**P087, a lactococcal phage with a morphogenesis module similar to an *Enterococcus faecalis* prophage**

Manuela Villion\(^1,2\), Marie-Christine Chopin\(^3\), Hélène Deveau\(^1,2\), S. Dusko Ehrlich\(^3\), Sylvain Moineau\(^{1,2,4*}\), and Alain Chopin\(^3\).

\(^1\)Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Groupe de Recherche en Ecologie Buccale (GREB), Faculté de Médecine Dentaire, 2420 rue de la Terrasse, Université Laval, Québec, Québec G1V 0A6, Canada,

\(^2\)Laboratoire de Génétique Microbienne, INRA, 78352 Jouy-en-Josas, France

\(^4\)Félix d’Hérelle Reference Center for Bacterial Viruses, 2420 rue de la Terrasse, Université Laval, Québec, Québec, G1V 0A6

* Corresponding author. Mailing address: Groupe de recherche en écologie buccale, Faculté de médecine dentaire, 2420 rue de la Terrasse, Université Laval, Quebec City, Quebec, Canada, G1V 0A6

Phone: 418-656-3712. Fax: 418-656-2861. E-mail: Sylvain.Moineau@bcm.ulaval.ca.
Abstract

The virulent lactococcal phage P087 was isolated from a dairy environment in 1978. This phage was then recognized as the reference member for one of the ten phage groups currently known to infect *Lactococcus lactis* strains. The double-stranded DNA genome of this *Siphoviridae* phage is composed of 66,074 bp and is circularly permuted. Five tRNA and 88 ORFs were found within an uncommon genome architecture. Eleven structural proteins were also identified through SDS-PAGE and LC-MS/MS analyses. Of note, 11 translated ORFs from the structural module of phage P087 have identities to gene products found in a prophage located in the genome of *Enterococcus faecalis* V583. The alignment of both genomic sequences suggests that DNA exchanges could occur between these two phages which are infecting low G+C bacteria found in similar ecological niches.
Introduction

Bacteriophages are now widely recognized as the most abundant microorganisms on the planet and they are arguably also the most diverse (Whitman et al., 1998; Wommack and Colwell, 2000). Such natural variation is a reflection of the array of hosts available to them as well as their high rate of adaptive evolution when facing selective pressure. In fact, in most environments, a large pool of phages and their respective bacterial hosts are involved in continuous cycles of co-evolution where phage-resistant host mutants help to preserve bacterial cell lineages while counter-resistant phage mutants emerge and threaten the new bacterial strains (Emond and Moineau, 2007). Genome recombination is one of the processes by which novel phages with unique characteristics are developed (Labrie and Moineau, 2007). In fact, it is the field of large-scale genome sequencing that led to unprecedented insights into phage and bacterial co-evolution (Ackermann and Kropinski, 2007). Viral metagenomic studies have also revealed that the overall gene pool of viruses is still vastly untapped (Edwards and Rohwer, 2005).

*Lactococcus lactis* is one of the relatively few Gram-positive bacterial species that has been domesticated by humans and extensively used for food fermentation processes (Moineau et al., 2002). Unfortunately, these bacteria are often susceptible to phage attacks, particularly during large-scale milk fermentation (Emond and Moineau, 2007). Due to their negative impacts on food bioprocesses and on the quality of fermented products, lactococcal phages are among the most studied bacterial viruses. Over 700 hundred lactococcal phage isolates have been reported in the literature while several other isolates are stored in company and university laboratories (Ackermann and Kropinski, 2007).
Basic research on lactococcal phages led to the development of practical phage classification schemes which were used to design effective strategies to limit phage propagation within manufacturing factories (Jarvis et al., 1991). The latest classification of \( L. \text{ lactis} \) phages recognized 10 morphologically and genetically distinct groups (Deveau et al., 2006), including 8 belonging to the \( \text{Siphoviridae} \) family (non-contractile tail) and 2 to the \( \text{Podoviridae} \) family (short tail). This grouping was based on electron microscopic observations, DNA-DNA hybridizations, and comparative genome analyses. Siphophages belonging to three groups (936, c2, and P335) have been the most scrutinized because they are the main causes of milk fermentation failures worldwide (Moineau et al., 1992; Josephsen et al., 1994; Bissonnette et al., 2000). For example, among the 25 wild-type lactococcal phage genomes for which a complete sequence is currently available, 13 belong to P335-like phage group, 6 belong to the 936 and 2 belong to the c2.

Members of the seven other lactococcal phage groups are infrequently found in the dairy industry or in ecological niches associated with lactococci. Nonetheless, members of these phage groups can still lead to fermentation collapses. With the aims of understanding how phage genomes are related to each other and what evolutionary mechanisms shaped the lactococcal phage population (Hatfull, 2008), the genome of phage members representing 4 of the less studied \( L. \text{ lactis} \) phage groups were recently analyzed. This included the virulent siphophages Q54 (Fortier et al., 2006) and 1706 (Garneau et al., 2008) as well as the virulent podophages KSY1 (Chopin et al., 2007) and assephi28, a member of the P034 group (Kotsonis et al., 2008). Thus, at least one complete genome sequence is available for 7 of the 10 recognized lactococcal phage groups. Here, we report the complete genome sequence and analysis of phage P087, a virulent representative of the lactococcal phage group that bears its name.
**Results and Discussion**

Phage P087 was isolated in Germany from a dairy environment in 1978 (Braun et al., 1989). Electron microscopy revealed that it belonged to the *Siphoviridae* family as for the majority of the lactococcal phages. Its noncontractile tail is 163 nm in length and 14-16 nm in width while its isometric capsid is approximately 59 nm in diameter (Fig. 1). It differs from other lactococcal phages through its complex baseplate with no terminal fiber (Deveau et al., 2006).

**Microbiological characterization**

We performed a host range analysis and phage P087 was able to infect the following *L. lactis* strains: SMQ-384 (also named C10), SMQ-385 (also named ML8 strain), NCK203, and SMQ-86. The latter two *L. lactis* strains are also sensitive to several lactococcal phages belonging to the P335 group (Hill et al., 1990; Moineau et al., 1992; Emond et al., 1997). Interestingly, it was recently shown that phage 1706 can also infect the same four strains (Garneau et al., 2008).

Phage P087 was not able to infect a set of 42 industrial *L. lactis* strains. Taken altogether, the host range of P087 is limited, but it can infect *L. lactis* strains that are sensitive to phages from at least two other genetically-distinct groups.

To determine whether P087 was sensitive to abortive infection mechanisms (Abi), we introduced a high copy vector (pNZ123) expressing AbiK (Emond et al., 1997), AbiQ (Emond et al., 1998) and AbiT (Bouchard et al., 2002) into *L. lactis* SMQ-384, and the resulting transformants were challenged with P087. Of the three mechanisms, only the AbiQ system was very effective against P087 (EOP of $\leq 10^{-8}$). Thus, natural means are already available to curtail the multiplication of this virulent phage.
A single contig with overlapping ends was obtained, suggesting that P087 extremities are circularly permuted. After removing the duplicated sequence, a sequence of 60,074 bp was obtained. This gives the second largest sequenced genome for a lactococcal phage, after KSY1 (Chopin et al., 2007). Most known **Siphoviridae** phages with genome of more than 60 kb and infecting Gram-positive bacteria are mycophages, but their genome have a GC content of 57.3% to 69% (Pedulla et al., 2003). The GC content of the lactococcal P087 genome is 34.4 % which is within the range of its **L. lactis** hosts (35.3 %) (Bolotin et al., 2001; Makarova et al., 2006; Wegmann et al., 2007), and other lactococcal phages (Hejnowicz et al., 2008).

The circular permutation of P087 genome was further analyzed by cloning and sequencing the ends. The sequences obtained were apparently random, with no common features (not shown). The terminal redundancy length could not be determined from the sequencing data. The positions of the ends were uniformly scattered throughout the phage genome, indicating a highly circularly permuted genome, characteristic of **pac**-type phages.

**Bioinformatic analysis and P087 genome organization**

Eighty-eight open reading frames were deduced from the complete genomic sequence. They were all oriented in the same direction and 22 of them were overlapping (Table 1 and Fig. 2). The smallest gene preceded by an adequate ribosome binding site (RBS) complementary to the end of **L. lactis** 16S rRNA would encode a protein of only 29 amino acids (gp44). The sizes of the remaining gene products varied from 33 (gp39) to 1309 amino acids (gp73). As observed for other phages, the P087 genome was highly compact, with only 5% of the genome having no
coding function. The longest non-coding region (about 600 bp) was found upstream of orf1. Globally, P087 genes were clustered in a typical temporal expression pattern. As seen for other dairy phages, the leftward half of the genome (orf1-orf55) contained genes involved in DNA metabolism, and expected to be expressed early in the phage infection cycle. The rightward half of the genome (orf56-orf88) contained genes encoding either structural proteins, or proteins involved in phage assembly or cell lysis (Fig. 2). These genes would likely be expressed later in the phage lytic cycle. No lysogeny module (or remnant) was found in the P087 genome, confirming its virulent nature. We could not find another siphophage with an identical genome architecture in databases. However, some similarities were observed with the recently described *Pseudomonas* phage YuA (Ceyssens et al., 2008). Bioinformatic analyses revealed that 32 gene products out of 88 (36%) were similar to known proteins in the GenBank database (Altschul et al., 1997) and in the ACLAME database, which is specific to viruses (Leplae et al., 2004). Of those, ten P087 gene products share significant homology to other lactococcal phage or host genome proteins available in public databases, although not necessarily giving the best hits.

**P087 DNA replication and metabolism**

Gp2 was found to contain a conserved domain from the helicase superfamily (HELICc: helicase superfamily C-terminal domain), a domain associated with DEXDc, DEAD-, and DEAH-boxes. Thus, gp2/helicase probably plays a role in DNA replication or transcription. The DNA polymerase of phage P087 can be encoded by orf4, as it deduced gene product has 27% identity with the polymerase of *Lactobacillus salivarius* phage SalI. Gp15 could encode a methylase homologous to the C-5 cytosine-specific DNA methylase family (Cheng, 1995) and be involved in the methylation of P087 genome to avoid the negative effect of host R/M systems.
Gp19 contains the motif CXXC (CMQC) and would be involved in reducing ribonucleotides while Gp23 possesses a peptidase domain (peptidase T family). We also identified a cysteine synthase (gp31, family CysK) involved in the amino acid transport and metabolism suggesting that P087 acquired a host pyrimidine biosynthesis function, as was the case for the myophage T4 (Drake and Kreuzer, 1994), and the large phages infecting *Pseudomonas aeruginosa* (Hertveldt et al., 2005). It was shown for coliphage T4 that the presence of such a gene helps phage growth (Drake and Kreuzer, 1994). To our knowledge, this is the first time that such a gene has been found in a lactococcal phage. Gp53, which could be a DNA polymerase, shares 45% identity with a putative polymerase found in the virulent lactococcal phage Q54 (Fortier et al., 2006) and gp77 from *Listeria* phage A511 (Klumpp et al., 2008). Gp54 contains a ParB-like nuclease domain in its N-terminus and may be involved in DNA binding and cleavage of single-stranded DNA (Johnson et al., 1999). Gp54 has also 40% identity with a methyltransferase from the temperate streptococcal phage EJ-1. The identity was limited to the N-terminal part, which corresponds to the ParB-like domain (data not shown). Gp62 is likely a protease/scaffold protein because it shows 34% identity with the N-terminus of the CLP-protease of *Lactobacillus reuteri* F275.

*The structural proteome of P087 and other genes coding for morphogenetic proteins*

Proteomic analysis of phage P087 led to the identification of eleven structural proteins (Fig. 3). The three most abundant (major) structural proteins were easily identifiable by SDS-PAGE and were in agreement with a previously published protein profile of P087 (Braun et al., 1989). The smallest of these three bands represented a protein containing a bacterial Ig-like domain found in cellular surface proteins (gp59), while the other two, identified as gp63 and gp70 are likely the major capsid (MCP) and the major tail (MTP) proteins. Gp73 has homology
to the tail tape measure protein of staphylococcal phage phiNM3, while a segment of gp86 shares similarity with the studied neck passage structure (NPS) of *L. lactis* phage TP901-1 (Johnsen et al., 1995; Vegge et al., 2006). A 547-bp region in the middle part of the *nps* gene was highly conserved (82% identity) between TP901-1 (P335 group) and P087. However, it is worth mentioning that gp86 is 274 amino acids larger than the NPS protein of TP901-1. The receptor binding protein (RBP) of phage P087 could be gp78 as it shares 26% identity with the RBP from *L. lactis* phage SL3 (936 group). The identity was limited to the C-terminus, which corresponds to the head domain involved in host recognition (Spinelli et al., 2006; Tremblay et al., 2006).

**P087 lysis genes**

Gp85 has the hallmarks of an endolysin as it contains an amidase domain and is highly homologous to several lactococcal phage endolysins. Interestingly, gp75 may also be an endolysin as it is 29% identical to an endolysin found in *Enterococcus faecalis*. The presence of two endolysins was also predicted in the lactococcal phage KSY1 (Chopin et al., 2007) and in other double-stranded DNA phages (Wang et al., 2000). As reported for KSY1 and ascphi28 (lactococcal P034 group), no holin gene could be detected in the genome of P087. However, gp83 possesses several holin structural features such as two transmembrane domains (MEMSAT program; Jones, 2007), a short hydrophobic N-terminus, and a highly charged C-terminus (Young and Blasi, 1995). Usually, the holin gene directly precedes an endolysin gene in siphophage genomes. However, a short *orf* is found between *orf83* (holin gene) and *orf85* (endolysin gene). Interestingly, the lysis module was located within the morphogenesis module of P087. This organization was previously only observed in the genome of lactococcal phages 1706 (Garneau et al., 2008) and KSY1 (Chopin et al., 2007) as well as in mycophages (Hatfull, 2008).
**Presence of tRNA in P087 genome**

Bioinformatic searches revealed the presence of five tRNAs clustered together at the 3’ end of the P087 genome, namely tRNA\(^\text{Asn}\) (recognizing the codon AAC), tRNA\(^\text{Asp}\) (GAC), tRNA\(^\text{Cys}\) (UGC), tRNA\(^\text{Pro}\) (CCA), and tRNA\(^\text{Thr}\) (ACA). Transfer RNAs are the only translation-associated genes usually found in phages, particularly those with a large genome (Bailly-Bechet et al., 2007).

Three tRNAs were also identified in lactococcal phage KSY1 but they were separated from each other by several short orfs (Chopin et al., 2007). To date, this is the highest number of tRNA genes found in a lactococcal phage genome. However, up to 14 tRNA genes were recognized in the genome (131,573 bp) of the virulent *Lactobacillus plantarum* myophage LP65 (Chibani-Chennoufi et al., 2004). These tRNA genes were likely acquired by recombination with other phages or host DNA (Weinbauer, 2004). Transfer RNAs found in virulent phage genomes tend to correspond to highly used codons, leading to an enhanced translational efficiency (Bailly-Bechet et al., 2007). Indeed, the frequency of usage per thousand codons corresponding to four (of the five) tRNAs was significantly higher in phage P087 than in *L. lactis* strains (Table 2). Moreover, the genes coding for three most abundant structural proteins of P087 (gp59, gp63, and gp70) appears to have more codons recognized by those tRNA found in P087 genome (at least 2/5). However, it should be noted that the genomic sequence is not available for any of P087 host strains thereby, limiting the interpretation.

**Similarity to the sequence of a putative Enterococcus faecalis V583 prophage**

Many structural proteins of P087 shows identities with proteins deduced from the genome of *Enterococcus faecalis* V583, a clinical vancomycin-resistant isolate (Paulsen et al., 2003).
Analysis of the complete genomic sequence of strain V583 revealed seven mobile regions that could be linked to prophages (Paulsen et al., 2003). One of these regions (coordinates EF2084 to EF2145) codes for several proteins sharing between 20 to 45% identities with P087 structural proteins (gp77 to gp54, Table 1, Fig. 2). This specific V583 region corresponds to the enterococcal prophage 05 (Lepage et al., 2006), which is located downstream of the 3’end of tRNA-Thr2 and flanked by a 15-bp repeat corresponding to its attL/R attachment site (Fig. 2). Interestingly, \textit{E. faecalis} prophage 05 is present only in clinical isolates and is apparently absent in \textit{E. faecalis} strains isolated from dairy foods. This enterococcal prophage could contribute to the adaptation of some \textit{E. faecalis} strains to a specific ecological niche (Lepage et al. 2006).

\textit{E. faecalis} is a low GC Gram positive bacterium found in the mammalian gastrointestinal tract but also in soil, water, and foods (Klare et al., 2001). It is tempting to speculate that the virulent \textit{Lactococcus lactis} phage P087 is derived, at least in part, from a phage infecting another low GC Gram positive bacteria. Alternatively, the \textit{E. faecalis} prophage 05 could be derived from a dairy phage. In fact, DNA exchanges between \textit{Lactococcus} and \textit{Enterococcus} have been observed previously. For example, a multi-antibiotic resistance plasmid (pK214) isolated from a \textit{Lactococcus} strain found in raw milk cheese could be transferred to an \textit{E. faecalis} strain (Perreten et al., 1997). Similarly, another plasmid (pRE25) could be transferred by conjugation from an \textit{E. faecalis} strain to a \textit{L. lactis} strain, confirming a molecular communication between these two bacterial species (Teuber et al., 2003).

In conclusion, we have described the reference member of the 8\textsuperscript{th} lactococcal phage group. Genome analysis of P087 revealed a mosaic structure made up of modules that come from disparate origins. The proteomic similarities with an enterococcal prophage coupled with the possible acquisition of a receptor-binding protein (gp78) from another lactococcal phage (936), suggests a mechanism for the emergence of P087.
Materials and Methods

Microbiological assays

Phage P087 and its host were obtained from the Félix d’Hérelle Reference Center for Bacterial Viruses (www.phage.ulaval.ca). The host, *Lactococcus lactis* C10, was grown at 30°C in M17 broth supplemented with 0.5% glucose (GM17). For phage amplification, phage and host were incubated at room temperature in the presence of 10 mM CaCl$_2$. When needed, glycine (0.5%) was added to the top agar to increase plaque size and facilitate phage enumeration (Lillehaug, 1997). Phage lysates were concentrated with polyethylene glycol (PEG) and purified on a discontinuous-step CsCl gradient (Sambrook and Russell, 2001). Phages were stained with 2% uranyl acetate and observed using a JEOL 1230 transmission electron microscope at 80 kV as described elsewhere (Deveau et al., 2006). To measure the efficacy of phage defense mechanisms, the EOP was calculated by dividing the phage titer for the tested *L. lactis* strain by the titer for the sensitive strain (*L. lactis* SMQ384 containing pNZ123; De Vos, 1987). The tested strains were SMQ-384 transformed with pNZ123 containing, either *abiK* (pSRQ823; Emond et al., 1997), *abiT* (pED209; Bouchard et al., 2002) or *abiQ* (pSRQ928; Emond et al., 1998).

DNA sequencing

The DNA of phage P087 was isolated from CsCl-purified phages as reported elsewhere (Chibani Azaiez et al., 1998). The complete genomic sequence was determined as previously described (Chopin et al., 2007). Approximately 1000 reads were assembled, achieving 8.6-fold coverage and resulting in a single circular contig. To identify genome ends, terminal genome
fragments were cloned and sequenced. P087 genomic DNA was treated with T4 polymerase (New England Biolabs). Blunted genomes were ligated into the SmaI-digested cloning vector pUC19. The genomic DNA/pUC19 ligation reaction was then digested with EcoRI and HindIII and self-ligated. pUC19 has unique EcoRI and HindIII sites while these enzymes have 0 and 30 restriction sites in the genome of P087, respectively. The ligation mixture was transformed into *Escherichia coli* TG1 competent cells. Transformants were picked at random and P087 DNA cloned fragments were PCR-amplified using oligonucleotides complementary to pUC19 sequence. The sequence of fragments cloned into 17 independent end clones was determined.

**Orf prediction and annotation**

Open reading frame and tRNA searches were carried out as described elsewhere (Chopin et al., 2007; Garneau et al., 2008). Predictions were visually inspected and the putative ribosome-binding sites were verified using the 3'-end of *L. lactis* IL1403 16S rRNA (Bolotin et al., 2001). The presence of conserved domains was investigated on NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), within the CDD database. The isoelectric point and molecular mass were determined using http://ca.expasy.org/tools/pi_tool.html. The translated ORF products were compared with known protein sequences using BLASTP (Altschul et al., 1997) and the non-redundant public GenBank database. Blast searches were also done using the ACLAME database of clustered viral proteins maintained at Service de Conformation de Macromolécules Biologiques et de Bioinformatique, Université Libre de Bruxelles (http://aclame.ulb.ac.be/) (Leplae et al., 2004). The frequency usage was calculated with the countcodon program from the web site: http://www.kazusa.or.jp/codon/cgi-bin/countcodon.cgi.
Analysis of phage P087 structural proteins

One liter of phage lysate was PEG-concentrated, purified on a discontinuous CsCl gradient, and on a one-step CsCl gradient. Ultracentrifugation was performed using a Beckman SW41 Ti rotor at 35,000 rpm for 3 h. The second ultracentrifugation was performed using a Beckman NVT65 rotor at 60,000 rpm for 17 h. The phage preparation (8 x 10^{11} PFU/ml) was then dialyzed against phage buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 mM MgSO_4) and analyzed for structural proteins by standard Tris-glycine 12% SDS-polyacrylamide gel electrophoresis (PAGE). Samples were mixed with 4X sample loading buffer and boiled for 5 min before loading. Protein bands were detected by Coomassie blue staining. The bands were cut out of the gel, digested with trypsin, and identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Genome Quebec Innovation Centre, McGill University).

Accession number

The sequence data was deposited in EMBL/Genbank/DDBJ databases under accession number FJ429185.

Acknowledgments

We are grateful to Marie-Eve Dupuis, Audrey Fleury, and Steve Labrie for their technical help in phage assays, and to Barbara-Ann Conway for editorial assistance. This study was funded, in part, by a strategic grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.M.
References


Legends of the Figures

Figure 1: Electron micrograph of lactococcal phage P087.

Figure 2: Genomic organization of the virulent *Lactococcus lactis* phage P087 compared to *Enterococcus faecalis* V583 prophage 05. Prophage 05 was previously identified by Paulsen et al. (2003) and the *att* sites by Lepage et al. (2006). Each putative ORF is represented by an arrow. The putative functions of the corresponding gene products are indicated above (for P087) or below (for prophage 05) the arrows. Bold arrows indicate structural proteins identified by LC-MS/MS (see Fig 3). Blue arrows linked by grey shadows indicate ORFs for which translated products share identity. The percentage of identity is also indicated. The scale above the map is in base pairs.

Figure 3: Analysis of structural proteins of lactococcal phage P087. Panel A, Coomassie blue staining of a 12% SDS-PAGE gel showing the structural proteins of phage P087. The bands extracted and identified by LC-MS/MS are numbered on the right side of the gel. On the left side, the 7-175 kDa Broad Range Marker (New England BioLabs) was used to estimate protein molecular mass. Panel B, Identification of the structural proteins of P087 identified in Panel A.
<table>
<thead>
<tr>
<th>ORF</th>
<th>Position</th>
<th>Start</th>
<th>End</th>
<th>Size (a.)</th>
<th>MM (KD)</th>
<th>pl</th>
<th>Predicted function</th>
<th>Best homologs in databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp1</td>
<td>3416</td>
<td>4942</td>
<td>508</td>
<td>37.2</td>
<td>8.1</td>
<td></td>
<td>Hypothetical protein at173 [Artocarpus siamensis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp4</td>
<td>5675</td>
<td>7642</td>
<td>655</td>
<td>75.4</td>
<td>5.2</td>
<td></td>
<td>DNA polymerase [Lactobacillus salivarius]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp6</td>
<td>4479</td>
<td>8974</td>
<td>231</td>
<td>26.5</td>
<td>5.5</td>
<td></td>
<td>Meta-Ag-lactase superfamily homolog [Bacillus subtilis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp7</td>
<td>4963</td>
<td>8676</td>
<td>137</td>
<td>16.3</td>
<td>5.6</td>
<td></td>
<td>Hypothetical protein CLOELP_01819 [Clostridium leptum DSM753]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp11</td>
<td>1147</td>
<td>1211</td>
<td>212</td>
<td>24.5</td>
<td>8.9</td>
<td></td>
<td>Putative 5s RNA methyltransferase [Bacillus subtilis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp13</td>
<td>13128</td>
<td>13487</td>
<td>119</td>
<td>14.2</td>
<td>4.4</td>
<td></td>
<td>C-terminal part of Lysin [Enterococcus faecalis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp14</td>
<td>1384</td>
<td>13720</td>
<td>78</td>
<td>9.1</td>
<td>5.5</td>
<td></td>
<td>Glycosyl hydrolase related [Lactococcus lactis SK11]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp23</td>
<td>16562</td>
<td>16908</td>
<td>348</td>
<td>39.6</td>
<td>4.4</td>
<td></td>
<td>Hypothetical protein at123 [Bacillus subtilis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp31</td>
<td>17772</td>
<td>18071</td>
<td>317</td>
<td>8.1</td>
<td>5.3</td>
<td></td>
<td>Cytosine synthase [E. coli]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp32</td>
<td>1871</td>
<td>1901</td>
<td>97</td>
<td>11.6</td>
<td>6.3</td>
<td></td>
<td>p12 putative [Lactococcus lactis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp42</td>
<td>23330</td>
<td>23439</td>
<td>109</td>
<td>43.4</td>
<td>9.7</td>
<td></td>
<td>p13 putative [Vibrio cholerae]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp48</td>
<td>26300</td>
<td>26791</td>
<td>163</td>
<td>18.6</td>
<td>4.3</td>
<td></td>
<td>Isocitryl-CoA-RNA synthetase [Methanobrevibacter smithii]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp53</td>
<td>28880</td>
<td>29324</td>
<td>414</td>
<td>4.7</td>
<td>5.3</td>
<td></td>
<td>DNA polymerase [Bacillus subtilis]</td>
<td>Unique (100%)</td>
</tr>
<tr>
<td>gp54</td>
<td>25234</td>
<td>26848</td>
<td>174</td>
<td>19.9</td>
<td>5.2</td>
<td></td>
<td>C-terminal part of Lysin [Enterococcus faecalis]</td>
<td>Unique (100%)</td>
</tr>
</tbody>
</table>

**TABLE 1. ORFs deduced from P087 genome sequence for which gene products present a putative function or have homologs in databases**

- **ORF**
- **Position**
- **Size**
- **MM**
- **pl**
- **Predicted function**
- **Best homologs in databases**
- **Identical/overall (%)**
- **Size (a.)**
- **E-value**
- **Accession number**

---

**gp1**
- **Position**: 3416
- **Start**: 4942
- **End**: 508
- **Size (a.)**: 37.2
- **MM (KD)**: 8.1
- **pl**: Hypothetical protein at173 [Artocarpus siamensis]
- **Best homologs in databases**: Unique (100%)

**gp4**
- **Position**: 5675
- **Start**: 7642
- **End**: 655
- **Size (a.)**: 75.4
- **MM (KD)**: 5.2
- **pl**: DNA polymerase [Lactobacillus salivarius]
- **Best homologs in databases**: Unique (100%)

**gp6**
- **Position**: 4479
- **Start**: 8974
- **End**: 231
- **Size (a.)**: 26.5
- **MM (KD)**: 5.5
- **pl**: Meta-Ag-lactase superfamily homolog [Bacillus subtilis]
- **Best homologs in databases**: Unique (100%)

**gp7**
- **Position**: 4963
- **Start**: 8676
- **End**: 137
- **Size (a.)**: 16.3
- **MM (KD)**: 5.6
- **pl**: Hypothetical protein CLOELP_01819 [Clostridium leptum DSM753]
- **Best homologs in databases**: Unique (100%)

**gp11**
- **Position**: 1147
- **Start**: 1211
- **End**: 212
- **Size (a.)**: 24.5
- **MM (KD)**: 8.9
- **pl**: Putative 5s RNA methyltransferase [Bacillus subtilis]
- **Best homologs in databases**: Unique (100%)

**gp13**
- **Position**: 13128
- **Start**: 13487
- **End**: 119
- **Size (a.)**: 14.2
- **MM (KD)**: 4.4
- **pl**: C-terminal part of Lysin [Enterococcus faecalis]
- **Best homologs in databases**: Unique (100%)

**gp14**
- **Position**: 1384
- **Start**: 13720
- **End**: 78
- **Size (a.)**: 9.1
- **MM (KD)**: 5.5
- **pl**: Glycosyl hydrolase related [Lactococcus lactis SK11]
- **Best homologs in databases**: Unique (100%)

**gp23**
- **Position**: 16562
- **Start**: 16908
- **End**: 348
- **Size (a.)**: 39.6
- **MM (KD)**: 4.4
- **pl**: Hypothetical protein at123 [Bacillus subtilis]
- **Best homologs in databases**: Unique (100%)

**gp31**
- **Position**: 17772
- **Start**: 18071
- **End**: 317
- **Size (a.)**: 8.1
- **MM (KD)**: 5.3
- **pl**: Cytosine synthase [E. coli]
- **Best homologs in databases**: Unique (100%)

**gp32**
- **Position**: 18718
- **Start**: 1901
- **End**: 97
- **Size (a.)**: 11.6
- **MM (KD)**: 6.3
- **pl**: p12 putative [Lactococcus lactis]
- **Best homologs in databases**: Unique (100%)

**gp42**
- **Position**: 23330
- **Start**: 23439
- **End**: 109
- **Size (a.)**: 43.4
- **MM (KD)**: 9.7
- **pl**: p13 putative [Vibrio cholerae]
- **Best homologs in databases**: Unique (100%)

**gp48**
- **Position**: 26300
- **Start**: 26791
- **End**: 163
- **Size (a.)**: 18.6
- **MM (KD)**: 4.3
- **pl**: Isocitryl-CoA-RNA synthetase [Methanobrevibacter smithii]
- **Best homologs in databases**: Unique (100%)

**gp53**
- **Position**: 28880
- **Start**: 29324
- **End**: 414
- **Size (a.)**: 4.7
- **MM (KD)**: 5.3
- **pl**: DNA polymerase [Bacillus subtilis]
- **Best homologs in databases**: Unique (100%)

**gp54**
- **Position**: 25234
- **Start**: 26848
- **End**: 174
- **Size (a.)**: 19.9
- **MM (KD)**: 5.2
- **pl**: C-terminal part of Lysin [Enterococcus faecalis]
- **Best homologs in databases**: Unique (100%)
The putative prophage 05 codons from *E. faecalis* V583 were calculated from positions 2004910 (start of EF2084) to 2048146 (stop of EF2145).

<table>
<thead>
<tr>
<th>tRNA</th>
<th>Codon</th>
<th>Amino Acid</th>
<th>P087</th>
<th>L. lactis MG1363</th>
<th>L. lactis IL1403</th>
<th>L. lactis SK11</th>
<th>E. faecalis V583</th>
<th>E. faecalis prophage 05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro</td>
<td>CCA</td>
<td>P</td>
<td>17.2</td>
<td>15.0</td>
<td>15.5</td>
<td>14.5</td>
<td>16.2</td>
<td>13.4</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
<td>T</td>
<td>28.6</td>
<td>22.1</td>
<td>22.9</td>
<td>22.1</td>
<td>24.8</td>
<td>23.5</td>
</tr>
<tr>
<td>Asn</td>
<td>AAC</td>
<td>N</td>
<td>20.3</td>
<td>10.7</td>
<td>10.6</td>
<td>11.1</td>
<td>14.0</td>
<td>14.9</td>
</tr>
<tr>
<td>Asp</td>
<td>GAC</td>
<td>D</td>
<td>23.4</td>
<td>14.7</td>
<td>13.9</td>
<td>14.4</td>
<td>12.5</td>
<td>13.3</td>
</tr>
<tr>
<td>Cys</td>
<td>UGC</td>
<td>C</td>
<td>1.7</td>
<td>1.1</td>
<td>1.0</td>
<td>1.1</td>
<td>1.7</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Total number of codons 18442 739646 667611 696252 963629 13623
Figure 1

Click here to download Figure: Figure 1.ppt

Figure 1
Figure 3

Click here to download Figure: Figure 3.ppt

<table>
<thead>
<tr>
<th>Band</th>
<th>Gene product</th>
<th>Molecular mass (KDa)</th>
<th>SDS-PAGE</th>
<th>Calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gp73</td>
<td>132</td>
<td>136.3</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>gp86</td>
<td>110</td>
<td>103.2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>gp57</td>
<td>54</td>
<td>56.4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>gp70</td>
<td>53</td>
<td>48.5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>gp63</td>
<td>42</td>
<td>42.5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>gp88</td>
<td>40</td>
<td>34.2</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>gp87</td>
<td>32</td>
<td>31.0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>gp76</td>
<td>31</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>gp80</td>
<td>29</td>
<td>29.1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>gp64</td>
<td>28</td>
<td>28.8</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>gp59</td>
<td>18</td>
<td>18.2</td>
<td></td>
</tr>
</tbody>
</table>