Arch Virol (2003) 148: 1051–1070
DOI 10.1007/s00705-003-0008-x

Archives of Virology
Printed in Austria

Analysis of the morphogenetic cluster and genome of the temperate *Lactobacillus casei* bacteriophage A2

P. García[1], **V. Ladero**[1], and **J. E. Suárez**[1,2]

[1]Area de Microbiología, Facultad de Medicina, Universidad de Oviedo, Julián Clavería s. n., Oviedo, Spain
[2]Instituto de Productos Lácteos de Asturias (CSIC), Villaviciosa, Spain

Received November 29, 2002; accepted January 14, 2003
Published online March 21, 2003

Summary. The genes that encode the morphogenetic proteins of bacteriophage A2 are clustered and expressed as a single operon which originates a late transcript of more than 20 kb. This DNA stretch is analyzed in the context of the whole phage genome, which presents the following peculiarities: a) the head presents two major proteins that share their NH_2 termini, i.e.: both are translated from a single gene (*orf5*), b) these two proteins suffer a proteolytic maturation process before being incorporated into the capsid, rendering a 123 NH_2 terminal putative polypeptide that is postulated to be the scaffolding protein of the phage, c) similar maturation processes occur at the portal and tail length determination proteins, having all in common a Pho-Pho-Arg↓ sequence (where Pho stands for any hydrophobic amino acid) at the processing point, d) the genes encoding the subunits of the terminase (*orf61* and *orf2*) are separated by the cohesive ends, e) two genes that might mediate lysogenic conversion (*orf19* and *orf22*) have been identified and f) the genome presents a dispensable region (which covers at least 10 *orfs*, as judged from analysis of deletion mutants) that might be involved in maintaining its size between the packaging limits of the capsid.

Introduction

In a couple of recent reviews new phylogeny based schemes for bacteriophage classification have been proposed [6, 50]. These proposals are a consequence of the fast growing number of phages whose genome sequence is completely known, which allows evolutive analysis of their complete nucleotide and amino acid sequences and permit mapping of gene clusters in relation to the whole genome and prediction of the secondary structures of proteins judged to have similar functions.

As a consequence of their interest, slightly more than 100 phage genomes have been completely sequenced [50], being especially well represented those infecting *Escherichia coli* and lactic acid bacteria (LAB). In this last case, the interest on phages is a consequence of their role as a threat to food fermentation [1, 5, 11]. Phages infect the starter bacteria slowing the production of lactic acid, which results in organoleptic defects and in shortening of the consumption period of the final product. The problem is specially worrying for dairies due to the high volumes of milk that are processed at the same time, to the relatively low number of proficient bacterial strains used and to the production of aerosols, which act as vehicles for phage dispersion. As a consequence, out of the 18 LAB-phage genomes completely known, 17 infect organisms used as dairy starters, i.e.: six infect *Streptococcus thermophilus*, nine develop on *Lactococcus lactis* and two grow on thermophilic lactobacilli [13, 32].

The temperate bacteriophage A2 was isolated from the whey of a failed cheese fermentation, on *Lactobacillus casei* 393 [25]. A2 probably belongs to the family *Siphoviridae* being composed by an isometric head with a diameter of 60 nm and a tail 280 nm long and 12 nm wide. The tail is ended by a basal plate with a protruding spike of 28 nm (J. L. Caso, personal communication). The early functions of the phage have been partially characterised; it is known for instance that, upon infection, two divergently oriented promoters are constitutively expressed. One of them, P_L, governs expression of the lytic cycle repressor gene, *cI*, while the other, P_R, directs expression of the *cro* repressor gene and of the replication module of the phage [20, 34, 35, 39]. Downstream of *cI* lies the gene encoding the integrase that mediates the recombination between the adjacent *attP* phage region and an *attB* sequence located at the end of a $tRNA^{Leu}$ gene [2]. Besides, the small terminase subunit and the *cos* ends of the phage have been characterised [19]. The sum of these DNA regions account for some 17 kb, i.e.: about 1/3 of the genome.

In this paper the structural proteins of A2 are characterized as a step towards location of their determinants in the genome of the phage and determination of the post-translational modifications suffered by some of them upon integration in the virions. In this context, a detailed description of the phage genome is included (a preliminary comparison of the A2 genome with those of other phages has appeared recently) [48].

Materials and methods

Bacterial strains, plasmids and bacteriophages

Lactobacillus casei ATCC 393 was used to propagate bacteriophage A2 on liquid MRS (Oxoid) supplemented with 10 mM $CaCl_2$ and 10 mM $MgSO_4$ (MCM) at 30 °C, without aeration. Plaque enumeration was on solid MCM (1.5% agar) covered by semisolid medium (0.7% agar). *Escherichia coli* XL1-Blue was used as recipient for pUC18 plasmid constructions [57]. It was propagated in 2xTY broth or 2xTY broth solidified with 1.5% (w/v) agar [51] at 37 °C under agitation. Ampicillin (100 μg/ml), IPTG (0.5 mM) and X-Gal (40 μg/ml) were used when required. Deletion mutants of phage A2 were obtained as previously described [35] through repeated treatments of phage suspensions with 10 mM sodium pyrophosphate.

Nucleic acid analysis techniques

General techniques such as DNA electrophoresis and fragment electroelution were performed by standard methods [51]. DNA related enzymes were obtained from Roche and used according to the supplier's instructions. Phage A2 DNA was extracted and purified as described previously [54]. The QIAGEN Plasmid Mini Kit was used for plasmid DNA extraction. *E. coli* was transformed by electroporation with a BioRad pulser apparatus [15]. PCR amplification was performed in a MiniCycler (MJ Research, Inc.) using the proofreading *Pwo* DNA Polymerase (Roche) according to the instructions of the suppliers. PCR products were purified using the QIAquick-spin PCR Purification Kit (QIAGEN). Large DNA templates were amplified with the Expand Long Template PCR System (Roche). RNA extraction and Northern blotting were performed as previously described [19].

DNA sequencing and analysis

A random library of the phage genome was established in pUC18 and DNA sequencing was determined on both strands by the PCR cycle sequencing method [41]. Gaps in the sequence were filled using PCR amplified segments or direct sequencing using the whole phage DNA as a template. The Genetics Computer Group sequence analysis package (University of Wisconsin, Madison, version 10.2) was used to assemble and analyze the sequences [14]. Nucleotide and predicted amino acid sequences were compared with those in the last versions of the data bases (GenBank, EMBL, PIR-protein, SWISS-PROT) using the BLAST program [45]. The nucleotide sequence reported here has been submitted to the EMBL Nucleotide Sequence Data Libraries under Accession Number AJ251789.

Purification of A2 virions and structural protein analysis

Phage A2 was purified through isopycnic centrifugation in continuous CsCl gradients [51]. Disruption of A2 virions and structural protein analysis was performed by heating CsCl-purified phage suspensions at 70 °C for 15 min. After incubation at 37 °C with DNaseI, the samples were centrifuged for 90 min through linear 13 ml 5% to 30% glycerol gradients at 35,000 rpm. Fractions were collected from the bottom of the tubes and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and electron microscopy. N-terminal amino acid sequencing was performed by automated Edman degradation.

Electron microscopy

Drops of the samples were applied onto Formvar-coated copper grids (300 mesh) and negatively stained with 2% uranyl acetate. Micrographs were taken in a JEOL 2000 EXII electron microscope at an acceleration of 80 kV with Agfa Scientic film plates.

Results

Structural proteins of A2 virions

Seven protein bands were visualized after SDS-PAGE of CsCl purified A2 particles (Fig. 1, lane 2). Their apparent sizes ranged from 26 to about 150 kDa. Out of them, three, with apparent sizes of 26, 35 and 42 kDa, were more abundant than the rest. The bands were subjected to NH_2-terminal amino-acid sequencing. Noticeably, the two major bands of 35 and 42 kDa (B and C), presented the same amino ends, indicating that they were probably the products of translation of a single gene.

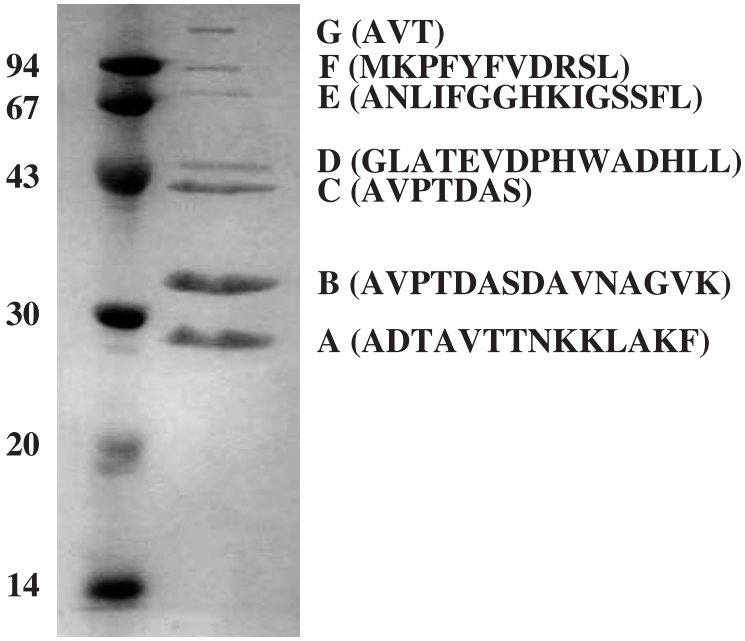

Fig. 1. Proteins from purified A2 virions separated by SDS-PAGE (15%) and stained with Coomassie brilliant blue (lane 2). Protein size standards are indicated in kDa (lane 1). N-terminal amino acid sequences of the proteins detected are shown

Besides, only band F conserved its formyl-methionine. These data were used to locate the corresponding genes in the A2 genome (see below).

To ascribe some of the structural proteins to the capsid or to the tail of the virion, purified A2 particles (Fig. 2A) were dissociated by heating and centrifuged through a glycerol gradient, which yielded enriched preparations of heads (Fig. 2B) and tails (Fig. 2C). Analysis by SDS-PAGE showed an increase of bands B and C in head enriched preparations with respect to band A, which was virtually absent (Fig. 2D, lanes 3 and 4). These proportions were reversed in tail enriched fractions (Fig. 2D, lanes 5 and 6), suggesting that the proteins included in bands B and C were the most abundant components of the head while A would be the major protein of the tail.

Analysis of the Lactobacillus casei *bacteriophage A2 genome*

The total circular length of the genome was determined to be 43,411 bp and its total G + C content is 44.8%, a value well in agreement with the 46% reported for its host *Lactobacillus casei* [27]. A total of 61 *orf*s were identified (Fig. 3) by using the following criteria: the *orf* should encode more than 50 codons, begin with an AUG or GUG start codon, be preceded by an identifiable ribosomal binding site (RBS) located at the appropriate distance to the initiation codon or the possibility of translational coupling existed. Out of the 61 *orf*s, 55 read from left to right and six read from right to left. These six *orf*s were mainly located in the early

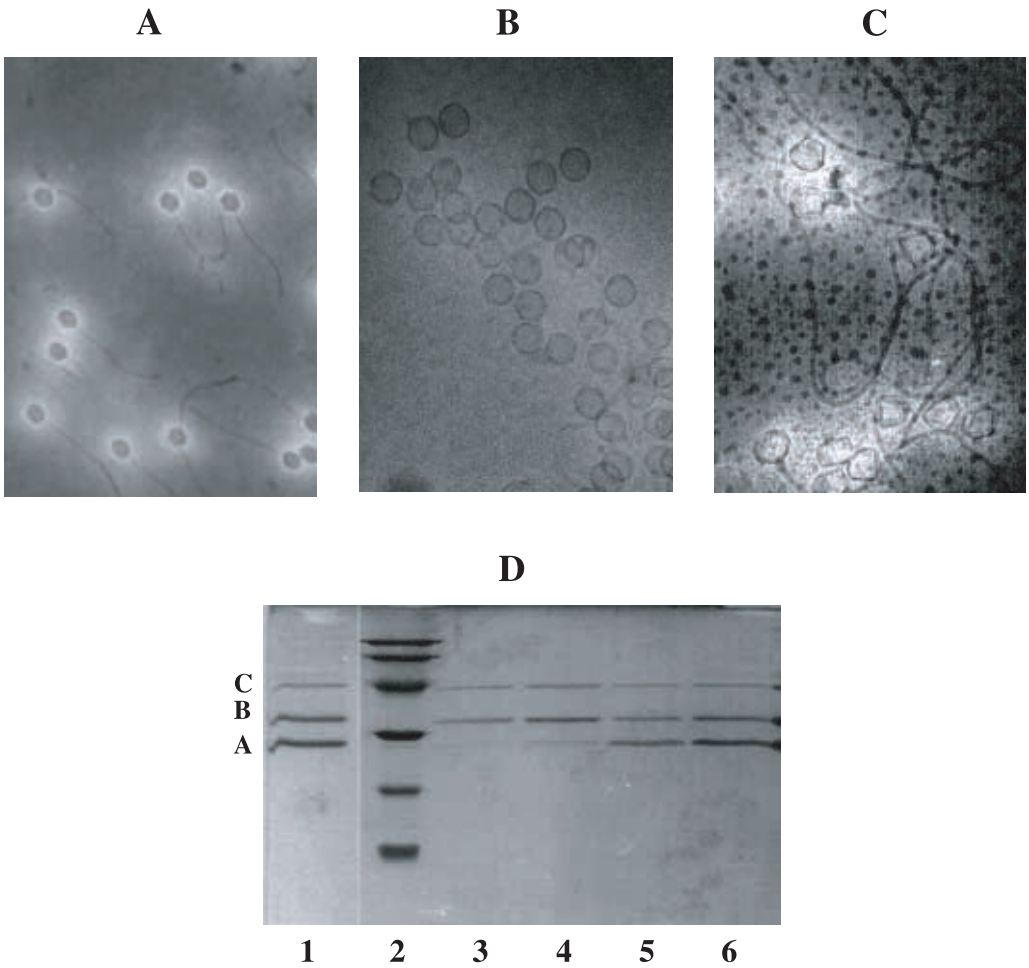

Fig. 2. Purification of bacteriophage A2 heads and tails. **A** micrographs of negatively stained virions, **B** head enriched and **C** tail enriched fractions obtained from a glycerol gradient. **D** SDS-PAGE of the fractions. *1*: complete capsids; *3* and *4*: head enriched fractions; *5* and *6*: tail enriched fractions; *2*: size standards (values are as in Fig. 1)

gene region which is placed in the centre of the phage genome. The point of divergence is located between *orf23* and *orf24*, which correspond to the lambda *cI* and *cro* homologues. In the intergenic region where the *orf*s converged (between *orf19* and *orf20*) a putative transcriptional terminator and the phage attachment site (*attP*) were identified [2]. Other putative transcriptional terminators were detected upstream of the lysis cassette and downstream of the replication module. The genome was closely packaged, indicative of transcription through polycistronic mRNAs and in some cases, several *orf*s overlapped, suggesting that they might be translationally coupled.

The deduced amino acid sequences of all the ORFs were compared with protein sequences from available databases. Significant homologies are shown in Table 1 and are discussed below:

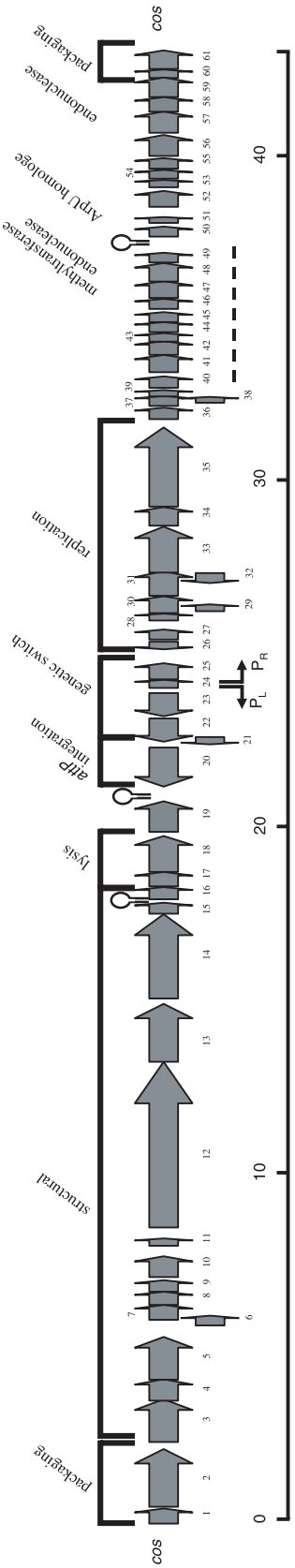

Fig. 3. Genome organization of bacteriophage A2. The 61 *orf*s are correlatively numbered, indicated by arrows proportional to their length and pointing towards their direction of transcription. The functional clusters are indicated on top of the scheme. The transcription terminators are indicated by loops while the previously characterized P_L and P_R early promoters are shown with bent arrows. The dashed line below the map indicates the dispensable stretch (under laboratory conditions) of the phage genome as determined through the analysis of several mutants showing 1 to 3 kb deletions in that region [35]

Table 1. General features of the open reading frames predicted from the DNA sequence of the bacteriophage A2 and comparison with those stored in databases

orf/previous name	Molecular size (kDa)	Predicted start position	Predicted stop position	Putative RBS (capital letters indicate matches with the 5′-end of *L. casei* 16S rRNA) (42)	ORF homology	Organisms matched	% Identity	Accession number
1/Orf4	17	102	557	auuAGGAGGUgaagugggaaaug	ORF2	*L. casei* PL-1	96 (78)	S73384
					ORF1	*S. aureus* PVL	35 (127)	AB009866
					ORF38	*L. lactis* bIL309	27 (120)	AF323670
					ORF40	*L. lactis* bIL286	25 (120)	AF323669
					putative terminase	*L. lactis* BK5-T	25 (120)	AF176025
					ORF161	*S. thermophilus* ϕSfi19	31 (72)	AF115102
2/Orf5	64.2	600	2291	uggcaaaauuuAAAGauccaaug	terminase large subunit			
					ORF2	*S. aureus* PVL	34 (564)	AB009866
					terminase	*L. lactis* bIL309	31 (551)	AF323670
					protein 34	*L. lactis* prophage pi1	31 (551)	AE006282
					ORF623	*S. thermophilus* ϕSfi21	30 (548)	AF115103
					ORF623	*S. thermophilus* ϕSfi19	30 (550)	AF115102
					terminase	*L. lactis* BK5-T	30 (552)	AF176025
3	46.1	2500	3753	aacgAAgGGAGGUgagauuaaug	portal	*S. aureus* PVL	28 (399)	AB009866
					portal	*L. lactis* bIL309	27 (341)	AF323670
					ORF397	*L. gasseri* ϕadh	24 (384)	AJ131519
					ORF24	*S. thermophilus* 7201	26 (374)	AF145054
					Portal	*S. thermophilus* DT1	26 (329)	AF085222
					Portal	*E. coli* HK97	19 (298)	AF069529
4	23.6	3707	4336	auugAAgGGAGGUgauacuaaug	ORF26	*Bacillus* ϕ105	48 (189)	AB016282
					ORF5a	*S. aureus* PVL	40 (209)	AB009866
					protease	*L. lactis* bIL309	40 (154)	AF323670
					protease	*L. lactis* bIL312	37 (172)	AF323673
					protease	*E. coli* O157:H7	35 (170)	AE005328
					ORF35	*Streptomyces* ϕC31	32 (191)	AJ006589
5	43.6	4378	5580	aaaauAAGGAGGgucacuagaug	major head			
					capsid	*L. lactis* bIL285	32 (416)	AF323668
					capsid	*S. aureus* PVL	22 (412)	AB009866
					capsid	*L. lactis* bIL309	27 (382)	AF323660
					ORF27	*Bacillus* ϕ105	23 (360)	AB016282
					head	*L. gasseri* ϕadh	24 (339)	AJ131519
					head	*L. lactis* bIL286	25 (221)	AF323669
6	11.2	5908	6207	aacaguuucAGGUucugaaaaug	DNA-packaging			
					ORF106	*S. thermophilus* ϕSfi21	27 (84)	AF115103
7	16.9	6086	6526	ugaagcaAguGaGGaagacuaug	head-tail joining			
					ORF10	*S. aureus* PVL	26 (108)	AB009866
8	14	6526	6912	acggcAAAGGcGGUgagcugaug	head-tail joining			
					ORF11	*S. aureus* PVL	34 (97)	AB044554
					ORF32	*Bacillus* ϕ105	30 (101)	AB016282
9	13.1	6912	7256	uugaaggAAGGcGGgauguaaug	head-tail joining			
					ORF12	*S. aureus* PVL	34 (64)	AB009866
10	21.7	7331	7939	ccuAAAGGAGGauuuuuaauaug	major tail			
					ORF34	*Bacillus* ϕ105	31 (124)	AB016282
					ORF13	*S. aureus* PVL	23 (142)	AB009866
11	8.5	8253	8474	uauaccuAAGGAGaUuaagcaug				

(*continued***)**

Table 1 (*continued*)

orf/previous name	Molecular size (kDa)	Predicted start position	Predicted stop position	Putative RBS (capital letters indicate matches with the 5′-end of *L. casei* 16S rRNA) (42)	ORF homology	Organisms matched	% Identity	Accession number
12	173	8787	13652	ucuaugAAAGGAGGUaaaacaug	tape measure			
					ORF1374	*S. aureus* ϕSLT	27 (567)	AB045978
					ORF15	*S. aureus* PVL	25 (536)	AB044554
					tail protein	*L. lactis* bIL286	29 (474)	AF323669
					TMP	*L. lactis* TP901-1	29 (429)	AF304433
					ORF48	*L. lactis* Tuc2009	28 (429)	AF109874
					tail protein	*E. coli* O157:H7	25 (528)	AP002559A
13	65.5	13653	15432	gAAGaAGGgggauuuucuaagug	ORF16	*L. lactis* BK5-T	30 (116)	AF176025
					ORF53	*L. lactis* bIL286	33 (105)	AF323669
					ORF52	*L. lactis* bIL309	33 (105)	AF323670
					ORF512	*S. thermophilus* ϕSfi11	30 (70)	AF158600
14	90	15507	17978	ucuuggcAGGAGGcauggcuaug	host interacting			
					ORF18	*S. thermophilus* DT2	26 (468)	AF348739
					ORF1291	*S. thermophilus* ϕSfi19	26 (456)	AF115102
					ORF1276	*S. thermophilus* ϕSfi21	26 (456)	AF115103
					ORF18	*S. thermophilus* DT1	25 (462)	AF085222
					ORF38	*S. thermophilus* 7201	30 (309)	AF145054
					ORF695	*S. thermophilus* ϕSfi11	38 (175)	AF158600
15	11.6	17988	18311	cgcaguuuAGGAGGcuguuaug				
16	13.2	18416	18766	gcaaauuAcAGGcaGUgacuaug	ORF5	Dp-1	38 (63)	Z93946
17	14.8	18780	19211	aaacugAAAGGAaugucaguaug	holin	*L. lactis* Tuc2009	45 (77)	AF109874
					holin	*L. lactis* TP901-1	45 (77)	AF304433
					LysA	*L. lactis* ϕLC3	44 (77)	AF066865
					ORF87	*S. thermophilus* ϕSfi21	48 (31)	AF115103
					holin	*S. thermophilus* ϕSfi11	48 (31)	AF158600
					holin	*S. thermophilus* ϕSfi19	48 (31)	AF115102
18	37.4	19213	20265	uagugAGGAGGUaaaauaguaug	lys	*Lactobacillus* PL-1	90 (35)	AJ251790
					enterolysinA	*E. faecalis*	66 (151)	AF249740
					lysin	*L. lactis* r1t	42 (189)	U38906
					lysin	*L. monocytogenes* A511	35 (203)	X85010
					autolysin	*S. pneumoniae*	41 (104)	AE007483
19	33.2	20388	21290	ugaAgAGGAGGUgguucaaaaug				
20/Int	43.8	22876	21721	uuacAcAGGAGGgcaauaauaug	integrase	*Lactobacillus* ϕg1e	33 (333)	X90510
					integrase	*L. oenos* 10MC	34 (337)	U77495
					integrase	*L. lactis* bIL309	27 (380)	AF323670
					ORF359	*S. thermophilus* ϕSfi21	29 (376)	AF115103
					ORF1	*S. thermophilus* ϕO1205	29 (376)	U88974
					integrase	*L. lactis* BK5-T	26 (388)	AF176025
21/Xis	8.8	23182	22952	cggcucucGAauUcuagaaaaug				
22/OrfA	24.4	23720	23043	gauuuuuuuGGAGGggaauaaug	ORFtxe3	*C. difficile*	56 (103)	U25132
					ORFdtxA	*C. difficile*	55 (103)	X57190
23/CI	25.2	24453	23779	ucauucuuGAGGUgagauacaug	cI-like	*L. lactis* prophage pi3	56 (79)	AE006373
					repressor	*L. lactis* bIL286	55 (79)	AF323669
					ORF127	*S. thermophilus* ϕSfi21	35 (111)	AF115103
					ORF221	*L. lactis* mv4	24 (219)	AF182207
24/Cro	9.1	24615	24860	aaacgAAAGGAGGcaauccaaug	repressor	*L. lactis* bIL286	59 (76)	AF323669
					cro-like	*L. lactis* prophage pi3	59 (76)	AE006373
					ORF75	*S. thermophilus* ϕSfi21	43 (76)	AF115103
25/Ant	17.8	24857	25339	gaacgAAAGGgGaUgacgguaug	antirepressor	*L. innocua* A118	53 (158)	AL596163
					ORF38	*L. lactis* BK5-T	49 (147)	AF176025
					ORF5	*L. lactis* r1t	46 (147)	U38906
					antirepressor	*L. lactis* bIL309	40 (171)	AF323670

(*continued*)

Table 1 (*continued*)

orf/previous name	Molecular size (kDa)	Predicted start position	Predicted stop position	Putative RBS (capital letters indicate matches with the 5′-end of *L. casei* 16S rRNA) (42)	ORF homology	Organisms matched	% Identity	Accession number
26/Orf70r	7.7	25957	25745	cccgAAAGGAGGUgauugauaug				
27/Orf96	10.5	26023	26313	acaggAAAGGAGGaaaugccaug	ORF117	*Lactobacillus* ϕg1e	25 (70)	X98106
					SA1799	*S. aureus* ϕN315	27 (70)	AP003135
					ORF111b	*L. lactis* ul36.2	30 (71)	AF212847
					ORF6	*L. lactis* r1t	32 (71)	U38906
28/Orf67	7.4	26591	26794	gaaacuAAGGgGaugacgugaug				
29/Orf163	17.9	26812	27303	ugauagAAAGGAaacuuauuaug	gp46	*L. monocytogenes* PSA	38 (153)	AJ312240
					ORF157	*S. thermophilus* ϕSfi19	30 (158)	AF158600
					ORF157	*S. thermophilus* ϕSfi21	29 (158)	AF115103
					ORF8	*S. thermophilus* ϕO1205	29 (156)	U88974
					ORF45	*L. lactis* BK5-T	30 (160)	AF176025
					ORF169	*L. lactis* ϕ31	30 (160)	AJ292531
30/Orf69	7.6	26870	27079	uuuAuuaGAacUagcugaagaug				
31/Orf240	26.4	27313	28035	ugauuaaguAGGAGGaaaucaug	ORF233	*S. thermophilus* ϕSfi19	55 (234)	AF158601
					ORF9	*S. thermophilus* ϕO1205	54 (234)	U88974
					ORF233	*S. thermophilus* ϕSfi21	54 (234)	AF115103
					ORF233	*S. thermophilus* ϕSfi11	54 (234)	AF158600
					ORF32	*S. thermophilus* DT1	52 (234)	AF085222
					ORF46	*L. lactis* BK5-T	51 (232)	AF176025
					NTP-binding	*L. gasseri* ϕadh	40 (214)	AJ131519
32/Orf93r	10.2	27965	27684	caaacuuAAaGAGGUccucaaug				
33/Orf455	50	27998	29365	uaaguuuGGAGGUgacagugaug	helicase	*L. innocua* clip11262	45 (415)	AL596173
					helicase	*L. monocytogenes* PSA	46 (415)	AJ312240
					helicase	*L. gasseri* ϕadh	44 (450)	AJ131519
					helicase	*L. lactis* ϕ31	40 (447)	AJ292531
					ORF10	*S. thermophilus* ϕO1205	40 (446)	U88974
					ORF443	*S. thermophilus* ϕSfi21	40 (446)	AF115103
34/Orf180	19.8	29362	29904	agaaAAcGGAGGaaacaagcaug	ORF175	*L . gasseri* ϕadh	41 (175)	AJ131519
					gp49	*L. monocytogenes* PSA	35 (157)	AJ312240
					protein 11	*S. thermophilus* ϕO1205	27 (153)	U88974
					ORF34	*S. thermophilus* DT1	28 (153)	AF085222
					ORF4	*L. lactis* ϕ31	30 (173)	AJ292531
					ORF151	*S. thermophilus* ϕSfi19	26 (153)	AF115102
					ORF151	*S. thermophilus* ϕSfi11	26 (153)	AF158601
					ORF124	*S. thermophilus* ϕSfi21	25 (153)	AF115103
					ORF10	*Bacillus* ϕ105	27 (175)	AB016282
					SSB	*B. halodurans*	57 (19)	AP001520
35/Orf770	84.7	29923	32235	aggagcgcgAGGuGGUgaauaug	primase	*L. gasseri* ϕadh	49 (763)	AJ131519
					primase	*L. monocytogenes* PSA	50 (739)	AJ312240
					ORF11	*Bacillus* ϕ105	37 (773)	AB016282
					gp9a	*Streptomyces* ϕC31	39 (349)	AJ006589
					primase	*S. thermophilus* ϕO1205	22 (330)	U88974
					ORF508	*S. thermophilus* ϕSfi19	22 (330)	AF115102
36	12.9	32494	32847	ucccguugaAAAaGAaaggaaug				

(*continued*)

Table 1 (*continued*)

orf/previous name	Molecular size (kDa)	Predicted start position	Predicted stop position	Putative RBS (capital letters indicate matches with the 5′-end of *L. casei* 16S rRNA) (42)	ORF homology	Organisms matched	% Identity	Accession number
37	10.7	32877	33170	acgaaauucAAucAGaaauuaug	gp51	*L. monocytogenes* PSA	48 (94)	AJ312240
					ORF38	*S. thermophilus* DT1	51 (89)	AF085222
					ORF15	*S. thermophilus* ϕO1205	51 (86)	U88974
					ORF7	*L. lactis* ϕ31	47 (90)	AJ292531
					ORF106	*S. thermophilus* ϕSfi19	47 (90)	AF115102
					ORF90	*L. gasseri* ϕadh	49 (53)	AJ131519
38	6.5	32978	33157	auugcuGGAccaccgaaaggaug				
39	6.8	33161	33349	aagcuAuuGGAGGcuaacaaaug				
40	12.4	33405	33740	uggaaagAAGAGGacucgaaaug	ORF5	*L. lactis* bIL286	55 (40)	AF323669
41	19	33874	34380	aucaaAcAGGAGGacgaaaaaug	gp51	*L. innocua* A118	27 (84)	
42	15.3	34370	34777	gaaAgAaGAGGggacugacgaug				
43	10.2	34774	35037	acaggAgAGGuGGUaaagcuaug				
44	10.9	35054	35341	gugaaAcuGGAGGcagagaaaug				
45	10.7	35338	35625	uggaaAAuaGAGGcggagaaaug	ORF29	*L. lactis* bIL285	29 (41)	AF323668
46	10.8	35746	36036	acaauugguGAGuUcugcaaaug				
47	19.2	36047	36535	aaaagaauAgAGGcGGagaaaug	ORF161	*Lactobacillus* ϕg1e	59 (160)	X98106
					methyltransf	*L. innocua* A118	55 (159)	AJ242593
					methyltransf	*S. enterica*	45 (160)	AL627268
48	20.6	36532	37083	auucagaaAgAGGcGGagaaaug	e11	*L. lactis* bIL170	36 (157)	AF009630
					endodeoxy ribonuclease	*B. subtilis* ϕE	29 (172)	U04813
					ORF168	*L. lactis* LL-H	44 (99)	L37351
					endodeoxy ribonuclease	*B. subtilis* ϕSP82	26 (185)	U04812
					intron ORF	*B. subtilis* ϕSPO1	30 (166)	M37686
					ORF6	*L. lactis* ϕ31	37 (170)	AJ292531
49	10.7	37080	37367	gagcuAcuGGAGGgaaaacaaug				
50	11.7	37702	38163	aaacaaauuauaccauaagagug	ArpU	*E. hirae*	32 (123)	Z50854
					ORF24	*S. thermophilus* ϕO1205	32 (125)	U88974
					gp137	*S. thermophilus* ϕSfi11	29 (124)	AF158600
					ORF147	*Lactobacillus* ϕg1e	25 (127)	X98106
51	6.3	38251	38430	ccgcguuuuuccuccGAGccaug				
52	18.3	38725	39198	gcauugAuaccGGuuaagcaaug				
53	8.2	39294	39518	accAAGaAaaauguccCgauaug				
54	9.4	39555	39809	auaauuucGGAGGcgauggcaug				
55	13.3	39833	40186	aauuuAcGGAGGcgaguagaug				
56	23.2	40213	40824	cagaacaauAuGAaucauugaug				
57	23.8	40887	41510	uucaacgcguaAAccGAacgaug	e11	*L. lactis* bIL170	39 (171)	AF009630
					ORF6	*L. lactis* ϕ31	35 (172)	AJ292531
					intron ORF	*B. subtilis* ϕSPO1	35 (175)	M37686
					ORF36.1	*B. subtilis* SPP1	42 (123)	X67865
					endodeoxy ribonuclease	*B. subtilis* ϕE	30 (186)	U04813
58	15.2	41514	41936	ucgaguuGGcGGuguaagucaug				
59/Orf1	21.4	41949	42530	uaucuuAGGAGGaauuacacaug				
60/Orf2	14.4	42511	42882	gGGAccUgaucucuuuaagaaug	ORF63	*S. aureus* PVL	41 (60)	AB044554
					ORF170	*L. gasseri* ϕadh	43 (66)	AJ131519
					ORF104b	*S. aureus* ϕSLT	40 (64)	AB045978
					ORF63	*L. lactis* BK5-T	41 (62)	AF176025
					ORF39	*L. lactis* bIL286	34 (76)	AF323669
61/gp3	16.7	42879	43313	ucauauuguuccAAGcGaagaug	gp54 terminase	*L. monocytogenes* PSA	39 (56)	AJ312240
					ORF1	*L. casei* PL-1	97 (71)	S73384

Packaging cluster

It may be composed by *orf61*, *orf1*, *orf2* and possibly *orf60* and *orf59* (there might be a transcriptional coupling between *orf59*, *orf60* and *orf61*). The gene product of *orf61* has been experimentally proven to be the small subunit of the terminase enzyme [19] while the product encoded by *orf2* showed homology with putative large terminase subunits of other phages (Table 1). In this cluster it is noticeable that the gene encoding the small terminase subunit (*orf61*) is located to the other side of the *cos* ends with respect to *orf1* and *orf2*, which is uncommon. Curiously, *orf60*, *orf1* and *orf2* share about 35% identity with the corresponding genes of PVL, a phage that infects *Staphylococcus aureus*, while no counterpart is found in that phage for *orf61*. In PVL, the homologue of *orf1* is considered to encode the small subunit of the phage terminase while in A2 *orf61* has been demonstrated to encode such function. Consequently, it might be interesting to determine whether A2 presents two genes, *orf61* and *orf1* with redundant activities. The *cos* sequence consists in a single-stranded extension of 13 nucleotides at the 3′ termini of the linear genome, being bisected by an axis of hyphenated twofold rotational symmetry [19].

Head morphogenetic cluster

Out of the three *orfs* included in this cluster, the products of *orf3* and *orf5* are present in the A2 virion. The first matches the NH_2 terminal sequence of band D (Fig. 1) from the fifth amino-acid onwards, i.e. it is proteolytically processed during or after capsid formation to render gp3. This protein may correspond to the portal protein of phage A2, based on its position in the genome and its sequence similarity to the portal protein of coliphage HK97 and to several others, putatively ascribed to that function (Table 1). An internal stretch of the translation product of *orf5*, starting at residue 124, corresponded to the amino terminus rendered by bands B and C (Fig. 1), i.e.: the major head proteins of the phage, indicating that they suffered proteolytic processing. These proteins have been named gp5A and gp5B respectively. The 123 residues polypeptide that might result from the processing presents characteristics typical of scaffolding proteins. Finally, the product of *orf4* presents homology with proteases of different phages.

The neck region

It includes from *orf6* to *orf9*. They have been ascribed this function because of their position between the clusters of the head and tail genes. They also present partial homology (about 30%) with PVL genes postulated to encode the head-tail joining proteins of that phage and are probably involved in a common function because *orf6* and *orf7* have 120 bp in common (although in different frames) while *orf7*, *orf8* and *orf9* overlap their stop and start triplets, suggesting that there exist a translation coupling of all of them.

The tail morphogenetic cluster

The first of the genes included in this cluster, *orf10*, corresponds to the major tail protein (band A in Fig. 1), which is devoid of just the f-met residue. Besides, other three virion proteins (bands G, E and F in Fig. 1) are encoded in this cluster by *orf12*, *orf13* and *orf14* respectively. The gp12 product appears to be the tape measure protein, i.e.: the template over which the other tail proteins polymerize [29, 30]. The mature form of gp12 is devoid of its 172 N-terminal amino acid residues, being the size of the processed polypeptide of 154 kDa. The precise role of gp13 could not be deduced from sequence comparisons (Table 1). Finally, gp14 may be the protein that recognizes the host receptor both because it presents 13 collagen like motifs (Gly-X-Y) in its second half, which is typical of these proteins [52] and because it shares amino acid sequences along its entire length with gp18 of the DT series of *S. thermophilus* phages (Table 1), which has been demonstrated to determine their host specificity [17]. The cluster also comprises *orf11* and *orf15*, which encode small proteins of about 10 kDa that do not present good matches with any of the proteins included in the databases. The respective stop and start triplets of *orf12* and *orf13* are in immediate vicinity, suggesting that they might be translationally coupled. Finally, the start codon of *orf13* is GUG. Downstream of *orf15* there exist a stem loop followed by a row of thymidines that may act as a transcription terminator.

The lysis cassette

In this part of the genome three *orfs* have been identified. Out of them, *orf17* and *orf18* may be the phage holin and lysin respectively, based on sequence homology (Table 1) and on the lethality that they induce on *E. coli* when cloned behind an inducible promoter (unpublished data). However, no role could be ascribed to *orf16*.

Orf19

This *orf* is puzzling. It does not present homology to any of the sequences of the data bases, is not closely related to previous genes, is the last one transcribed from left to right and the stretch of DNA into which it is located was the only one of the whole genome that resulted to be unclonable under any circumstances. Besides, it is 35% G + C, which might indicate a horizontal transfer from another organism. Its position in the gene map of the phage points towards a gene involved in lysogenic conversion, but this extreme has not been addressed. It is followed by a transcription terminator structure that may act as such for the converging genes encoding the integration of the phage into the host genome as well.

The integration cassette

It includes *orf21* and *orf20* plus the *attP* attachment site of the phage into the genome of its hosts. The functionality of *orf20* as the phage integrase plus the role

of the *attP* sequence have been already described [2, 3], but the suggested function of *orf21* as the phage excisionase could not be demonstrated (M. A. Alvarez and J. E. Suarez, unpublished data).

The genetic switch

It is composed by two sets of genes that read in opposite directions. The first set comprises *orf23*, which encodes the *cI* like repressor of the lytic cycle [20, 35] and *orf22*, which predicts a protein whose carboxi terminal half presents a substantial degree of homology (55%) with the second half of the product of gene *tcdC* from the pathogenicity island of *Clostridium difficile* [23].

The second set of genes involved in the genetic switch of the phage comprises *orf24* and probably *orf25*. The first encodes the functional homologue of lambda Cro and its role as repressor of the lysogenic P_L promoter has been proven [33, 34]. The second gene might encode an antirepressor that would be responsible for inactivation of CI, based on its sequence similarity to putative antirepressors from different phages of low G + C Gram positive bacteria (Table 1). However, gp25 might be a defective protein; while its homologues are about 255 residues long, it only has 160 amino acids, due to a stop codon placed in that position (this particular stretch has been sequenced several times and even using directly phage DNA as template). On the other hand, gp25 presents the single stranded DNA binding motif of the BRO (baculovirus repeated orf) proteins, some of which appear to participate in DNA replication [58]. Given the position of *orf25* immediately upstream of the phage replication module, its participation in this function can not be ruled out.

Replication module

It has been already described [39]. In short, it putatively comprises from *orf26* to the *ori* sequence that is located immediately downstream of *orf35*. Although most deduced proteins have counterparts in other bacteriophages (Table 1), only three can be ascribed to replication functions, namely gp33, which presents helicase typical motifs, gp34, that might act as single-stranded DNA binding protein while gp35 might be the phage primase. In the same report the *ori* sequence was shown to act as a *bona fide* origin of replication upon supply of phage encoded replication factors.

The extreme right region

It comprises about 10 kb characterized by the presence of multiple *orf*s that originate mostly small proteins of unknown function (18 would be 15 kDa or less while the other seven have between this figure and 23 kDa). Many of these proteins seem to be dispensable under laboratory conditions since deletion mutants lacking different DNA stretches that go from *orf40* to *orf49* do not show any special phenotype (Fig. 3) [35]. However, informed guesses may be formulated with

respect to some of the *orf*s. For example, the product of *orf47* is quite similar to N-4 cytosine-specific methyltransferases from the *Listeria* bacteriophage A118 and *Neisseria gonorrheae* (M.*Lmo*A118I and M.*Ngo*MXV, respectively) [7, 49]. The hypothetical gp47 might methylate the newly synthesized viral DNA at specific residues. Similarly, *orf48* and *orf57* seem to encode endodeoxyribonucleases frequently associated to self-splicing group I introns [22] since their deduced products present endonuclease motifs. When compared, these two hypothetical enzymes show 36% identity and 53% similarity, which are in the same range than the values obtained when each of them is compared with related proteins from other bacteriophages, thus excluding a recent duplication event in A2. Finally, *orf50* encodes a protein that presents a relatively high degree of homology with ArpU, which has been proposed to control the export of muramidase-2, a peptidoglycan hydrolase found in the wall of *Enterococcus hirae* [12]. Although similar genes have been found in the genomes of other bacteriophages (Table 1), the meaning of this protein in their biology is beyond our present understanding.

Transcription analyses of the morphogenetic region of bacteriophage A2

Total RNA was isolated from *L. casei* at various times post A2 infection. Northern blotting using probes internal to *orf5*, *orf10*, *orf12* and *orf14*, rendered production of a single transcript in excess of 20 kb, already reported as specific for *orf61* [19], from 60 min post-infection onwards. However, a probe internal to *orf17* rendered a much shorter transcript (unpublished data) indicating that the transcriptional terminator located between *orf15* and *orf16* is probably functional. This implies that the transcript covers a total of at least 16 *orf*s in a stretch of about 19 kb (from *orf61* to *orf15*) and possibly more towards its 5′ end, due to the overlapping of *orf61* with *orf60* and of this with *orf59*. It is additionally deduced that the morphogenetic transcript crosses through the sealed cohesive ends of the intracellular phage.

Discussion

The bacteriophage A2 presents a long tail of 280 nm. In lambda it has been found that the product of gene H acts as a template in determining the length of the tail. During this process gpH suffers a proteolytic cleavage that reduces its size from 92,292 Da to about 80,000 Da (a reduction of about 13.4%) [30]. Six monomers of this processed form assemble in a rod-like fashion, so that the total length of gpH* determines the length of the tail [29]. In A2, only the product of *orf12* would be long enough to act as the tape measure protein of the tail and is processed (from 173,050 to 154,236 Da, a 10.9% reduction in size), being the longest virion protein. Finally, it is related to the characterized tape measure protein of phage TP901-1 [46]. If we assume that, as established for lambda, six monomers assemble in each particle, we can calculate the abundance of other virion proteins simply by comparing the densitometry scans of the bands visualized in SDS-PAGE

Table 2. Number of copies per virion of analogous proteins from A2 and lambda[a]

A2/λ	Function	Copies/particle A2	Copies/particle λ
gp12/gpH	Tape measure	6	6
gp14/gpJ	Tail fiber	9	3
gp13/?		15	–
gp3/gpB	Portal	12	12
gp5B/-	Major head	112	–
gp5A/gpE	Major head	380	405
gp10/gpV	Major tail	650	192

[a]Data from [9]

A2 proteins, separated as indicated in Fig. 1, were scanned and their densities compared to that of gp12, correcting for their respective sizes and assuming that 6 copies of gp12 were present per phage particle. The data shown are the mean of the values obtained from three independent gels. The actual values never differed by more than 15% from the figures shown

gels and correcting the values by the sizes of the corresponding polypeptides (Table 2).

The major tail protein is the most abundant of those identified in phage A2, with about 650 copies per particle. The tail of A2 is made out of about 72 rings, indicating that each might be constituted by 9 monomers of gp10. This value conforms to the three fold symmetry found for the tails of phage lambda, in which about 32 rings and 6 monomers per ring have been reported [9] and of phage ΦCbK, whose tail presents about 78 rings with three monomers in each of them [43].

Of interest are also the two major head proteins, which share their amino termini indicating that both are derived from *orf5*. It appears that the A2 virion contains about 380 monomers of the smallest form (gp5A) and 112 of the biggest (gp5B). Since the size of gp5A, as deduced from SDS-PAGE, conforms to the expected value of the *orf5* translation product after processing, it was deduced that gp5B results from translation of a longer message as a consequence of a reading through the stop codon of *orf5*. Consistently with this, 3′ of *orf5* there is quite a long intergenic segment whose translation in a particular frame presents a 38% identity to the translation product of a region that lies behind gene 10 of phage T3, which encodes its coat protein. In this case, during translation the ribosomes undertake a -1 frameshift close to the stop codon in about 10% of the transcripts, which results in a polypeptide (10B) 86 amino-acids longer than the "canonical" one (10A) [10]. Preliminary evidence indicates that this is the case as well for A2, resulting in a gp5B product that is 85 residues longer than gp5A, being this difference consistent with their respective sizes as observed by SDS-PAGE.

Furthermore, gp5A and B are devoid of their amino terminal 123 residues polypeptide. This hypothetical peptide presents characteristics typical of scaffolding proteins such as a N-terminal end rich in charged residues and an α-helix structure which includes a leucine zipper that allows the formation of a coiled–coil

superstructure, that is essential for stabilization of the pro-head components, as it occurs with the scaffolding protein of P22 [44].

The other three (minor) virion proteins identified after SDS-PAGE corresponded to gp3, gp13 and gp14. The first was postulated to be the portal protein due to the location of *orf3* in the phage genome and to its partial homology with the portal protein of phage HK97. In addition, it appears that each particle contains 12 copies of gp3, which coincides with the number found in the portals of all phages in which this detail is known [8, 31]. Furthermore, gp3 is proteolytically processed as it happens with the portal protein (gpB) of lambda [55].

The A2 virions contain about 9 copies of gp14. The presence of several collagen-like motifs links this protein to the host recognition protein of phage DT1 [17]. This points towards a location of gp14 in the tail tip as it happens for gpJ of lambda although, there the number of monomers per particle is reduced to three [9].

Analysis of the cleavage regions of the portal, major-head and tape measure proteins revealed that all have in common the sequence Pho-Pho-Arg↓ (where Pho stands for hydrophobic amino-acids). This is reminiscent of the case of T4, where a phage encoded protease recognizes the sequence Pho-Pho-Glu↓ [4], i.e.: two hydrophobic amino-acids followed by a charged residue (although of an opposite polarity). On the other hand, the protease that fulfils processing of the major head protein of several phages was found to be encoded by a determinant located immediately upstream of the corresponding gene [16, 36]. Since *orf4* occupies such a position in the A2 genome and its hypothetical product presents clear sequence similarities to proteases of different phages, we propose that gp4 would be the processing enzyme of the morphogenetic proteins in this phage.

To the centre of the A2 genome lie *orf19* and *orf22* that might be involved in lysogen conversion. The first one, *orf19*, is located between the lysin gene and the *attP* sequence. This position is occupied by virulence related genes in phages that infect pathogenic Gram positive bacteria, such as the temperate PVL bacteriophage of *Staphylococcus aureus*, which harbours two genes that encode the Panton–Valentine leukocidin activity [28]. Although the putative product of *orf19* does not have matches in the data bases, its low G + C content and the lethality it induces upon multicopy cloning might suggest a horizontal transfer of the gene to A2 and the acquisition by the phage of a potentially toxic activity.

The second half of gp22 is related to the carboxi terminal end of the *tcdC* product, which is encoded in the pathogenicity island of virulent *C. difficile* strains [23] and has been postulated to regulate expression of the adjacent enterotoxin A and citolysin B genes [26]. However, it has been suggested [37] that gp22 might be involved in superinfection exclusion of hetero-immune phages, based on the location of its determinant in the A2 genome and in the presence of two transmembrane helices in its amino terminus. This suggests that *orf22* might be a chimaeric gene of unknown function. However, *orf22* is expressed during the lysogenic cycle [33], thus fulfilling one of the requisites needed for a lysogen conversion gene. Furthermore, most of gp22 is predicted to lie to the outside of

the plasma membrane which, coupled to its low pI value (4.3), might be taken as indications of its proposed role as a DNA injection blockage protein. Another interesting feature of the A2 genome is the presence of a DNA stretch, from *orf40* to *orf49*, which is dispensable, at least under laboratory conditions. This is amazing, given the high gene density in the genome as a whole. A possible role for this stretch of DNA might be related to the limits in the amount of DNA that can be packaged in the phage head. If this were the case, this segment would allow the genome to reach the lower packaging limit of the phage and, upon acquisition of a useful gene, a deletion mutant having lost part of this zone would have a selective advantage over its siblings due to the strong negative selection that is usually observed when the packaging upper limit is approached [18, 53].

Transcription analysis of the morphogenetic region revealed that all genes are included into a single operon. This is a common situation; it has been reported for lambda and P22 [40, 56] and even for the unrelated phage Φ29 [38]. This might be a simple way of getting coordinated expression. However, not all proteins arising from the translation of a single polycistronic transcript will be needed in equivalent amounts. This may be solved by controlling the efficiency of recognition of the Shine-Dalgarno sequences by the ribosomes. In A2 all genes that originate virion proteins present SD sequences matching quite well the 5′-end of the *L. casei* 16S rRNA (Table 1) [42] while others, presumed to be needed in smaller amounts, such as the terminase subunits, are only scarcely similar to it. Complementarily, translational coupling, which is frequent in A2, may help to maintain constant the relative amounts of two proteins [21].

Comparison of the A2 genome to those of other phages reveals that: i) no extensive homology is observed with the *Lactobacillus* phages sequenced to date with the possible exception of the replication module of Φadh, ii) it is most related to the *S. aureus* phage PVL (although this similarity is restricted to the morphogenetic cluster) and iii) it is a new member of the Sfi21 group of phages (based on its general organization and sequence similarity of their respective morphogenetic, integration and part of the replication clusters [6, 48]. However, PVL has been recently classified as belonging to the TP901-like monophyletic cluster, while Sfi21 is located in a separate, polyphyletic cluster [50]. It might be thus deduced that A2 represents a bridge between these two phage groups, being a further example of the mosaic organization that results from the frequent interexchange of genetic information observed among dsDNA phages [26]. This mechanism of variability constitutes a huge evolutionary alternative to the mutational versatility of RNA viruses and the sexual processes of cellular organisms, by which new variants of dsDNA bacteriophages might arise, justifying in part their parasitism to virtually all kinds of prokaryotic organisms, including *Archea* [47].

Acknowledgements

These studies were supported by grants ALI99-0821, BIO2000-3621 and BMC2002-0638 from CICYT (cosponsored by the Plan FEDER from the European Union). V. L. was the recipient of a FPI (Spanish Ministry of Education) fellowship.

References

1. Accolas JP, Peigney C, Limsowtin GKY, Cluzel JP, Sechaud L (1994) Lutte contre les bacteriophages dans l' industrie laitiere. In: de Roissart H, Luquet FM (ed) Bactéries lactiques vol. 1. Lorica, Uriage France, pp 473–492
2. Alvarez MA, Herrero M, Suárez JE (1998) The site-specific recombination system of the *Lactobacillus* species bacteriophage A2 integrates in gram-positive and gram-negative bacteria. Virology 250: 185–193
3. Alvarez MA, Rodriguez A, Suárez JE (1999) Stable expression of the *Lactobacillus casei* bacteriophage A2 repressor blocks phage propagation during milk fermentation. J Appl Microbiol 86: 812–816
4. Black L, Showe M (1983) Morphogenesis of the T4 head. In: Mathews E, Kutter E, Mosig G, Berget P (ed) Bacteriophage T4. ASM Publications, Washington, pp 219–245
5. Brussow H (2001) Phages of dairy bacteria. Ann Rev Microbiol 55: 283–303
6. Brussow H, Desiere F (2001) Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol Microbiol 39: 213–222
7. Bujnicki JM, Radlinska M (2001) Cloning and characterization of M.LmoA118I, a novel DNA:m4C methyltransferase from the *Listeria monocytogenes* phage A118, a close homolog of M.NgoMXV. Acta Microbiol Pol 50: 155–160
8. Carazo JM, Santisteban A, Carrascosa JL (1985) Three-dimensional reconstruction of bacteriophage phi 29 neck particles at 2×2 nm resolution. J Mol Biol 183: 79–88
9. Casjens S, Hendrix R (1988) Control mechanisms in dsDNA bacteriophage assembly, In: Calendar R (ed) The bacteriophages, vol 1. Plenum Press, New York, pp 15–91
10. Condreay JP, Wright SE, Molineux LJ (1989) Nucleotide sequence and complementation studies of the gene 10 region of bacteriophage T3. J Mol Biol 207: 555–561
11. Daly C, Fitzgerald GF, Davis R (1996) Biotechnology of lactic acid bacteria with special reference to bacteriophage resistance. Antonie van Leeuwenhoek 70: 99–110
12. del Mar Lleo M, Fontana R, Solioz M (1995) Identification of a gene (arpU) controlling muramidase-2 export in *Enterococcus hirae*. J Bacteriol 177: 5912–5917
13. Desiere F, Lucchini S, Canchaya C, Ventura M, Brüssow H (2002) Comparative genomics of phages and prophages in lactic acid bacteria. Antonie van Leeuwenhoek 82: 73–91
14. Devereux J, Haeberil P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395
15. Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *Escherichia coli* by high voltage electroporation. Nucleic Acids Res 16: 6127–6145
16. Duda RL, Martincic K, Xie Z, Hendrix RW (1995) Bacteriophage HK97 head assembly. FEMS Microbiol Rev 17: 41–46
17. Duplessis M, Moineau S (2001) Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. Mol Microbiol 41: 325–336
18. Feiss M, Fisher RA, Crayton MA, Egner C (1977) Packaging of the bacteriophage lambda chromosome: effect of chromosome length. Virology 77: 281–293
19. García P, Alonso JC, Suárez JE (1997) Molecular characterization of the *cos* region of the *Lactobacillus casei* bacteriophage A2. Gene product 3, gp3, specifically binds to its downstream *cos* region. Mol Microbiol 23: 505–514
20. García P, Ladero V, Alonso JC, Suárez JE (1999) Cooperative interaction of CI protein regulates lysogeny of *Lactobacillus casei* by bacteriophage A2. J Virol 73: 3920–3929
21. Georgopoulos C, Tilly K, Casjens S (1983) Lambdoid phage head assembly. In: Hendrix R, Roberts J, Stahl F, Weisberg R (ed) Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp 279–304

22. Goodrich-Blair H, Shub DA (1994) The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. Nucleic Acids Res 22: 3715–3721
23. Hammond GA, Johnson JL (1995) The toxigenic element of *Clostridium difficile* strain VPI 10463. Microb Pathogenesis 19: 203–213
24. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF (1999) Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proc Natl Acad Sci USA 96: 2192–2197
25. Herrero M, de los Reyes-Gavilán CG, Caso JL, Suárez JE (1994) Characterization of ϕ393-A2, a bacteriophage that infects *Lactobacillus casei*. Microbiology 140: 2585–2590
26. Hundsberger T, Braun V, Weidmann M, Leukel P, Sauerborn M, von Eichel-Streiber C (1997) Transcription analysis of the genes tcdA-E of the pathogenicity locus of *Clostridium difficile*. Eur J Biochem 244: 735–742
27. Kandler O, Scillinger U, Weis N (1986) *Lactobacillus*. In: Snearth PHA, Mair NS, Sharpe ME, Holt JG (ed) Bergey's Manual of Systematic Bacteriology, vol 2. Williams & Wilkins, Baltimore, pp 1209–1234
28. Kaneko J, Kimura T, Narita S, Tomita T, Kamio Y (1998) Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton–Valentine leukocidin genes. Gene 215: 57–67
29. Katsura I (1987) Determination of bacteriophage λ tail length by a protein ruler. Nature 327: 73–75
30. Katsura I, Hendrix RW (1984) Length determination in bacteriophage lambda tails. Cell 39: 691–698
31. Kochan J, Carrascosa J, Murialdo H (1984) Bacteriophage lambda preconnectors: purification and structure. J Mol Biol 174: 433–447
32. 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
33. Ladero V, García P, Alonso JC, Suárez JE (1999) A2 Cro, the lysogenic cycle repressor, specifically binds to the genetic switch region of *Lactobacillus casei* bacteriophage A2. Virology 262: 220–229
34. Ladero V, García P, Alonso JC, Suárez JE (2002) Interaction of the Cro repressor with the lysis/lysogeny switch of the *Lactobacillus casei* temperate bacteriophage A2. J Gen Virol 83 2891–2895
35. Ladero V, García P, Bascaran V, Herrero M, Alvarez MA, Suárez JE (1998) Identification of the repressor-encoding gene of the *Lactobacillus* bacteriophage A2. J Bacteriol 180: 3474–3476
36. Martín AC, López R, García P (1998) Pneumococcal bacteriophage Cp-1 encodes its own protease essential for phage maturation. J Virol 72: 3491–3494
37. McGrath S, Fitzgerald GF, van Sinderen D (2002) Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. Mol Microbiol 43: 509–520
38. Mellado RP, Barthelemy I, Salas M (1986) In vivo transcription of bacteriophage phi 29 DNA early and late promoter sequences. J Mol Biol 191: 191–197
39. Moscoso M, Suárez JE (2000) Characterization of the DNA replication module of bacteriophage A2 and use of its origin of replication as a defense against infection during milk fermentation by *Lactobacillus casei*. Virology 273: 101–111
40. Murialdo H, Siminovitch L (1972) The morphogenesis of bacteriophage lambda. IV. Identification of gene products and control of the expression of the morphogenetic information. Virology 48: 785–823

41. Murray V (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res 17: 8889
42. Nakagawa T, Mukai H, Shimada M, Fujino K, Katou I (1995) Method for detecting bacterium of genus *Lactobacillus*. Patent number JP 1995051100 – A10.
43. Papadopoulos S, Smith PR (1982) The structure of the tail of the bacteriophage phi CbK. J Ultras Res 80: 62–70
44. Parker MH, Prevelige PE (1998) Electrostatic interactions drive scaffolding/coat protein binding and procapsid maturation in bacteriophage P22. Virology 250: 337–349
45. Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448
46. 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
47. Pfister P, Wasserfallen A, Stettler R, Leisinger T (1998) Molecular analysis of *Methanobacterium* phage psiM2. Mol Microbiol 30: 233–244
48. Proux C, van Sinderen D, Suárez JE, García P, Ladero V, Fitzgerald GF, Desiere F, Brüssow H (2002) The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J Bacteriol 184: 6026–6036
49. Radlinska M, Bujnicki JM, Piekarowicz A (1999) Structural characterization of two tandemly arranged DNA methyltransferase genes from *Neisseria gonorrhoeae* MS11: N4-cytosine specific M.*Ngo*MXV and nonfunctional 5-cytosine-type M.*Ngo*Morf2P. Proteins 37: 717–728
50. Rohwer F, Edwards R (2002) The phage proteomic tree: a genome-based taxonomy for phage. J Bacteriol 184: 4529–4535
51. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor
52. Smith MCM, Burns N, Sayers JR, Sorrell JA, Casjens SR, Hendrix RW (1998) Bacteriophage collagen. Science 279: 1834
53. Suárez JE, Chater KF (1980) DNA cloning in *Streptomyces*: a bifunctional replicon comprising pBR322 inserted into a *Streptomyces* phage. Nature 286: 527–529
54. Suárez JE, Chater KF (1981) Development of a DNA cloning system in *Streptomyces* using phage vectors. Ciencia Biologica 6: 99–110
55. Walker JE, Auffret AD, Carne A, Gurnett A, Hanisch P, Hill D, Saraste M (1982) Solid-phase sequence analysis of polypeptides eluted from polyacrylamide gels. An aid to interpretation of DNA sequences exemplified by the *Escherichia coli unc* operon and bacteriophage lambda. Eur J Biochem 123: 253–260
56. Weinstock G, Riggs P, Botstein D (1980) Genetics of bacteriophage P22. III. The late operon. Virology 106: 82–91
57. Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–119
58. Zemskov EA, Kang W, Maeda S (2000) Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. J Virol 74: 6784–6789

Author's address: Juan E. Suárez, Area de Microbiología, Facultad de Medicina, Universidad de Oviedo, Julián Clavería s. n., E-33006 Oviedo, Spain; e-mail: evaristo@correo.uniovi.es