Genomic organization of lactic acid bacteria phages

C Schouler

Laboratoire de génétique microbienne, Inra, Domaine de Vilvert, 78352 Jouy-en-Josas, France

Summary — The infection of lactic acid bacteria starter cultures by bacteriophages is one of the most common causes of slow or incomplete fermentation in the dairy industry. Given the magnitude of the problem and its economic impact, considerable research has been undertaken on lactic acid bacteria phages. In recent years, numerous projects for the determination of the complete nucleotide sequence of lactobacilli and lactococci phage genomes were carried out. Functions have been assigned to the products of several open reading frames. The genome of the phages has a modular organization. It is divided in several regions, each involved in one step of the phage life cycle: integration/excision, replication, production of structural proteins, assembly, packaging of the DNA and release of the phage outside the cell after the host lysis.

lactic acid bacteria / bacteriophage / genome organization

INTRODUCTION

The infection of lactic acid bacteria by bacteriophages is of considerable economic importance in the dairy industry. It results in decreased acid production and impaired product quality (Lawrence, 1978). Given the importance of the problem, considerable research has been conducted on lactic acid bacteria bacteriophages, especially on phages infecting lactococci and, in some degree, on those attacking lactobacilli. Initial studies most commonly provided data on
morphological, serological and physiological characteristics. Combined with the comparative DNA studies, these results led to a consistent classification of phages infecting lactococci in 12 different groups (Jarvis et al, 1991). DNA homology analyses have also been used to determine the relationships between Lactobacillus phages. Those of Lactobacillus delbrueckii subsp bulgaricus and lactis were classified in four groups (Sechaud et al, 1988), while all Lactobacillus helveticus phages examined so far were assembled in a single homology group (Lahbib-Mansais et al, 1988).

In recent years, increasing research efforts have been devoted to the molecular biology of lactic phages. The first requirement for understanding phage biology is to determine the organization of the genome. Studies on gene function and gene expression would subsequently provide knowledge on phage development. This will lead, in particular, to a better understanding of the regulation of the lysogeny and of the synthesis, assembly and release of virion particles. The phages so far examined possess double stranded linear DNA and have either 3' staggered cohesive ends or circularly permuted ends with terminal redundancy. The G+C content of their genome reflects that of the host. Thus, lactococcal phages have a G+C value ranging around 36% (Schouler et al, 1994), while that for Lactobacillus delbrueckii phages has been reported to range near 50% (Mikkonen and Alatossava, 1994).

The origin of new lactococcal bacteriophages in the dairy industry has been the subject of speculation for many years (Huggins and Sandine, 1977). It has been suggested that lytic phages may evolve from one another by exchange of DNA modules or that lysogenic starter cultures may be the source of new virulent phages (Jarvis, 1989). Sequence analysis of related phages provided some knowledge on phage evolution.

PHAGES OF LACTOBACILLI

Characterization of phages

Only a few Lactobacillus phages have been characterized in detail. For most of them, the only data available at the molecular level consist of the physical map. The nucleotide sequence of some genes or genomic elements has been determined: an insertion sequence in phage FFSV (Shimizu-Kadota et al, 1985), the attachment site (attP) (Raya et al, 1992) and the integrase gene of phage Ŧad (Fremaux et al, 1993), the cohesive ends of phage PL-1 (Nakashima et al, 1994) and the lysis genes of phage Ŧadh (Henrich et al, 1995). The characterization of genetic elements required for the integration of phage Ŧadh into the Lactobacillus gasseri chromosome led to the construction of an integration vector (Raya et al, 1992).

The best studied are the Lactobacillus delbrueckii subsp temperate phage mv4 and the virulent phage LL-H. Both are closely related and belong to the same homology group (Mata et al, 1986). The 36 kb genome of phage mv4 has been mapped physically and its DNA is circularly permuted (Lahbib-Mansais et al, 1992). Several genes have been characterized, such as those involved in cell lysis (Dupont, unpublished results, GenBank Z26590), the genes encoding structural proteins (Vasala et al, 1993) and the genetic elements required for site-specific integration (attP, integrase, excisionase) (Dupont et al, 1995). The segments sequenced cover about 23% of the mv4 phage genome.

It has been shown that integration of the phage mv4 DNA in the host chromosome occurs at the tRNA_{Ser} gene but an intact gene is preserved (Dupont et al, 1995). A non-replicative vector based on phage integration elements has been constructed and was shown to integrate into the tRNA_{Ser}
gene of *Lactobacillus plantarum* (Dupont et al, 1995).

Phage LL-H has a circularly permuted and terminally redundant double stranded DNA genome (Trautwetter et al, 1986). Its length is 34.6 kb with terminal repeats of 2.8 ± 0.2 kb (Forsman and Alatossava, 1991). The genes encoding the structural proteins (Mikkonen and Alatossava, 1994), the terminase (Mikkonen and Alatossava, 1995) and the lysin (Alatossava, personal communication) have been characterized. This sequenced segment represents about 67% of the genome. A group 1 intron has been discovered in the gene encoding the large subunit of the terminase (Mikkonen and Alatossava, 1995).

**Comparison studies**

Extensive studies on the homology between phages LL-H and mv4 have been conducted. Firstly, DNA hybridization studies between these two phages and some others have shown that they are closely related and form one group of DNA homology (Mata et al, 1986). Comparison at the nucleotide level of the cluster encoding structural proteins for both phages revealed that they are very similar (Vasala et al, 1993). More interestingly, it has been found that the virulent phage LL-H genome has a partially deleted integrase gene homologous to the 3' end of the mv4 integrase gene. In the LL-H genome also some traces of site-specific integration elements remain. These data suggest that the virulent phage LL-H and the temperate phage mv4 could have evolved from a common temperate ancestor (Mikkonen and Dupont, personal communication). It was previously shown that the virulent phage ΦFSV has been derived from the prophage ΦFSW, in *Lactobacillus casei* strain S-1, by acquisition of a new ISL1 insertion sequence (Shimizu-Kadota et al, 1985).

Comparison studies of two highly homologous phages of *Lactobacillus delbrueckii* subsp *lactis*, LL-H and LL-K, revealed the existence of an 1.5 kb insertion in the genome of the phage LL-K, called KIS-element (Forsman and Alatossava, 1994). This element, composed of two putative open reading frames flanked by perfect 19 bp direct repeats, is absent in the phage LL-H genome which retained only one copy of the repeats.

These results indicate that related lactobacilli phages, like lambdoid phages (Campbell, 1994) are able to evolve by exchange, insertion or deletion of modules.

**PHAGES OF LACTOCOCCI**

Much more data are available on phages active against lactococci than on those active against lactobacilli. Lactococcal phages have been classified in 12 groups based on morphology, serological studies and DNA homology (Jarvis et al, 1991). The majority of phages encountered in dairy plants belong to three groups. In France, 48% of the phages analyzed were virulent small isometric-headed (P008 type phage), 29% were virulent prolate-headed (c6a type phage) and 21% were either virulent or temperate small isometric-headed (P335 type phage) (Prevots et al, 1990). This distribution is grossly the same in New Zealand (Jarvis, 1977) and in Ireland (Casey et al, 1993). In contrast, in Germany, 60% of the phages isolated were prolate-headed (c6a type phage) and 40% were small isometric-headed (P008 and P335 type phages) (Lembke et al, 1980). The best characterized phages belong to one of these three prevailing groups of homology.

**P335 phage species**

These phages are isometric-headed and have a genome size ranging around 36 kb.
This group of homology is the only one, in the lactococci phage classification, which contains both virulent and temperate phages. The phage genome possesses cohesive ends or circularly permuted ends with terminal redundancy (Jarvis et al, 1991). Extensive studies have been carried out on the characterization of the temperate phages essentially with the aim to develop genetic tools, such as chromosomal integration systems and to have a better understanding of the mechanisms governing the conversion between the lytic cycle and the lysogenization step. The complete genome sequence of phage r1t (van Sinderen et al, 1995) and Tuc2009 (Daly and Fitzgerald, personal communication) was determined. In contrast, few data are available on virulent phages.

The genetic elements required for phage integration into the host chromosome (the attachment site which is the site of recombination between the phage and the host chromosome, and the integrase which catalyzes the recombination event) have been identified for several phages. The attP sequence, 5'-nCnCATG-3', and the integrase from phages Tuc2009 (van de Guchte et al, 1994a), φLC3 (Lillehaug and Birkeland, 1993) and r1t (van Sinderen et al, 1995) are almost identical. In contrast, the TP901-1 genome possesses a different attP sequence, 5'-TCAAT-3', (Christiansen et al, 1994), suggesting an integration mechanism different from the one used by Tuc2009, φLC3 and r1t. The genetic elements required for phage integration have been used to develop integrative vectors (van de Guchte et al, 1994a).

Maintenance of lysogeny depends on the synthesis of a repressor which blocks transcription of the lytic genes. A gene coding for a protein homologous to lambda phage repressor CI has been identified in Tuc2009 (van de Guchte et al, 1994b). A highly homologous gene, rro, identified in phage r1t (van Sinderen et al, 1995), was shown to encode a repressor of the lytic genes (Nauta et al, 1995). The repressor protein Rro binds to three operator sequences of 21 bp which repress transcription of the lytic genes (Nauta et al, 1995). These data have been used to develop an inducible gene expression vector (Nauta et al, 1995).

Some genes involved in the lytic cycle have also been identified by homology with sequences from the databases: proteins involved in DNA replication, a dUTPase (van Sinderen et al, 1995). Genes coding for structural proteins were identified by homology with the N terminal sequence of phage proteins isolated on SDS-PAGE electrophoresis (Arendt et al, 1994; van Sinderen et al, 1995). The functions of the genes involved in cell lysis have been confirmed by expression in *E. coli* (Arendt et al, 1994; Birkeland, 1994).

The presence of terminally redundant ends in the Tuc2009 genome (Arendt et al, 1994) suggests a headful packaging mechanism. φLC3 and r1t have instead identical 3' staggered cohesive ends (Lillehaug et al, 1991; van Sinderen et al, 1995). Both types of known phage packaging mechanisms seem therefore to be present in this family of lactococcal phages. A transducing vector based on the cohesive end region of phage φLC3 has been constructed (Birkeland and Holo, 1993).

Finally, a site-specific DNA endonuclease, homologous to those found in group I introns was identified in phage r1t (van Sinderen et al, 1995).

The only study on virulent P335 phages concerns the presumed origin of replication of the phage φ50 (Hill et al, 1990). It is constituted of long direct repeats. This locus is able to affect φ50 DNA replication in trans, and when cloned in a plasmid, the resulted recombinant has an increased copy number during infection by the phage.
However, studies on both types (temperate and virulent) of phages are of particular interest due to the fact they may clarify their relationships.

**P008 phage species**

Phages related to P008 are virulent small isometric-headed with a genome size ranging around 30 kb and cohesive ends (Jarvis et al., 1991). Only few genes from this type have been cloned and sequenced. A conserved region has been found in the genomes of several of these phages (Kim and Batt, 1991a). An open reading frame has been identified within the region, and analysis of the deduced protein sequence encoded by the gene indicated that it has a high lysin content and that it shares homology with eucaryotic translation initiation factor. It was suggested that this conserved protein may have a role in either regulating phage replication or specifying expression of its own gene (Kim and Batt, 1991a). Several genes encoding structural proteins from phage F4-1 have been characterized (Chung et al., 1991; Kim and Batt, 1991b). The sequence of the region surrounding the cohesive ends of phage sk1 has been determined (Chandry et al., 1994a). It presents structural features that may be involved in recognition by the terminase. The ΦUS3 lysin gene has been identified and its product, contrary to the other lysins characterized to date, seems to be an amidase instead of a muramidase (Platteeuw and de Vos, 1992).

Transcriptional studies on phage sk1 established that transcription is temporally regulated into three distinct phases concerning the early, middle and late stage of infection (Beresford et al., 1993; Chandry et al., 1994b). Sequence analysis of the biL66 middle expressed region revealed an operon formed of four open reading frames and transcribed 10 min after infection (Bidnenko et al., 1995). Characterization of the biL66 middle promoter has been reported. It is constituted of an extended −10 promoter sequence consensus and no consensus −35 region was observed (Bidnenko et al., 1995). This observation suggests the possibility that transcription could be initiated from this promoter by an activator. Identification of such kind a factor could be useful for the design of a controlled expression vector and for the understanding of the temporal regulation of transcription.

**c6a phage species**

Phages related to c6a are virulent prolate-headed with a genome size ranging around 22 kb and cohesive ends (Jarvis et al., 1991). Transcriptional studies have been done on phage c2 (Beresford et al., 1993). Two families of time dependent transcripts have been detected at early and late stage of infection. Until recently, the complete nucleotide sequence was available only for phage biL67 (Schouler et al., 1994). Thirty-seven open reading frames, organized in two clusters, were identified. Functions were assigned to the putative products of six of them. These were a DNA polymerase, a protein involved in recombination, a lysin, a terminase subunit, a structural protein and a holin. By Northern analysis, transcripts corresponding to the region including the gene encoding lysin were detected 20 min after infection (data not shown) and the cluster was referred to as the late region. Lysis of the host cell (IL1403, Chopin et al., 1984) occurs within 70 min after infection. In comparison with the amino acid sequence of known proteins, the second late gene has been identified as the gene encoding the lysin. It shares very high homology, up to 95%, with the lysin gene of ΦVML3 (Shearman et al., 1989), P001 (Hertwig, 1990) and c2 (Ward et al., 1993). It seems that the start codon for all these genes is...
not an AUG but an AUA (encoding isoleucine) in phages c2 and VML3, and an AUC (encoding isoleucine) in P001 and bIL67. AUA has already been found as a start codon in E coli (Belin et al., 1979), and recently, AUC has also been described as a start codon (Chalut and Egly, 1995). It has been suggested that the use of a rare start codon could be a way to decrease translation of lysin during phage multiplication (Schouler et al., 1994). Low levels of the VML3 lysin were produced even under the inducible T7 promoter (Shearman et al., 1994).

The last late gene located near the cos site, orf37, has all the structural features defining a holin (Schouler et al., 1994). Holins are membrane proteins which form pores to allow access of the lysin to the peptidoglycan substrate (Young, 1992). Orf37, subcloned in E coli, is able to complement a bacteriophage λ defective for its holin gene (λSam) (data not shown). This experiment gives the first indication that the gene really encodes a holin. The location of the bIL67 holin gene, 12.6 kb downstream of the lysin gene is very unusual. It resembles the organization of T7 (Dunn and Studier, 1983), while other holin genes map immediately upstream of the lysin gene. In any case, this location seems to be conserved for phages related to c6a, since by homology research, it has also been found in phage P6 (Ermel et al., 1994) P001 and c2 (Perrin et al., unpublished results, GenBank L37090 and L37091).

The recent publication of the complete nucleotide sequence of phage c2 (Polzin, personal communication) will allow a more detailed comparison of c6a phage species genome organization and evolution.

CONCLUSION

Accumulated data arising from the characterization of the lactobacilli and lactococci phage genomes provided information on their global organization. Functions have been assigned to the products of several genes. Even though many functions remain unknown; these studies provide some information on the phage life cycle, such as temporal expression of the genes, genetic control between lytic and lysogen cycle and release of the virion capsid outside the host. Bacteriophages represent a very dynamic population which is constantly evolving to accommodate changes. Some studies have given information on how the phages could evolve by acquisition of modules through recombination to overcome certain phage resistance mechanisms. Phage Φ50 acquired the methylase gene Lai from the R/M region of pTR2030 to circumvent the defense mechanisms of lactococci (Alatossava and Klaenhammer, 1991; Hill et al., 1991). Interestingly, van Sinderen et al (1995) reported homologies between Φ50 and r11 at the borders of insertion of the methylase gene in the phage Φ50 genome. Phage ul36 acquired a region from the chromosome of the infected host to overcome an abortive infection mechanism encoded on conjugative plasmid pTN20. This recombination led to the formation of a new lytic phage, ul37, resistant to the abortive infection mechanism, but, also with different morphology and origin of replication from that of the original phage (Moineau et al., 1994). These two examples of DNA cassette exchange between the infecting and the resident DNA are of particular interest. Elucidation of the recombination event should provide insight into the process of phage evolution and would lead to the elaboration of more performing starter cultures.

A short sequence could act as a recombination hot spot as it was found for r11 (van Sinderen et al., 1995). Two direct repeats of 25 and 31 bp have been found in bIL67 (Schouler et al., 1994). These repeats flank a cluster of genes, suggesting an exchangeable module.
Evidence is now available on the emergence of lytic phages from a lysogenic ancestor phage (Mikkonen and Dupont, personal communication).

The presence of a group I intron was also reported in the genome of unrelated phages (Mikkonen and Alatossava, 1995; van Sinderen et al, 1995). Such mobile introns are phylogenetically diverse, being found in both procaryotic and eucaryotic genomes (Derbyshire et al, 1995).

Phage encoded traits on cloned fragments may interfere with the normal lytic cycle by overproduction or titration of essential regulatory signals necessary for phage propagation (Hill et al, 1990). Another possible approach is the use of antisense mRNA to inhibit expression of essential phage genes (Kim and Batt, 1991c; Chung et al, 1992). We could then expect that characterization of the phage genome will allow the design of new strategies to prevent phage infection.

ACKNOWLEDGMENTS

I am indebted to L Dupont, P Ritzenhaler, M Mikkonen, T Alatossava, A Nauta, D van Sinderen, J Kok and G Venema for results communicated before publication. I also thank C Anagnostopoulos and MC Chopin for critical reading of the manuscript.

REFERENCES


Chalut C, Egly JM (1995) AUC is used as a start codon in Escherichia coli. Gene 156, 43-45


Chung DK, Chung SK, Batt CA (1992) Antisense RNA directed against the major capsid protein of Lactococcus lactis subsp cremoris bacteriophage 4-1 confers partial resistance to the host. Appl Microbiol Biotechnol 37, 79-83


coccus lactis subsp lactis bacteriophages. Genes 98, 95-100

Kim SG, Batt CA (1991b) Nucleotide sequence and deletion analysis of a gene coding for a structural protein of Lactococcus lactis bacteriophage F4-1. Food Microbiol 8, 27-36


Bacteriophage of lactic acid bacteria


Raya RR, Fremaux C, De Antoni GL, Klaenhammer TR (1992) Site-specific integration of the temperate bacteriophage φadh into the Lactobacillus gasseri chromosome and molecular characterization of the phage (attP) and bacterial (attB) attachment sites. J Bacteriol 174, 5584-5592


