TABLE 1. Comparative analysis of the genes from *S. pneumoniae* phage MM1 with proteins included in the databases

Gene	Nucleotide		No. of amino acids encoded	Function	Best match (% amino acid identity)	$\log_{10} E$	Comment and/or further matches (% amino acid identity)
	Start	Stop					
int	1230	103	375	Integrase			See reference 12
orf1	1699	1349	116	—[a]	*L. lactis* phage TPW22 (76)	−15	*S. thermophilus* phage O1205 (55); *S. pyogenes* phage 370.1 (42)
orf2	2093	1713	126	—	*S. agalactiae* phage λSa2 (54)	−26	*S. thermophilus* phage Sfi21 (43); *S. pyogenes* phages ϕNIH1.1 (43) and 315.1 (38)
cI	2477	2115	120	Repressor	*S. pyogenes* phage ϕ speA (51)	−18	Many phage repressors: 315.1, 315.4, NIH1.1, PSA, PS1, bIL310, TPW22, ϕgle
orf4	2774	2965	63	Cro-like repressor	*Lactobacillus gasseri* phage ϕadh (45)	−8	*S. pyogenes* phages 315.4 (44) and 315.6 (42); *L. lactis* prophage PS2 (42)
orf5	3140	3406	88	Replication	*S. pyogenes* phage 315.4 (39)	−8	
orf6	3638	3904	88	—			
orf7	3985	4509	174	—			
orf8	4561	5013	150	—			
orf9	5019	5726	235	Recombination	*S. thermophilus* phage 7201 (39)	−25	*S. pyogenes* phages ϕNIH1.1, ϕ SDA, and 315.4 (24)
orf10	5729	6721	330	DNA binding	*S. thermophilus* phage 7201 (38)	−35	*L. lactis* phage ul36 (26); *S. pyogenes* phage 370.3 (27)
orf11	6743	6919	58	—	*S. pyogenes* phage similar to ϕ370.2 (39)	−5	
orf12	6909	7454	181	single-stranded-DNA binding	*S. pneumoniae* phage VO1 (98)	−73	*S. agalactiae* Ssb-3 (64); *S. pyogenes* phages 315.4 and ϕNIH1.1 (63); *S. thermophilus* phage 7201 (60); several SSB proteins from gram-positive bacteria
orf13	7480	8637	385	C5-MTase (α subunit)	*S. agalactiae* phage λSa2 (62)	−74	*S. pneumoniae* transposon Tn*5252* (55), *S. pneumoniae* type 4 (53), and many other modification methylases from phages and bacteria (30–35)
orf14	7631	8056	141	C5-MTase (β subunit)	*S. agalactiae* phage λSa2 (62)	−28	*L. lactis* AAM27270 (48); *S. agalactiae* (44); *S. pneumoniae* transposon Tn*5252* (43); *S. pneumoniae* type 4 (41); *Lactococcus* phage 4268 (46); *E. coli Eco*HK31I polypeptide β (35)
orf15	8672	8875	67	—			
orf16	8929	9363	144	—	*S. pyogenes* phage 370.1 (60)	−27	
orf17	9365	10102	245	—	*S. pyogenes* phage 370.1 (61)	−55	
orf18	10119	10331	70	—	*S. pyogenes* phage 370.1 (64)	−15	
orf19	10342	10743	133	—	*L. lactis* phage TP901-1 (46)	−20	Several *L. lactis* phages: ϕ31.1 (45), ul36.1 (45), bIL309 (38), pi1 (37); *L. monocytogenes* phage A118 (32)
orf20	10745	10942	68	—			
orf21	11000	11317	105	—	*Lactobacillus* phage ϕadh (36)	−6	
orf22	11319	11816	165	—	*S. thermophilus* phage O1205 (39)	−11	*S. pyogenes* phage 370.3 (36); *S. thermophilus* phages Sfi11 (37) and Sfi19 (32); *L. monocytogenes* phage A118 (34)
orf23	11816	12103	95	—			
orf24	12103	12276	57	—			
orf25	12291	12710	139	—	*S. pyogenes* phage ϕ speA (52)	−23	*L. lactis* phage Tuc2009 (37); *S. agalactiae* phage λSa1 (43); *S. pyogenes* phage ϕNIH1.1 (33)
orf26	12770	13012	80	—	*S. pyogenes* phage ϕ SDA (63)	−17	*S. agalactiae* phage λSa2 (69); *S. pyogenes* phages 315.4, 315.5, and ϕNIH1.1 (63); *L. monocytogenes* phage PSA (41); *Staphylococcus aureus* phage ϕ 12 (44)
orf27	13009	13401	130	—	*S. pneumoniae* SP1142 (54)	−20	*S. pyogenes* phages ϕ speA and 315.1 (27)
orf28	13467	14561	364	Nuclease	Mycobacteriophage L5 gp1 (36)	−25	Mycobacteriophage D29 (36); *S. pyogenes* phages ϕ SDA, 315.4, ϕNIH1.1 (50), and 370.1 (47)
orf29	14542	15288	248	—			
orf30	15492	15947	151	—	*S. pyogenes* phage 370.1 (38)	−9	*S. pyogenes* phages 315.4, ϕ SDA, and ϕNIH1.1 (36)

Continued on following page

TABLE 1—*Continued*

Gene	Nucleotide		No. of amino acids encoded	Function	Best match (% amino acid identity)	$\log_{10} E$	Comment and/or further matches (% amino acid identity)
	Start	Stop					
orf31	15937	17247	436	Terminase large subunit	*L. monocytogenes* phage A118 (53)	−96	Many phage terminases: ϕ gle, SPP1, Lj771, 370.3, PBSX
orf32	17260	18828	522	Minor capsid protein	*S. pyogenes* phage 315.4 (62)	−70	*S. pyogenes* phages ϕNIH1.1, 370.1, ϕ SDA (62); *Lactobacillus* phage ϕ gle (43)
orf33	18761	18988	75	—			
orf34	19031	20182	383	Minor capsid protein	*S. pyogenes* phage ϕNIH1.1 (49)	−71	*S. pyogenes* phage 315.4 and 370.1 (50); *L. monocytogenes* phage A118 (32)
orf35	20321	20884	187	Scaffolding	*L. monocytogenes* phage A118 (37)	−11	*Lactobacillus* phages ϕ gle (33) and LL-H (30); other phages (SPP1, mv4)
orf36	20902	21789	295	Major capsid protein	*L. lactis* phage ul36 (37)	−29	N-terminal amino acid sequence determination (see text)
orf37	21793	22026	77	—			
orf38	22069	22461	130	—	*S. pyogenes* phage 370.1 (51)	−20	*S. pyogenes* phages 315.4, ϕ SDA (50), and ϕNIH1.1 (48)
orf39	22451	22822	123	Minor capsid protein	*S. pyogenes* phage ϕ SDA (42)	−12	*S. pyogenes* phages 315.4 and 370.1 (42)
orf40	22822	23166	114	Minor capsid protein	*S. pyogenes* phage 370.1 (36)	−12	*S. pyogenes* phage 315.4 (36); *Lactobacillus* phages ϕ gle (39) and LL-H (35); *L. monocytogenes* phage A118 (29)
orf41	23166	23573	135	Minor capsid protein	*S. pyogenes* phage 315.4 (44)	−21	*S. pyogenes* phages 370.1 and ϕ SDA (44); *Lactobacillus* phage ϕ gle (31)
orf42	23570	24019	149	Tail protein	*S. pyogenes* phage ϕ SDA (73)	−45	*S. pyogenes* phages 315.4, ϕNIH1.1 (72), and 370.1 (71); *L. monocytogenes* phage A118 (44); N-terminal amino acid sequence determination (see text)
orf43	24084	24572	162	—	*S. pyogenes* phage ϕ SDA (54)	−25	*S. pyogenes* phages 370.1, 315.4, and ϕNIH1.1 (54)
orf44	24585	25154	189	—	*S. pyogenes* phage 370.1 (52)	−33	*S. pyogenes* phages ϕNIH1.1, 315.4 (52), and ϕ SDA (51); *Lactobacillus* phage ϕ gle (30); *L. monocytogenes* phage A118 (33)
orf45	25147	28428	1,093	Tail protein	*S. pyogenes* phage 315.4 (59)	−170	*S. pyogenes* phages ϕNIH1.1,ϕ SDA, and 370.1 (59); *L. lactis* phage ul36 (32); *S. thermophilus* phages Sfi11 (21) and O1205 (21)
orf46	28425	29939	504	Minor structural protein	*S. thermophilus* phage O1205 (43)	−79	Several *S. thermophilus* phages: Sfi11, 7201, Sfi19, and Sfi21; *S. pyogenes* phage 315.5 (43); *S. agalactiae* phage λSa1 (33)
orf47	29941	35910	1,989	Antireceptor	*S. thermophilus* phage Sfi21 (35)	−77	Several *S. thermophilus* phages: DT2, MD2, Sfi11, and O1205; *S. pyogenes* phage 315.5 (32); *S. agalactiae* phage λSa1 (32)
orf48	35921	37999	692	Minor tail protein	*S. pneumoniae* phage Dp-1 (69)	−147	Several *S. thermophilus* phages: Sfi11, O1205, DT1, and 7201; *S. pyogenes* phage 315.5 (37); *S. agalactiae* phage λSa1 (33)
orf49	38053	38322	89	—	*S. pneumoniae* phage VO1 (96)	−28	*S. pneumoniae* phage Dp-1 (62)
orf50	38332	38748	138	Holin	*S. pneumoniae* phage VO1 (91)	−48	*Bacillus subtilis* phages GA-1 (33), B103, ϕ29, and pZA (22)
orf51	38752	39084	110	Holin/antiholin?	*S. pneumoniae* phage VO1 (80)	−29	*S. mitis* phage SM1 (75); *S. agalactiae* phages λSa2 (67) and λSa1 (48); *S. pyogenes* phage 315.5 (54)
mml	39088	40044	318	Lytic amidase	*S. pneumoniae* phage HB-3 (97)	−141	*S. pneumoniae* LytA (89) and other phage and bacterial choline-binding proteins

[a] —, function unknown.

very similar to the SP1142 gene (62% identity) from the genome of *S. pneumoniae* strain TIGR4 (35). SP1142 is part of a 10.5-kb cluster of 19 contiguous ORFs (from SP1129 to SP1147) that is absent in the pneumococcal R6 genome and that likely corresponds to a phage remnant. As examples, SP1129 potentially codes for a protein 42% identical to the integrase of the phage 370.4 from *Streptococcus pyogenes*, whereas the products of genes SP1130, SP1131, and SP1134 were 27, 52, and 42% identical, respectively, to the products of ORFs from the *Lactococcus lactis* prophage ps3 (data not shown).

orf31 encodes a 436-aa protein exhibiting 53% identity with the large subunit of the terminase identified in *Listeria monocytogenes* phage A118 (18). The product of *orf30* is a 151-aa protein that is 38% identical to the protein encoded by an ORF from the *S. pyogenes* phage 370.1 and that may represent the small subunit of the terminase (Table 1). Comparison of the gene organization of the MM1 genome (Fig. 1) with those of several *Siphoviridae* phage DNAs (6) suggested that the structural cluster includes 17 genes (from *orf32* to *orf48*) that should participate in the formation of the head and tail of the MM1 phage (Table 1). The *orf35* gene product is a protein of 187 aa that has significant similarity (37% identity) to the putative scaffold protein of *L. monocytogenes* phage A118; *orf35* is located immediately upstream of the major capsid protein gene (*orf33*), which is characteristic of genes encoding scaffolding proteins (13). Sequencing the N-terminal amino acids of the major structural MM1 protein yielded M-P-S-N-Q-N, and sequencing the second-largest protein band produced M-T-R-Q-K-N, corresponding to Orf36 and Orf42, respectively. The large Orf47 (1,989 aa) may be the protein recognizing the phage receptor at the pneumococcal surface. Antireceptor proteins from *Siphoviridae* phages infecting low-G+C-content gram-positive bacteria usually contain repeated G-X-Y motifs at their C moieties (21). This motif appears to be characteristic of tropocollagen molecules, and its biological function is to provide elasticity and resistance.

orf50 codes for a protein of 138 aa, and preliminary cloning experiments with *Escherichia coli* suggest that this protein corresponds to the holin of phage MM1 (M. Gindreau, R. López, and P. García, unpublished observations). Orf51 is 75% identical to the predicted holin from the *Streptococcus mitis* temperate phage SM1 recently isolated (1). Finally, the *mml* gene codes for a protein of 318 aa highly similar (97% identity) to the lytic amidase characterized in phage HB-3 of *S. pneumoniae* (27). It has been proposed that holin is the protein that provokes unspecific lesions into the cytoplasmic membrane that allow the murein hydrolase to escape and hydrolyze the cell wall (37). Interestingly, the G+C content of the *mml* gene (47.2%) is noticeably higher than the average content of MM1 genome (38.4%), suggesting a possible acquisition of this gene by horizontal transfer since the G+C content of a whole genome is characteristic for a given species or group (33). Differences between the codon usage of a gene and the codon bias of the host organism are additional criteria for identifying horizontal transfers. Table 2 shows the codon usage of *mml* compared to those of the MM1 and *S. pneumoniae* genomes. The codon usages of MM1 and *S. pneumoniae* were similar, whereas that of *mml* was noticeably different. Thus, in *mml* the codons CGC, AAC, GAC, AUC, and AAG are the most frequently used for Arg, Asn, Asp, Ile, and Lys, respectively, whereas in phage MM1 the codons primary utilized for the same amino acids are AGA, AAU, GAU, AUU, and AAA, respectively. The most frequent codon used for Arg in *S. pneumoniae* is CGU. Table 2 also shows that the codon usage for *orf14* differs from the average usage for MM1 genes although this was not evident for *orf13*.

TABLE 2. Partial codon usage for the MM1 genes *orf13*, *orf14*, and *mml*

Amino acid	Codon	Codon usage[a] for:				
		MM1	*Spn*	*orf13*	*orf14*	*mml*
Arg	AGG	0.12	0.06	0.22 (5)	0.57 (4)	0.00 (0)
	AGA	0.33	0.19	0.48 (11)	0.00 (0)	0.18 (2)
	CGG	0.04	0.05	0.04 (1)	0.00 (0)	0.09 (1)
	CGA	0.14	0.12	0.17 (4)	0.14 (1)	0.09 (1)
	CGU	0.26	0.43	0.04 (1)	0.29 (2)	0.18 (2)
	CGC	0.11	0.16	0.04 (1)	0.00 (0)	0.45 (5)
Asn	AAU	0.62	0.67	0.48 (10)	0.17 (1)	0.37 (7)
	AAC	0.38	0.33	0.52 (11)	0.83 (5)	0.63 (12)
Asp	GAU	0.72	0.67	0.70 (14)	0.38 (3)	0.37 (10)
	GAC	0.28	0.33	0.30 (6)	0.62 (5)	0.63 (17)
Gln	CAG	0.29	0.34	0.26 (5)	0.56 (5)	0.50 (5)
	CAA	0.71	0.66	0.74 (14)	0.44 (4)	0.50 (5)
Ile	AUA	0.19	0.10	0.23 (7)	0.33 (2)	0.00 (0)
	AUU	0.51	0.55	0.47 (14)	0.17 (1)	0.27 (3)
	AUC	0.31	0.35	0.30 (9)	0.50 (3)	0.73 (8)
Lys	AAG	0.29	0.36	0.38 (9)	0.29 (2)	0.62 (13)
	AAA	0.71	0.64	0.62 (15)	0.71 (5)	0.38 (8)

[a] Data for *Spn* were compiled from reference 35 (http://www.tigr.org/tigr-scripts/CMR2/codon_tables.spl). Numbers in parentheses are the numbers of amino acid residues present in the corresponding protein.

MM1 is the first temperate phage infecting *S. pneumoniae* for which the complete nucleotide sequence has been determined and whose genome has been analyzed in detail. These data are a starting point to carry out further studies on clinical isolates of the pneumococcus, where as many as 75% of the isolates have been reported to be lysogenic (26), since phage conversion might play an important role in the evolution of many pathogenic bacteria (5). The modular organization of the MM1 genome is similar to those of other temperate streptococcal phages, where genes belonging to the lysogeny cluster were the only genes carried in the opposite DNA strand of the phage genome (4). Twenty-six out of the 53 proteins deduced from the MM1 DNA sequence have significant similarities to products of ORFs reported in the databases (Table 1), and this enabled ascription of putative functions to some of them on the grounds of homologies to defined proteins. Most of these proteins belong to temperate phages of the *Siphoviridae* family infecting gram-positive bacteria and, particularly, to *S. pyogenes* phages, namely, 370.1, ϕ SDA, ϕNIH1.1, 315.1, 315.4, and 315.5 (Fig. 1). DNAs from nearly all of these phages were found when the genomes of three virulent *S. pyogenes* strains were sequenced (2, 8, 15, 30). The noticeable similarity between genes of phages infecting *S. pyogenes* and *S. pneumoniae* suggests a frequent genetic interchange between both species or a recent divergence from a common ancestor phage. From

the evolutionary viewpoint the MM1 genome appears to be organized in at least two different regions (Fig. 1). The first one (from *int* to *orf45*) has many similarities in sequence and organization with the *S. pyogenes* phages except phage 315.5, whereas the right part of the MM1 prophage (with the noticeable exception of *mml*), that is, from *orf46* to *orf51*, is closely related to phages 315.5 and λSa1 from *S. pyogenes* and *S. agalactiae* (34), respectively. As shown above, the *mml* gene encoding the lytic enzyme of the phage might have been introduced into the MM1 genome by horizontal transfer.

Still-unresolved issues are the impact of phages on the evolution of host genomes and the links between phages of pathogenic and nonpathogenic strains in gram-positive bacteria exhibiting similar organizations; these issues raise questions as to the contribution of the phages to survival in different environments (4, 6, 23). Since there is a large incidence of lysogeny among clinical strains of *S. pneumoniae* (26), the report of the first complete genome of a temperate pneumococcal phage of this bacterium provides an important tool to facilitate the study of potential virulence genes carried by pneumococcal viruses that might infect different species colonizing the same habitat. Studies in progress in our laboratory will expand our observations on the importance of prophages for shaping the natural population of *S. pneumoniae* as well as on the evolutionary diversification of the bacterial host.

Nucleotide sequence accession number. The MM1 genome sequence has been deposited in the EMBL, GenBank, and DDBJ databases and appears under accession no. AJ302074.

We thank H. Brüssow for critical reading of the manuscript and for helpful suggestions, A. Fenoll for providing the 949 lysogenic strain, M. Rejas for electron microscopy preparation, and E. Cano and M. Carrasco for technical assistance.

This work was supported by grants from the Dirección General de Investigación Científica y Técnica (BCM2000-1002) and from Programa de Grupos Estratégicos de la Comunidad Autónoma de Madrid.

REFERENCES

1. **Bensing, B. A., I. R. Siboo, and P. M. Sullam.** 2001. Proteins PblA and PblB of *Streptococcus mitis*, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. Infect. Immun. **69:**6186–6192.
2. **Beres, S. B., G. L. Sylva, K. D. Barbian, B. Lei, J. S. Hoff, N. D. Mammarella, M.-Y. Liu, J. C. Smoot, S. F. Porcella, L. D. Parkins, D. S. Campbell, T. M. Smith, J. K. McCormick, D. Y. Leung, P. M. Schlievert, and J. M. Musser.** 2002. Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. USA **99:**10078–10083.
3. **Bernheimer, H. P.** 1979. Lysogenic pneumococci and their bacteriophages. J. Bacteriol. **138:**618–624.
4. **Brüssow, H.** 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. **55:**283–303.
5. **Cheetham, B. F., and M. E. Katz.** 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. **18:**201–208.
6. **Desiere, F., W. M. McShan, D. van Sinderen, J. J. Ferretti, and H. Brüssow.** 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.
7. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
8. **Ferretti, J. J., W. M. McShan, D. Ajdic, D. J. Savic, K. Lyon, C. Primeaux, S. Sezate, A. N. Suvorov, S. Kenton, H. S. Lai, S. P. Lin, Y. Qian, H. G. Jia, F. Z. Najar, Q. Ren, H. Zhu, L. Song, J. White, X. Yuan, S. W. Clifton, B. A. Roe, and R. McLaughlin.** 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc. Natl. Acad. Sci. USA **98:**4658–4663.
9. **García, E., A. Gómez, C. Ronda, C. Escarmís, and R. López.** 1983. Pneumococcal bacteriophage Cp-1 contains a protein bound to the 5′ termini of its DNA. Virology **128:**92–104.
10. **García, P., J. M. Hermoso, E. García, J. L. García, and R. López.** 1986. Formation of a covalent complex between the terminal protein of pneumococcal bacteriophage Cp-1 and 5′-dAMP. J. Virol. **58:**31–35.
11. **García, P., A. C. Martín, and R. López.** 1997. Bacteriophages of *Streptococcus pneumoniae*: a molecular approach. Microb. Drug Resist. **3:**165–176.
12. **Gindreau, E., R. López, and P. García.** 2000. MM1, a temperate bacteriophage of the 23F Spanish/USA multiresistant epidemic clone of *Streptococcus pneumoniae*: structural analysis of the site-specific integration system. J. Virol. **74:**7803–7813.
13. **Hendrix, R. W., and R. L. Duda.** 1998. Bacteriophage HK97 head assembly: a protein bullet. Adv. Virus Res. **50:**235–288.
14. **Hoskins, J., W. E. Alborn, J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D.-J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoje, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P.-M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. Zhao, C. A. Zook, R. H. Baltz, R. Jaskunas, P. R. J. Rosteck, P. L. Skatrud, and J. I. Glass.** 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J. Bacteriol. **183:**5709–5717.
15. **Ikebe, T., A. Wada, Y. Inagaki, K. Sugama, R. Suzuki, D. Tanaka, A. Tamaru, Y. Fujinaga, Y. Abe, Y. Shimizu, and H. Watanabe.** 2002. Dissemination of the phage-associated novel superantigen gene *speL* in recent invasive and noninvasive *Streptococcus pyogenes* M3/T3 isolates in Japan. Infect. Immun. **70:**3227–3233.
16. **Jongeneel, C. V., J. Bouvier, and A. Bairoch.** 1968. A unique signature identifies a family of zinc-dependent metallopeptidases. FEBS Lett. **242:**211–214.
17. **Lacks, S., and R. D. Hotchkiss.** 1960. A study of the genetic material determining an enzyme activity in *Pneumococcus*. Biochim. Biophys. Acta **39:**508–517.
18. **Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar.** 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol. Microbiol. **35:**324–340.
19. **López, R., M. P. González, E. García, J. L. García, and P. García.** 2000. Biological roles of two new murein hydrolases of *Streptococcus pneumoniae* representing examples of module shuffling. Res. Microbiol. **151:**437–443.
20. **Lucchini, S., F. Desiere, and H. Brüssow.** 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J. Virol. **73:**8647–8656.
21. **Lucchini, S., F. Desiere, and H. Brüssow.** 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage ϕ Sfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. Virology **246:**63–73.
22. **McDonnell, M., C. Ronda-Laín, and A. Tomasz.** 1975. "Diplophage": a bacteriophage of *Diplococcus pneumoniae*. Virology **63:**577–582.
23. **McGee, L. K., K. P. Klugman, and A. Tomasz.** 2000. Serotypes and clones of antibiotic-resistant pneumococci, p. 375–379. *In* A. Tomasz (ed.), *Streptococcus pneumoniae*: molecular biology and mechanism of disease. Mary Ann Liebert, Inc., Larchmont, N.Y.
24. **Muñoz, R., T. J. Coffey, M. Daniels, C. G. Dowson, G. Laible, J. Casal, R. Hakenbeck, M. Jacobs, J. M. Musser, B. G. Spratt, and A. Tomasz.** 1991. Intercontinental spread of a multiresistant clone of serotype 23F *Streptococcus pneumoniae*. J. Infect. Dis. **164:**302–306.
25. **Obregón, V., P. García, R. López, and J. L. García.** VO1, a temperate bacteriophage of the type 19A multiresistant epidemic 8249 strain of *Streptococcus pneumoniae*: analysis of variability of lytic and putative C5 methyltransferase genes. Microb. Drug Resist., in press.
26. **Ramirez, M., E. Severina, and A. Tomasz.** 1999. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. J. Bacteriol. **181:**3618–3625.
27. **Romero, A., R. López, and P. García.** 1990. Sequence of the *Streptococcus pneumoniae* bacteriophage HB-3 amidase reveals high homology with the major host autolysin. J. Bacteriol. **172:**5064–5070.
28. **Romero, A., R. López, R. Lurz, and P. García.** 1990. Temperate bacteriophages of *Streptococcus pneumoniae* that contain protein covalently linked to the 5′ ends of their DNA. J. Virol. **64:**5149–5155.
29. **Sampath, J., and M. N. Vijayakumar.** 1998. Identification of a DNA cytosine methyltransferase gene in conjugative transposon Tn*5252*. Plasmid **39:**63–76.
30. **Smoot, J. C., K. D. Barbian, J. J. Van Gompel, L. M. Smoot, M. S. Chaussee, G. L. Sylva, D. E. Sturdevant, S. M. Ricklefs, S. F. Porcella, L. D. Parkins, S. B. Beres, D. S. Campbell, T. M. Smith, Q. Zhang, V. Kapur, J. A. Daly, L. G. Veasy, and J. M. Musser.** 2002. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. USA **99:**4668–4673.
31. **Speicher, D. W.** 1994. Methods and strategies for the sequence analysis of proteins on PVDF membranes. Methods **6:**262–273.
32. **Stanley, E., L. Walsh, A. van der Zwet, G. F. Fitgerald, and D. van Sinderen.** 2000. Identification of four loci isolated from two *Streptococcus thermophilus* phage genomes responsible for mediating bacteriophage resistance. FEMS Microbiol. Lett. **182:**271–277.
33. **Sueoka, N.** 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85:**2653–2657.
34. **Tettelin, H., V. Masignani, M. J. Cieslewicz, J. A. Eisen, S. Peterson, M. R.**

Wessels, I. T. Paulsen, K. E. Nelson, I. Margarit, T. D. Read, L. C. Madoff, A. M. Wolf, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, R. T. DeBoy, A. S. Durkin, J. F. Kolonay, R. Madupu, M. R. Lewis, D. Radune, N. B. Fedorova, D. Scanlan, H. Khouri, S. Mulligan, H. A. Carty, R. T. Cline, S. E. Van Aken, J. Gill, M. Scarselli, M. Mora, E. T. Iacobini, C. Brettoni, G. Galli, M. Mariani, F. Vegni, D. Maione, D. Rinaudo, R. Rappuoli, J. L. Telford, D. L. Kasper, G. Grandi, and C. M. Fraser. 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. USA **99:**12391–12396.

35. **Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelber, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser.** 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science **293:**498–506.
36. **Tiraby, J. G., E. Tiraby, and M. S. Fox.** 1975. Pneumococcal bacteriophages. Virology **68:**566–569.
37. **Wang, I.-N., D. L. Smith, and R. Young.** 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. **54:**799–825.