Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 348 (2005) 45–54

GENE
SECTION
FUNCTIONAL GENOMICS

www.elsevier.com/locate/gene

Sequence analysis of the *Lactobacillus plantarum* bacteriophage ΦJL-1☆

Z. Lu[a], E. Altermann[b], F. Breidt[b,c,*], P. Predki[d], H.P. Fleming[e], T.R. Klaenhammer[b]

[a]*Duke University Medical Center, Durham, NC 27710, United States*
[b]*Department of Food Science, NC State University, Raleigh, NC 27695-7624, United States*
[c]*U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS), Raleigh, NC 27695-7624, United States*
[d]*Protometrix, Inc., Branford, CT 06405, United States*
[e]*Retired from USDA-ARS and NC State University, United States*

Received 9 August 2004; received in revised form 16 November 2004; accepted 27 December 2004

Received by D.L. Court

Abstract

The complete genomic sequence of a *Lactobacillus plantarum* virulent phage ΦJL-1 was determined. The phage possesses a linear, double-stranded, DNA genome consisting of 36,677 bp with a G+C content of 39.36%. A total of 52 possible open reading frames (ORFs) were identified. According to N-terminal amino acid sequencing and bioinformatic analyses, proven or putative functions were assigned to 21 ORFs (41%), including 5 structural protein genes. The ΦJL-1 genome shows functionally related genes clustered together in a genome structure composed of modules for DNA replication, DNA packaging, head and tail morphogenesis, and lysis. This type of modular genomic organization was similar to several other phages infecting lactic acid bacteria. The structural gene maps revealed that the order of the head and tail genes is highly conserved among the genomes of several *Siphoviridae* phages, allowing the assignment of probable functions to certain uncharacterized ORFs from phage ΦJL-1 and other *Siphoviridae* phages.

Keywords: Bacteriophage; ΦJL-1; Sequence analysis; Structural proteins; *Lactobacillus plantarum*

1. Introduction

Lactic acid bacteria (LAB) are used as starter cultures in the production of various fermented foods. Bacteriophage (phage) infection of LAB has been a major problem in the dairy industry, causing slow fermentation or complete starter failure and, thus, economic losses. Due to their economical importance, dairy phages (mainly *Lactococcus lactis* or *Streptococcus thermophilus* phages isolated from dairy industry or products) became the most thoroughly studied phage group in the database (Brüssow, 2001). *Lactobacillus* phages received much less attention, probably because they are less of a practical problem in the fermentation industries (Altermann et al., 1999).

Currently, over 20 LAB phage genome sequences are available in databases. Most of these sequences are from lactococcal or streptococcal phages. A few are from *Lactobacillus* phages, including *Lactobacillus plantarum*

Abbreviations: LAB, lactic acid bacteria; MRS, DeMan Rogosa Sharpe; ORF, open reading frame; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; DNA, deoxyribonucleic acid; RBS, ribosome binding site; NTP, nucleotide triphosphate; gp, gene product; pI, isolectric point; aa, amino acid.

☆ Paper no. FSR03-07 of the Journal Series of the Department of Food Science, NC State University, Raleigh, NC 27695-7624. Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or North Carolina Agricultural Research Service, nor does it imply approval to the exclusion of other products that may be suitable.

* Corresponding author. U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS), Raleigh, NC 27695-7624, United States. Tel.: +1 919 515 2979; fax: +1 919 513 0180.

E-mail address: breidt@ncsu.edu (F. Breidt).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.12.052

phage ϕgle (Kodaira et al., 1997), *Lactobacillus gasseri* phage ϕadh (Altermann et al., 1999), *Lactobacillus delbrueckii* ssp. *lactis* phage LL-H (Mikkonen et al., 1994), and *Lactobacillus casei* phage A2 (Proux et al., 2002). These available phage sequences have provided significant information on the biology of the individual phages. Many new insights have been revealed from the detailed and comparative analysis of these sequences in areas of phage evolution, genetic diversity, horizontal/vertical gene transfer, module similarity, and lytic/lysogenic cycles (Brøndsted et al., 2001; Brüssow and Desiere, 2001; Desiere et al., 1999; Desiere et al., 2000; Lucchini et al., 1999; Mahanivong et al., 2001). Compared with other organisms, the total number of phage sequences in the database is small. More phage genome sequences from a diverse array of phages and comparative sequence analysis are needed to elaborate upon a sequence-based theory and to improve our understanding of these viruses and their interaction with their hosts.

In the United States, commercial vegetable fermentations are usually natural fermentations without addition of starter cultures (Fleming et al., 1995). With the increasing interest in reducing waste brine disposal, low-salt fermentation is currently being developed. This will require greater control of the non-lactic flora and is likely to involve the use of starter cultures. *L. plantarum* BI7 and its derivative, MU45 (deficient in malolactate fermenting ability), have been evaluated as starter cultures for controlled cucumber fermentations and as biocontrol microorganisms for minimally processed vegetable products in the USDA-ARS Food Fermentation Laboratory. Since vegetable fermentation systems are not sterile, the starter cultures may be susceptible to infection by phages naturally present in these environments. A virulent bacteriophage, ΦJL-1 (active against both *L. plantarum* BI7 and MU45), was recently isolated from a commercial cucumber fermentation. Some of its biological properties were described previously (Lu et al., 2003). The phage has an isometric head, a long non-contractile tail, and belongs to morphotype B1 within the *Siphoviridae* family. Tail fibers were not observed. Phage ΦJL-1 has a linear, double-stranded, DNA genome of 36.7 kb. SDS–PAGE revealed the presence of six structural proteins. Using *L. plantarum* MU45 as a host, the phage ΦJL-1 had an average burst size of 22 and a latent period of 35 min. Little is known about the genetic content, organization, or functions of genes in ΦJL-1. A better understanding of the genetics and biological properties of the *Lactobacillus* phage ΦJL-1 is fundamental to the understanding of phage–host interactions and possibly to the development of phage-control strategies for controlled vegetable fermentations and biocontrol systems using *L. plantarum* BI7 or MU45.

The objectives of this study were to determine and analyze the complete genome sequence of the *L. plantarum* phage ΦJL-1, to identify the structural genes (including the major head and tail protein genes), and to explore the genomic organization of the phage.

2. Materials and methods

2.1. Bacterial strain, phage, and media

L. plantarum MU45 was grown in MRS broth (Difco Laboratories, Detroit, MI) at 30 °C. Phage ΦJL-1 was propagated on *L. plantarum* MU45 in MRS medium supplemented with 10 mM $CaCl_2$ at 30 °C (Lu et al., 2003).

2.2. Purification of ΦJL-1 and isolation of phage DNA

Phage ΦJL-1 particles were concentrated from 1 L of phage lysate by PEG precipitation and then resuspended in 6 mL of 10 mM Tris–HCl buffer (pH 7.4). The phage suspension was intentionally vortexed (Daigger Vortex-Genie 2, A. Daigger and Company, Inc., Vernon Hills, IL) at the highest speed for 2 min in an attempt to generate defected phages. The mixture of intact and defected phage particles was separated and purified by CsCl density gradient centrifugation at 600,000×*g* for 6 h at 15 °C (Sorvall micro-ultra-centrifuge with rotor S100AT6, RC-M150 GX, Sorvall, Newtown, CT). Two visible bands, consisting of intact and defective phages, respectively, were collected separately, and dialyzed against 3 L of 10 mM Tris–HCl buffer (pH 7.4). Phage DNA was isolated as described by Lu et al. (2003).

2.3. Electron microscopy

CsCl-purified phage samples were negatively stained with 2% (w/v) aqueous uranyl acetate (pH 4.0) on a carbon-coated grid and examined by transmission electron microscopy (JEOL JEM-100S, Japan Electronics and Optics Laboratory, Tokyo, Japan) at an accelerating voltage of 80 kV. Electron micrographs were taken at a magnification of 50,000× and printed at 85,000× (V. Knowlton, Center for Electron Microscopy, NC State University, Raleigh, NC).

2.4. Sequence and analysis of ΦJL-1 DNA

DNA sequencing was carried out at the Department of Energy Joint Genome Institute (JGI) sequencing facility (Walnut Creek, CA) and Davis Sequencing (Davis, CA) using shotgun cloning and primer walking sequencing strategies. Sequence annotation was performed using the Global Annotation of Multiplexed On-site Blasted DNA Sequences software package (Altermann and Klaenhammer, 2003). Briefly, five ORFs were manually determined based on several criteria (see Results and discussion). Based on these ORFs, a training model was built using build-icm provided in the glimmer package to identify the remaining

ORFs in the ΦJL-1 genome. The identified ORFs were analyzed with BlastP using the non-redundant database from NCBI (May 2004). These results were used to establish the automated computer annotation and to generate the flatfile databases. Sequence alignments were performed using Clone Manager 6, Plasmid Map Enhancer v. 3 (Scientific Educational Software, Durham, NC), ClustalX (Thompson et al., 1997), and ClustalW (Thompson et al., 1994). The Molecular BioComputing Suite (Muller et al., 2001) was used to calculate the molecular mass and isoelectric point of predicted proteins of ΦJL-1 and other LAB phages currently available in databases. Transmembrane domains were predicted by the TMHMM program (http://www.cbs.dtu.dk/services/TMHMM-2.0; Sonnhammer et al., 1998). The complete genomic sequence of the *L. plantarum* phage ΦJL-1 was deposited in the GenBank database under the accession number AY236756.

2.5. Restriction enzyme analyses

The phage DNA was digested with restriction endonucleases (*Pst*I, *Nhe*I, and *Nru*I) according to the supplier's recommendations (Promega, Madison, WI). The DNA fragments were separated by agarose (1%) gel electrophoresis and visualized under UV-light (300 nm).

2.6. Analysis of structural proteins

SDS–PAGE was performed as described by Lu et al. (2003). CsCl-purified phages in SDS–PAGE sample buffer were heated in a boiling water bath for 10 min and then applied to a NuPAGE precast gradient minigel (4–12% Bis–Tris, Invitrogen Corporation, Carlsbad, CA). After electrophoresis, the SDS–PAGE separated proteins were transferred by electroblotting onto a polyvinylidene difluoride (PVDF) membrane (0.2 μm pore size) in a transfer buffer containing 12 mM Tris base, 96 mM glycine, and 10% methanol (v/v) according to the protocol provided by Invitrogen. After transfer, the PVDF membrane was stained with 1% Amido Black. Visible protein bands were excised, and their N-terminal amino acid (10 or 11 residuals) sequences were determined by Andrew Brauer (ProSeq, Inc., Boxford, MA), using an Applied Biosystems 494 Protein Sequencer. Based on the N-terminal amino acid sequence of each protein, the corresponding open reading frame was identified in the ΦJL-1 genome. A broad range protein marker (Mark 12, Invitrogen) and pre-stained, multicolor, molecular mass markers (Invitrogen) were used to estimate the molecular weights of the phage structural proteins.

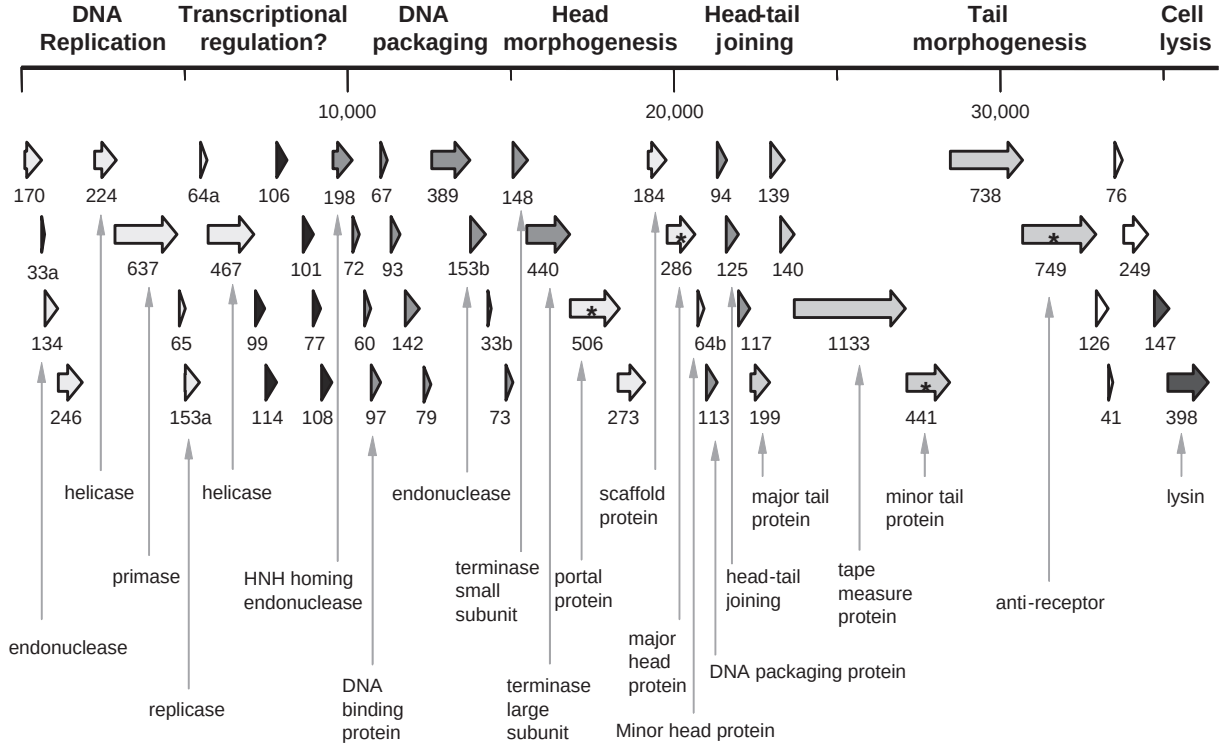


Fig. 1. Modular genomic organization of the phage ΦJL-1 genome. The double-stranded DNA is shown as a thick black line. The forward arrows represent ORFs predicted from the genomic sequence. The direction of the arrows corresponds to the direction of transcription. The numbers below the arrows represent the ORF designation and size (in aa). Proven or putative functions of individual ORFs (see Table 1) are indicated by gray upward arrows. The experimentally determined structural genes are indicated with asterisks inside arrows. The proposed functional modules are indicated at the top of the figure. ORFs belonging to the same module are shown in the same pattern of arrows. The regulation module is based on an extrapolation from other *Siphoviridae* phages. Thus, ORFs represented by dark arrows may or may not belong to the regulation module. ORFs represented by open (no fill) arrows have not been assigned to any functional module.

3. Results and discussion

3.1. Complete nucleotide sequence and genomic organization of ΦJL-1

The complete nucleotide sequence of ΦJL-1 was determined by combining shotgun cloning with a primer walking sequencing strategy. Phage ΦJL-1 has a linear, double-stranded, DNA genome consisting of 36,677 bp with a G+C content of 39.36%, which is lower than that (43.1%) of *L. plantarum* phage ϕgle (Kodaira et al., 1997), but slightly higher than that (35.3%) of *L. gasseri* phage ϕadh (Altermann et al., 1999). The G+C content of the phage host (*L. plantarum* MU45) was not determined. Bioinformatic analysis of the ΦJL-1 genome revealed 52 possible ORFs based on several criteria: (i) the ORF begins with either an ATG, GTG, or TTG, and ends with either TGA, TAG, or TAA; (ii) the ORF contains at least 30 codons; (iii) the codon usage of the ΦJL-1 OFRs was determined by the training model (data not shown); and (iv) with a few exceptions, the ORF is preceded by an identifiable ribosomal binding site (RBS). The ORFs were named according to the number of amino acids (aa) in the deduced proteins. All ORFs were oriented in the same direction (Fig. 1). Of the 52 ORFs, 6 are initiated with the start codon TTG, 5 with GTG, and 41 with ATG. A potential RBS, complementary with the 3′ ends of 16S rRNAs of various bacteria, can be identified upstream of most of the 52 ORFs. In most of the sites listed in Table 1, the core consensus sequence (AGGAGG) of RBS from the *L. delbruckii* ssp. *lactis* phage LL-H (Mikkonen et al., 1994) is highly conserved. An in silico restriction site analysis of the nucleotide sequence agreed well with the experimentally determined restriction pattern (data not shown).

The deduced amino acid sequences of all the ORFs were compared with the non-redundant database provided by NCBI using gapped BlastP, and subsequently analyzed using hmmer2.2g and the Pfam databases for global and local alignment models. A total of 29 (56%) ORFs showed homologies with previously characterized genes in databases. In most cases, homologies were found to phages infecting gram-positive bacteria, primarily LAB (Table 1).

Table 1
Selected ORFs and genetic features of the *Lactobacillus plantarum* phage MJL-1[a]

				Predicted product				
			Putative RBS[b] and start codon				Database search results[c]	
ORF[h]	Start	End[d]	3′-AUCUUUCCUCCACUAGGUC…[e]	Size [aa]	Mass[f] [kDa]	pI[g]	Predicted function	Organism matched
134	722	1123	ccAGAAgGGAaGcGtaaataatg	134	14.9	9.5	endonuclease (ORF48)	*Lactobacillus casei* phage A2
224	2242	2913	TtGAAAGGtGaTGtTtttaaatg	224	24.5	5.8	Helicase (NTP-binding)	*L. plantarum* phage ϕgle
637	2858	4768	TAaAAAGcAatcaATtaAaaatg	637	72.2	5.5	DNA primase /helicase	*S. pyogenes* phage 315.6
153a	5014	5472	ActcAAGGAGGaaATtaAaaatg	153	17.3	4.8	Replicase	Beet yellows virus
467	5728	7128	ttattgtagggAGAAAtagtatg	467	53.1	8.6	Helicase	*L. casei* phage A2
198	9550	10143	ttAaattAGGAGGaatcgtAatg	198	22.6	9.2	HNH homing endonuclease	*Lactococcus lactis* phage bIL170
97	10716	11006	TttatAGGAGaaaATaaAaaatg	97	11.1	4.8	DNA binding	*S. thermophilus* phage Sfi18
153b	13757	14215	gggAattAGGAGtgGcgaCtatg	153	17.6	9.5	Endonuclease	*L. casei* phage A2
148	15075	15518	tggtAGGAGGTGtataaGccttg	148	17.0	8.1	Terminase, small subunit	*Bacillus subtilis* PBSX prophage
440	15484	16803	tgatagtcAAGtAGtcGtgaatg	440	51.0	8.4	Terminase, large subunit	*S. pyogenes* M1GAS
506	16817	18334	aggActAtAGGAGGcttagCatg	506	57.7	4.6	Minor head protein Portal protein	*S. thermophilus* phage Sfi11
184	19210	19761	gtcggGAtAGGAGGAttaCCatg	184	20.1	4.7	Scaffold protein	*S. thermophilus* phage Sfi11
286	19784	20641	AAaAAAcGAGGTttaaaAttatg	286	30.4	4.6	Major head protein	*Lc. lactis* phage ul36
64b	20722	20913	gcgAtactcGTaATattaccgtg	64	6.0	3.6	Minor head protein	*Yersinia pestis* CO92 phage
113	20985	21323	tacgaAAAGGAaGTGATtaaatg	113	12.7	4.6	DNA packaging	*Lc. lactis* phage TP901-1
125	21600	21974	GAAAGtgacggtgttaatctgtg	125	14.1	9.1	Head to tail joining	*Lc. lactis* phage TP901-1
199	22340	22936	caAtttAAGGAGGatAaaacatg	199	21.6	4.2	Major tail protein	*B. subtilis* phage SPP1
1133	23689	27087	atcAcGGAGGTGAataatatatg	1133	112.1	9.6	Tape measure protein	*Lc. lactis* phage TP901-1
441	27125	28447	gaattAAAGcctGccAgtgtatg	441	49.7	5.4	Minor tail protein	*L. casei* phage A2
749	30671	32917	aTAtAAAGGtGGTaATgtAGatg	749	82.4	4.8	Minor tail protein Host specificity	*Lc. lactis* phage bIL170
398	35167	36360	accagcAAcGGAGGaatagtatg	398	43.6	9.7	Lysin	*O. oenos* phage 10MC

[a] See the text for details.
[b] The sequence shown includes the immediate upstream 20 nucleotides of the putative start codon. The nucleotide that is complementary to the one found at the 3′ end of the 16S rRNA is shown in uppercase letters.
[c] Database searches based on homologies of deduced amino acid sequences were performed with gapped BlastP algorithm.
[d] The end position does not include the stop codon.
[e] The nucleotides complementary to the 3′ end of the 16S rRNA of *L. delbruckii* (3′-AUCUUUCCUCCACUAGGUC…; Mikkonen et al., 1994).
[f] Molecular weights were calculated with Molecular Biocomputing Suite (Muller et al., 2001).
[g] Isoelectric points were calculated with Molecular Biocomputing Suite.
[h] ORFs were designated according to the number of amino acids of the corresponding coded proteins.

Eight ORFs displayed homologies to unknown functions in the databases. Only those ORFs for which a putative function could be attributed are discussed below.

3.2. Genes involved in DNA replication

The predicted protein product of ORF134 shows similarity to the putative endodeoxyribonuclease from *L. casei* phage A2 (Table 1; Fig. 1). Downstream of ORF134 is ORF224 whose derived protein product is homologous to the DNA helicase or NTP-binding protein from *L. plantarum* phage ϕgle (Kodaira et al., 1997), suggesting that the putative gene products (gp) might be involved in DNA synthesis. The derived protein products of ORF637 and ORF153a exhibit sequence similarities (identified by PSI-Blast searching) to a putative DNA primase from *Streptococcus pyogenes* phage 315.6 and to a replicase from beet yellows virus, respectively, suggesting that these two proteins may be required for DNA replication. The predicted gene product from ORF467 exhibits high similarity with the putative helicase from *L. casei* phage A2 (Proux et al., 2002), suggesting that the protein is probably involved in DNA synthesis.

3.3. Genes involved in phage DNA packaging

ORF440 resembles the gene encoding the putative large terminase subunit found in the *S. pyogenes* prophage M1 GAS (Table 1; Ferretti et al., 2001) and in *Lactobacillus johnsonii* prophage Lj771 (Desiere et al., 2000). The product derived from ORF148 exhibited homology to the small terminase subunit from *Bacillus subtilis* prophage (Krogh et al., 1996). These results suggested that the two proteins (gpORF440 and gpORF148) are likely to be involved in phage DNA packaging. In tailed phages, the small terminase subunit is responsible for specific DNA binding, and the large terminase subunit mediates the cleavage of phage DNA into genome units and prohead binding. Generally, the DNA-interaction sites (*pac* or *cos*) of the terminases are located within or close to the structural genes (Black, 1989). This is also true in ΦJL-1 (Fig. 1). After restriction of ΦJL-1 DNA with different restriction enzymes, heating the restricted DNA to 65 °C prior to electrophoresis did not alter the banding pattern (Fig. 2), suggesting that ΦJL-1 DNA has no *cos* site. Therefore, ΦJL-1 is likely to utilize the *pac* mechanism of DNA packaging.

The protein specified by ORF97 exhibits high similarity with the putative DNA binding protein from *S. thermophilus* phage Sfi18 (Table 1). The position of ORF97 is near terminase genes (Fig. 1). These features suggested that gpORF97 may also be involved in DNA packaging. The gene product of ORF198 shares homology (33% overall identity) with the HNH homing endonuclease of *L. lactis* phage bIL170 (Crutz-Le Coq et al., 2002). HNH homing endonucleases confer mobility to their own genes or to host intervening sequences, either an intron or intein, by

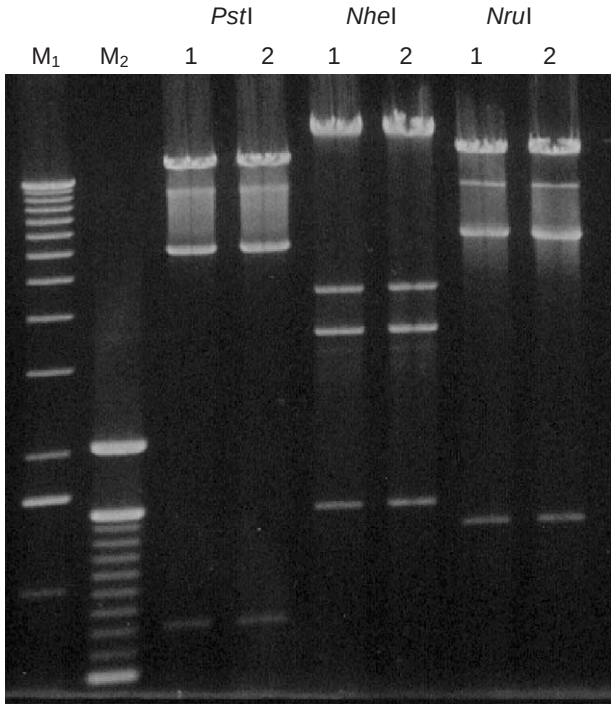


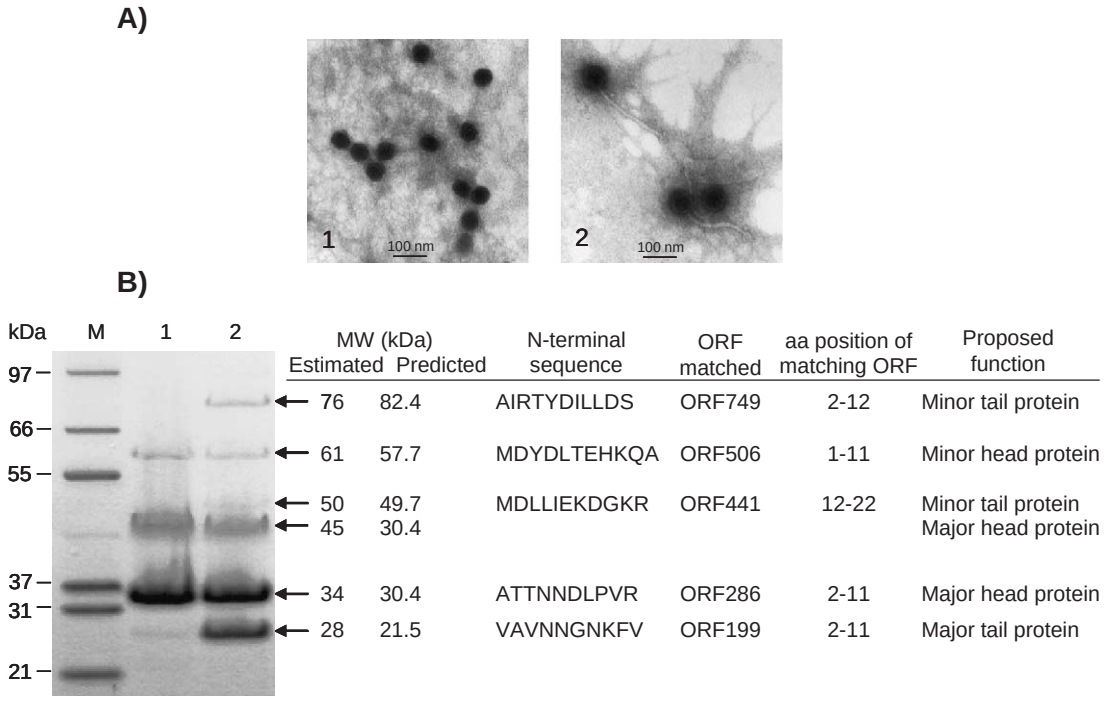
Fig. 2. Restriction analysis of ΦJL-1 DNA. The phage DNA was digested with *Pst*I, *Nhe*I, or *Nru*I. Lanes M1 and M2: 1 kb and 100 bp DNA ladders; 1 and 2 indicated that the digests were unheated and heated prior to electrophoresis, respectively.

catalyzing a highly specific, double-strand break in a cognate allele lacking the intervening sequence (Chevalier and Stoddard, 2001). These endonucleases can be found as free-standing ORFs between genes or encoded within introns or inteins. The function of HNH homing endonucleases in the phage cycle and/or the reason for their maintenance in such compact phage genomes is intriguing (Crutz-Le Coq et al., 2002). However, based on the position of ORF198 in the genome (close to terminase genes), it is likely that the HNH homing endonuclease is involved in DNA packaging. Similarly, the protein specified by ORF153b, showing high similarity with the putative endodeoxyribonuclease from *L. casei* phage A2, is likely to be involved in DNA packaging.

3.4. Experimentally determined structural proteins of ΦJL-1

During purification of phage ΦJL-1, several bands appeared in the CsCl density gradients. Samples from two phage-containing bands were individually analyzed by electron microscopy and SDS–PAGE. The electron micrograph (Fig. 3A) shows that the main band contained intact phage particles (Fig. 3A2) and the lower band contained only phage heads (tail-less phage particles, Fig. 3A1). The morphology (with an average length of about 182 nm) of the intact phage particles was consistent with the initial description of ΦJL-1 in our previous study (Lu et al., 2003).

SDS–PAGE of the phage head sample revealed the presence of three head proteins, whereas SDS–PAGE of the



MW (kDa) Estimated	MW (kDa) Predicted	N-terminal sequence	ORF matched	aa position of matching ORF	Proposed function
76	82.4	AIRTYDILLDS	ORF749	2-12	Minor tail protein
61	57.7	MDYDLTEHKQA	ORF506	1-11	Minor head protein
50	49.7	MDLLIEKDGKR	ORF441	12-22	Minor tail protein
45	30.4				Major head protein
34	30.4	ATTNNDLPVR	ORF286	2-11	Major head protein
28	21.5	VAVNNGNKFV	ORF199	2-11	Major tail protein

Fig. 3. Analysis of the phage ΦJL-1 structural proteins. After purification of ΦJL-1 by CsCl density gradient centrifugation, intact phage particles from the main band (2) and phage heads from the lower band (1) were analyzed by electron microscopy (A) and 4–12% SDS–PAGE (B). Lane M, molecular mass marker; lane 1, proteins from ΦJL-1 heads; lane 2, proteins from ΦJL-1 particles. The N-terminal amino acid (aa) sequences of the five structural proteins were determined. The position of the N-terminal amino acid sequence in the corresponding ΦJL-1 ORF and its proposed function are indicated.

intact phage sample showed six structural proteins, including the three head proteins (Fig. 3B), suggesting that the three proteins absent in the phage head sample were tail proteins. In order to identify the corresponding ORFs in ΦJL-1 DNA, the six protein bands from the intact phage sample were transferred to a PVDF membrane, and the N-terminal sequences were determined (Fig. 3B).

The three head proteins present in both intact and defective phage samples have molecular weights (MW) of 34 kDa, 45 kDa, and 61 kDa, respectively, as estimated by SDS–PAGE (Fig. 3B). The 61-kDa protein appeared to be a minor head protein as it was much less abundant than the other two head proteins (Fig. 3B). The first 11 amino acids (MDYDLTEHKQA) of this protein (61 kDa) exactly matched amino acids 1 to 11 of ΦJL-1 gpORF506 with predicted molecular mass of 57.7 kDa. This minor head protein showed very strong similarity to the putative portal protein of *S. thermophilus* phage Sfi11 (Table 1). The 34-kDa and 45-kDa proteins appear to be major head proteins according to their abundance shown in Fig. 3B. They shared identical N-terminal sequence (ATTNNDLPVR), which perfectly matched the residues 2 to 11 of the ΦJL-1 gpORF286. The 34-kDa protein observed on the SDS–PAGE corresponded to the predicted molecular mass of gpORF35 (30.4 kDa). However, the observed 45-kDa protein had a much higher MW than the predicted value (30.4 kDa) of gpORF286. A 2D-gel electrophoresis prior to N-terminal sequencing may be needed to determine the attribution of this protein. The first methionine residue was absent in the two major head proteins, which was in accordance with the rule that the N-terminal methionine is generally processed when the second amino acid residue is alanine (Ben-Bassat et al., 1987). Processing of the initiation methionine during protein maturation has been observed in many phages and occurs via the host methionine aminopeptidase activity (Lowther and Mathews, 2000; Mahanivong et al., 2001). The product deduced from ORF286 exhibits a noticeable sequence similarity (37% or 43% overall identity) with the experimentally determined major head protein of *L. lactis* phage ul36 (Table 1) or with the hypothetical protein (gpORF36) of *Streptococcus pneumoniae* phage MM1 (Fig. 4). This bioinformatic link suggested that gpORF36 of *S. pneumoniae* phage MM1 may also be a major head protein.

The three tail proteins which were only present in the intact phage sample had observed MW of 28 kDa, 50 kDa, or 76 kDa, respectively (Fig. 3). The observed 28-kDa protein appears to be a major tail protein according to its abundance shown in SDS–PAGE (Fig. 3). The first 10 amino acids of the protein were VAVNNGNKFV. This sequence was identical to residues 2 to 11 (except residue 8) of the ΦJL-1 ORF199. Residue 8 of ORF199 was V instead of N. The discrepancy may reflect an error in the N-terminal sequencing because re-sequencing ORF199 gave the same nucleotide sequence. The N-terminal methionine is not present in the mature 28-kDa protein. The predicted MW (21.5 kDa) of gpORF199 is lower than the observed (28 kDa). The major tail protein displayed a strong homology to gpORF21 of *B. subtilis* phage SPP1 (Table 1).

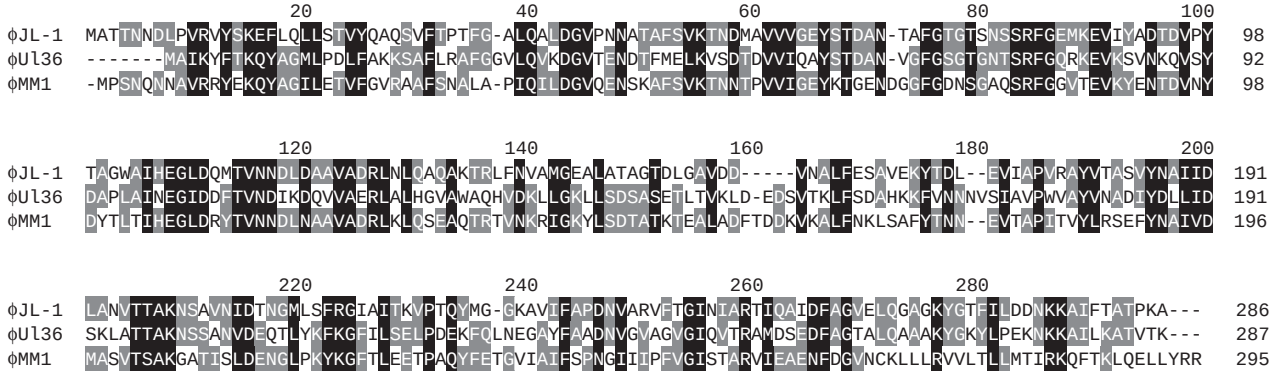


Fig. 4. Multiple sequence alignment of the major head proteins from *Lactobacillus plantarum* phage ΦJL-1 and *Lactococcus lactis* phage ul36 with ORF36 of *Streptococcus pneumoniae* phage MM1. Residue numbers of the proteins are given on the right. Perfectly conserved residues are highlighted in black boxes. Residues that are conserved in 2 of the aligned sequences are shaded in gray. Numbers refer to the amino acid position.

The N-terminal peptide sequence (MDLLIEKDGKR) of the 50-kDa protein revealed on the SDS–PAGE matched amino acid positions position 12 to 22 of the predicted gpORF441, suggesting that the 11 amino acids may actually not belong to this protein, or the protein was proteolytically processed during maturation. Processing of the gpORF441 (49.7 kDa) at position 12 predicts a protein with MW of 48.5 kDa, slightly lower than the observed (50 kDa). Similar proteolytic cleavage of N-terminal amino acids during phage morphogenesis has also been observed in other LAB phages such as *L. lactis* phage BK5-T (Mahanivong et al., 2001), *S. thermophilus* phages (Desiere et al., 1998), and *L. gasseri* ϕadh (Altermann et al., 1999). The 50-kDa protein is probably a minor tail protein because it was much less abundant than the other two tail proteins (Fig. 3B). The result from the database search showed that this minor tail protein exhibits a weak sequence similarity with a tail component protein from *L. casei* phage A2 (Table 1).

N-terminal sequence analysis of another tail protein (76 kDa, observed from SDS–PAGE) revealed the sequence AIRTYDILLDS, which is identical to amino acids 2–12 of the protein (82.4 kDa) derived from ORF749. Again, the N-terminal methionine was absent in the mature protein. This protein, apart from its function as a tail protein, may be also responsible for host specificity because it shows homology with putative anti-receptors from several dairy phages (Brøndsted et al., 2001; Desiere et al., 1999; Lucchini et al., 1999), as well as with ORF112 of *L. lactis* phage bIL170 (Crutz-Le Coq et al., 2002), which was possibly involved in host range determination.

3.5. Genes involved in host lysis

The predicted protein from ORF398 shares a strong sequence similarity (57% overall identity) with the lysin of *Oenococcus* (previously *Leuconostoc*) *oenos* phage 10MC (Table 1) and lysin from several other phages (data not shown). In addition, ORF147, which is located immediately upstream of the lysin gene, exhibits a strong similarity with the gene (also immediately upstream of lysin gene) from *Oenococcus oenos* phage 10MC. The product of ORF147 is predicted to contain a transmembrane domain (Fig. 5). These features are characteristics of the λ S holin (Gründling et al., 2000a,b) and other phage holins (Wang et al.,

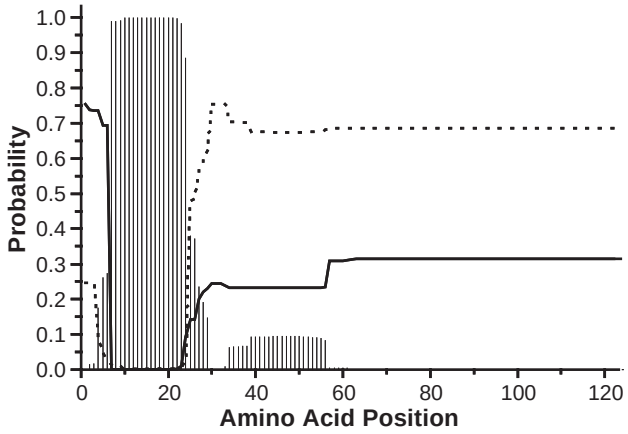


Fig. 5. Prediction of the transmembrane domain and posterior probabilities in the putative holin of ΦJL-1 using the TMHMM program. Vertical bars represent the transmembrane domain; solid and dotted lines indicate protein regions located in the cytoplasma or the periplasma, respectively. Overall architecture of the putative holin protein is represented by the two horizontal lines, featuring the same pattern/location scheme.

2000), implying that ORF147 is very likely to be the holin gene of ΦJL-1.

3.6. Other ΦJL-1 genes with putative functions

Generally, the genes located between the major head and major tail genes are involved in formation and connection of the head and tail structures and in DNA packaging (Brøndsted et al., 2001). In this region of the ΦJL-1 genome (Fig. 1), ORF125 and ORF113, respectively, showed homology to head–tail joining and DNA packaging proteins from *L. lactis* phage TP901-1 (Table 1), suggesting that ORF125 and ORF113 products may be involved in phage assembly.

ORF1133 is the longest ORF in the ΦJL-1 genome. The predicted product of this ORF showed strong sequence similarity to the tail tape measure protein of *L. lactis* phage TP901-1, suggesting that the protein may be responsible for determining phage tail length. In phage lambda, the tail tape measure protein is used as a template for tail polymerization and remains inside the tail tube (Katsura and Hendrix, 1984).

3.7. DNA packaging and structural gene map of ΦJL-1

A gene map of ΦJL-1, displaying the predicted structural module, is aligned with corresponding genome sections of five other *Siphoviridae* LAB phages (Fig. 6), including two *Lactobacillus* phages (φgle, and φadh), two lactococcal phages (TP901-1, ul36), and one streptococcal phage (Sfi21). The alignment demonstrated that these phages share a highly conserved structural gene order, supporting the notion that the structural gene order is highly conserved among *Siphoviridae* phages (Lucchini et al., 1998, 1999). In addition, the size (in aa or kDa) and isoelectric point (pI) value of structural proteins with the same function appear to be relatively conserved. These physical properties can be useful for extrapolating and predicting gene functions within the structural module for

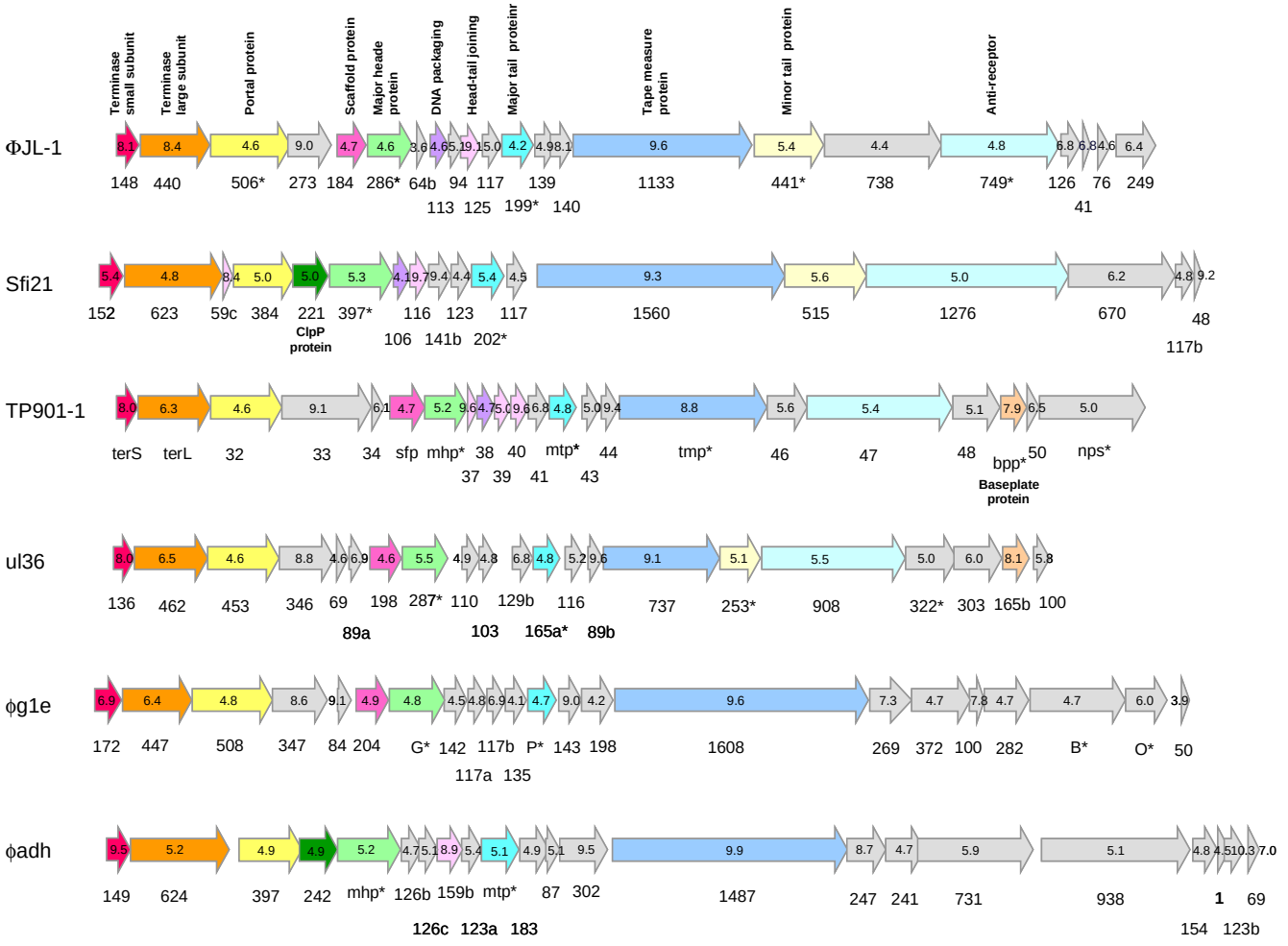


Fig. 6. Comparison of the partial genome of phage ΦJL-1 with other *Siphovaridae* LAB phages. The numbers below the maps refer to ORFs of ΦJL-1 (Table 1), Sfi21 (Desiere et al., 1999), TP901-1 (Brøndsted et al., 2001), ul36 (Labrie and Moineau, 2002), φgle (Kodaira et al., 1997), and φadh (Altermann et al., 1999). Sizes of individual ORFs are reflected by the lengths of the arrows. The pI values predicted for individual proteins are indicated inside the arrows representing the respective ORFs. Corresponding genes are indicated with the same color code. ORFs whose structural function has been experimentally verified are indicated by asterisks.

closely or distantly related phages even showing little or no regions of sequence homology. The alignment of the ΦJL-1 structural gene map with other LAB phages (Fig. 6) predicts that the experimentally determined minor head protein (gpORF506) of ΦJL-1 may be also a portal protein because the gene location (immediately downstream of large subunit terminase). In addition, the size (506 aa) and pI value (4.6) of the gene product were very similar to a portal protein from Sfi21, TP901-1, ul36, and ϕgle. This prediction was supported by the database search result (Table 1) as discussed above. Portal proteins are generally responsible for forming the entrance to the head during DNA packaging and determining the amount of DNA to be packaged (Dube et al., 1993). Similarly, we predict that uncharacterized ORF397 of ϕadh is probably a portal protein as well. The derived protein from ORF184 shows limited sequence similarity with the putative protein from *S. thermophilus* phage Sfi11 (Table 1). However, the size (184 aa) of the protein, its pI value (4.7) and its relative position in the gene map were very similar to putative scaffold proteins of several LAB phages, including ul36, TP901-1, ϕgle (Fig. 6), and Sfi11 (Table 1). These features supported the prediction that gpORF184 might have a scaffolding function. In the ϕadh genome, the function of ORF159b is unknown, but its physical properties (the location in the gene map, and the size and pI value of the gene product) are similar to putative head–tail joining protein gene from phages ΦJL-1, Sfi21, and TP901-1 (Fig. 6). Thus, ORF159b of ϕadh may encode a head–tail joining protein. As mentioned earlier, the derived product from ORF36 of S. *pneumoniae* phage MM1 (NCBI accession no. NC003050) shared a striking sequence similarity (Fig. 4) with the major head protein from ΦJL-1 and ul36. The structural gene map of phage MM1 (data not shown) reveals that immediate upstream of ORF36 is a gene encoding a putative scaffold protein. Downstream of ORF36 are several small (77–130 aa) protein genes, including two small putative minor capsid protein genes (114–123aa). Furthermore, the deduced protein of ORF36 consists of 295 aa and has pI value of 5.7. These features are very similar to those of the major head protein from ΦJL-1, TP901-1, ul36, and ϕgle (Fig. 6), strongly supporting our prediction that ORF36 of phage MM1 may encode a major head protein.

3.8. Functional modules and genomic organization in ΦJL-1

In many phages the genes encoding related biological functions are clustered. Analysis of the proven and putative gene functions and the locations of individual ORFs from phage ΦJL-1 reveal that the phage genome is highly modular, with functionally related genes clustered together. Thus, the following functional modules are proposed and indicated in Fig. 1: DNA replication, transcription regulation, DNA packaging, head morphogenesis, head–tail joining, tail morphogenesis, and cell lysis.

The DNA replication module (Fig. 1) consists of genes encoding a putative endodeoxyribonuclease, two putative helicases, a putative primase, a putative replicase, and a few other ORFs between or nearby these genes. The packaging module contains ORFs encoding putative HNH homing endonuclease, DNA binding protein, endonuclease, terminase subunits, and several other proteins in this region (Fig. 1). The head morphogenesis module includes ORFs encoding the experimentally determined minor head protein (also a putative portal protein), a putative scaffold protein, the experimentally identified major head protein, and a putative minor head protein, and gpORF273 (Fig. 1). The head–tail joining module starts with ORF113 encoding a DNA packaging protein, followed by ORF94 and the gene encoding the head–tail joining protein, and ends with ORF117. The tail morphogenesis module includes genes encoding three experimentally identified tail proteins (one major and two minor tail proteins) including the putative tape measure protein and anti-receptor (Fig. 1). The cell lysis module consists of a putative lysin gene and ORF147 which is suspected to be a holin gene. Besides these modules, a regulation module was also assigned in Fig. 1, based upon extrapolations from other *Siphoviridae* phages (Brøndsted et al., 2001; Brüssow, 2001; Brüssow and Desiere, 2001; Stanley et al., 1997). Further analysis and experimental evidence are needed to confirm these modules.

Notably, genes involved in packaging of the genome into the phage head are immediately followed by structural modules. Head genes are clustered together and precede the tail genes, which are also clustered together. These genes are followed by a gene cluster required for lysis of the host. Although a few ORFs have not been assigned to any functional module due to lack of information regarding the biological functions of the encoded genes, the overall organization of functional modules within ΦJL-1 revealed a striking correlation with those observed in many other *Siphoviridae* LAB phages, such as the virulent phage ul36 (Labrie and Moineau, 2002), and temperate phages TP901-1 (Brøndsted et al., 2001), ϕgle (Kodaira et al., 1997), ϕadh (Altermann et al., 1999), and O1205 (Brüssow and Desiere, 2001). No remnants of a lysogeny module were found in the ΦJL-1 genome.

Further studies on gene structure, transcription, and functions in phage ΦJL-1 are needed for better understanding the biology of the phage and potentially assist the development of phage-control strategies in vegetable fermentations relying on *L. plantarum* starter cultures.

Acknowledgements

This work was supported in part by a research grant from Pickle Packers International, Inc., St. Charles, IL, and by the U.S. Department of Energy, Office of Biological and Environmental Research Joint Genome Institute (JGI), together with the University of California, Lawrence Liver-

more National Laboratory (under contract no. W-7405-ENG-48), Lawrence Berkeley National Laboratory (under contract no. DE-AC03 76SF00098), and Los Alamos National Laboratory (under contract no. W-7405-ENG-36). We thank Dr. Stephanie Stilwagen of JGI, Mr. Jaime Lizarraga of Davis Sequencing, Davis, CA, for assistance in phage sequencing, and Ms. Dora Toler, for excellent secretarial assistance.

References

Altermann, E., Klaenhammer, T.R., 2003. GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. OMICS 7, 161–169.

Altermann, E., Klein, J.R., Henrich, B., 1999. Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage ϕadh. Gene 236, 333–346.

Ben-Bassat, A., Bauer, K., Chang, S.Y., Myambo, K., Boosman, A., Chang, S., 1987. Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. J. Bacteriol. 169, 751–757.

Black, L.W., 1989. DNA packaging in dsDNA bacteriophages. Annu. Rev. Microbiol. 43, 267–292.

Brøndsted, L., Ostergaard, S., Pedersen, M., Hammer, K., Vogensen, F.K., 2001. Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. Virology 283, 93–109.

Brüssow, H., 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55, 283–303.

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

Chevalier, B.S., Stoddard, B.L., 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. 29, 3757–3774.

Crutz-Le Coq, A.M., Cesselin, B., Commissaire, J., Anba, J., 2002. Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. Microbiology 148, 985–1001.

Desiere, F., Lucchini, S., Brüssow, H., 1998. Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. Virology 241, 345–356.

Desiere, F., Lucchini, S., Brüssow, H., 1999. Comparative sequence analysis of the DNA packaging, head and tail morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage Sfi21. Virology 260, 244–253.

Desiere, F., Pridmore, R.D., Brüssow, H., 2000. Comparative genomics of the late gene cluster from *Lactobacillus* phages. Virology 275, 294–305.

Dube, P., Tavares, P., Lurz, R., van Heel, M., 1993. The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. EMBO J. 12, 1303–1309.

Ferretti, J.J., et al., 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc. Natl. Acad. Sci. U. S. A. 98, 4658–4663.

Fleming, H.P., Kyung, K.H., Breidt Jr., F., 1995. Vegetable fermentations. In: Rehm, H.J., Reed, G. (Eds.), Biotechnology, 2nd ed. Enzymes, Biomass, Food and Feed, vol. 9. VCH Publishers, Inc., New York, pp. 629–661.

Gründling, A., Bläsi, U., Young, R., 2000a. Biochemical and genetic evidence for three transmembrane domains in the class I holin, λ S. J. Biol. Chem. 275, 769–777.

Gründling, A., Bläsi, U., Young, R., 2000b. Genetic and biochemical analysis of dimmer and oligomer interactions of the λ S holin. J. Bacteriol. 182, 6082–6090.

Katsura, I., Hendrix, R.W., 1984. Length determination in bacteriophage lambda tails. Cell 39, 691–698.

Kodaira, K.I., et al., 1997. Genome structure of the *Lactobacillus* temperate phage ϕgle: the whole genome sequence and the putative promoter/repressor system. Gene 187, 45–53.

Krogh, S., O'Reilly, M., Nolan, N., Devine, K.M., 1996. The phage-like element PBSX and part of the skin element, which are resident at different positions on the *B. subtilis* chromosome, are highly homologous. Microbiology 142, 2031–2040.

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.

Lowther, W.T., Mathews, B.W., 2000. Structure and function of the methionine aminopeptidases. Biochim. Biophys. Acta 1477, 157–167.

Lu, Z., Breidt Jr., F., Fleming, H.P., Altermann, E., Klaenhammer, T.R., 2003. Isolation and characterization of a *Lactobacillus plantarum* bacteriophage, ΦJL-1, from a cucumber fermentation. Int. J. Food Microbiol. 84, 225–235.

Lucchini, S., Desiere, F., Brüssow, H., 1998. The structural gene module in *Streptococcus thermophilus* bacteriophage ϕSfi1 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. Virology 246, 63–73.

Lucchini, S., Desiere, F., Brüssow, H., 1999. Comparative genomics of *Steptococcus thermophilus* phage species supports a modular evolution theory. J. Virol. 73, 8647–8656.

Mahanivong, C., Boyce, J.D., Davidson, B.E., Hillier, A.J., 2001. Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. Appl. Environ. Microbiol. 67, 3564–3576.

Mikkonen, M., Vuoristo, J., Alatossava, T., 1994. Ribosome binding site consensus sequence of *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H. FEMS Microbiol. Lett. 116, 315–320.

Muller, P.Y., Studer, E., Miserez, A.R., 2001. Molecular biocomputing suite: a word processor add-in for the analysis and manipulation of nucleic acid and protein sequence data. BioTechniques 31, 1306–1313.

Proux, C., Sinderen, D., van Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., 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.

Sonnhammer, E.L.L., von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow, J., Littlejohn, T., Major, F., Lathrop, R., Sankoff, D., Sensen, C. (Eds.), Proc. Sixth Intern. Conf. on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA, pp. 175–182.

Stanley, E., Fitzgerald, G.F., Le Marrec, C., Fayard, B., van Sinderen, D., 1997. Sequence analysis and characterization of ϕO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. Microbiology 143, 3417–3429.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Wang, I.-N., Smith, D.L., Young, R., 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54, 799–825.