

Abteilung Mikrobiologie
Zentralinstitut für Ernährungs- und Lebensmittelforschung Weihenstephan
Technische Universität München

Comparative genomics of *Listeria* bacteriophages

JULIA DORSCHT

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung eines akademischen Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. D. Langosch
Prüfer der Dissertation: 1. Univ.-Prof. Dr. S. Scherer
2. Univ.-Prof. Dr. M. J. Loessner
Eidgenössische Technische Hochschule Zürich, Schweiz

Die Dissertation wurde am 20.11.2006 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 24.01.2007 angenommen.

TABLE OF CONTENT

ABBREVIATIONS

SUMMARY **1**

ZUSAMMENFASSUNG **3**

I. INTRODUCTION **5**

1. The genus *Listeria* 5

1.1. Microbiology and taxonomy 5

1.2. Pathogenesis of listeriosis 5

1.3. Intracellular life cycle of *Listeria monocytogenes* 6

2. Bacteriophages 8

2.1. Historical Sketch 8

2.2. Phage Taxonomy 9

2.3. Bacteriophage proliferation 10

2.3.1. Lytic life cycle 10

2.3.2. Lysogenic life cycle 11

2.4. *Listeria* phages 11

2.4.1. General information 11

2.4.2. *Listeria* phage applications 12

3. Impact of bacteriophage genomics 14

4. Aims of this work 15

II. MATERIALS & METHODS **17**

1. Materials 17

1.1. Strains 17

1.2. Bacteriophages 17

1.3. Plasmids 18

1.4. Media 18

1.5. Buffers 19

1.6. Enzymes 20

1.7. Kits 20

2. Methods 20

2.1. Phage propagation 20

2.2. Phage titres 21

2.3. Phage purification 21

2.4. DNA extraction from bacteriophages	22
2.5. DNA extraction from <i>Listeria</i>	22
2.6. DNA precipitation	22
2.7. Agarose gel electrophoresis	22
2.8. Recover DNA fragments from agarose gels	23
2.9. Polymerase chain reaction (PCR)	23
2.10. Purification of PCR products	24
2.11. Ligation	24
2.12. Electrotransformation	25
2.13. Cloning and nucleotide sequencing	25
2.14. Determination of the phage attachment site <i>attP</i>	26
2.14.1. Lysogenisation of <i>Listeria</i> strains	26
2.14.2. Inverse PCR	27
2.15. Phage induction from lysogens by UV radiation	27
2.16. Protein analysis	27
2.17. Mass spectrometry (MS), peptide mass fingerprinting	29
2.18. Bioinformatics	29
2.19. Nucleotide sequence accession numbers	30
III. RESULTS	31
1. Genome analysis of the small bacteriophages	31
1.1. Results of nucleotide sequencing: general genome features	31
1.2. The genome of B025 features cohesive ends	32
1.3. Identification of ORFs and functional assignments	33
1.4. Genomes are organised into functional modules	37
1.5. P35 lacks the module of lysogeny control	37
1.6. Identification of the attachment sites	38
1.7. Protein homologies	41
1.8. Comparative genomics	42
2. Genome analysis of the virulent Myovirus A511	46
2.1. General features of the genome	46
2.2. Functional assignments of the predicted gene products	46
2.3. Genomic organisation of A511	47
2.4. Comparative genomic analysis	48

3. Protein profiles	53
IV. DISCUSSION	59
1. Comparative genomics and phage relationships	59
2. Protein profiles indicate programmed translational frameshifting	67
3. Site specific integration and attachment sites	69
4. Clues on <i>Listeria</i> phage evolution	71
V. REFERENCES	74
VI. PUBLICATIONS	85
VI. APPENDIX	86
Bioinformatical results and protein similarities of <i>Listeria</i> bacteriophages A006, A500, B025, B054, P35 and A511	
Table VI.1. A006	86
Table VI.2. A500	89
Table VI.3. B025	92
Table VI.4. B054	94
Table VI.5. P35	97
Table VI.6. A511	99
ACKNOWLEDGEMENTS	105

ABBREVIATIONS

aa	Amino acid
Arg	Arginine
bp	Base pairs
C-terminal	Carboxy-terminal
Cps	Capsid protein
CsCl	Caesium chloride
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP)
dsDNA	double stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
EtOH	Ethanol
Fig.	Figure
x g	relative centrifugal force
GP	Gene product
GDP	Guanosine di-phosphate
GTP	Guanosine tri-phosphate
kb	kilo base pairs
kDa	kilo Dalton
Lys	Lysine
Met	Methionine
MS	Mass spectrometry
MW	Molecular weight
nt	nucleotides
N-terminal	Amino-terminal
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamid gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pfu	Plaque-forming units

pI	Isoelectric point
REA	Restriction enzyme analysis
rpm	rounds per minute
RT	Room temperature
SDS	Sodium dodecyl sulphate
ssDNA	single stranded DNA
sv	Serovar
Tmp	Tail tape measure protein
tRNA	transfer ribonucleic acid
Tsh	Tail shaft protein
Tris	Tris(hydroxymethyl)aminomethane
U	Unit
v/v	volume/volume
w/v	weight/volume

SUMMARY

Molecular information on bacteriophage genomes provides insight into phage biology and evolution and is of special interest with regard to the development of applications and molecular tools in research and industry. In this study, the dsDNA genomes of six tailed *Listeria* bacteriophages, A006, A500, A511, B025, B054, and P35 featuring different morphotypes and host ranges have been sequenced and computationally analysed. Phages A006, A500, B054, and P35 featured terminally redundant genomes. The genome of B025 exhibited single stranded 3' overhanging ends of ten nucleotides. In consistency with the G+C content of their host bacteria, the *Listeria* phage genomes featured low G+C contents of 35-36 mol% in average, except for P35 with a G+C content of 40.8 mol%. Of the investigated phages with non-contractile tails (*Siphoviridae*), P35 possessed the smallest genome of 35.8 kb, whereas A006 and A500 featured a unit genome size of 38.1 kb and 38.9 kb, respectively. B025 had the largest genome with 42.7 kb. The phages with contractile tails (*Myoviridae*) revealed larger genomes: the temperate phage B054 featured 48.2 kb, and the virulent phage A511 exhibited 134.5 kb. In general, functional assignments to predicted gene products, which were based on similarity to known proteins, indicated genomic arrangement in functional modules. The temperate phage genomes comprised three life-cycle specific gene clusters, the "early" lytic genes (DNA transcription, replication and modification), the "late genes" (DNA packaging enzymes, structural components and cell lysis system) and a module for the regulation of lysogeny, which also contains the phage attachment site *attP*. The correspondent bacterial integration sites *attB* for A006, A500, and B025 were located at the 3'-termini of tRNA genes, whereas B054 was shown to integrate into the 3'-terminus of the gene encoding the translation elongation factor EF-Ts. In P35, the lysogeny related functions are completely absent, which supposedly benefits the phage in avoiding homoimmunity suppression during infection of prophage encoding bacterial cells.

The virulent phage A511 encoded all required factors for a host-independent DNA replication. A cluster of sixteen tRNA genes mainly represented codons of high frequency and may be important in avoiding bottle necks in large scale synthesis of the virion.

Analysis of the structural proteins by peptide mass fingerprinting allowed a correlation between predicted gene products and protein bands from the profiles and indicated translational frameshifts during protein synthesis of the major capsid and tail proteins in A118 and A500.

Comparative genomic analysis demonstrated the mosaic nature of the genomes investigated and supports the modular evolution theory. Close genetic relationship was indicated between A006, A500 and the published genome of A118 on the one hand, and between B025 and PSA on the other hand. Both groups revealed similarity to several prophages of *L. monocytogenes* and *L. innocua*. B054 clearly resembled one of the prophages annotated in the genome of *L. innocua* Clip11262. The virulent bacteriophages A511 and P35 seem to be more distantly related to the temperate ones. The genome of A511 revealed high synteny to the recently published genome of the virulent *Listeria* phage P100. But there was little protein similarity to the other *Listeria* phages investigated this study. Predicted proteins mainly showed similarity to virulent *Staphylococcus* phages. P35 encoded only two gene products of low similarity to a *L. innocua* prophage. Instead, many of the “late genes” were found to be similar to gene products of an *Enterococcus faecalis* prophage.

Altogether, this comparative genomic analysis is important for our understanding of the relationships of *Listeria* phages, and provides a good basis for application-oriented research as well as further investigations concerning molecular phage biology.

ZUSAMMENFASSUNG

Molekularbiologische Erkenntnisse über Phagengenome ermöglichen Einblicke in die Biologie und Evolution der Bakteriophagen. Zudem sind sie hinsichtlich der Entwicklung neuer Anwendungsmöglichkeiten und molekularbiologischer Werkzeuge für Forschung und Industrie von besonderem Interesse. Im Rahmen dieser Arbeit wurden die Genome von sechs *Listeria*-Bakteriophagen unterschiedlicher Morphotypen und verschiedener Wirtsbereiche vollständig sequenziert und bioinformatisch analysiert. Dabei handelte es sich dsDNA-Moleküle der Phagen A006, A500, A511, B025, B054 und P35. Die Phagen A006, A500, B054 und P35 besitzen terminal redundante Genome. Dagegen weist die Genomstruktur von B025 einzelsträngige, 3'-überhängende Enden mit einer Länge von zehn Nukleotiden auf. In Übereinstimmung mit dem G+C Gehalt ihrer Wirtsbakterien lag der G+C Gehalt der Phagengenome durchschnittlich zwischen 35-36 mol%. Mit einem deutlich höheren Wert von 40,8 mol% stellte P35 eine Ausnahme dar. Von den untersuchten Phagen mit nicht-kontraktile Schwänzen aus der Familie der *Siphoviridae* enthielt P35 das kleinste Genom mit 35,8 kb. A006 und A500 lagen im mittleren Bereich mit 38,1 kb bzw. 38,9 kb, während bei B025 die längste Genomsequenz von 42,7 kb identifiziert wurde. Aus der Familie der *Myoviridae*, der Phagen mit kontraktile Schwänzen, ergab die Sequenzierung des temperenten Phagen B054 eine Genomlänge von 48,2 kb, im Gegensatz zu dem großen virulenten Phagen A511 mit einem Genom von 134,5 kb. Insgesamt wurden 517 offene Leseraster identifiziert. Ähnlichkeiten zu bekannten Proteinen lieferten Hinweise darauf, dass die Gene in den Phagengenomen in funktionellen Clustern angeordnet sind, die bei den temperenten Phagen den verschiedenen Stadien eines Infektionszyklus entsprechen. Dabei handelt es sich um die „frühen“ lytischen Gene (Transkription, Replikation und Modifikation der Phagen-DNA), die „späten“ Gene (Strukturproteine der Phagenpartikel, sowie Enzyme für die DNA-Verpackung und für die Lyse der Wirtszellen) und eine Genregion zur Regulation der Lysogenie, die auch eine Erkennungssequenz zur Integration des Phagengenoms ins Wirtschromosom beinhaltet (*attP*). Die zugehörigen Integrationsstellen der Phagen A006, A500 und B025 in ihren jeweiligen Wirtsstämmen (*attB*) sind am 3'-Ende bestimmter tRNA-Gene lokalisiert. B054 hingegen integriert in das 3'-Ende des Gens für den Translationsfaktor EF-Ts.

Im Genom von P35 ist keine Lysogenie-Region vorhanden. Der Vorteil liegt offensichtlich in einem erweiterten Wirtsspektrum, da P35 dadurch auch zur Infektion lysogener Zellen mit homologen Prophagen befähigt ist.

Der virulente Phage A511 codiert alle Faktoren, die für eine wirtsunabhängige Replikation seines Genoms notwendig sind. Die 16 tRNA Gene repräsentieren überwiegend besonders häufige Codons und dienen möglicherweise zur Vermeidung von Engpässen bei der Synthese neuer Phagenproteine.

Die Analyse von Strukturproteinen mittels Massenspektrometrie ermöglichte die Zuordnung definierter Banden der Proteinprofile zu bioinformatisch vorhergesagten Peptiden. Außerdem lieferte sie in diesem Zusammenhang Hinweise auf translationale Leserasterverschiebungen bei der Synthese der Hauptkapsid- und Hauptschwanzproteine von A118 und A500.

Der direkte Vergleich der Phagen Genome und -proteome verdeutlichte die mosaikartige Zusammensetzung der Genomsequenzen und unterstützt somit die Theorie der „modularen Evolution“ der Phagen. Zudem deutet dies auf eine nahe Verwandtschaft der Phagen A006 und A500 mit A118 einerseits, sowie der Phagen B025 und PSA andererseits hin. Beide Gruppen zeigten auch Ähnlichkeiten zu verschiedenen Prophagen von *L. monocytogenes* und *L. innocua*. B054 ist nahezu identisch mit einem Prophagen von *L. innocua* Clip11262. Die virulenten Phagen A511 und P35 scheinen dagegen nur entfernt mit den temperenten verwandt zu sein. Zwar zeichnete sich A511 durch hohe Kolinearität zum Genom des virulenten *Listeria* Phagen P100 aus, besaß aber kaum Ähnlichkeiten zu den anderen untersuchten Phagen. Die vorhergesagten Proteine wiesen vor allem Ähnlichkeiten zu virulenten *Staphylococcus*-Phagen auf. Nur zwei Genprodukte von P35 zeigten schwache Ähnlichkeit zu einem Prophagen von *L. innocua*, während viele der „späten“ Gene einem Prophagen von *Enterococcus faecalis* ähnlich sind.

Insgesamt vermittelte die vergleichende Genomanalyse einen Einblick in die phylogenetische Verwandtschaft der bisher untersuchten *Listeria*-Phagen und stellt eine gute Grundlage für die anwendungsorientierte Forschung sowie weiterführende Untersuchungen zur Molekularbiologie der Bakteriophagen dar.

I. INTRODUCTION

1. The genus *Listeria*

1.1. Microbiology and taxonomy

Species of the genus *Listeria* are mesophilic, facultative anaerobic, non spore forming gram positive rods of 0.4 by 1 to 1.5 μm length. They are ubiquitously distributed in nature and have been isolated from soil, water, wastewater and plants, but can also be found in the gut of animals and humans (Weis and Seeliger, 1975; Bille and Doyle, 1991). The natural habitat of these bacteria is thought to be decomposing plant matter, in which they live as saprophytes. At growth temperatures of 10-25°C, *Listeria* are motile by a few peritrichous flagella whereas synthesis of flagellin is reduced at 37°C (Peel et al., 1988). *Listeria* are able to grow at a wide range of temperature (0°C to 50°C) and pH (5.5 to 9.5) and even in the presence of high concentration of sodium chloride of up to 10% (Junttila et al., 1988; Cole et al., 1990).

The genus *Listeria* belongs to the *Clostridium* subbranch, together with *Staphylococcus*, *Streptococcus*, *Lactobacillus* and *Brochothrix*. This phylogenetic position of *Listeria* is consistent with the low G+C content of the DNA of about 38%, the lack of mycolic acid and the presence of lipoteichoic acids (Bille and Doyle, 1991). Since 1992, the genus *Listeria* includes the six species *L. monocytogenes*, *L. innocua*, *L. ivanovii*, *L. seeligeri*, *L. grayi* and *L. welshimeri*. Two of them, *L. monocytogenes* and *L. ivanovii*, are facultative intracellular pathogens. The infectious disease caused by these bacteria is named listeriosis.

1.2. Pathogenesis of listeriosis

Only a few years after the discovery of *L. monocytogenes* in laboratory animals in 1924, it became apparent that the disease also affects humans. *L. monocytogenes* is the major pathogenic species causing listeriosis in both animals and human, whereas only few cases of human infections caused by *L. ivanovii* have been reported (Cummins et al., 1994; McLauchlin et al., 2004).

Since the organisms are widely disseminated in the rural environment, they often contaminate raw materials used for industrially processed foods. Due to their ability to grow under refrigeration temperatures and to tolerate low pH and high salt concentrations, *Listeria* are well equipped to survive food-processing technologies and storage conditions and therefore

can become endemic in food processing environments. The foods most frequently implicated are soft cheese and dairy products from raw milk, raw meat, sausages, smoked fish, seafood and salads (Pini and Gilbert, 1988; Trussel, 1989; Farber and Peterkin, 1991; Ben Embarek, 1994; Dalton et al., 1997).

L. monocytogenes is the causative agent of serious epidemic and sporadic listeriosis in human. The incidence of this illness is very low, usually about two to eight sporadic cases annually per million population in Europe and the United States (Farber and Peterkin, 1991; Vazquez-Boland et al., 2001b). However, the risk of infection is markedly increased among immunocompromised patients, newborns, pregnant women and elderly people, and is associated with a mortality rate of about 20 to 30% or even higher, despite antibiotic treatment (McLauchlin, 1990a; McLauchlin, 1990b; Schuchat et al, 1991). During the course of food-borne infections, the bacteria are capable to penetrate the mucosa of the intestine. Once in the bloodstream, the organisms spread to liver and spleen and may also cross the blood-brain barrier and reach brain tissue. The listerial infection most frequently reported in non-pregnant adults affects the central nervous system (CNS), concerning 55-70% of all cases. The infection normally develops a meningoencephalitis accompanied by severe changes in consciousness and movement disorders (Vazquez-Boland et al., 2001b). But *L. monocytogenes* also can cause a number of other clinical syndromes, including sepsis, gastroenteritis and local infections of liver, spleen or endocardium (Dalton et al., 1997; Lorber, 1997; Vazquez-Boland et al., 2001b).

In pregnant women the infections are usually asymptomatic in the mother or may appear as flu-like symptoms. However, *Listeria* may gain access to the fetus by penetration of the placental barrier causing abortion or fetal death (Lorber, 1997; Vazquez-Boland et al., 2001b).

1.3. Intracellular life cycle of *L. monocytogenes*

L. monocytogenes is able to enter eukaryotic cells, proliferates in the cytoplasm and spreads from one host cell to another without any contact to the extracellular matrix, thus escaping attacks by the host immune system. The intracellular life cycle of *L. monocytogenes* is displayed in Figure 1.

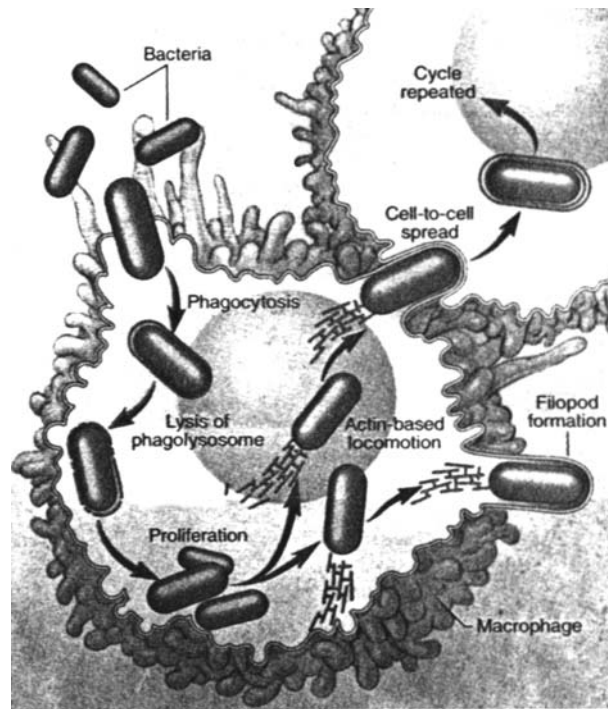


Fig. 1. Life cycle of *L. monocytogenes* in host cells (Southwick and Purich, 1996).

Infection occurs via contaminated food in the intestines where the bacteria adhere to epithelial cells and induce their own phagocytosis. Following absorption into the host cell, the bacteria are enclosed in a vacuole called phagolysosome (Cossart and Lecuit, 1998). By expression of phospholipases (PlcA and PlcB) and the pore-forming toxin listeriolysin O (LLO) *Listeria* escape from the phagolysosome and enter the cytoplasm where they replicate with a generation time of one hour. To invade and infect adjacent host cells, *Listeria* has evolved an efficient mechanism that couples actin polymerisation to intracellular movement. The protein ActA induces assembly of actin filaments, which rearrange into structures resembling comet tails at one distal end of the bacterial cells and allow the pathogens to migrate nondirectionally (Domann et al., 1992). The motile bacteria eventually reach the cytoplasmic membrane and form filopodia or protrusions that can be ingested by an adjacent cell, an event that is called cell-to-cell spread. When the pathogen escapes from the secondary double membrane vacuole, the proliferation cycle begins anew.

As in other pathogenic bacteria, most of the virulence genes of *Listeria* are organised into discrete genetic units known as pathogenicity islands. The virulence factors mentioned above

are responsible for key steps of *L. monocytogenes* intracellular parasitism and are physically linked together with other virulence genes in a 9-kb chromosomal island referred to as LIPI 1 (for *Listeria* pathogenicity island 1). Of course, pathogenesis of listeriosis is complex and requires many other important genes and factors that contribute to the virulence of *L. monocytogenes* during the passage through the hostile environment of the gut and the diverse steps of infection. Detailed characterisation and information on *Listeria* pathogenesis, virulence factors and regulation of virulence gene expression are described elsewhere (Cossart and Lecuit, 1998; Vazquez-Boland et al., 2001a; Cossart, 2002; Dussurget et al., 2002; Sleator et al., 2005).

2. Bacteriophages

Bacteriophages (also known as phages) are viruses that infect bacterial cells. For effective multiplication they depend on the energy production and biosynthetic activity of their host bacteria. Therefore, they can be considered as obligate intracellular parasites (Klaus et al., 1992).

2.1. Historical sketch

Bacteriophages were discovered twice, in 1915 by F.W. Twort and in 1917 by F.D. d’Herelle. Twort observed some colonies of *Staphylococcus* becoming transparent after long times of incubation. Even after filtration the material was still potent in changing colony morphology, when it was passed on normal cultures of *Staphylococcus*. This phenomenon came to be known as “glassy transformation”. Whereas Twort was uncertain about the causative agent of his discovery, F. D’Herelle described plaques on *Shigella* cultures, local regions of bacterial cell lysis induced by viruses. He claimed an invisible “microbe” as origin of these plaques, an obligate bacteriophage (deduced from the Greek word for “bacteria eater”) (Duckworth, 1976). D’Herelle’s publication initiated great interest in phage research driven by the hope to use these bacterial killers in the treatment of bacterial infections. Although the advent of antibiotics in the 1930s resulted in decline of interest in phage therapy (at least in western medical research), bacteriophages became the main model system in molecular biology. Phage studies have made historical contributions to molecular biology. The discovery of DNA as the hereditary molecule by Hershey and Chase (1952) was the basis for this new field of research. Bacteriophage research is now undergoing a renaissance, in which the focus

moves from examination of molecular mechanisms to phage biology itself. The revival of interest is particularly due to the recognition of the importance of phages in bacterial pathogenesis and evolution, their role in natural ecosystems and considering the antibiotic-resistance crisis their potential in treating bacterial infections (Campbell, 2003b; McGrath et al., 2004).

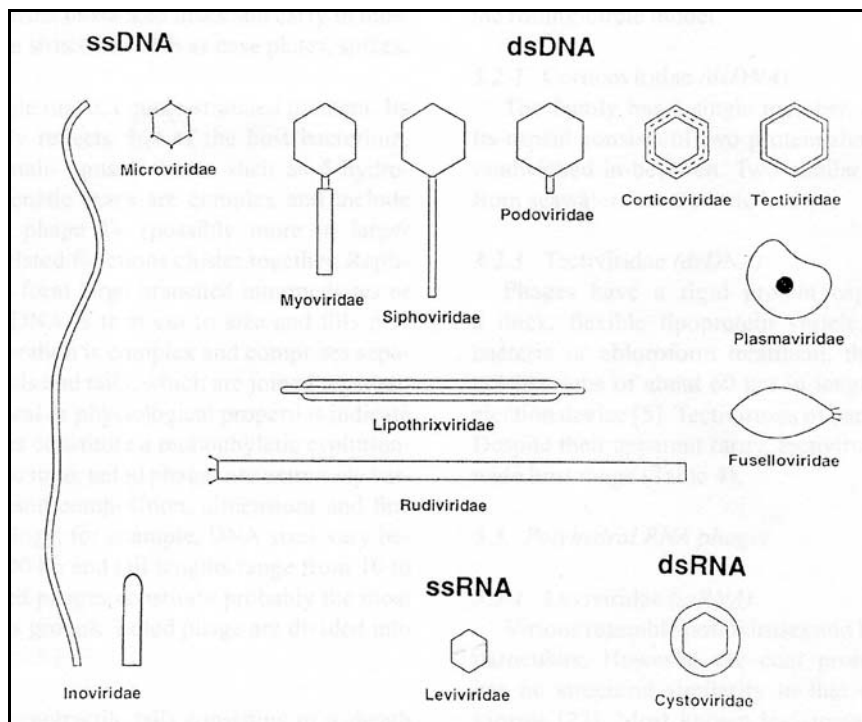


Fig. 2. Classification of bacteriophages according to the ICTV. Schematic representation of the major phage groups and families (Ackermann, 2003).

2.2. Phage taxonomy

Currently, the International Committee on Taxonomy of Viruses (ICTV) classifies bacteriophages into 13 families according to their type of nucleic acid (double stranded or single stranded RNA or DNA) and morphological features (Ackermann, 2003), (www.virustaxonomyonline.com). The largest group of bacteriophages (about 96%) belongs to the order of *Caudovirales*, tailed phages with isometric capsid and dsDNA, which are subdivided into three large, phylogenetically related families according to tail morphology:

Myoviridae have long contractile tails, *Siphoviridae* possess long non-contractile tails, which are often flexible, and *Podoviridae* feature short tails. The minor part of bacteriophages is polyhedral, filamentous or pleomorphic and a few types have lipid containing envelopes. A schematic representation of the major phage groups is given in Figure 2.

2.3. Bacteriophage proliferation

Viruses have evolved several mechanisms for proliferation in their host cell. Their life cycles can be lytic, lysogenic or chronic infections. A chronic infection occurs, when progeny virions are constantly released from the host by protrusion without destroying the cell. This is the case for some filamentous phages (Klaus et al., 1992; Ackermann, 2003). But the majority of bacteriophages separate in either of lytic or lysogenic life-cycle. Virulent phages are only able to proliferate in a lytic life-cycle, in which the host cells are destroyed. Temperate phages may additionally switch into a lysogenic life cycle, in which the phage genome integrates and replicates together with the host genome.

2.3.1. Lytic life cycle

Four main steps of a life cycle are common to all bacteriophages: adsorption, injection of nucleic acid, replication and virion assembly, and phage release.

Bacteriophages meet their hosts passively by incidence. After an initial contact, the phage reversibly adsorbs to a defined cell surface structure. Then, irreversible binding occurs between the phage and a specific receptor on the host cell, such as cell wall components or outer membrane proteins. Following penetration of the cell wall, the phage releases its genome into the host cell, whereas the capsid remains outside the cell. Transcription and replication of the phage genome occurs via biosynthetic machinery of the host. New phage particles assemble in the host cytoplasm. The progeny virions are released after maturation by sudden lysis of the bacteria, which is effected by the disruption of the cell envelope. In the tailed phages, lysis is accomplished by a dual lysis system consisting of holins, which perforate the plasma membrane, and endolysins, which attack the peptidoglycan of the cell wall (Wang et al., 2000; Young et al., 2000).

2.3.2. Lysogenic life cycle

While the lytic cycle is the only mode of reproduction for virulent phages, the temperate phages have an alternative life cycle, in which a fraction of the infected host cells survive the infection and harbour the phage genome in a quiescent form. In this state, the phage genome usually is integrated into the host chromosome as a prophage and replicates along with the bacterial DNA during cell division, thus becoming transmitted to all daughter cells. Some prophages also replicate autonomously as a plasmid (Klaus et al., 1992).

Host cells that harbour a prophage are known as lysogens and are resistant against superinfection by the same phage (homoimmunity) but not by heterologous phages. It is possible for a cell to carry more than one prophage, a state that is termed polylysogeny. Only a few gene products with regulatory functions are expressed during prophage status. They inhibit transcription of lytic functions of the residing prophage and can also act on newly arrived DNA thereby conferring immunity against any superinfecting phage to which the repressor may bind (Klaus et al., 1992; Birge, 1994).

Although very stable, the lysogenic state is not permanent. It is possible for the prophage to revert to the vegetative state and to enter the lytic life cycle at different phage-specific rates. DNA damaging agents such as UV radiation or Mitomicin C can induce lysogenic strains to re-enter the lytic cycle (Loessner et al., 1991).

2.4. *Listeria* phages

2.4.1. General information

Since the initial discovery of phages specific for *Listeria* species in 1945 (Schultz, 1945), about 400 *Listeria* phages have been isolated from environmental sources or lysogenic strains. Examinations of more than 120 *Listeria* phages by electron microscopy revealed two different morphotypes of tailed phages in the order of *Caudovirales*. The majority represented morphotype B1 with isometric capsids and long non-contractile tails, which are often flexible (Rocourt, 1986; Zink and Loessner, 1992; Loessner et al., 1994a; Loessner et al., 1994b). They belong to the *Siphoviridae* family. The five species representing this family are distinguished between each other on the basis of their tail lengths. They differ also in their host ranges that tend to be rather limited. Only a few phages described so far feature contractile tails and isometric capsids (morphotype A1), and thus belong to the *Myoviridae* family (Rocourt, 1986; Loessner, 1994a; Zink et al., 1995).

Listeria bacteriophages are strictly genus specific. Temperate phages show a quite narrow host range and recognize bacteria of individual serovar groups (Loessner and Busse, 1990; Loessner, 1991; van der Mee-Marquet et al., 1997). Here, teichoic acid associated N-acetylglucosamine and rhamnose or galactose have been identified as phage receptors in *L. monocytogenes* serotype 1/2 and serotype 4 strains, respectively (Wendlinger et al., 1996; Tran et al., 1999). In contrast, virulent phages like A511 can infect several species and serovars and are therefore supposed to use serovar-independent surface structures, e.g. the listerial peptidoglycan as primary receptor (Wendlinger et al., 1996).

Many *Listeria* phages have been investigated by electron microscopy, and their structural proteins were used for protein profiling (Loessner et al., 1994a; Loessner et al., 1994b; Zink and Loessner, 1992), but there has been little information about *Listeria* phages on genome level. Up to now, only two genome sequences of the temperate bacteriophages PSA and A118 have been sequenced and analysed, and revealed interesting data concerning overall genome organisation, integration sites, DNA packaging mechanism and translational frame-shifting in synthesis of structural proteins (Loessner et al., 2000; Zimmer et al., 2003). Moreover, a 10kb portion of the structural genes of A511 has been determined, encompassing the *cps* (capsid) and *tsh* (major tail protein) genes (Loessner and Scherer, 1995). The latest database entry concerning *Listeria* phages represents the complete genome sequence of the virulent Myovirus P100 that was published very recently and exhibited high similarity to A511 (Carlton et al., 2005).

2.4.2. *Listeria* phage applications

Phage research not only provides an insight into phage biology and its molecular mechanisms, but also enables the development of applications for the study, manipulation, detection or elimination of their bacterial hosts. Some interesting examples for the use of *Listeria* bacteriophages or their components in research laboratories and food industry are described below.

Phage typing

The ability to differentiate bacteria beyond species level is essential for identifying and tracking outbreaks of infectious diseases. Cases of human listeriosis are mainly caused by only three of the twelve known serovars of *L. monocytogenes*, 4b, 1/2a and 1/2b, which account for more than 90% of listerial infections (Vazquez-Boland et al., 2001b). *Listeria*

phages are useful for further subtyping of *Listeria* strains in epidemiological investigations concerning outbreaks of listeriosis. The application benefits from the different host spectra of a set of bacteriophages, resulting in distinct lysis patterns of the strains investigated (Loessner and Busse, 1990; Loessner, 1991; Lebek et al., 1993; Zink et al., 1994). According to a study on evaluation of an improved phage set for *Listeria* typing, about 90% of all the strains tested were typable (van der Mee-Marquet et al., 1997).

Detection of viable *Listeria* cells

Due to their genus specificity, bacteriophages represent appropriate agents for the detection of *Listeria* contaminants in foods. The construction of a reporter phage A511::luxAB that expresses bacterial luciferase genes during infection of *Listeria*-cells facilitates the detection of the infected bacteria via measuring of emitted bioluminescence (Loessner et al., 1996). The utilisation of this reporter phage in a simple, inexpensive and reliable assay was proven for large scale screening of *L. monocytogenes* in foods and environmental samples (Loessner et al., 1997).

Bio-disinfection and bio-control

Listeria phages were studied with respect to bio-disinfection and bio-control of *L. monocytogenes* during food processing and storage. For that purpose, lytic bacteriophages were applied on vegetables, fruits and high-risk-foods like red-smear cheese (Leverentz et al., 2003; Carlton et al., 2005). Significant reduction of bacterial populations by two to five log units or even complete eradication of viable bacteria cells were observed using phage particles only, or in combination with a bacteriocin (Leverentz et al., 2001; Carlton et al., 2005). Oral toxicity studies in rats gave no hints for any potential risk of bacteriophages used as food additives (Carlton et al., 2005).

An approach for biological control of *L. monocytogenes* in diary products, e.g. during a cheese ripening process, could be achieved by the application of lactic acid bacteria expressing recombinant phage endolysin. Because of the stringent substrate specificity, the *Listeria* phage endolysins only hydrolyse listerial cell walls. Gaeng and co-workers (2000) introduced the phage lysin gene *plyA511* fused to an N-terminal signal peptide into the starter organism *Lactococcus lactis*, which can be employed for the fermentation of milk. Secretion of active enzyme into the surrounding media was indicated by efficient lysis of an indicator strain.

Molecular biology and biotechnology

Biotechnological applications are rather based on single phage components and utilise bacteriophages as molecular tool boxes. For example, the purified recombinant phage endolysins can be applied for rapid lysis of bacterial cells that facilitates preparation of genomic DNA from *Listeria* (Loessner et al., 1995b).

Another interesting possibility is the use of recombinant C-terminal cell wall binding domains (CBD) of phage endolysins. Fusions of CBDs with fluorescent labels such as GFP allow specific detection of *Listeria* cells in mixed bacterial populations (Loessner et al., 2002). Due to their high affinity, specific binding capacity, CBD polypeptides can also be used for the immobilisation of host cells to solid surfaces. For example, magnetic beads coated with CBDs provide the opportunity to develop useful applications, such as recovering *Listeria* cells from contaminated foods (Loessner, 2005).

Dietrich and co-workers constructed a suicide *L. monocytogenes* strain that was engineered to lyse upon entry into the host cell due to the production of a phage lysin (Dietrich et al., 1998). Thus, the intracellular pathogens were used for *in vitro* transfer of plasmid DNA into mammalian phagocytic and non-phagocytic cells.

Site specific integration vectors facilitate genetic manipulation and systematic strain construction in by inserting at precise locations in the bacterial chromosome. For this purpose, Lauer and co-workers cloned two distinct integrase genes and their associated phage integration sites into low-copy-number plasmids (Lauer et al., 2002). The shuttle vectors replicate in *E. coli* and can be transferred into *L. monocytogenes*.

3. Impact of bacteriophage genomics

Apart from the development of molecular applications for research and industry, bacteriophage genomics also deliver insight into the great genetic diversity of the whole phage population and contribute to a better understanding of evolutionary processes. The viruses are associated in a very ancient group and are supposed to have evolved before the separation of the three contemporary domains of life (Hendrix et al., 2000; Hendrix, 2002). However, the remarkable diversity of their genomes makes it difficult to identify long range evolutionary relationships. According to the current hypothesis, dsDNA tailed phage genomes are mosaically related to each other and gain access to a common gene pool (Hendrix et al., 2003). Genetic reassortment of phage DNA take place by horizontal exchange of sequences

via homologous and non-homologous recombination, supposedly resulting in a web-like phylogenetic tree. Though, considering a global phage population of about 10^{30} particles (Wommack and Colwell, 2000), the number of about 300 complete genome sequences in the databases is still too low to allow an estimation of the true dimensions of phage evolution.

One of the most interesting and important aspects of bacteriophage genomics is probably how phages contribute to bacterial population dynamics and bacterial virulence. In fact, a great number of phages encoding virulence factors were discovered long ago, e.g. the neurotoxin of *Clostridium botulinum*, the diphtheria toxin and the cholera toxin (Brussow et al., 2004). Bacteriophages play a critical role in bacterial biology, diversity and evolution. Proceeding genome research revealed high incidence of phage related sequences in bacterial genomes, indicating that bacterial and phage genomes were coevolving. Phages are a major source of horizontally transferred DNA in bacteria, and prophage DNA accounts in many bacteria for important inter-strain genetic variability (Brussow and Hendrix, 2002).

Lysogeny is widespread among *Listeria*. The complete genome sequences of *L. innocua* and four *L. monocytogenes* strains including food isolates associated with human listeriosis revealed the presence of monocins and several cryptic and functional prophages that constitute up to 7.6% of the bacterial genomes (Glaser et al., 2001; Nelson et al, 2004). Although general synteny was demonstrated for the genomes of *L. monocytogenes* strains, a significant portion of sequence variability consisted in phage insertions. Investigations about prophage contributions to population dynamics of *Salmonella* cocultures suggest that prophages can improve the competitive fitness of the lysogenic host strain due to spontaneous phage release that destroys bacteria from rival strains (Bossi et al., 2003). Thus, this kind of selection pressure also results in diversification and generation of new strains by lysogenic conversion. In case of *Listeria*, the potential influence of prophages on their host strains, like phenotypic variation or provision of selective benefits, has not been elucidated so far. To gain insight in bacteriophage-host interactions as well as molecular phage biology much more information is needed concerning *Listeria* phage genomes.

4. Aims of this work

In order to gather essential information on *Listeria* phages on molecular level, the major aim of this work was to sequence and computationally analyse in detail the genomes of six

Listeria bacteriophages. These were bacteriophages A006, A500, A511, B025, B054, and P35, which differ significantly in morphology and host range as described below.

Listeria phages A500 and A006 were both classified as the same species 2671 of the *Siphoviridae* family. A006 infects *L. monocytogenes* serovar 1/2 strains, while A500 only proliferates on a few serovar 4b strains (Loessner and Busse, 1990; Loessner, 1991). Analysis of the structural proteins of A500 and A006 revealed differences in virion morphology (Loessner et al., 1994b). On the basis of these data as well as DNA hybridisation patterns, A006 was proposed to represent a new species.

B025 and B054 are both phages of the non-pathogenic species *L. innocua*, but are members of different phage families. In contrast to B025 featuring a non-contractile tail, B054 is a Myovirus that revealed an unusual tail contraction mechanism away from the capsid towards the base plate (Zink and Loessner, 1992).

P35 is a temperate-sensitive *L. monocytogenes* phage isolated from silage and is only capable to grow at room temperature (Hodgson, 2000). It belongs to the *Siphoviridae*, and was proposed as a new species. In contrast to other temperate bacteriophages, P35 featured a broader host range and proliferated in 75% of all serovar 1/2 strains tested by Hodgson.

A511 is one of the few virulent *Listeria* phages that have been isolated so far. It belongs to the family of *Myoviridae* and has a broad host range infecting about 95% of all *L. monocytogenes* strains tested (Loessner, 1991; Zink and Loessner, 1992).

In addition to the prediction of open reading frames (ORFs) and the screening for protein homologies in the databases, profiling of the structural proteins and analysis by peptide mass fingerprinting should be performed. In consideration of the potential influence of prophages on the host bacteria and the utility in development of molecular tools, bacterial and phage attachment sites were supposed to be determined for the temperate phages.

Comparative analyses of the nucleotide sequences and the predicted gene products were supposed to deliver insight into the degree of relatedness between the *Listeria* phages and were thought to be interesting regarding phage evolution.

II. MATERIALS & METHODS

1. Materials

1.1. Strains

All *Listeria* strains were taken from the Weihenstephan *Listeria* Strain Collection (WSLC) and were used for phage propagation and as indicator strains or for the generation of lysogenised host strains. *E. coli* strains were used for cloning of the genomic libraries of *Listeria* phages. Detailed information on the genotypes of the *E. coli* strains is provided by the suppliers. The strains were kept on agar plates at 4°C and were streaked on fresh media every six weeks.

Species	Type strain	Source/strain
<i>Listeria monocytogenes</i> sv 1/2b	ATCC 19112	WSLC 1001
<i>Listeria monocytogenes</i> sv 4b	ATCC 23074	WSLC 1042
<i>Listeria ivanovii</i> sv 5	SLCC 4769	WSLC 3009
<i>Escherichia coli</i> DH5 α MCR		Invitrogen
<i>Escherichia coli</i> XL1 Blue MRF ⁺		Stratagene
<i>Escherichia coli</i> TOP10		Invitrogen

1.2. Bacteriophages

Stocks of *Listeria* bacteriophages (in SM buffer) were stored at 4°C.

Phage	Species	Morphotype	Source	Reference	Propagating strain
A006	2671	Siphovirus	lysogenic strain WSLC 1006	(Loessner, 1991)	WSLC 1001
A500	2671	Siphovirus	ATCC 23074	(Loessner and Busse, 1990)	WSLC 1042
A511	A511	Myovirus	Sewage	(Loessner and Busse, 1990)	WSLC 3009
B025	2685	Siphovirus	lysogenic strain WSLC 2025	(Loessner and Busse, 1990)	WSLC 3009
B054	4286	Myovirus	lysogenic strain WSLC 2054	(Zink and Loessner, 1992)	WSLC 3009
P35	P35	Siphovirus	Silage	(Hodgson, 2000)	WSLC 1001

1.3. Plasmids

The genomic library of bacteriophage B054 was cloned into the vector pBluescript II SK (+/-) and transformed into the strains *E. coli* DH5 α MCR and *E. coli* XL1 Blue MRF', respectively. The vector pCR 4Blunt-TOPO in combination with the strain *E. coli* TOP10 was used for cloning the genomic library of phage B025.

pBluescript II SK (+/-)	3.0 kb cloning vector, Amp ^r (Stratagene)
pCR 4Blunt-TOPO	3.9 kb cloning vector, Amp ^r (Invitrogen)

1.4. Media

All media were prepared using demineralised water, and pH values were adjusted before autoclaving. *Listeria* strains were cultivated on Tryptose media (Merck KGaA, Darmstadt, Germany) at 30°C. Luria Bertani (LB) was used for the cultivation of *E. coli* at 37°C. Recombinant strains carrying plasmids were kept on selective media containing ampicillin at a final concentration of 100 μ g/ml.

BHI (Brain heart infusion) pH 7.3 – 7.4

BHI (Merck, KGaA, Darmstadt, Germany) was used half strength of the concentration recommended by the manufacturer to avoid browning of the media due to the Maillard reaction during autoclaving.

LB (Luria Bertani) broth and agar pH 7.3 – 7.4

Tryptone	10 g/l
Yeast extract	5 g/l
NaCl	8 g/l

To prepare solid media, 1.4% (w/v) agar were added.

Tryptose broth and agar pH 7.3 – 7.4

Tryptose broth (Merck KGaA, Darmstadt, Germany) was prepared according to the manufacturer's instructions. For the preparation of plates, 1.4% (w/v) agar were added.

After autoclaving, the media were supplemented with 1.25 mM CaCl₂ of a sterile stock solution (1 M). The addition of Ca⁺⁺ promotes binding of bacteriophages to the bacterial cell wall.

1.5. Buffers

All buffers or solutions were prepared using MilliQ water.

Lysisbuffer, pH 8.0

Tris/HCl	10 mM
EDTA	1 mM
SDS	1 % (w/v)

Resuspension buffer, pH 7.0

Sucrose	20 % (w/v)
NaPO ₄	10 mM

SM buffer, pH 7.4

Tris/HCl pH 7.4	50 mM
NaCl	100 mM
MgSO ₄	9 mM

TAE 50x, pH 8.0

Tris	2 M
EDTA	50 mM
Acetic acid	1 M

TE buffer, pH 7.5

Tris/HCl, pH 8.0	10 mM
EDTA, pH 8.0	1 mM

1.6. Enzymes

Enzyme	Source
Lysozyme	Sigma
Proteinase K	Roth
Restriction endonucleases	MBI Fermentas, New England Biolabs
Ribonuclease A	Sigma
Shrimp alkaline phosphatase (SAP)	MBI Fermentas
Taq polymerase	Qiagen
T4-ligase	Roche

1.7. Kits

Kits	Manufacturer
QIAquick Gel Extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
GenElute™ Plasmid Miniprep	Sigma-Aldrich
TOPO® shotgun subcloning kit	Invitrogen

2. Methods

2.1. Phage propagation

Bacteriophages were propagated using the soft agar-layer technique (plating method) or liquid culture method slightly modified according to standard protocols described for phage λ (Sambrook and Russel, 2001). Phage infections were incubated at 30°C except for infections with P35, which were performed at room temperature (RT) due to the temperature sensitivity of this phage.

Phage propagation on agar plates

For the plating method, dilutions of phage stock solutions were added to 4 ml of molten soft agar (40°C) inoculated with 150 μ l of log-phase culture of the propagating *Listeria* strain. The mixture was poured onto Tryptose agar plates and incubated over night at 30°C or at RT

(using P35), respectively. Confluent lysed plates only were used to generate high titre phage stocks. Soft agar layers were covered with 9 ml of SM buffer, scraped off the plates and transferred into sterile centrifuge tubes (Beckmann, JA-14, 250 ml). Incubation for 2 hours at 4°C allowed the phages to diffuse from agar into the buffer. Afterwards the solution was centrifuged (20 min, 8,000 rpm) and sterile filtered.

Phage propagation in liquid culture

Liquid culture-method was used for large scale propagation of virulent bacteriophages A511 and P35. 600 ml of 0.5x BHI broth were inoculated with host strain from an over night culture to a final concentration of 5×10^7 cells/ml. Then phages were added to a final concentration of approximately 10^5 pfu/ml. The mixture was incubated at 30°C (or RT in case of P35) for 2-3 hours until the optical density reached at least $OD_{600} = 0.1$. Additional phages were added to a final concentration of 5×10^7 pfu/ml, and the culture was incubated for another 3 hours at 30°C or over night at RT until lysis of the cells was clearly visible and optical density dropped towards $OD_{600} \approx 0$. To remove cellular debris and prevent growth of phage resistant cells the culture was centrifuged (15 min, 6,000 x g, 4°C) and the supernatant was subsequently sterile filtered.

2.2. Phage titres

Phage titres quantified in plaque forming units (pfu/ml) were determined using the soft agar layer technique (see above). Aliquots of host strains in molten soft agar were infected with 100 µl of serial tenfold dilutions of the phage stock, poured on plates and incubated over night at 30°C or RT (infections with P35).

2.3. Phage purification

Phages from high titre stock solutions were precipitated using standard protocols for purification of phage λ (Sambrook and Russel, 2001) by adding 1 M NaCl and 10% (w/v) PEG 8000. After an overnight incubation at 4°C with slight stirring during the first 2 hours of incubation, phages were collected by centrifugation (10,000 x g, 20 min, 4°C). Pellets were carefully resuspended in 5 ml SM buffer and further purified using CsCl density gradient centrifugation (densities: 1.2, 1.3, 1.4, 1.45, 1.5 and 1.7 g/cm³) at 85,000 x g for 16 hours at 4°C. Intact phage particles accumulated in bluish band at the interface between 1.45 g/cm³

and 1.5 g/cm³ CsCl layers and were removed with a syringe and a hypodermic needle (0.9 x 4.0 mm) by puncturing the side of the tube. CsCl was removed from purified phage particles by dialysis in flexible, autoclaved cellulose ester tubes against 2000 ml SM buffer for 2 hours and a second dialysis over night at 4°C.

2.4. DNA extraction from bacteriophages

Putative RNA contaminations from bacterial hosts were degraded by addition of 10 µg RNaseA per ml purified phage solution and incubation at RT for 15 minutes. Phage proteins were then degraded by incubation at 50°C (1 hour) after addition of 20 mM EDTA, proteinase K (50 µg/ml) and 0.5% (w/v) SDS (final concentrations). DNA was extracted successively with phenol, phenol/chloroform and chloroform according to standard protocols, and finally ethanol precipitated.

2.5. DNA extraction from *Listeria*

Cells from 10 ml overnight culture were harvested by centrifugation and resuspended in 2 ml resuspension buffer. After addition of 100 µg/ml lysozyme, cells were incubated for 1 hour at 37°C (gentle shaking a few times). Eight ml lysis buffer and 1 mg proteinase K were added before incubating the mixture for 2 hours at 37°C or 1 hour at 50°C, respectively. DNA was extracted according to standard protocols using Phenol, Phenol/Chloroform and Chloroform.

2.6. DNA precipitation

Samples of DNA were precipitated by addition of 0.3 M sodium acetate (pH 5.2) and 2-2.5 volumes of 100% EtOH. The mixture was incubated at RT for 30 min. DNA was then pelleted by centrifugation (13,000 x g at 4°C), washed twice with 70% EtOH and then left to air-dry before it was resuspended in ½ TE buffer. Alternatively, DNA was precipitated by adding 0.8 volume of isopropanol.

2.7. Agarose gel electrophoresis

DNA fragments were separated by agarose gel electrophoresis in gels containing 0.8% agarose. The agarose was dissolved in 1x TAE by heating in a microwave. The hot agarose

solution was poured into a gel tray with slot formers (Perfect Blue MiniS or MiniM, PeQlab). DNA samples were mixed with 1/5 volume of 6x loading buffer (MBI Fermentas) and pipetted into the slots. GeneRuler™ DNA Ladder Mix or GeneRuler™ 1 kb DNA ladder (MBI Fermentas) were used as molecular weight markers. Electrophoresis was performed at 70-90 V. Afterwards gels were stained with EtBr (0.5 µg/ml MilliQ water) for 15-30 min. Bands were visualized by UV illumination using a Kodak Image Station (Kodak).

Table 1. Fragment sizes of the molecular marker used in agarose gel electrophoresis.

Marker	Fragment sizes [bp]
DNA Ladder Mix (MBI Fermentas)	100, 200, 300, 400, 500, 600, 700, 800, 900, 1031, 1200, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000, 10000
1 kb DNA Ladder (MBI Fermentas)	250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000, 10000

2.8. Recover DNA fragments from agarose gels

Isolation of DNA fragments from agarose gels was performed using the QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions. DNA bands were excised from the gel, and the agarose was dissolved with the help of a NaJO₄ containing buffer. The purified DNA was eluted from spin columns with TE buffer.

2.9. Polymerase chain reaction

Polymerase chain reaction (PCR) is a method for amplification of DNA fragments *in vitro* (Saiki et al., 1988). Oligonucleotides used for PCR and sequencing reactions were obtained from MWG Biotech (Ebersberg, Germany). PCR amplifications were performed with Primus Advanced Thermocycler (Peqlab, Erlangen, Germany). Reactions were carried out in 100 µl volumes within 0.2 ml tubes containing template DNA, oligonucleotide primers, a mix of deoxyribonucleotide triphosphate (dATP, dCTP, dGTP, dTTP), reaction buffer and DNA polymerase (Qiagen). The standard PCR protocol initiates with a denaturing step at 94°C for 5 min or 15 min respectively, when a hotstart polymerase was used. Amplification of the PCR product was performed in 30-35 cycles followed by a final extension step. The annealing temperature (52°C-60°C) and the extension time were adjusted to the melting temperature T_m

of the primers and the size of the expected PCR product, respectively. The standard PCR reaction mix and protocol used are summarized below.

PCR reaction mix

Template DNA	20-50 ng
PCR reaction buffer (10x)	10 μ l
dNTP (5 mM)	2-5 μ l
Primers (10 μ M) each	5 μ l
Taq DNA polymerase	1.25 U
Sterile MilliQ water	ad 100 μ l

PCR standard protocol

94°C 5 min
[52°C 30s
72°C 1 min
94°C 30s] x 35
52°C 2 min
72°C 5 min

2.10. Purification of PCR products

PCR products were purified before sequencing using spin columns from the QIAquick PCR Purification Kit (Qiagen) according to the instructions of the manufacturer. The DNA fragments were eluted using TE buffer.

2.11. Ligation

Ligations of DNA fragments were performed adding 1 unit of T4 Ligase (Roche) and 1/10 volumes of 10x reaction buffer. The mixture was then incubated at 15°C over night or at RT for 4 hours.

2.12. Electrotransformation

Transformation of *E. coli* was performed by electroporation as described by Dower and co-workers using a Biorad Gene Pulser and sterile electroporation cuvettes (Dower et al., 1988).

Preparation of competent cells

For the preparation of competent cells, 500 ml LB were inoculated with 1 ml of an overnight culture of *E. coli* and incubated at 37°C shaking at 150 rpm until the optical density reached $OD_{600} = 0.5 - 0.6$. Cells were harvested by centrifugation (8,000 x g, 4°C; 10 min) and the pellets were stored on ice. Following three washing steps with 500 ml 5% (v/v) glycerol (4°C), 250 ml 5% (v/v) glycerol (4°C) and 10 ml 10% (v/v) glycerol (4°C), the pellets were resuspended in 1 ml 10% glycerol (4°C) and stored at -70°C in aliquots of 40 µl.

Electroporation

Ligation reactions or plasmids were dialysed against MilliQ water on nitrocellulose membranes (0.025µm pore size, MILLIPORE) to remove ions for a low conductivity. One aliquot of competent cells (40 µl) and 10-50 ng plasmid were pipetted into an electroporation cuvette on ice and electroporated under following conditions: resistance 200 Ω, capacity 25 µFD, voltage 2.5 kV. Subsequently, 960 µl of prewarmed LB medium were added directly to the cells, which were then incubated for regeneration at 37°C for 1 hour. 100 µl aliquots of the sample and its tenfold dilution were plated on LB agar supplemented with ampicillin and incubated at 37°C for 16 hours.

2.13. Cloning and nucleotide sequencing

Phage genomic DNA was extracted and purified as described above (section 2.4). The construction of the genomic library of B054 was performed essentially as described previously (Loessner et al., 2000). In short, limited digestion with restriction enzyme Tsp509 I (New England Biolabs) was performed and fragments of 1-2 kb length were ligated into the vector pBluescript II SK (+/-) and transformed into *E. coli* DH5α MCR. Clones containing plasmid with inserts were identified by blue-white screening on LB supplemented with X-gal and ampicillin. Plasmids purified from small scale cultures (GenElute Plasmid Miniprep, Sigma-Aldrich) were digested with *PauI* and 45 clones carrying inserts of various sizes were selected for sequencing.

The genomic library of B025 was constructed using TOPO[®] shotgun subcloning kit (Invitrogen) according to the instructions of the manufacturer. Genomic DNA was sheared by sonication and fragments of 1-2 kb were blunt-end repaired with a mixture of T4 DNA Polymerase and Klenow DNA Polymerase before cloning into the vector pCR[®] 4Blunt-TOPO[®] and transformation into *E. coli* TOP10. Again plasmid-bearing cells were identified by blue-white screening and plasmids were selected for sequencing after digestion with *EcoRI*.

For the construction of the phage DNA libraries of A006, A500, A511 and P35, DNA was mechanically sheared using a Nebulizer (GATC Biotech AG). Fragments of the desired size (1-1.5 kb) were recovered from agarose gels and ligated blunt end into pBluescript II SK. Ligation products were transformed into *E. coli* DH5 α MCR via electroporation. Blue-white screening on Xgal-containing agar plates was used to identify plasmid-bearing clones.

Genomic libraries of all bacteriophages were sequenced by GATC Biotech AG. The obtained nucleotide sequences were edited and aligned into contigs using the software Vector NTI Suite 8 or Hibio/DNasis (version 2.1). Gaps between the contigs were closed by primer walking strategy using purified genomic phage DNA as templates. Direct sequencing of chromosomal DNA was performed by Sequiserve (Vaterstetten, Germany). Specific primers were derived from the sequences of the contigs. Sequences of terminally redundant genomes were completed by upstream and downstream sequencing the single last contig until an overlap of the left and right ends of the DNA molecule was encountered.

In case of B025, distinct chain termination signals were generated at the ends of the molecules, i.e. the single stranded 3'-overhangs (cos-sites). The genome sequence of B025 was determined by sequencing a PCR-product spanning the cos-site. To prepare the template DNA, overhanging ends of the DNA-molecules were ligated using T4-ligase. The PCR-product then was obtained with primers complementary to sequences upstream and downstream of the cos-site.

2.14. Determination of phage attachment sites *attP*

2.14.1 Lysogenisation of *Listeria* strains

Bacteria were lysogenised by superinfection with bacteriophages using the agar plate method (see 2.1. phage propagation). Confluently lysed plates were incubated at RT for 3 days until colonies of phage resistant bacteria were visible. These lysogenic strains were purified from

free phage particles by 2 times serial streaking on agar plates. Then, lysogeny of the strain was proved by phage induction through UV radiation and a PCR reaction using phage specific primers.

2.14.2 Inverse PCR

Purified genomic DNA of lysogenised *Listeria* strains was used as template for the identification of the attachment site via inverse PCR (Ochman et al., 1988). The phage attachment site *attP* was expected to be located in a non-coding region immediately downstream of the integrase gene (*int*) on the bacteriophage genome. Restriction endonucleases with cutting sites in a distance of about 1,000 bp up- and downstream of this region were chosen for a complete digestion of 400 ng bacterial DNA of the lysogenic host in a reaction volume of 10 µl. The obtained fragments were treated with T4 DNA ligase (Roche) in a sample volume of 100 µl to obtain self-ligated circular molecules. Divergent primers located in areas nearby the restriction sites on the phage genomes were designed and used for the PCR amplification of the ligated fragments. PCR reactions were performed using 40 ng DNA as template (= 10 µl of ligation sample) and an extension time of 3 min. Purified PCR products were sequenced using the same primers. The obtained nucleotide sequences were compared to the *Listeria* genome sequences available at NCBI using BlastN to identify the location of the bacterial attachment sites *attB*.

2.15. Phage induction from lysogens by UV radiation

10 ml Tryptose broth were inoculated with 1 ml of an overnight culture of prophage containing *Listeria* strain (lysogen) and were incubated for 2 h at 30°C. The culture was then poured into a sterile Petri dish and was irradiated with UV (256 nm) for 60-90 s to induce lytic phage propagation. Following 2 h incubation at 30°C in the dark without shaking, bacterial cells and cell debris were removed by centrifugation and sterile filtration of the supernatant. Existence of phage particles was proven by formation of plaques on sensitive host strains (plating method).

2.16. Protein analysis

Purification of structural proteins was performed as described earlier (Zink and Loessner, 1992; Loessner et al., 1994b). Phage particles withdrawn from CsCl gradients were carefully

diluted about 10-fold with SM buffer and were recovered by high speed centrifugation in conical tubes using a Beckmann swing rotor SW28.1 (25,000 rpm, 2 hours, 15°C). Protein pellets were resuspended in small volumes of MilliQ, mixed with an equal volume of 2x reducing loading buffer and denatured by heating (10 min, 95°C). Protein samples were stored at -20°C.

SDS-PAGE

Denatured proteins were separated by SDS-PAGE on discontinuous horizontal precast gels (ExcelGel SDS gradient 8-18%, Amersham) according to Laemmli (1970). Electrophoresis was performed according to manufacturer's instructions on a horizontal system (2217 Ultrophor Electrofocusing unit, Pharmacia). Gels were run at 12°C under the following conditions: 50 mA, 35 W and constant voltage of 200 V for 20 min and 600 V for another 60 min. Dalton Mark VII-L (Sigma) was used as molecular weight marker. Gels were dyed for 30 min in prewarmed (60°C) Coomassie stain and then the background dye was removed by two or three washing steps with prewarmed wash solution. Finally, protein bands were fixed by placing the gel in a preserve solution for 30 min.

SDS reducing loading buffer

10 µl/ml of the reducing agent DTT (1µg/ml Stock solution) were added per ml buffer just before use.

2.5 ml	Tris/Cl (1.25 M, pH 6.8)
2.0 g	SDS
15 mg	Bromphenol Blue
ad 100 ml	MilliQ water

Coomassie R350 dye solution (1 liter)

Coomassie R-350	4 tablets
Methanol	25 ml
Acetic acid	10 ml

Wash solution

Methanol	25% (v/v)
Acetic acid	10% (v/v)

Preserve solution

Methanol	25% (v/v)
Acetic acid	10% (v/v)
Glycerol	10% (v/v)

2.17. Mass spectrometry (MS), peptide mass fingerprinting

Results from peptide mass fingerprinting were used to assign bands from protein profiles (SDS-PAGE) to gene products predicted by computer analyses. For these experiments, prominent protein bands were excised from the SDS-gels. Corresponding bands from several gels were pooled and stored at 4°C in tubes containing a drop of Ethanol (70%). The in-gel digestion with trypsin was carried out overnight, at an enzyme concentration of 12.5 ng/μl and in 50 mM ammonium bicarbonate buffer (pH 8.0) at 37°C. Subsequent elution was performed in two steps using 0.1% trifluoroacetic acid/acetonitrile (2:3) during 20 hours. Afterwards, eluted supernatants were pooled and vacuum air-dried. The peptides were dissolved in 5% formic acid.

Matrix-assisted laser-desorption ionization mass spectrometry (MALDI-MS) was carried out as described earlier (Zimmer et al., 2003). Analysis was performed using a Bruker Reflex III TOF mass spectrometer with a nitrogen UV laser ($\lambda_{\text{max}} = 337\text{nm}$) and a dual channel plate detector (Bruker Daltonik, Bremen). For recording of the spectra, samples of 1 μl were mixed with an equal volume of saturated, freshly prepared solution of α -cyano-4-hydroxy-cinnamic acid (CHCA) in acetonitrile/H₂O (2:1) with 0.1% trifluoroacetic acid. An acceleration voltage of 20kV was used, and the detector voltage was adjusted to 1.7 kV.

2.18. Bioinformatics

The programmes Hibio (Hitachi), DNASIS (version 2.1), Vector NTI Suite 8 (InforMax) and diverse tools from the HUSAR Sequence analysis package (<http://genome.dkfz-heidelberg.de>) were used for sequence assembly and analyses of nucleotide and amino acid sequences.

Similarity searches in the databases were performed using the BLAST algorithms (Altschul et al., 1997) available in the HUSAR Sequence analysis package or through NCBI (<http://www.ncbi.nlm.nih.gov>). Open reading frames (ORFs) were predicted with Artemis available from <http://www.sanger.ac.uk/Software/Artemis> (Rutherford et al., 2000).

Deduced peptide sequences were scanned for functional protein domains or motifs using InterProScan available from the [www Service](http://www.ebi.ac.uk/InterProScan/) at the European Bioinformatics Institute (<http://www.ebi.ac.uk/InterProScan/>) (Zdobnov and Apweiler, 2001; Zdobnov et al., 2002) and CD-Search (Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2005).

Prediction of tRNA genes in the genome of phage A511 was performed using tRNA-Scan SE (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE/>) (Lowe and Eddy, 1997).

ClustalW was used for evolutionary analysis in order to generate a phylogenetic tree. The programme is available at <http://www.ebi.ac.uk/clustalw/>, the [www Service](http://www.ebi.ac.uk/) at the European Bioinformatics Institute (Thompson et al., 1994)

2.19. Nucleotide sequence accession numbers

The nucleotide sequences of the annotated genomes reported in this work will be accessible in the EMBL and GenBank databases under following accession numbers:

A006	DQ003642
A500	DQ003637
A511	DQ003638
B025	DQ003639
B054	DQ003640
P35	DQ003641

III. RESULTS

For reasons of clearness and due to a multitude of common features, the results from sequencing and computer analysis are summarised for all of the analysed genomes of the small bacteriophages. Particular characteristics of each genome are described in separate paragraphs. The genome of the large virulent Myovirus A511 revealed little similarity to the small genomes analysed, and it is treated of in a separate section. The last section comprises the results from protein analysis of all *Listeria* phages investigated.

1. Genome analysis of the small bacteriophages

1.1. Results of nucleotide sequencing: general genome features

The genomes of *Listeria* phages A006, A500, B025, B054, and P35 were sequenced by a shotgun approach and subsequent primer walking strategy. Resulting nucleotide sequences were then further analysed by bioinformatical tools. Restriction maps predicted *in silico* were in agreement with patterns observed by restriction enzyme analysis (REA), indicating correct contig assembly (data not shown).

All *Listeria* phages feature dsDNA with an average G+C content of 35-36 mol%. These values are comparable to those reported for *Listeria* phages A118 and PSA, respectively (Loessner et al. 2002; Zimmer et al., 2003), but are slightly lower than those described for different *L. monocytogenes* strains (38-39 mol%) and *L. innocua* Clip11262 (37 mol%) (Glaser et al., 2001; Nelson et al., 2004). In contrast to these findings, P35 exhibits an average G+C content of 40.8 mol%.

Of all the *Siphoviridae* investigated in this study, P35 has the smallest genome with 35.8 kb, whereas A006 and A500 possess an average genome size (38.1 kb and 38.9 kb, respectively). The largest genome with 42.7 kb is that of B025. The genome of the small temperate Myovirus B054 features 48.2 kb. Genomic features and morphological characteristics of all *Listeria* bacteriophages investigated are summarised in Table 2.

Table 2. General features of the *Listeria* phages investigated. Morphotypes, capsid diameters and tail lengths were determined by Zink and Loessner (1992) via electron microscopy. The genome structure of A511 could not be clearly identified, yet.

	A006	A500	B025	B054	P35	A511
Morphotype	B1	B1	B1	A1	B1	A1
Capsid diameter	62 nm	62 nm	63 nm	64 nm	58 nm	80 nm
Tail length	280 nm	274 nm	252 nm	244 nm	110 nm	180 nm
Genome size	38124 bp	38867 bp	42653 bp	48172 bp	35822 bp	134494 bp
Genome structure	terminally redundant	terminally redundant	cos-site	terminally redundant	terminally redundant	unknown
Predicted ORFs	62	64	65	80	56	190
Start codons						
ATG	54	56	56	73	54	148
GTG	4	5	4	4	0	16
TTG	4	3	5	3	2	26
G+C content	35.5 mol%	36.7 mol%	35.1 mol%	36.2 mol%	40.8 mol%	36.1 mol%
Attachment-site	tRNA-Arg	tRNA-Lys	tRNA-Arg	EF-Ts	none	none
Life cycle	temperate	temperate	temperate	temperate	virulent	virulent

Abbreviations: Arginine (Arg), Lysine (Lys), Translation elongation factor Ts (EF-Ts).

1.2. The genome of B025 features cohesive ends

Listeria phages A006, A500, B054, and P35 were supposed to have terminally redundant genomes. In contrast, nucleotide sequence of B025 revealed the presence of single stranded cohesive ends designated cos-site or *cosN*, which were indicated by typical break offs in sequence signals (data not shown). Another hint for the presence of a cos-site was the alteration of the restriction pattern in REA after heat treatment of the samples followed by sudden cool down on ice directly before loading the agarose gel. During heat treatment at 50°C the annealed overhanging ends detached, and therefore bands containing the cos-sites dissociated into two smaller fragments (Figure 3). Restriction enzyme analysis (REA) of the linear molecules resulted in n+1 numbers of fragments corresponding to n cutting sites for an endonuclease. The *cosN* was determined by ligating the ends of the genomic DNA molecules before sequencing, and revealed single stranded 3'-end overlaps of ten nucleotides (CGGTGTGGGG).

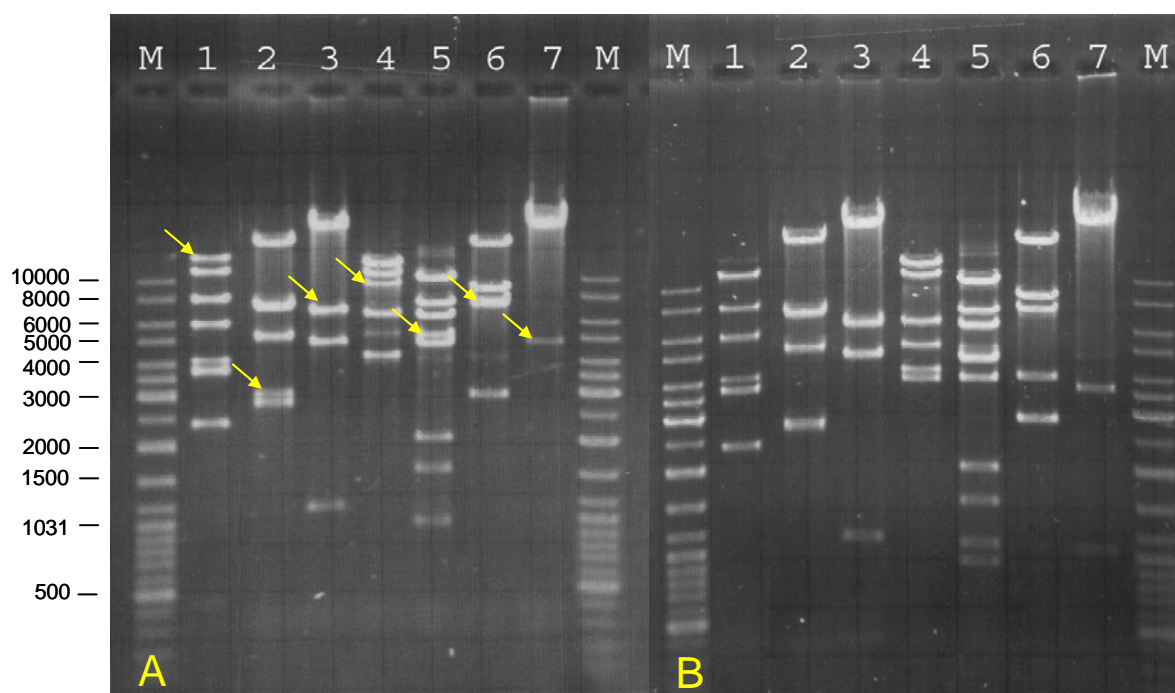


Fig. 3. Restriction enzyme analysis of the B025 genome. The presence of cohesive ends was indicated by a change of the restriction pattern, when probes were heated before loading the gel. Gels were run without heating the probes (A) and after incubation at 50°C (B). Fragments containing the cos-site are marked with arrows in A. Lanes from left to right: DNA ladder Mix (M), *BclI* (1), *SnaBI* (2), *EcoRV* (3), *HpaI* (4), *NdeI* (5), *NheI* (6), *PvuI* (7). The difference of the lanes 3 between panel (A) and (B) is not clearly visible, since melting the cohesive ends of the indicated band (6777 bp) generated one fragment of similar size (6414 bp) and a small one of 363 bp.

1.3. Identification of Open Reading Frames and functional assignments

Prediction of open reading frames (ORFs) of a minimum size of 30 codons was performed using Artemis software (www.sanger.ac.uk). ORFs starting with one of the possible initiation codons ATG, GTG or TTG and preceded by a recognizable ribosome binding site of the consensus sequence 5'-AGGAGGTG- 3' (Glaser et al., 2001) were considered as potential protein coding sequences. The majority of all putative ORFs initiates translation with the ATG start codon, whereas about eight ORFs on average in a genome start with alternative codons GTG or TTG. In the P35 genome, only two ORFs were identified starting with TTG. Generally, the average length of an ORF is about 200 – 218 codons and the number of

predicted ORFs per genome reflects the genome size. P35 has the smallest genome that encodes 56 putative gene products, whereas 80 ORFs have been predicted for the large genome of Myovirus B054.

Predicted peptide sequences deduced from putative ORFs were screened for similarities with proteins from the databases using BlastP. Significant database matches and preliminary functional assignments are listed in Tables VI.1-VI.5 (Appendix). In general, a large proportion of the predicted gene products showed no obvious biological function. But the majority of them at least revealed some homologies to proteins from the databases. Definitive functions were assigned to a significant number of proteins encoded in the “late” gene cluster (RESULTS, section 1.4).

To give an overview on the most important protein functions, fundamental results concerning all genomes investigated are defined below.

The genomic maps displayed in Figure 5 were arranged in a way that the first two ORFs represent the small and the large subunit (TerS and TerL) of the **terminase**, an enzyme required to pack dsDNA into the viral procapsids. Replication of phage DNA leads to accumulation of concatemers, which are processed to mature DNA monomers during packaging. Translocation of the DNA into the procapsids is an ATP dependent process reviewed elsewhere (Klaus et al., 1992; Fujisawa and Morita, 1997). In short, packaging occurs by the recognition of specific sequences called *cos* or *pac* followed by specific or non-sequence specific cleavage of the concatemer. Translocation starts from the generated end, and the DNA is cleaved for a second time at a specific or non-specific site, when the prohead is filled. In general, the genome structure is linked to the packaging mechanism. Site specific staggered cleavage at *cosN* leads to unique DNA molecules of exactly 100% of the genome sequence featuring cohesive single stranded terminal ends. This type of DNA molecules is packed into capsids of B025. Mature capsids of A006, A500, B054, and P35 supposedly contain terminally redundant and circularly permuted DNA molecules of >100% of the complete genome length. These result from non-specific initiation cleavage near a *pac* recognition site on concatemeric DNA molecules followed by a non-specific cleavage when the prohead shell is filled with DNA.

Major and minor **head proteins** and the **scaffold protein** are important structural components required for proper morphogenesis of the icosahedral capsid. The **portal protein** is part of a circular protein complex that connects the capsid and the phage tail (Klaus et al., 1992). The portal complex serves as starting point for head assembly and as docking site for DNA

packaging enzymes (Kutter and Sulakvelidze, 2005). During genome packaging or DNA injection into the host, the DNA passes the portal.

The long flexible tail tubes of *Siphoviridae* are composed of many copies of a **major tail protein (Tsh)** that is built up on a initiator complex to which one or more tail fibres are attached. The exact tail length is determined by a protein referred to as **tape measure protein (Tmp)**, whose length (amino acids) is proportional to the tail length of the virion. Tail formation and the mechanisms of length determination have been analysed in detail for bacteriophage λ (Katsura, 1990; Kutter and Sulakvelidze, 2005).

Base plate and **tail fibres** are variable constructs of the tail tip that mediate adhesion to the surface of the bacterial host and enzymatic degradation of the peptidoglycan (Kutter and Sulakvelidze, 2005). Assembly of the complex contractile tail of Myovirus T4 begins with formation of the baseplate and proceeds with polymerisation of the tail tube and the tail sheath. During attachment to the cell surface the baseplate undergoes a conformational change and initiates contraction of the tail sheath (Rossmann et al., 2004).

Holin and endolysin, the two components of the lysis cassette, are important factors that allow the progeny phages to escape from the host cell at the end of a lytic cycle. Phage **lysins** are peptidoglycan hydrolases, cell wall degrading enzymes that are composed of two functional domains. The C-terminus mediates binding to the substrate and is therefore called cell wall binding domain (CBD) (Loessner et al., 2002), whereas the N-terminal domain displays the catalytic activity (Loessner, 2005; Bernhardt et al., 2002). During assembly of the structural components of the progeny virions, **holin** proteins accumulate in the cytoplasmic membrane of the bacterial cell. They assemble into oligomers at a genetically determined time point, causing a collapse of the membrane potential and permeabilisation of the membrane (Wang et al., 2000). This way, the endolysin activity that accumulated in the cytosol gains access to the murein. The presence of a dual start motif as described for holins of λ and other lambdoid phages (Bonovich and Young, 1991; Wang et al., 2000; Young et al., 2000) is evident for the holins of *Listeria* phages A500 and A118 (Vukov et al., 2000; Vukov et al., 2003). In case of Hol500, two start codons are separated by the codons methionine and lysine. The function of both, the long and the short holin transcripts were tested in λ - Δ S genetic background and revealed distinct lytic activities (Vukov et al., 2000). In contrast, inactivation of the two start codons in Hol118 variants showed no significant influence on lysis timing. Instead, translation initiation from a third start codon resulted in a holin product, which was

devoid of its first transmembrane domain and acted as a functional inhibitor of the native Hol118 (Vukov et al., 2003).

Phage integrases are enzymes that mediate site specific recombination between two DNA recognition sequences, the phage attachment site *attP* and the bacterial attachment site *attB* (Groth and Calos, 2004). This reaction is essential to enter the alternative, lysogenic life-cycle, in which the phage genome integrates into the bacterial chromosome. The integrases predicted for the *Listeria* phages A006, A500, B025, and B054 were analysed by InterProScan and are supposed to be tyrosine recombinases, which utilise a catalytic tyrosine to mediate strand cleavage. Interestingly, within the “early” genes of B054 the gene product GP64 represents another putative integrase.

Homology search in the protein databases also revealed similarities to **transcriptional regulators, repressor** and **antirepressor proteins** that control gene expression during lysogeny or lytic proliferation, respectively. The genomes of the temperate phages A006, A500, B025, and B054 encode pairs of adjacent repressor proteins with opposite transcription direction. These are located at the border of the lytic genes and the lysogeny related genes, upstream of the integrase genes. Protein domain analysis revealed similarity to the helix-turn-helix motifs of phage λ repressor CI and antirepressor Cro, the functions of which are characterised elsewhere (Dodd et al., 2001; Shearwin et al., 2002; Dodd et al., 2004; Svenningsen et al., 2005).

Bacteriophage genomes encode important factors for replication, recombination and modification of genomic DNA; these are **DNA-polymerase, helicase, recombinase, or methylase** proteins, respectively. Single-stranded DNA-binding proteins (**SSB**) stabilise single-stranded DNA-intermediates during genome replication or recombination (Pedre et al., 1994).

In P35, gene product GP25 shows similarity to a **HNH homing endonuclease**, an enzyme that is involved in DNA rearrangements by catalysing sequence specific double-strand breaks and promoting the transfer of introns (Belfort and Roberts, 1997; Guhan and Muniyappa, 2003). Analysis of the putative protein GP25 using InterProScan revealed the presence of a protein domain, which is similar to a new type of sequence domain found in DNA-binding regions of homing endonucleases (Sitbon and Pietrokovski, 2003). However, this is the first putative endonucleases found in a *Listeria* phage genome, and a potential contribution to genomic variability of P35 has not been investigated so far.

1.4. Genomes are organised into functional modules

All protein coding regions on the phage genomes are closely packed with few intergenic spaces. Actually, start and stop codons of adjacent ORFs are often overlapping.

The observed database matches (as deduced from bioinformatical analysis) for bacteriophages A006, A500, B025, B054, and P35 revealed an organisation of the genomes into functional modules (Figure 5). These appear to be typical for temperate bacteriophages (Loessner et al., 2000; Zimmer et al., 2003; Brussow, 2001) and represent three phases of an infection cycle. The first rightward transcribed cluster comprises the “late genes”, which are required for assembly of the virion and cell lysis in the late phase of infection. It encodes structural proteins like head and tail components, enzymes required for DNA packaging and the lysis system that consists of a holin and an endolysin.

Usually, the first module is followed by a second cluster, which is largely transcribed in the opposite direction. It represents the region of lysogeny control that is transcriptionally active during the prophage status and includes lysogeny related functions like repressor proteins, the phage attachment site *attP* and an integrase.

The third module, again orientated rightwards, comprises the “early genes” and is involved in the early lytic state of infection, encoding products for replication, recombination and modification of the phage DNA.

1.5. P35 lacks the module of lysogeny control

The genome of P35 shows two striking differences in its organisation. First, the lysis cassette comprising holin and endolysin is interrupted by nine ORFs of unknown function. Secondly, the genome lacks the entire lysogeny module and all putative ORFs of P35 are transcribed in only one direction. No genes encoding an integrase or other factors required for the control of lysogeny are present. In conclusion, P35 is not a temperate bacteriophage, but a virulent one. These results are consistent with data published by Hodgson (2000) showing that general transduction with P35 occurs only infrequently, and results gained in this study indicating that P35 was not able to lysogenise its host strain.

1.6. Identification of the attachment sites

The phage attachment sites (*attP*) are always located in the short intergenic non-coding regions upstream of the integrase gene *int* and were identified via inverse PCR technique. Genomic DNA of the lysogenic strains *L. monocytogenes* 1001::A006, *L. monocytogenes* 1042:A500, *L. ivanovii* 3009::B025, and *L. ivanovii* 3009::B054 was used as template. PCR products covering the regions upstream and downstream of the supposed *attP* were sequenced, and the locations of the corresponding integration sites on the bacterial genomes (*attB*) were identified using BlastN (NCBI).

A006, B025, and B054 revealed integration sites with core sequences of similar length (16 bp and 17 bp, respectively), whereas the region of homology between *attP* of A500 and *attB* of its host strain *L. monocytogenes* sv 4b covered 45 bp. Recognition of two identical sequences of <50 bp is typical for members of the tyrosine recombinase family (Groth and Calos, 2004). In general, determination of the *attB* in the host strains revealed their location at the 3'-ends of tRNA-genes, which are single copy genes for one specific codon in each case. The exception is phage B054 that integrates into the 3'-end of a putative translation elongation factor. Alignments of the bacterial and phage attachment sites including the core sequences are shown in Figure 4.

A006 integrates into the 3'-end of a tRNA-Arg gene of *L. monocytogenes* EGDe. The adjacent genes lmo2466 and lmo2467 encode a hypothetical protein of unknown function and a putative chitin binding protein, respectively (Glaser et al., 2001).

The *attB* of **A500** was identified at the 3'-end of a tRNA-Lys gene in the genome of *L. monocytogenes* strain F2365 and contig 511 of the unfinished genome of strain H7858, respectively. In both strains, the tRNA genes are located downstream of a putative D-isomer specific 2-hydroxyacid dehydrogenase (Nelson et al., 2004).

B025 integrates into the 3'-end of a tRNA-Arg gene. Homology search was performed with the genome sequence of *L. innocua* Clip 11262, due to the fact that no complete genome sequence of *L. ivanovii* is available in the databases. The tRNA-Arg(2) (anticodon AGA) containing the *attB* was identified at nucleotides position 1,217,574 – 1,217,648. Interestingly, the *attP* of B025 is identical to the *attP* of the *Listeria* phage PSA (Lauer et al., 2002), and therefore both bacteriophages are assumed to use the same locus to reside as prophage in the host genome. In agreement to these findings, the predicted integrases of both bacteriophages show 95% amino acid identity. The *attB* in *L. innocua* is flanked by the genes

lin1204, whose gene product revealed similarity to internalins, and lin1203 that encodes a hypothetical protein of the phosphoesterase family (Glaser et al., 2001).

The putative *attB* of **B054** in *L. innocua* Clip11262 exactly reflects the 3'-terminus of a gene predicted to encode the translation elongation factor EF-Ts (lin1766) (Glaser et al., 2001). The core sequence of homology was identified at two positions in the genome of *L. innocua*, based upon a resident prophage that obviously recognised the same *attB*. Another hint is the identification of an integrase (lin1765) adjacent to *attB*. Downstream of the second location identified, the predicted gene lin1697 encodes a protein homolog of gp30 of A118 (Loessner et al., 2000).

Figure 4 (next page). Bacterial and phage attachment sites. The nucleotide sequences of bacteriophages and their host strains containing the attachment sites are aligned one below the other. Core sequences of the attachment sites are marked with grey boxes. Numbers indicate the position of the displayed fragment within the genome sequences. In chart A), B), and C) the attachment sites are located at the 3'-ends of tRNA genes, which are underlined, and the anticodon is marked in bold letters. Phage B054 (chart D) integrates into the 3'-end of the translation elongation factor EF-Ts (underlined), which exceeds the displayed fragment.

1.7. Protein homologies

The NCBI databases were scanned by BlastP to find protein similarities with the predicted peptide sequences, which provided insight into the relatedness to other bacteriophages. The results of the protein homology search are summarised in Tables VI.1-VI.5 (Appendix).

The proteomes of **A006** and **A500** revealed high localised similarity to *Listeria* phage A118 (Loessner et al., 2000). Homology between these phages comprises different functional clusters (Figure 5). The “early” gene products are almost identical in A006 and A118, whereas there is only one structural protein sharing limited amino acid similarity (GP12 of A006 and GP13 of A118, respectively). The other predicted gene products primarily resemble proteins of prophages annotated in the genome sequences of *L. monocytogenes* (strains H7858, F6854 and EGDe) and *L. innocua* Clip11262 (Glaser et al., 2001; Nelson et al., 2004). In contrast, A500 is highly similar to A118 with respect to the structural gene cluster, although the first database matches preferably indicate gene products that are A118-homologs of *L. innocua* prophages. Exceptions to this finding are the major head protein (GP6) and two minor proteins (GP21 and GP22) - probably base plate components - that revealed no similarity to A118. Even though eight more gene products matched database entries of A118 in the remaining part of the genome, the majority of predicted peptides resembles gene products of *L. innocua* and *L. monocytogenes* strains EGDe and F6854.

The overall genome sequence of the temperate Myovirus **B054** appears to be closely related to one of the *L. innocua* prophages. Most of the predicted gene products of B054, especially the structural components, revealed high similarity of up to 100% amino acid identity to gene products of the locus tags lin1700 – lin1765 annotated on the genome of *L. innocua* Clip11262 (Glaser et al., 2001). Interestingly, the second best homology hits of the “late” genes comprise predicted proteins of *Enterococcus faecalis* V583 (Paulsen et al., 2003). However, amino acid identities do not exceed 57%. Some putative gene products of the “early” module and the lysogeny control are similar to another prophage of *L. innocua* (lin1233 – lin1299). Only ten of a total of 80 predicted gene products revealed no hits to any protein sequence in the databases.

In contrast to the genomic colinearity found in B054 and a *L. innocua* prophage, the genome of **B025** appears to be very mosaic. Predicted peptide sequences are highly similar to proteins of *Listeria* phages A118 and PSA (Figure 5), and of *L. innocua* Clip11262 (locus tags lin2585 – lin2610) and *L. monocytogenes* H7858. Some of the structural components resemble hypothetical proteins of *Staphylococcus* phages 77 and A3 (Kwan et al., 2005), but amino

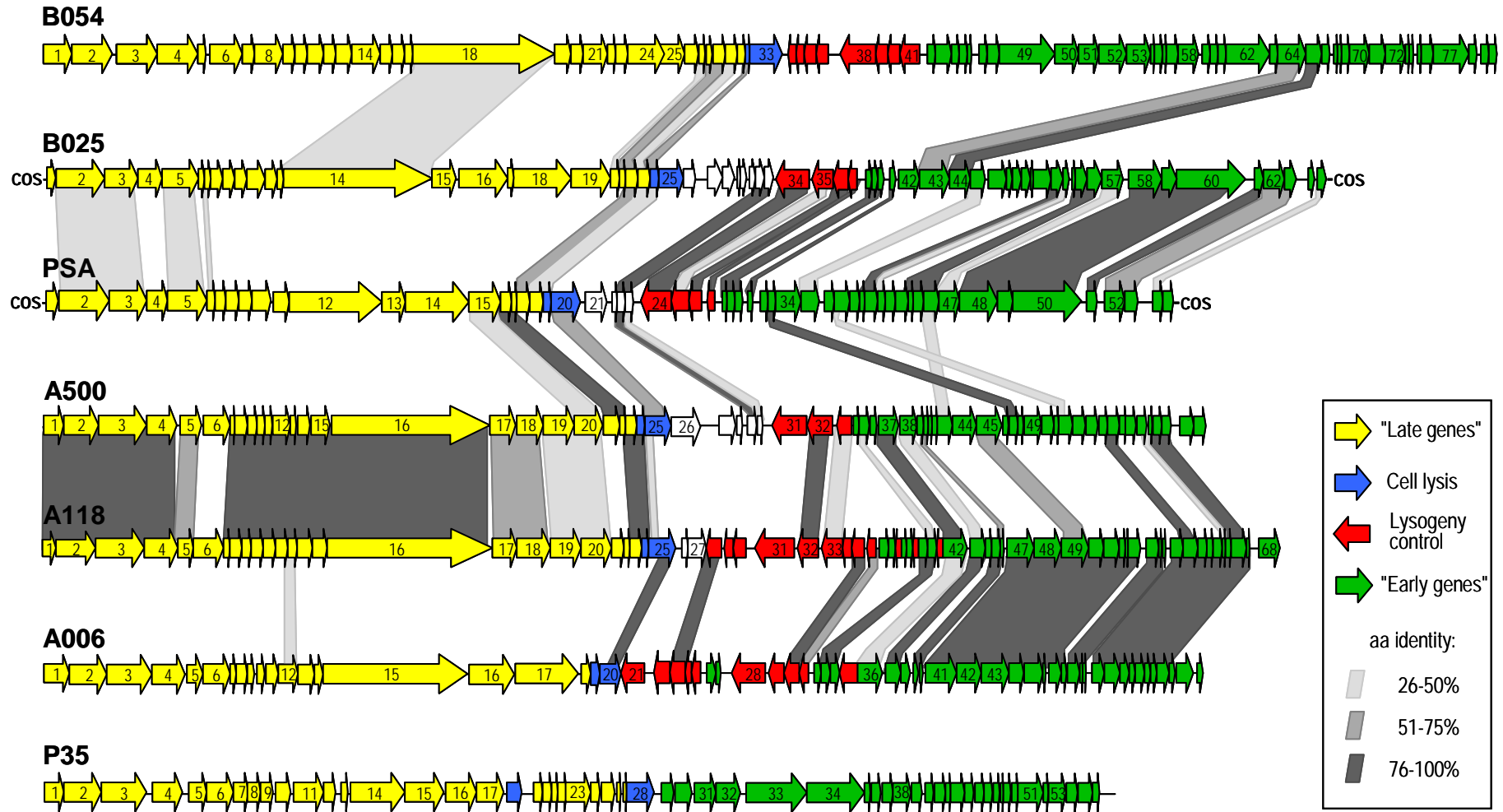
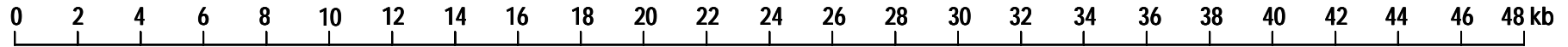
acid identities are low (average about 40% identity). Eleven of 65 predicted gene products show no hits to any protein in the databases.

Results from homology search are strikingly different for the virulent phage P35. The gene products revealed only weak similarity to any database entry. Actually, 33 of 56 polypeptides show no significant hits or no hits at all. Ten putative “late” gene products are slightly similar to proteins of *E. faecalis* V583, but there are also similarities to the yet unfinished genomes of *Clostridium thermocellum* and *Leuconostoc mesenteroides*. There are almost no similarities to other *Listeria* phages or prophages. Only two proteins, one of which was the phage lysin, weakly resemble proteins of *L. innocua* Clip11262. Thus, P35 appears to be a very distantly related member of the *Listeria* bacteriophages investigated to date.

1.8. Comparative genomics

To determine the relationship of the *Listeria* bacteriophages investigated, protein databases of all predicted gene products of each phage were created and screened for homologies using BlastP. The protein similarities identified are visualised as alignment of the genomic maps. One of the various possible arrangements including the previously published genomes of A118 and PSA is shown in Figure 5, which indicates protein similarity by grey shaded areas connecting the predicted similar ORFs. Light, medium and dark grey shaded linkages represent amino acid similarity of 26-50%, 51-75%, and 76-100%, respectively.

Fig. 5 (next page). Protein homologies of *Listeria* bacteriophages. Genomes are roughly drawn to scale and predicted ORFs are displayed as arrows, which also indicate transcription direction. The ORFs are numbered consecutively. Functional modules are marked in different colours: module of structural components (yellow), cell lysis (blue), lysogeny related functions (red), and “early” gene cluster (green). ORFs that could not be clearly assigned to a functional module are colourless. The region of lysogeny related functions includes the phage attachment site *attP* and the integrase. ORFs that revealed homology on protein level are directly linked by grey shadings according to the percentage of amino acid identity.



Protein similarities between the temperate *Listeria* phages are also reflected by a high degree of similarity on nucleotide level. The results of pairwise genomic alignments revealed large regions of nucleotide identity, which are displayed as Dotplot diagrams in Figure 6. The analysis was performed using the tools “Compare” and “Dotplot” provided by the HUSAR Sequence analysis package (settings: window-size 25, word size 20).

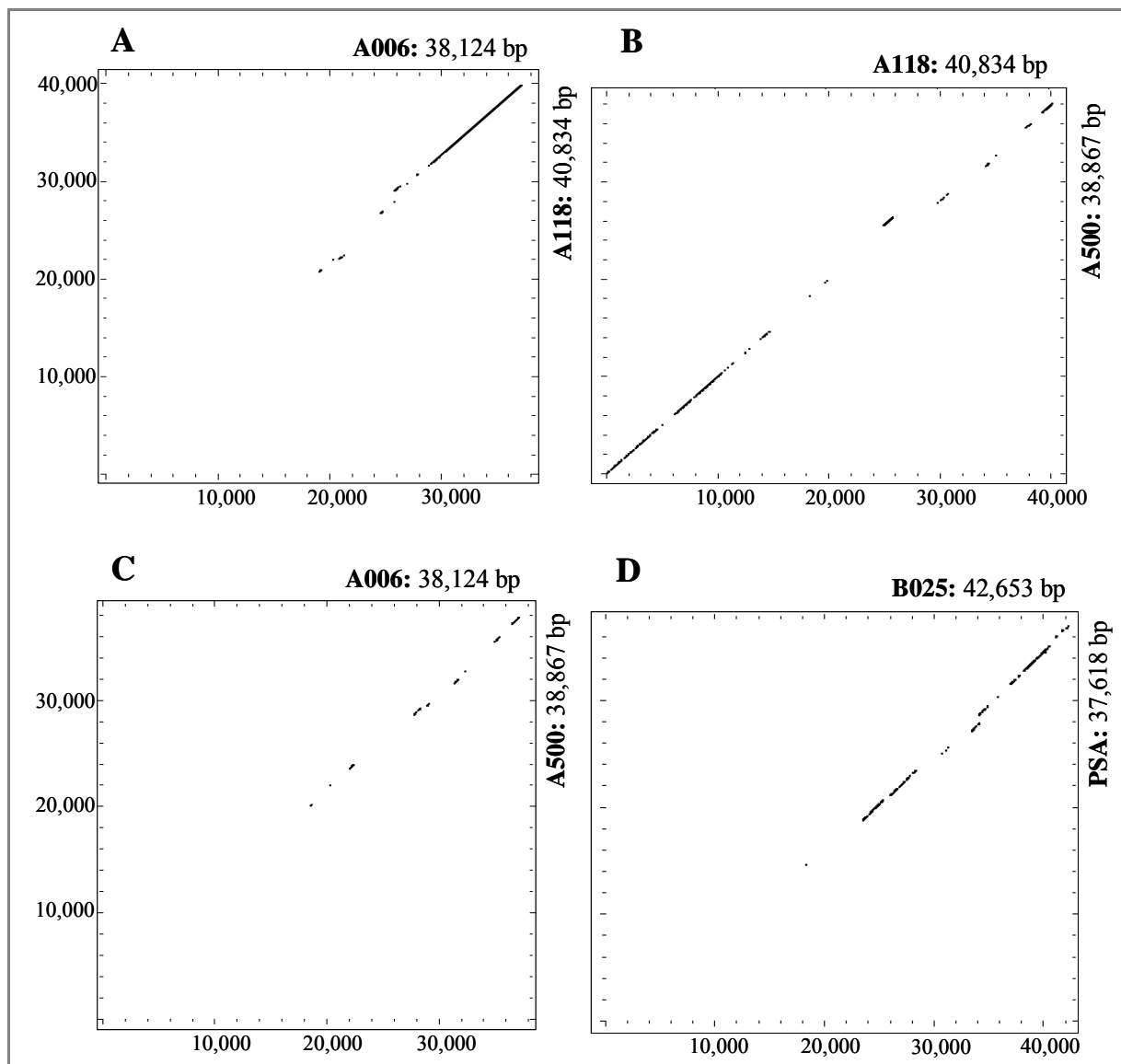


Fig. 6. Dotplots representing regions of nucleotide similarity between compared genome sequences. The scales on x- and y-axis indicate the number of base pairs. Sequence similarity was identified between A006 and A118 (panel A), A500 and A118 (panel B), A006 and A500 (panel C), B025 and PSA (panel D).

With respect to comparative analysis of the bacteriophages studied here, the most database matches were found between the proteomes of A006 and A500. They revealed fifteen common gene products of 35-98% amino acid identity. These gene products represent protein homologs of A118 and are predominantly encoded by the module of “early” genes. This kind of “indirect linkage” of homology is also the case for the relatedness between the other *Listeria* phages investigated. With few exceptions, similarity between A006, A500, B025, B054, and P35 could be observed in predicted gene products that show homology to prophages of *L. innocua* and *L. monocytogenes* strains or to *Listeria* phages A118 and PSA. According to this, protein similarity between B025 and A006 concerns ten gene products (30-100% amino acid identity) that are homologs of A118, PSA or *L. innocua* prophage. Phages B025 and B054, which are both capable to infect *L. innocua*, are similar with respect to seven gene products (Tmp, Hol, structural components and “early” proteins) that primarily resemble *L. innocua* prophages. In general, there are only few protein homologies between the *Listeria* phages investigated here that indicate relatedness.

In agreement with the results from protein homology search indicating a distant relatedness of P35 to database entries, direct comparison of the P35 proteome to the other *Listeria* phage proteomes revealed no significant similarities. Only two gene products of P35, the main capsid protein (GP6) and the endolysin (GP28) weakly resemble those of A006 and A500, respectively. Consequently, the large proportion of unknown proteins of P35 remains undefined.

To summarise it from an evolutionary point of view, A006, A118, and A500 form a closely related group, as is true for B025 and PSA. These two groups seem to be linked by only few homologies. B054, which appears almost identical to one of the putative *L. innocua* prophages, is more distantly related to these two groups mentioned above. P35 however, differs to a great extent and likely originated from a different lineage.

2. Genome analysis of the virulent Myovirus A511

2.1. General features of the genome

The genome of A511 was completely sequenced and analysed. It has a size of 134,493 bp. Altogether, 190 putative ORFs and a cluster of sixteen tRNA genes were predicted. Forty ORFs start with alternative initiation codons GTG or TTG (Table VI.6, Appendix).

The G+C content of 36.1 mol% is comparable to those reported for the temperate *Listeria* phages investigated in this study and previously (Loessner et al., 2000; Zimmer et al., 2003), and also resembles the G+C content of different *L. monocytogenes* strains (38-39 mol%) and *L. innocua* Clip11262 (37 mol%) (Glaser et al., 2001; Buchrieser et al., 2003; Nelson et al., 2004). An interesting feature of the genome is that it completely lacks GATC sites, and therefore cannot be digested by restriction enzymes that recognise this site, for example *Sau3A1*, *BamHI* or *PvuI*.

2.2. Functional assignments of the predicted gene products

The gene products of A511 predicted *in silico* were screened for homologies to the NCBI database entries. Altogether, 170 gene products (GP) revealed similarity to existing protein records, whereas 20 revealed no matches and therefore represent new entries in the database. On the basis of sequence similarities and protein domains search using InterProScan and CD-Search (Marchler-Bauer and Bryant, 2004; Marchler-Bauer et al., 2005), putative functional assignments were made for 36 gene products, which are summarised in Table VI.6 (Appendix). However, it is important to note that the identities specified below, with some exceptions only amount to about 30%.

Aside from the capsid (Cps), the tail sheath protein (Tsh), and the endolysin Ply511 (GP19) featuring N-Acetylmuramoyl-L-alanine amidase activity (Loessner et al., 1995a; Loessner and Scherer, 1995), seven other “late gene” functions have been assigned to predicted proteins. These are the terminase subunits (GP13 and GP14), the portal protein (GP23), a putative capsid protein (GP1), a putative base plate component (GP31) and two putative tail proteins (GP26 and GP32), of which GP26 presumably is the tail tape measure protein (Tmp). The protein functions of the “late gene” cluster are explained in RESULTS, section 1.3. Unexpectedly, no holin protein could be identified. GP73 revealed very low similarity to the

holin protein of *Bacillus* phage SPP1, but the distant localisation from plyA511 and the low protein similarity argue against this possibility.

The major part of the functional assignments comprises gene products involved in DNA replication, transcription and modification. Important replication enzymes are, for example, the predicted DNA polymerase subunits (GP66, GP67), DNA exonucleases (GP43, GP77), primase, helicase (GP40, GP42, GP46) and a DNA ligase (GP166). The role of these factors in the events at the DNA replication fork is outlined elsewhere (Madigen et al., 1997; Knippers, 2001).

The putative integration host factor IHF (GP65) is a member of the DNA binding protein family that binds and bends DNA, and is required in many cellular processes including transcription, site-specific recombination, and higher-order nucleoprotein complex assembly (Lynch et al., 2003). Further putative functional assignments to components involved in DNA transcription, modification and repair are a recombinase A (GP70), the ribonucleoside diphosphate reductase subunits alpha and beta (GP52 - GP54), an ATPase (GP87), a dUTPase (GP47), a replication protein (GP41), a sigma factor (GP72), a transcription factor (GP96), a *mom*-like protein involved in DNA modification (GP107) and a repressor homolog (GP175).

The putative thioredoxin protein encoded by GP57 catalyses protein crosslinking and is involved in posttranslational modification. PrsA, a phosphoribosylpyrophosphate synthetase (GP61) plays a role in nucleotide transport and metabolism.

2.3. Genomic organisation of A511

At the first glance, the overall genome structure of A511 seems to be arranged in two blocks of predicted ORFs with opposite transcription directions. However, functional assignments and comparative genomic analysis indicate that the ORFs are clustered into more than two functional modules. The genomic organisation of A511 in comparison to phages K and LP65 is displayed in Figure 8. The module of “late genes” comprising structural components, phage assembly proteins, DNA packaging enzymes and the endolysin Ply511 is supposed to be represented by *gp1-35*, including *cps* and *tsh*.

The second large functional gene cluster, the module of DNA replication and transcription, is represented by *gp40-72*. In addition to a number of proteins required for DNA transcription and modification (RESULTS, section 2.2), it encodes important protein functions required at a replication fork, such as the DNA polymerase subunits, DNA repair exonuclease, helicase,

primase and DNA binding proteins. The ability to constitute an “own” replisome enables A511 to replicate genomic DNA mostly independently from host functions. The phage-encoded putative sigma factor is probably used to modify the bacterial RNA polymerase core enzyme to recognise phage promoter regions, thus shifting gene expression from transcription of bacterial genes to phage genes.

The two functional modules mentioned above represent about half of the genome sequence. The second half supposedly contains at least one more module, which is indicated by a cluster of genes with opposite transcription direction (*gp147-188*). This genome region also contains a cluster of sixteen tRNA genes, three of which were predicted to be pseudogenes according to parameters applied by the programme tRNAscan-SE (Lowe and Eddy, 1997).

Analysis of the codon usage of A511 revealed that most of the phage encoded tRNAs and the dedicated amino acids, respectively, represent codons of high frequency. These are, for example, lysine (Lys), asparagine (Asn), threonine (Thr), and glutamate (Gln), which are frequent components of structural proteins. The codon usage in *Listeria* and A511 is similar in both, the phage and its host.

2.4. Comparative genomic analysis

Only 20 predicted gene products of A511 show no significant protein similarities, whereas 164 are similar to predicted proteins of *Listeria* phage P100 (Carlton et al., 2005). P100 has a genome of 131,384 bp, which is 3,110 bp smaller than that of A511, and it encodes 18 tRNA genes and 174 putative gene products. With few exceptions, the overall genome organisation appears consistent between these two *Listeria* phages (Table VI.6, Appendix): First, ten of the predicted ORFs of P100 (*gp76*, *gp103-105*, *gp108*, *gp118*, *gp120*, and *gp138-140*) have no homologues in A511. Second, a sequence stretch of about 5,000 bp in A511 (*gp105-120*) seems to be lacking in P100.

Altogether, protein homologies are high, with amino acid identities of 90-100%. A few gene products show lower similarities, between 70-90%. In addition, every tRNA gene of A511 is also present in P100, featuring the same anticodons. The direct alignment of the nucleotide sequences reflects these results and is visualized by a dotplot diagram in Figure 7. Identities on nucleotide level range between 91-99%.

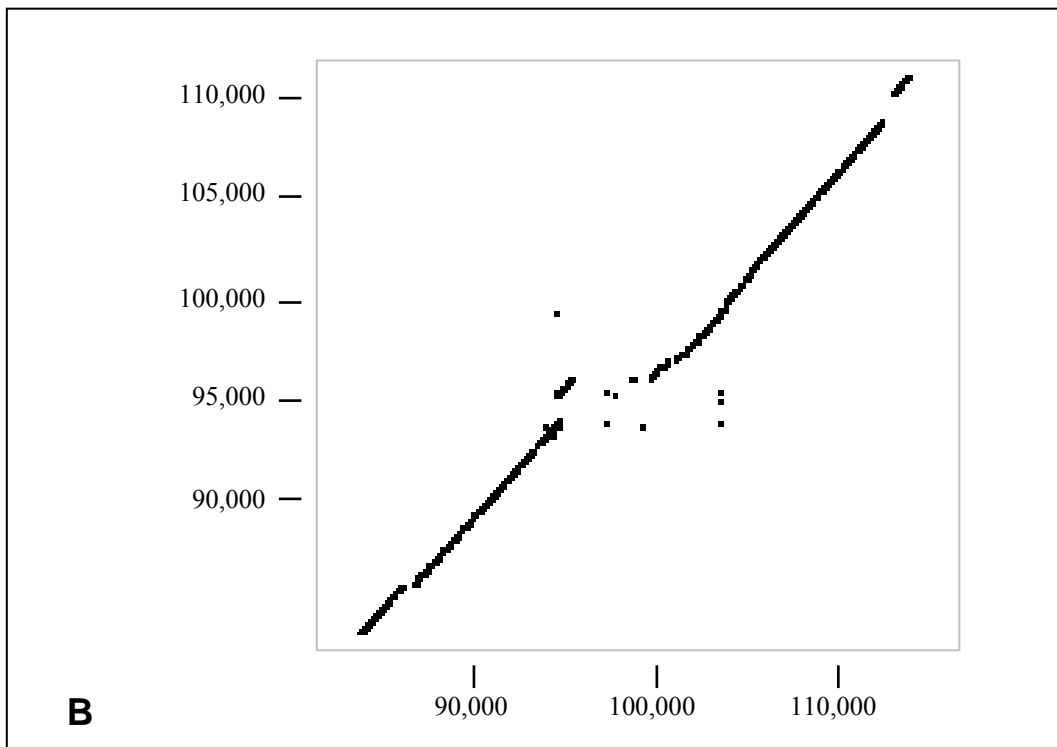
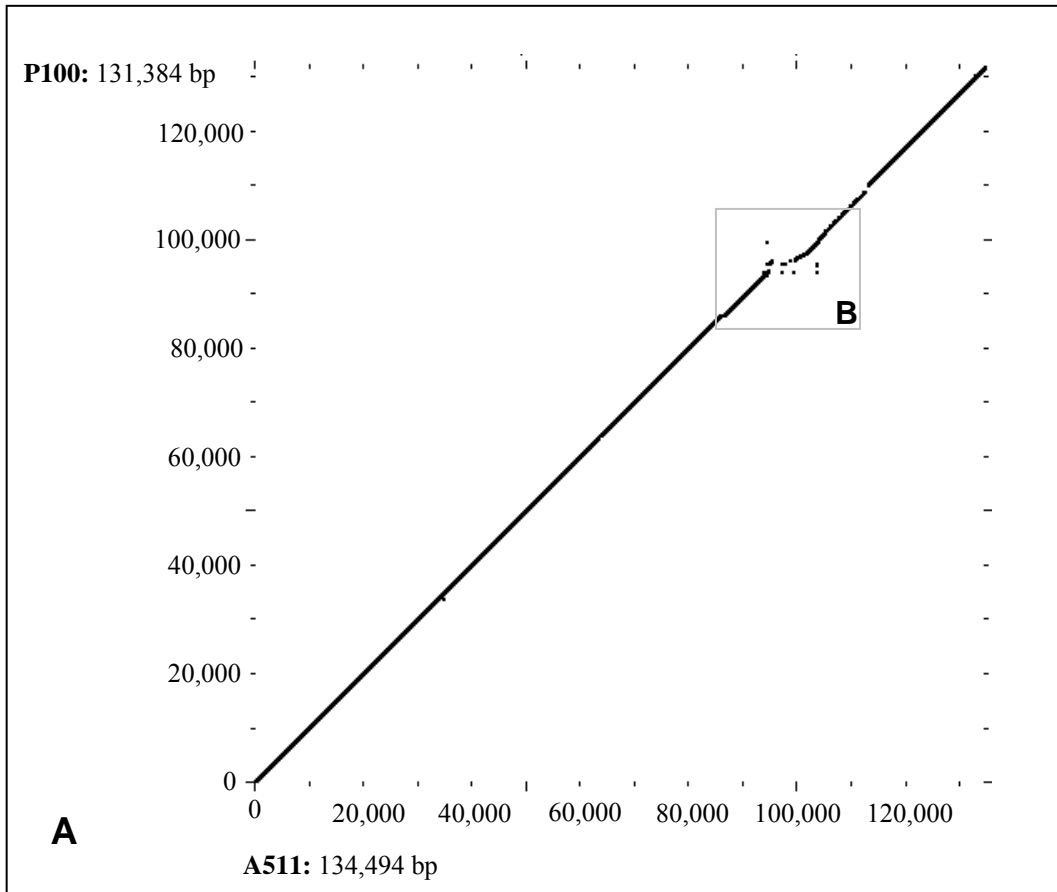


Fig. 7 (previous page). Dotplot matrix of direct genome comparison between A511 and P100. The diagram was created by the tools “Compare” and “Dotplot” available in the HUSAR Sequence analysis package using a window size of 25 nucleotides and a word size of 20 nucleotides. Regions of significant nucleotide similarity are dotted in the diagonal. X-axis: A511 (134,494 bp), y-axis: P100 (131,384 bp). The area showing differences in genomic organisation between A511 and P100 is zoomed and displayed in panel B.

Despite the outstanding homologies of A511 and P100, similarities of A511 gene products to proteins other than those of P100 are more interesting in consideration of evolutionary aspects and relationship between virulent phages of the *Myoviridae* family in general. Ninety predicted peptide sequences of A511 show other significant database matches. However, in most cases amino acid identity is rather low, between 20-60%. Only six gene products revealed more than 65% similarity, of which five are protein homologues from *L. innocua* or *L. monocytogenes*.

The major portion of protein homologies represents hits to *Staphylococcus* phage K (O'Flaherty et al., 2004), *Lactobacillus plantarum* bacteriophage LP65 (Chibani-Chennoufi et al., 2004), as well as *S. aureus* bacteriophages Twort and G1 (Kwan et al., 2005). An alignment of the genomic maps of A511 in comparison to LP65 and phage K is displayed in Figure 8. ORFs sharing protein similarity are linked via grey shaded bonds, which also indicate the percentage of amino acid identity of 20-40% (light), 41-60% (medium) and 61-66% (dark), respectively. LP65 and phage K were chosen for this alignment because they were the first Myoviruses, whose genomes had been completely sequenced and analysed in detail. In addition, they had revealed similarity to structural proteins of A511. Furthermore, phages G1 and Twort are clearly related to phage K. This is especially true for phage G1, which was proven to be 90% identical to phage K on nucleotide level (Kwan et al., 2005).

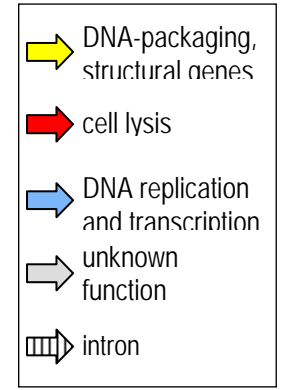
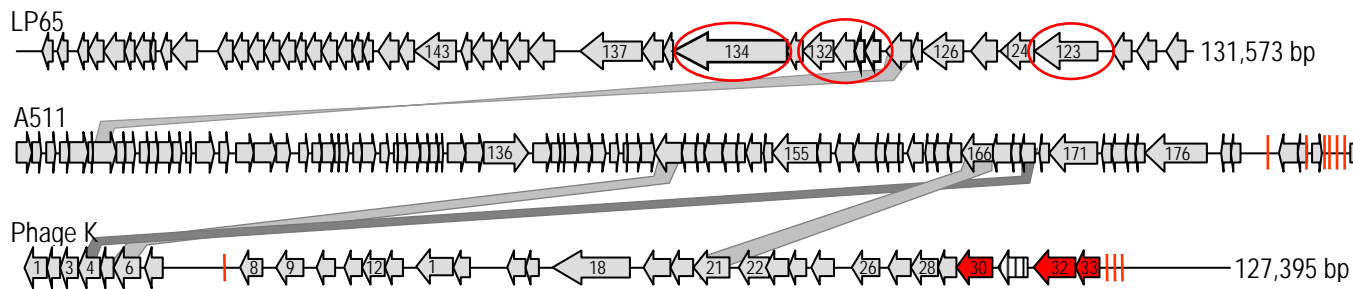
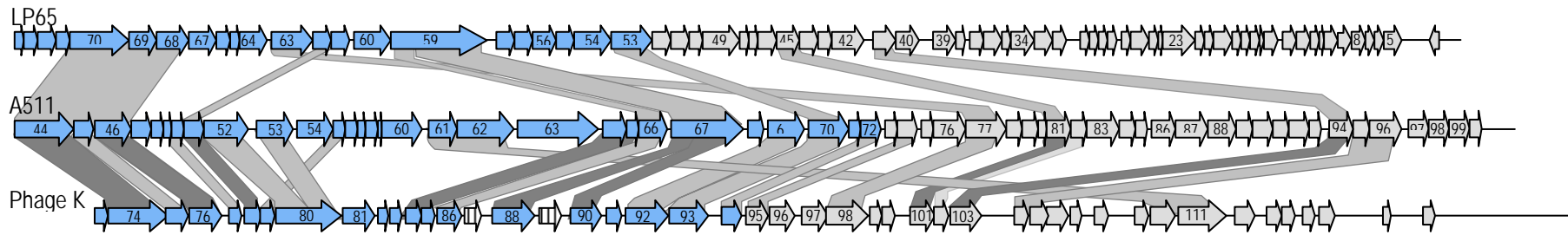
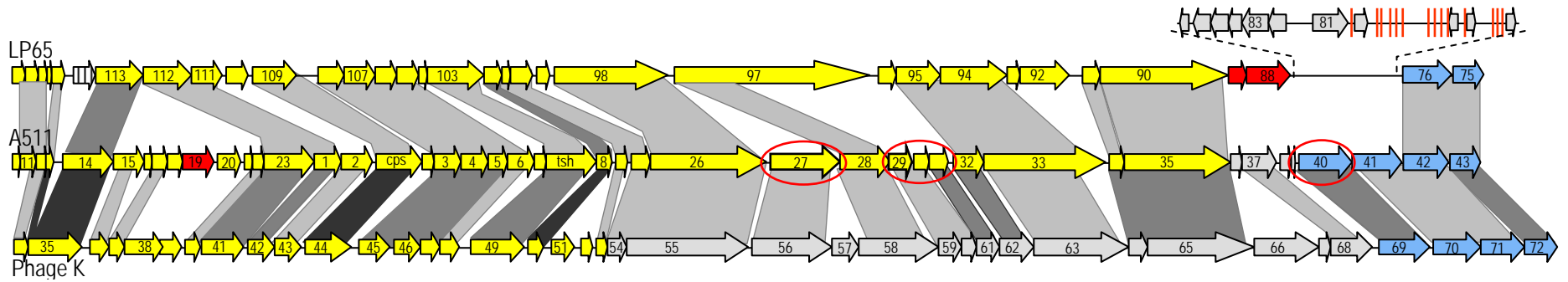
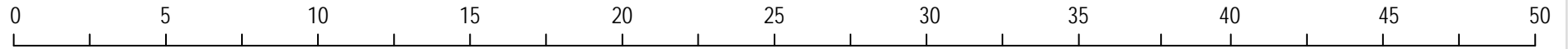


Fig. 8 (previous page). Genomic maps of phage A511, *Staphylococcus* phage K and *Lactobacillus plantarum* phage LP65. Predicted ORFs are represented by arrows drawn to scale and are numbered consecutively. Functional modules are coloured differently: structural components (yellow), cell lysis (red), DNA replication and transcription (blue). Red vertical bars indicate tRNA genes. ORFs sharing amino acid identity are linked by grey shaded bonds. By reason of clearness, some ORFs could not be linked with their homologues of A511 and are therefore marked with red circles. Pairs of homology are GP27 (A511) and orf134 (LP65), GP29-31 (A511) and orf129-132 (LP65), as well as GP40 (A511) and orf123 (LP65). (cp Table VI.6, Appendix).

According to the genomic arrangements of the predicted ORFs as well as to the detected amino acid identities, there is a particularly high degree of conservation concerning the structural genome regions of phages A511, K and LP65. Though, the overall genome organisation indicates significant genomic rearrangements, especially among functional gene clusters. For example, the cell lysis genes appear at the end of the structural module (LP65), within the structural module nearby the DNA packaging enzymes (A511) or at the beginning of a gene cluster with opposite transcription direction (phage K). Similarly, the tRNA genes differ in genome localisation. The most interesting rearrangements relate to four putative structural components of A511 (GP27 and GP29-31), including a tentative base plate protein similar to gene products of LP65 located in a gene cluster of opposite transcription direction of unknown function. The genes concerned are *orf129-132* and *orf134*, a putative cell wall hydrolase. Intron-associated HNH endonuclease genes annotated in phages K and LP65 have not been identified in A511.

With respect to overall genome organisation and protein similarities, A511 appears to be closer related to phage K than to LP65.

Comparative genomics between all the *Listeria* phages investigated in this study revealed six gene products of low protein similarity about 20-50% amino acid identity (Table 3). Five of these gene products encoded in the “early” gene cluster are of unknown function. The putative structural protein GP39 of A511 is similar to GP22 of A500, which is speculated to represent a component of the distal tail end (base plate of tail fibre).

Table 3. Protein similarities between A511 and the other *Listeria* phages investigated. Similarity of the gene products (GP) is represented as percentage of amino acid (aa) identity.

Gene product	Protein similarity	% amino acid identity
A511 GP39 (45 aa)	A500 GP22 (48 aa)	44%
A511 GP110 (68 aa)	B054 GP35 (67 aa)	22%
A511 GP136 (394 aa)	P35 GP38 (194 aa)	40% (C-terminus of A511 gp136)
A511 GP137 (141 aa)	P35 GP39 (143 aa)	22%
A511 GP148 (106 aa)	A006 GP51 (30 aa)	50% (N-terminus of A511 gp148)
A511 GP150 (175 aa)	B025 GP45 (187 aa)	33%
	A500 GP49 (202 aa)	27%
	A006 GP45 (186 aa)	24%

3. Protein profiles

Structural proteins, which assemble the virion, were purified and separated by horizontal SDS-polyacrylamide gelelectrophoresis (Figure 9 and 11). In addition, proteins of A118 were analysed likewise. Selected bands were excised and analysed by peptide mass fingerprinting, in order to assign predicted gene products to bands of the protein profile. In general, the protein profiles were in agreement with earlier results (Zink and Loessner, 1992).

According to the large, relative proportion of two major protein bands to the total amount of structural phage protein, these most prominent bands of the examined profiles were assigned to the major head (Cps) and the major tail protein (Tsh). The correlation of capsid proteins to upper major protein band in profiles from intact phages was shown by investigation of isolated capsids of four *Listeria*-phages by Zink and Loessner (1992).

In case of A006, A118, A500, and B054, Cps and/or Tsh were identified in more than a single band (Figure 9), indicating either a contamination of adjacent bands, frameshifting or posttranslational processing of the peptide sequence resulting either in a larger or smaller molecular weight. For example, peptide mass fingerprinting revealed fragments of Cps in three bands of the profile of **A006**. The molecular weight of Cps is predicted to be 32.2 kDa, and the putative Tsh (GP12) has 23.5 kDa. An N-terminal processing of 50 amino acids of the

Cps was assumed according to the MS results in band number 3, representing a calculated molecular weight of 26.4 kDa.

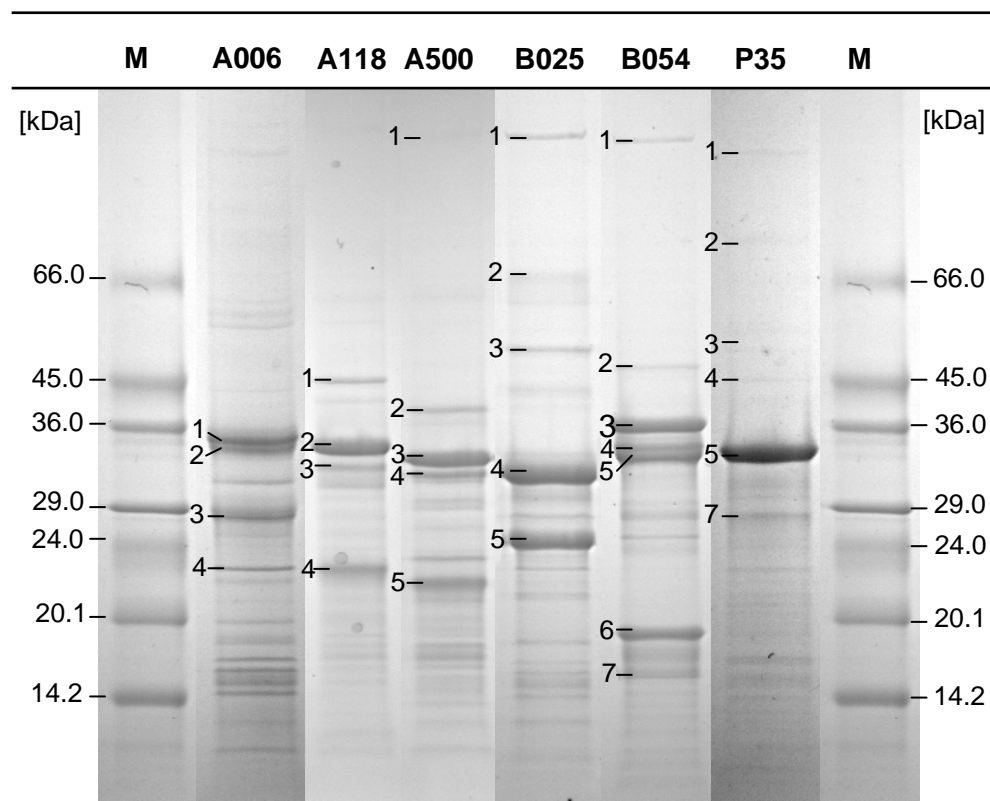


Fig. 9. SDS-PAGE of phage structural proteins.

Abbreviations: M molecular marker, Cps major capsid protein, GP gene product, Tsh major tail protein, Tmp tail tape measure protein. Detailed information on the annotated gene products of A006, A500, B025, B054, and P35 are listed in Tables VI.1- VI.5 (Appendix). Results from MS analysis are summarised for each protein profile as follows:

A006 Cps (1 and 2), truncated Cps (3), tsh (4).

A118 Cps and GP7 (1), Cps (2), truncated Cps (3), tsh (4).

A500 Tmp (1), GP18, GP19 and GP20 (2), cps (3), truncated cps and tsh (4), tsh (5).

B025 Tmp (1), GP18 (2), GP19 (3), mature cps (4), tsh (5).

B054 Tmp (1), GP3 (2), GP14 putative Tsh (3), GP8 putative Cps (4 and 5), GP7 (6), GP15 (7).

P35 GP14 putative Tmp (1 and 2), GP15 (3), GP16 (4), Cps and GP11 putative Tsh (5), GP17 (6).

Cps and Tsh in the profiles of A118 and A500, were found in bands of higher molecular weight than predicted by bioinformatical analysis. The larger proteins are thought to originate from translational frameshifts near the 3'-ends of the genes, overlapping the downstream reading frames (Zimmer et al., 2003). In the profile of **A118**, a mixture of peptide fragments of the adjacently encoded Cps and GP7 was identified by MS analysis, both of which are significantly smaller than the observed 45 kDa. Bioinformatical analysis indicated that the 5'-end of the reading frame of *gp7* overlaps the 3'-end of *cps* (= *gp6*) in the -1 frame. A potential short slippery sequence 5'-AAAA-3' is located nine nt upstream of the original *cps* stop codon at bp 6082. Ribosomal frameshifting also occurs during synthesis of Tsh. The protein was identified in band 4 at a molecular size of 23 kDa, though the predicted value was 15.7 kDa (Loessner et al., 2000). Computational analysis revealed a "shifty stop" at the 3'-end of *tsh* (bp 8181). Experimental evidence for the translational frameshifts in A118 and A500 were provided via MS analysis (Bielmann, unpublished data).

In case of **A500**, the full-length Cps was identified in band number 3 at slightly higher molecular weight than the predicted 29.7 kDa. Peptides indicating terminal processing of Cps were present in band 4. The predicted Tsh (15.9 kDa) ran at significant higher molecular weight about 22 kDa in the SDS-gel. The translational frameshifts in Cps and Tsh of phage A500 are shown in Figure 10. A possible slippery sequence (5'-GCGGGA-3') is supposed to facilitate a -1 frameshift resulting in a large head protein Cps-L that comprises the overlapping GP7 reading frame. The predicted Cps-L features a molecular weight of 35.3 kDa, which actually meets the size of the major band containing Cps in the profile. In contrast to Cps-L, the elongated tail protein Tsh-L of calculated 24 kDa molecular weight is supposedly produced by a +1 shift into the reading frame of *gp13*. The slippery site overlaps the stop codon (5'-CCCTGA-3') of the original *tsh* reading frame and resembles the situation described for Tsh-L of phage PSA, which has been designated as "shifty stop" (Zimmer et al., 2003).

In general, the genomic synteny of the "late" gene products shown for A118 and A500 is also reflected in the pattern of protein bands, which slightly differ in size and intensity. Interestingly, the second band in the A500 profile contained three proteins (GP18, GP19, GP20) of similar size (predicted molecular weights of 37.3 - 40.0 kDa), which are announced to represent putative tail or base plate components. The Tmp was assigned to band number 1 of the profile.

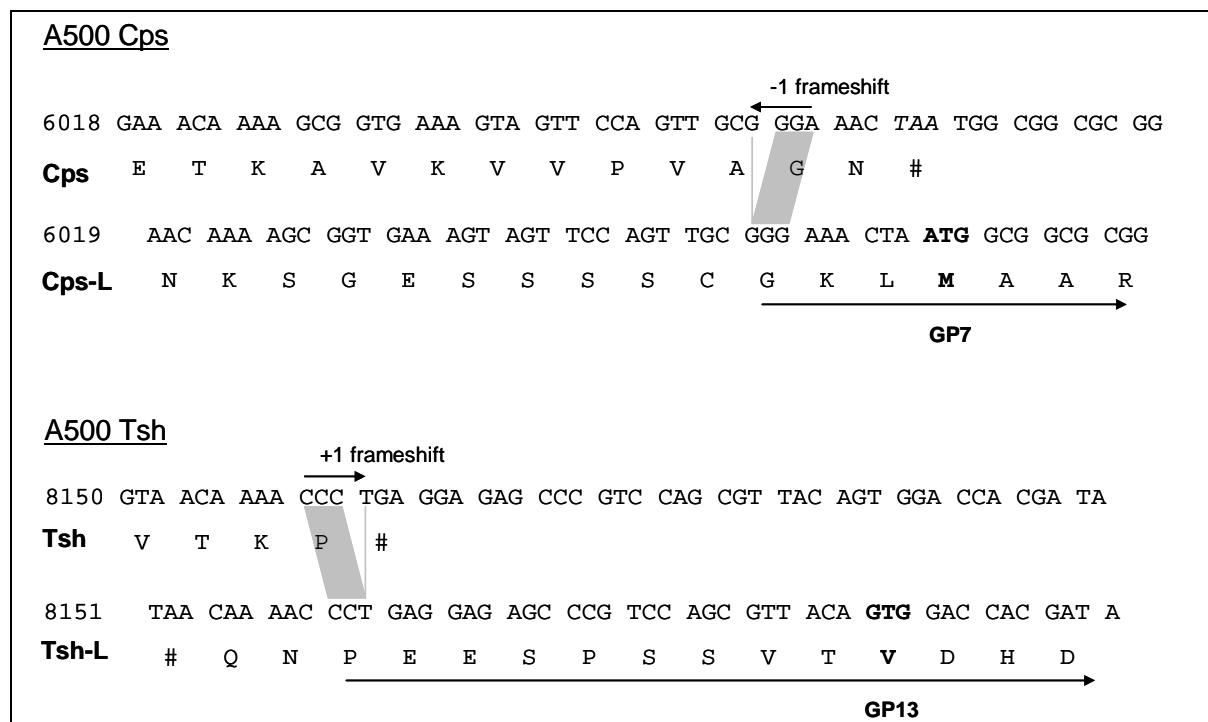


Fig. 10. Putative translational frameshifts in A500. The DNA sequence representing the overlapping reading frames of *cps* and *gp7*, *tsh* and *gp13*, respectively, and the deduced amino acid sequences are shown. The codons, where the ribosome possibly slips into an alternative reading frame leading to the synthesis of elongated polypeptides (Cps-L and Tsh-L), are marked by grey shaded boxes.

Six of the protein bands of Myovirus **B054** were assigned to gene products of unknown function. These comprised GP3, GP14, GP8, GP7, and GP15. On the basis of band intensity, the molecular size of the proteins and the genomic localisation of the assigned gene products, GP8 (band 4 and 5) and GP14 (band 3) were assumed to contain the Cps and Tsh, respectively. However, these assumptions could not be supported by indications derived from analysis by InterProScan. In contrast to the protein profiles of the Siphoviruses, the profile of Myovirus B054 exhibits a Tsh with a higher molecular weight than Cps. The first band of the profile represents the putative Tmp (GP18).

Because **P35** features a rather short tail (unpublished data, Table 1), a single major band is supposed to contain both, the Cps (32.9 kDa) and the putative Tsh GP11 (35 kDa). Bands number 1 and 2 were assigned to the same gene product (GP14), the putative tail tape measure protein (Tmp) of P35. The first fragment featuring a higher molecular mass of

estimated 140 kDa possibly contained a dimeric form of the 67.9 kDa monomeric tail protein. Further assignments were made to predicted gene products of unknown functions GP15 (band 3), GP16 (band 4) and GP17 (band 6).

N-terminal sequencing of Cps of **B025** was performed previously (Loessner et al., 1994b). According to this result, the peptide sequences of the mature Cps could be determined and revealed an N-terminal truncation of 81 amino acids. Therefore, the protein assigned to band 4 of the protein profile ran at lower molecular weight (about 33 kDa) as predicted (42.4 kDa). Gene products Tmp, GP18 and GP19 were assigned to bands 1 to 3, and the major tail protein Tsh was assigned to band 5.

Eight prominent bands of the **A511** protein profile (Figure 11) were analysed by peptide mass fingerprinting, three of which contained protein mixtures. To identify possible functions of the gene products, predicted peptide sequences were analysed by InterProScan. Band number 2 consisted of gene products featuring similar molecular weight GP35 (128.4 kDa) and GP26 (131.1 kDa). GP26 represents the putative Tmp, whereas the function of GP35 is unknown. A processed peptide of GP35, missing 110 N-terminal amino acids, was assigned to the band 3 in the profile.

Band 6 contained a protein mixture of GP31 (26.4 kDa), a putative base plate component, and GP74 (26.1 kDa), which is not encoded in the structural gene cluster. A processed peptide of GP74 was also identified in band 7, together with GP8 (15.7 kDa), which is a putative tail protein. Results from MS analysis concerning GP74 indicated a processing of 11 amino acids at the N-terminus and 60 amino acids at the C-terminus.

Cps and Tsh were clearly identified as the most prominent bands (Loessner and Scherer, 1995). Like in the profile of the Myovirus B054, Tsh (predicted 61.3 kDa) exhibited a larger molecular weight than Cps (51.4 kDa). Figure 11 shows the protein profile of A511 and putative functional assignments of the analysed proteins, as well as their probable location in the virion.

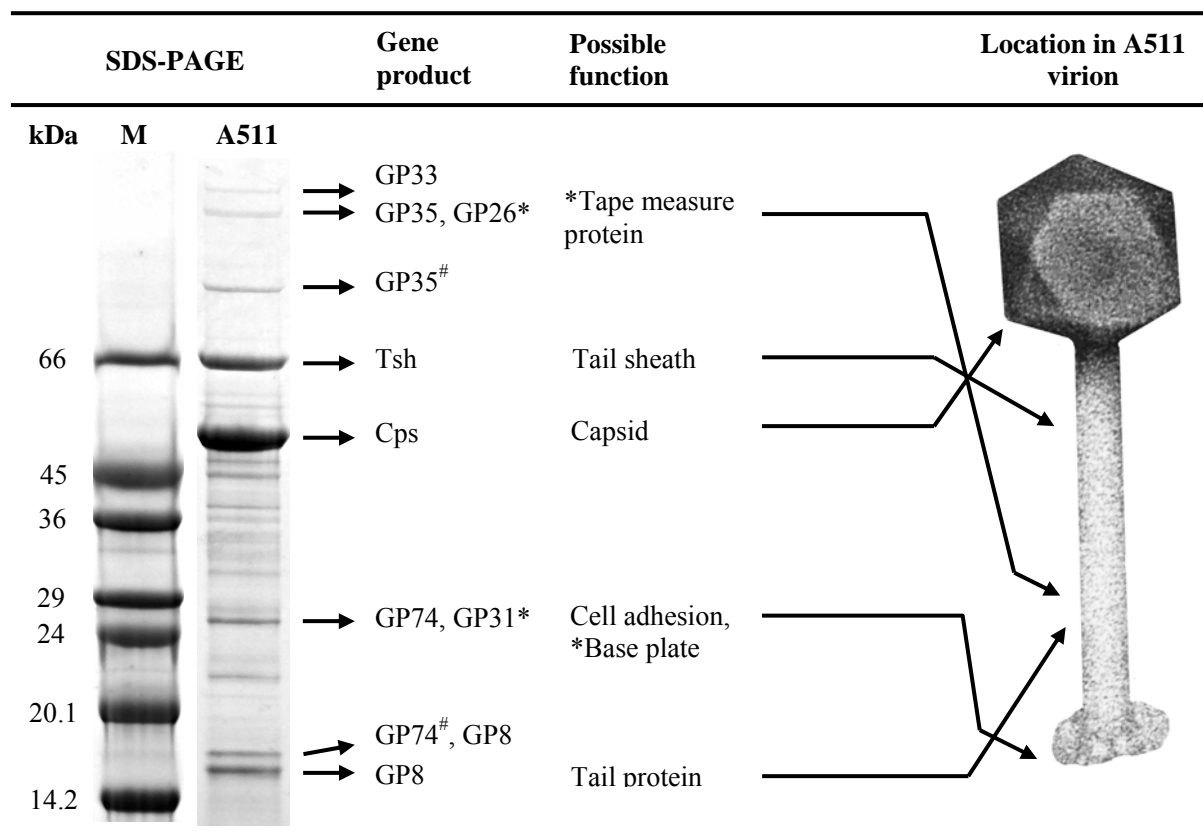


Fig. 11. Protein profile of A511, functional assignments of the gene products identified and their location in the virion. The functions assigned to the gene products (Table VI.6, Appendix) are based on results of BLAST and InterProScan. Gene products that were identified as terminally processed peptides are marked with [#] (see description in the text). An asterisk * indicates that the gene products constitute the minor part of a protein mixture. GP33 and GP35 are proteins of unknown functions.

IV. DISCUSSION

At the beginning of this study, detailed information on *Listeria* phage genomics was limited to the complete genome sequences of two temperate bacteriophages from the family of *Siphoviridae*, A118 and PSA (Zimmer et al., 2003; Loessner et al., 2000). Only very recently, the data pool was supplemented by the genome sequence of P100, a large virulent Myovirus with a broad host range (Carlton et al., 2005). The aim of this study was to gather new information on *Listeria* phages on molecular level, including genome sequencing, prediction of gene products, and determination of the attachment sites. In order to gain insight into relationship between the *Listeria* phages, comparative analyses of the nucleotide sequences were performed. Therefore, the genomes of six *Listeria* bacteriophages (A006, A500, A511, B025, B054, and P35) featuring different morphotypes, host ranges and life styles have been sequenced and computationally analysed.

All phage genomes investigated have in common a G+C content of 35-36 %, which is comparable to *Listeria* phages A118 and PSA (Zimmer et al., 2003; Loessner et al., 2000), but is slightly lower than those described for different *Listeria* strains. Somewhat different was P35, which exhibited an average G+C content of 40.8 %. This difference is not based on the lack of genes involved in lysogeny control, which seem to exhibit a G+C content that is 1-3 % lower than the remaining nucleotide sequence. The numbers of predicted ORFs per genome are proportional to the genome size, and the encoded gene products have an average length of approximately 200 amino acids. Functional assignments to predicted gene products revealed a genomic organisation of the annotated ORFs into life-cycle specific gene clusters (RESULTS, sections 1.4) that appear to be conserved in the temperate *Listeria* phages. The existence of functional modules was also described for many other phages of other Gram-positive, low G+C bacteria (Lazarevic et al., 1999; Brussow, 2001; Brussow and Desiere, 2001; Desiere et al., 2001; Zimmer et al., 2002; Obregon et al., 2003; Kwan et al., 2005). Within the “late” gene clusters, the basic arrangement of the structural genes appears to be highly conserved. This conservation is speculated to be due to strong sequential interactions during assembly of the virions, subsequent DNA packaging and particle maturation.

1. Comparative genomics and phage relationships

Following genome sequencing, a total of 517 ORFs have been predicted for all phage genomes investigated. In order to make functional assignments to the predicted gene products,

the deduced peptide sequences were screened for homologies to database entries. Pairwise comparison of the genome sequences and protein homology search provided insight into phage relationships. The most interesting findings are discussed below.

P35

In comparison to the other *Listeria* phages investigated, the virulent P35 strikingly differs in gene content and genome size, which is significantly smaller due to the absence of the lysogeny module, including an *attP* site and an integrase gene. The lytic life style seems to be advantageous in avoiding homoimmunity suppression, which enables the phage to infect a wider range of host strains including those that already contain prophages. This may be an explanation for the relatively broad host range of P35 compared to other *Listeria* phages of this morphotype. P35 was capable to proliferate in 75 % of all serotype 1/2 *L. monocytogenes* tested (Hodgson, 2000).

Considering the smaller genome size of P35 and the lack of a lysogeny module, one could speculate whether the phage might have been derived from a temperate ancestor. Actually, comparative genomics of *S. thermophilus* phages (Lucchini et al., 1999) and *Lactobacillus delbrückii* phages (Mikkonen et al, 1996), respectively, indicate that virulent phages conceivably derive from temperate phages by deletion and rearrangement events of the lysogeny module, which seems to be a recombination hot spot. Genetic instabilities of the temperate *Streptococcus thermophilus* bacteriophage Φ SFi21 were studied in detail by (Bruttin and Brussow (1996). Variants of the wild-type phage occurred with a frequency of 1% after a single passage, and upon serial passages the deletion mutants even replaced the wild-type. These investigations demonstrate that lytic phages can arise spontaneously from temperate phages, and that the lytic life style seems to have a short-term advantage, at least under specific growth conditions. However, in case of P35, there must have been several and precise step by step deletions for the removal of a large module, making this hypothesis rather notional.

An unusual finding concerning the overall genome organisation is that the genes encoding holin (*gp18*) and endolysin (*gp28*) are apparently separated by a stretch of 3,500 bp encoding nine putative gene products of unknown function. The reason for this atypical arrangement of the cell lysis components, which normally are located within a “lysis-cassette”, remains unclear. GP25, which is one of the gene products encoded in this genome region, revealed 31% amino acid identity to an HNH homing endonuclease (Table VI.5, Appendix). Even

though similarity was low, the conserved putative DNA-binding motif *NUMOD4* (nuclease-associated modular DNA-binding domain) of homing endonucleases and related proteins was identified by InterProScan. In general, “homing” is the transfer of a mobile intervening sequence, such as an intron or intein, initiated by an endonuclease that is encoded within the mobile sequence (Guhan and Muniyappa, 2003; Stoddard, 2005). The enzyme promotes its self-propagation without providing apparent function to the host. However, homing endonucleases are believed to play key roles in genome rearrangements and shuffling of protein domains, thus contributing to the evolution of genomes. In case of P35, GP25 could have been involved in disruption of the lysis cassette, but there is no experimental evidence for this hypothesis. Actually, homing endonucleases have not been identified in any *Listeria* phage genome, so far.

Concerning the relationship to other phages, the major portion of the predicted gene products of P35 revealed no protein homologies and therefore represents new database entries. Altogether, ten gene products of the “late” gene module resemble structural components of an *E. faecalis* V583 prophage. Interestingly, more than a quarter of the genome of *E. faecalis* V583, a vancomycin resistant clinical isolate, consists of mobile elements and foreign DNA, including seven probable integrated phage regions (Paulsen et al., 2003). Enterococci are obligate anaerobic Gram-positive bacteria that are primarily of faecal origin (Madigen et al., 1997). They are only distantly related to *Listeria*. Both genera belong to the class of Bacilli, but mainly occupy different ecological niches. The relatedness between the structural proteins of P35 and *E. faecalis* prophage points to a vertical passage of this module. However, the similarities are solely based on somewhat limited amino acid identities of 20-40%, whereas there is no significant homology on nucleotide level. These results hint on early divergence of the viral line represented by P35, because there was obviously no recent horizontal exchange. Although P35 exhibits a broad host spectrum of *Listeria* serovar 1/2 strains (Hodgson, 2000), the proteome surprisingly revealed no noticeable homology to other *Listeria* phages or prophages, except for the major capsid protein Cps (GP6) and the endolysin plyP35 (GP28). Cps features 26.8% amino acid identity to the major capsid protein GP34 of *L. monocytogenes* H7858 prophage, and GP28 has 38% identical amino acids with a phage lysin of *L. innocua*. These findings support the assumption that P35 is only distantly related to the other *Listeria* bacteriophages, and probably originated from a common ancestor in the evolutionary past. Due to its virulent life style, P35 largely misses the opportunity to recombine with prophages or other co-infecting phages. However, the incomplete collection

of genomic information in the databases is another aspect for the absence of homologies of P35 to the other *Listeria* phages. Considering the huge global phage population of at least 10^{30} particles and their great diversity, Rohwer and co-workers estimated that phage researchers sampled less than 0.0002% of the global phage metagenome (Rohwer, 2003). Incorporation of new genome sequences into the databases will foster our future knowledge about phage relationship.

B054

On the basis of electron microscopy investigations, B054 was classified into the family of *Myoviridae* featuring nonflexible contractile tails (Zink and Loessner, 1992). The authors discovered an unusual contraction mechanism of the tail sheath away from the capsid towards the base plate. In agreement with this morphotype, annotation of the genome revealed a large cluster of 29 putative structural genes in B054 (Figure 5, Table VI.4). The major part of the structural components is probably involved in assembly of the complex tail architecture. Database matches indicate nearly perfect synteny between the genome of B054 and a prophage of *L. innocua* Clip11262 annotated in segment 7/12 of the bacterial genome (Glaser et al. 2001). This specific *L. innocua* prophage was proven to be inducible by UV radiation in contrast to the other four prophages of *L. innocua*, which appeared to be non-functional (unpublished data). However, it is important to mention that the genomes of B054 and *L. innocua* prophage are not identical. Nine gene products of B054 revealed protein homology to another prophage of *L. innocua* located in segment 6/12, ten gene products showed no or no significant database hits, and one gene product resembled a hypothetical protein of *Bacillus clausii* KSM-K16. An example of genetic mosaicism within a single gene product represents the phage lysin (Ply) of B054, the two functional domains of which were combined from different origins. The C-terminal cell wall binding domain shows 77% amino acid identity to prophage endolysins of *L. innocua*, whereas the N-terminal domain resembles the corresponding protein of *Clostridium tetani*. The catalytic N-terminal domain featuring N-acetylmuramoyl-L-alanine amidase activity exhibits a lower amino acid similarity of 32%. The complete peptide sequence of the endolysin of B054 including C- and N-terminus is similar to that of PSA (47% amino acid identity), which likewise consists of functional domains of different origins (Zimmer et al., 2003). The authors speculate the assembly of different domains may result from horizontal gene transfer in a lysogenic cell. In addition to the natural occurring protein chimeras, cloning of recombinant endolysins that combine

varying functional C- and N-terminal domains may be useful for applications in food science, biotechnology and medicine.

B025

The genome of phage B025 features single stranded cohesive ends of ten nucleotides, in contrast to the other genomes investigated in this work, which are terminally redundant. During packaging of DNA, mature monomers are cleaved from the concatemer synthesised by a rolling circle mechanism. Progeny virions of *cos*-site phages contain genomes of unit-length, which are generated by introduction of staggered nicks at specific recognition sites designated *cosN* (Becker and Murialdo, 1990; Fujisawa and Morita, 1997). The region surrounding *cosN* is supposed to contain recognition sequences for the binding of terminase (*cosB*), integration host factor, and other phage- or host-encoded proteins required for initiation and termination cleavage. In phage λ , three direct repeat sequences located in *cosB* are important for terminase binding (Becker and Murialdo, 1990). Although the genome of B025 does not contain such repeat sequences, *cosN* is flanked by GC-rich runs of nucleotides: the sequence C₅G₃ is located 31 nt upstream, and G₂AG₃AG₃ is located 36 nt downstream of *cosN*, respectively. Similar strings of Gs and Cs were identified in *L. lactis* bacteriophage BK5-T and *C. perfringens* phage Φ 3626, respectively (Mahanivong et al., 2001; Zimmer et al., 2002). The occurrence of such sequences in genomes of low G+C content is unusual. Therefore, these sequences possibly constitute recognition sites that promote binding and activity of the terminase during DNA packaging, although this remains to be proven experimentally.

About half of the annotated gene products of B025 revealed amino acid similarity to PSA proteins. Most of them were found in the region of lysogeny control, and the “early” gene cluster. Consistency of the integrase genes and the attachment sites is discussed in section 3 of this chapter. The genome is strikingly mosaic, featuring several patches of diverse homologies, which revealed database hits to A118, *L. innocua* prophages, *S. aureus* phages 3A, 47, and 77 and also to *S. pyogenes* phage SSI-1. In this respect, B025 exemplifies the theory of modular evolution of phage genomes, which presumes DNA exchange between tailed phages by non-homologous (illegitimate) and homologous recombination, which occurs within phylogenetically related groups (Hendrix, 2002). Staphylococci and *Listeria* are grouped into the taxon *Bacillales*. However, phages are very genus-specific, and protein homologies of the concerning gene products are rather low. According to these indications,

the genetic exchange between common ancestors is likely to have taken place in the distant past.

A006

Bacteriophages A006, A500, and A118 have previously been classified as species 2671, based upon similar morphological features (Zink and Loessner, 1992). Later, results from investigations on DNA characteristics and protein analysis proposed A006 as a new species (Loessner et al., 1994b). Now that the genome sequences of these *Listeria* phages are available, they offer a clearer view of the actual grade of relatedness. With respect to analysis of the structural components, proposal of the new species A006 (Loessner et al., 1994b) is supported. Only GP12 revealed some similarity to the corresponding protein of A118 (GP13), whereas the rest of the structural module represents a patchwork of homologies to database entries of *L. monocytogenes* strains EGDe, F6854 and H7858. However, from a sequence-based point of view, A118 and A006 are clearly related; both featuring a large segment of “early” genes that are identical. Acquisition of this block of genes must have been a recent event of horizontal gene transfer since even the nucleotide sequences are almost 100% identical. In this respect, A006 is a representative example for the dilemma of phage classification according to the ICTV taxonomic system (Ackermann, 2003), which largely overlooks genomic and proteomic information. Instead, it is based on vertical transmission of genetic characteristics, and prompted proposals for a new system of phage taxonomy, which includes genomic and proteomic data, and at the same time largely maintains the classifications of the ICTV (Rohwer and Edwards, 2002; Nelson, 2004).

A500

A500 and A118 feature a strong homology of up to 99 % amino acid identity in the “late” gene cluster, except for four gene products. These comprise the major capsid protein (GP6), the endolysin (GP25), and two small gene products of unknown function (GP21, GP22), which might be putative tail fibres or base plate components. The latter two structural proteins are suggested to be involved in adhesion of the virion to the bacterial cell wall, which offers a potential explanation for the recognition of different host serovars. A118 is specific for *L. monocytogenes* serovar 1/2 strains, whereas A500 recognises serovar 4b strains (Loessner, 1991). Another hint is given by Yoichi and co-workers, who demonstrated that host specificities of bacteriophages are interchangeable by artificial alteration of a tail fibre protein

(Yoichi et al., 2005). Though the modules of late genes are colinear and highly homologous in A500 and A118 (Figure 5), there exists an even better database hit to a *L. innocua* prophage annotated in segment 1/12 of the strain Clip11262. Nine predicted gene products of A500 (Table VI.2, Appendix) are 100% identical to predicted proteins of this *L. innocua* prophage. In general, high similarity on protein level was also reflected by significant nucleotide homology, implicating quite recent horizontal exchange of DNA or relatively recent divergence from a common ancestor.

In contrast to the structural genes, the “early” gene cluster of A500 revealed a higher degree of mosaicism, and resembles prophages of *L. monocytogenes* strains EGDe and F6854, as well as *L. innocua*, again indicating recent horizontal gene transfer. Eight gene products of this cluster represent proteins homologous to A118, suggesting a more distant relatedness between A500 and A118 concerning these gene products.

A511

Comparative genomics and protein homologies of A511 indicated functional clustering of gene products in at least two large modules. Within the structural module, the phage endolysin appears to be located at an unusual position close to the large subunit of the DNA packaging enzyme terminase. The second large cluster represents the module of DNA replication and transcription comprising about 30 gene products, which encode the most important proteins required at the replication fork. The ability to constitute a phage-encoded replisome guarantees A511 a host-independent DNA replication. Modification of the bacterial RNA polymerase core enzyme to recognise phage promoter regions is probably achieved by a phage-encoded σ -factor (GP72), which shifts gene expression from transcription of bacterial genes to phage genes. The sixteen tRNAs encoded by A511 were initially thought to recognise codons that are relative minor in *Listeria* and more frequent in A511. This case was described for the tRNA genes of bacteriophage T4 (Miller et al., 2003a). However, analysis of the codon usage in A511 revealed that the tRNAs in general represent codons of high frequency (RESULTS, section 2.3). The dedicated amino acids are frequent constituents mainly of structural phage proteins and are also very common in *Listeria*. These findings indicate that the tRNAs encoded by A511 probably serve to avoid bottlenecks in “large scale” synthesis of virion components at the end of the multiplication phase. Availability of large amounts of tRNA facilitates the rapid synthesis of progeny viruses. On the other hand, there are also some codons that are rare in the entire phage coding sequences. Possibly, these

tRNAs are required in a few host strains or may enhance the expression of low-abundance proteins of A511.

A511 revealed striking sequence similarity to the genome of *Listeria* phage P100 (Carlton et al., 2005), which is reflected in protein homologies of 70-100 % amino acid identity and colinearity of the genome sequences. Therefore, P100 and A511, apparently represent closely related environmental isolates, and are the only large virulent *Listeria* bacteriophages of the *Myoviridae* family that have been investigated so far. The major difference is the size of the genomes: the DNA of P100 is 3,110 bp shorter than A511 and contains ten predicted ORFs with no counterpart in A511. In addition, the genome of A511 comprises an inserted segment of about 5,000 bp missing in P100. The genomic rearrangements are reflected in breaks of the diagonal, when the nucleotide sequences are compared directly in a Dotplot matrix (Figure 7). An unusual finding was that neither A511 nor P100 appear to encode a holin protein. Most of the large bacteriophages of both Gram-positive and Gram-negative bacteria require a holin, in combination with an endolysin, to escape from the host cell. Located close to the endolysin of A511 (GP19), the (unknown) GP21 at least meets the criteria of a putative holin, featuring three transmembrane domains that are typical for class I holins (Grundling et al., 2000). However, the characteristic two or three transmembrane domains apply to several predicted gene products distributed anywhere in the genome of A511, and therefore the putative holin function of GP21 remains to be verified. The only gene product exhibiting some similarity to a phage-encoded holin of *Bacillus subtilis* bacteriophage SPP1 (Alonso et al., 1997) was GP73, which is a putative “early” gene product. Apart from this counter-argument, there are even more reasons why GP73 seems to be implausible as a holin. First, only 20% amino acid identity appears too low to be significant, and secondly, no functional protein domains could be detected. In addition, a distant localisation of the functionally cooperating genes *ply* and *hol*, which in this case would be separated by 68,000 bp, seems rather unlikely.

Staphylococcus phage K (O'Flaherty et al., 2004) and *Lactobacillus plantarum* phage LP65 (Chibani-Chennoufi et al., 2004) were the first virulent bacteriophages of the *Myoviridae* family described that revealed similarity to structural proteins of A511 (Loessner and Scherer, 1995). With respect to the overall genome organisation and the protein homologies found, A511 is supposed to be closer related to phage K than to LP65. However, BLAST analysis between these phages revealed rather low amino acid identities of about 20-60%. Except for the late gene clusters, which were largely syntenic, the functional modules of the large virulent bacteriophages are not as well defined as the accurately organised genomes of the

temperate phages investigated in this study. Comparative genomics revealed clear rearrangements concerning the localisation of the cell lysis genes, the module of DNA replication and transcription, and the oppositional orientated cluster of unknown ORFs including the tRNA genes (Figure 8). The most peculiar reorganisation occurred in LP65: four structural proteins of A511 (GP27 and GP29-31) and a putative helicase (GP40) are located in the “leftwards” orientated genome part of LP65. Apparently, not all structural proteins necessarily have to be encoded within a single gene cluster. Separate clusters of tail genes also were described in T4-like phages (Miller et al., 2003a; Miller et al., 2003b). To conclude, the genomic organisation of the virulent bacteriophages seems to be more flexible in comparison to the temperate phages. Though, highly varied transcriptional organisation without single common genome architecture was also evident in temperate Mycobacteriophage genomes (Pedulla et al., 2003).

The colinearity found for the largest portion between the structural modules of A511, phage K and LP65 indicates that the conserved gene arrangement has proved itself for long evolutionary terms, and appears to be important for the generation of a functional virion. The reason for the genomic restructuring concerning the module of late genes remains unclear. However, these rearrangements point to illegitimate recombination events, and support the current theory of modular phage evolution (Hendrix, 2002; Casjens, 2005).

Homology search between the protein databases created from all predicted gene products of the phages investigated in this study revealed only few similarities between A511 and the other *Listeria* phages. These comprise five gene products of unknown functions encoded in the early gene cluster of A511, and another small protein (GP39), with some relatedness to A500 GP22, a putative tail fibre or base plate component. Protein similarities between the virulent and the temperate phages were expected to be low, because of their different life styles. Whereas lysogeny enables the temperate phages to recombine and exchange DNA fragments with prophages of the infected host cell at high frequency, horizontal gene transfer in virulent phages is supposed to be rare.

2. Protein profiles indicate programmed translational frameshifting

In order to correlate predicted structural gene products to protein profiles in SDS-gels prepared of purified phage particles, the most prominent bands were excised from the gels and analysed by peptide mass fingerprinting. The SDS-PAGE profiles were found to be in

agreement with those obtained by previous examinations of structural phage proteins (Zink and Loessner, 1992) except for slight differences concerning the intensity of some protein bands. Especially proteins of lower molecular weight of A511, A118 and B054 appeared to be less prominent in the gels produced in this study. In general, two major bands of a profile represented the major capsid protein Cps and the major tail protein Tsh. Except for the profiles of B054 and A511 representing phages of the *Myoviridae* family, Cps generally featured a higher molecular mass than Tsh.

MS analysis in some cases gave no clear results, and peptide fragments of two or three different predicted gene products could be identified in one examined band. A possible explanation could be the insufficient segregation of proteins with similar molecular size, such as GP35 and GP26 (band 2), or GP74 and GP31 (band 6), respectively, in the profile of A511 (Figure 11). Two-dimensional gel electrophoresis might be a possible solution to this problem, because it provides more discriminative power based on separation of proteins by isoelectric focusing in addition to molecular weight. For the same reason, this technique was applied in former investigations of structural *Listeria* phage proteins (Loessner et al, 1994b).

Unclear results of MS analysis can also arise from contaminations by protein “smear” within the lanes, which results from overloaded gels that had to be produced for the examinations of otherwise faint bands. Moreover, assignments of peptide fragments and predicted gene products to a certain degree varied dependently according the software settings of the used programme *Mascot Search* (Perkins et al., 1999).

Posttranslational modifications, e.g. processing and/or truncation of the N- or C-terminus, result in shortened peptide sequences and therefore a lower molecular weight. N-terminal truncations particularly concerned the major capsid proteins of phages A118, A006, A500, B025 (Figure 9), and A511 (Figure 11). In case of A511 and B025 posttranslational cleavage of 23 amino acids (A511) and 81 amino acids (B025) regarding the predicted peptide sequences of Cps became evident by N-terminal sequencing of these proteins (Loessner et al., 1994b). If processing is taken into account, the resulting calculated molecular weights of the mature Cps proteins of 48.7 kDa (A511) and 33 kDa (B054) are in agreement with their masses deduced from SDS-PAGE.

Cps and Tsh of A118 and A500, run at higher molecular weight in the SDS-gel than predicted by bioinformatical analysis of the genome sequences (RESULTS, section 3). The synthesis of the large proteins Cps-L and Tsh-L is considered to be due to translational frameshifting into small downstream overlapping reading frames, which in neither A500 nor A118 exhibit a

specific protein function. The results obtained with bioinformatic analysis indicating frameshifting in A118 and A500 were confirmed by MALDI-MS (Bielmann, unpublished data).

Bioinformatics indicated that in both cases a -1 shift at a slippery sequence G'GGA results in a larger Cps-L protein, whereas Tsh-L is synthesised by a +1 shift. Here, the slippery site CCC'U on the mRNA overlaps the stop codon UGA of the original *tsh* reading frame (Figure 10), which has been designated as “shifty stop” and was described for Tsh-L of phage PSA, (Zimmer et al., 2003). According to the authors, during the translation process the ribosome appears to slip from the prolin-encoding codon CCC into the overlapping prolin codon CCU. This event might be supported by the availability of the corresponding tRNAs. In the low G+C genome of *Listeria* the codon CCU is more frequently used than the rare CCC, the deficiency of which possibly promotes the ribosome to slip.

In general, translational frameshifts during synthesis of major structural proteins were described for several temperate bacteriophages (Levin et al., 1993; Weisberg et al., 1999; Christie et al., 2002; Zimmer, 2002, Zimmer et al., 2003). They were shown to occur with a low frequency of 4-10% and resulted in the production of an essential variant of a structural component sharing identical N-termini but featuring different C-terminal extensions. In contrast, the abundance of the C-terminally modified Cps-L and Tsh-L in A500 seems to be high, because they represent the major protein bands in the profile of structural proteins. In case of PSA, the variants of the major capsid protein Cps and Cps-L, respectively, are assumed to represent the pentameric and hexameric protein subunits of the virion head due to their correspondent ratios of small and large protein (Zimmer et al., 2003). Possibly, this is also true for A500, though the Cps/Cps-L ratio is reversed in comparison to PSA.

3. Site specific integration and attachment sites

In contrast to the genetic regulation of lysogeny, insertion of the viral DNA into the host chromosome appears to be secondary in establishing lysogeny. For example, prophages of the temperate coliphages P1 and N15 reside as plasmids in the host cytoplasm (Rybchin and Svarchevsky, 1999; Li and Austin, 2002). The temperate *Listeria* phages A006, A500, B025, and B054 investigated in this study use the mechanism of site specific recombination and integration of their genomes to reside as prophages within their host.

The host attachment sites of A006, A500, and B025 are located at the 3'-termini of tRNA genes, which are single copy genes for specific codons in each case. Sequencing indicated that they are reconstituted and therefore kept functional. The reconstitution is important to avoid an impairment of translation in the host. Integration close to the 3'-end of tRNA genes also has been shown for P4-related phages, whereas others like the lambdoid P22 integrate into the anticodon loop (Campbell, 2003a).

The bacterial attachment site of B054 is supposed to be located at the 3'-end of the translation elongation factor EF-Ts (lin1766) (Glaser et al., 2001). The elongation factor promotes the exchange of GDP to GTP at the complex EF-Tu-GDP resulting in regeneration of the active form of EF-Tu (Knippers, 2001). During the translation process, the activated elongation factor EF-Tu mediates binding of tRNA to the mRNA codon at the ribosome. In general, attachment sites located in conserved genome regions seem to be "preferable" for phages, because they constitute almost invariable sequences, and "guarantee" site specific integration.

The core sequence of the phage attachment site *attP* of B025 was shown to be identical to *attP* of PSA as described by Lauer and co-workers (Lauer et al., 2002). Both phages integrate into the 3'-terminus of a tRNA-Arg gene in *L. monocytogenes* EGDe. In agreement with this result, the predicted integrases of these bacteriophages revealed a high amino acid identity of 95%. But a high degree of homology does not necessarily implicate that the enzymes are fully functionally interchangeable. Examples are the related integrases of the well studied group of lambdoid phages. Phage λ and phage 434 insert at the same chromosomal site (Campbell, 1994) as do phage 21 and the defective element e14 (Wang et al., 1997). Whereas λ and 434 can cure a lysogen containing a heterologous prophage on superinfection, the integrases of 21 and e14 can not complement each other, because each has its own specific mechanism of site recognition. In summary, several integrase molecules are required to bind N-terminally to specific DNA sequences within the 240-bp *attP* DNA. The catalytic C-termini are positioned close to the crossover point due to supercoiling of the phage DNA and bending by the integration host factor (IHF). The actual crossover event takes place between *attB* and the core sequence of *attP*. Although PSA and B025 insert into the same core sequences, the surrounding nucleotide sequences of *attP* only share 91% nucleotide identity. Whether or not the phage integrases will complement each other and are interchangeable has not been investigated so far.

The integrases, which mediate site specific recombination, can generally be grouped into two major families, the tyrosine recombinases and serine recombinases (Groth and Calos, 2004).

Bioinformatics indicates that the predicted integrases of the *Listeria* phages investigated here are members of the tyrosine family, which utilise a catalytic tyrosine to mediate strand cleavage, and tend to recognise longer *attP* sequences. The integrases revealed some similarity to XerC and XerD, which are non-integrase family members of tyrosine recombinases, and recognise simple identical core sequences of about 50 bp or less (Groth and Calos, 2004). In contrast, A118 is the only *Listeria* phage featuring an integrase of the serine family, which recognises short regions of identity above 3 bp (Loessner et al, 2000).

4. Clues on *Listeria* phage evolution

According to the current theory of phage evolution (Casjens, 1992; Hendrix et al., 1999; Hendrix, 2003; Hendrix et al., 2003; Casjens, 2005), the global population of tailed bacteriophage genomes represents a common gene pool. The mosaic nature of the phage genomes is based upon horizontal exchange of sequences, by homologous and non-homologous recombination events that take place during infection. A non-homologous horizontal exchange is suggested to cause genomic rearrangements and to create new sequences, whereas a homologous recombination may play the major role in distribution of novel genes within a phage population multiplying on a common host (Hendrix, 2002).

In agreement with this theory, comparative analysis of the *Listeria* phage genomes revealed their mosaic nature, indicating a graded and reticulate relatedness between the *Listeria* phages investigated. Based on nucleotide and amino acid sequence homologies, A006, A500, and A118, as well as B025 and PSA form two different groups of close related phages. Both groups are considered to be linked via homologies to the non-inducible defective prophages of *L. innocua* Clip11262, with which the infecting phages could interact. In contrast, B054 shares few homologies to the other *Listeria* phages investigated, because it revealed high homologies to the only functional prophage of *L. innocua*. The genomic alignments of homologies mapped in Figure 5 represent one of various combinatory possibilities, but it visualises as best as possible the data discussed above. A suitable illustration summarising the degree of genomic or proteomic similarity between the *Listeria* phages investigated would be a kind of rootless family tree. Figure 12 represents a phylogram obtained by alignment of genome sequences via ClustalW including all temperate *Listeria* phages investigated so far. P35 is not included in the alignment, because it would be wrongly grouped with B054. The strong attraction of genome sequences that are very different from each other seems to be an

apparent problem of the programme, which was also recognised by Rohwer and Ewards (2002), who developed a sequence-based taxonomic system called “Phage Proteomic Tree”. The system is based on the predicted phage proteomes, and the problem of grouping distantly related sequences was resolved by imposing a penalty for every ORF that two genomes do not share with each other.

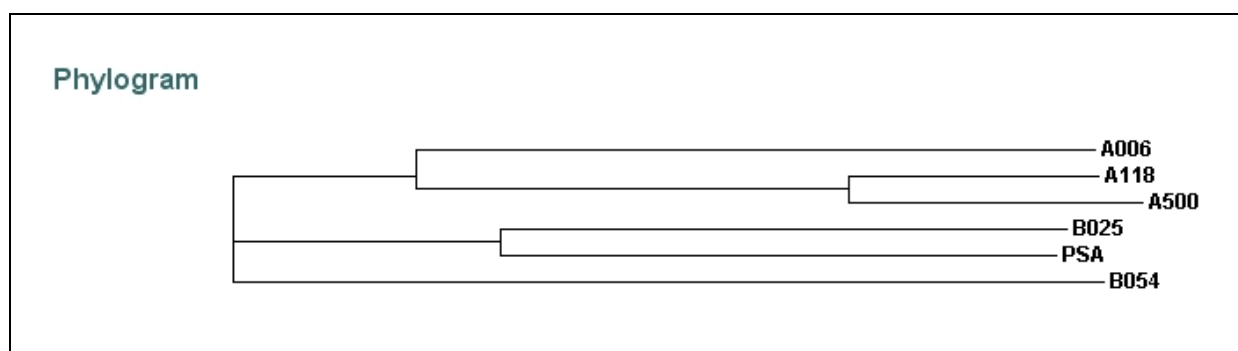


Fig. 12. Rootless phylogenetic tree of the temperate *Listeria* phages. Based on alignment of the genome sequences via ClustalW, the phages are assorted into two groups of close relationship, one of which includes A006, A118 and A500. The second group comprises the cos-site phages PSA and B025. B054 appears to be more distantly related.

A more complete but also more complicated picture of the actual genome alignment in Figure 5 or the phylogram in Figure 12 would arise by integrating the various prophages annotated in the genomes of *L. monocytogenes* strains EGDe, H7858, F6854, F2365, and *L. innocua* Clip11262. However, the virulent P35 still would be the odd one out due to the lack of protein homologies. At the present state of knowledge, P35 seems to be only distantly related to all the other *Listeria* phages investigated and probably represents a line of viruses that diverged at an early point from *Enterococcus* phages. Synteny of the overall genomic organisation solely indicates an ancient divergence of the ancestors (Brussow and Hendrix, 2002). It still remains unclear whether virulent phages gain only limited access to genetic exchange due to hydrolysis of the host DNA upon infection. This possibly implicates that the virulent phages form biological coherent groups with limited opportunities for horizontal gene transfer outside these groups (Hendrix, 2003). On the other hand, the presence of temperate phage genes on virulent phage genomes has been proven for A511 in this study, and was detected in

other virulent phages (Hendrix, 2003; Chibani-Chennoufi et al., 2004; Casjens, 2005) indicating that horizontal gene transfer between different groups actually works. After all, P35 might have originated from a temperate ancestor by deletion of the lysogeny module; even though this hypothesis seems rather unlikely.

To conclude, the only way to clear the unresolved questions on relationship between *Listeria* bacteriophages is to collect new sequence data, which will gradually complete our understanding of phylogeny and evolution of *Listeria* phages.

V. REFERENCES

- Ackermann, H.W.** (2003). Bacteriophage observations and evolution. *Res Microbiol* **154**, 245-251.
- Alonso, J.C., Luder, G., Stiege, A.C., Chai, S., Weise, F., and Trautner, T.A.** (1997). The complete nucleotide sequence and functional organization of *Bacillus subtilis* bacteriophage SPP1. *Gene* **204**, 201-212.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.
- Becker, A., and Murialdo, H.** (1990). Bacteriophage lambda DNA: the beginning of the end. *J Bacteriol* **172**, 2819-2824.
- Belfort, M., and Roberts, R.J.** (1997). Homing endonucleases: keeping the house in order. *Nucleic Acids Res* **25**, 3379-3388.
- Ben Embarek, P.K.** (1994). Presence, detection and growth of *Listeria monocytogenes* in seafoods: a review. *Int J Food Microbiol* **23**, 17-34.
- Bernhardt, T.G., Wang, I.N., Struck, D.K., and Young, R.** (2002). Breaking free: "protein antibiotics" and phage lysis. *Res Microbiol* **153**, 493-501.
- Bille, J., and Doyle, M.P.** (1991). *Listeria* and *Erysipelothrix*. In *Manual of Clinical Microbiology*, A. Balaous, W.J. Hausler, K.L. Herrman, H.D. Isenberg, and H.J. Shadomy, eds (Washington: American Society of Microbiology), pp. 1364.
- Birge, E.A.** (1994). *Bacterial and bacteriophage genetics*. (New York: Springer Verlag), pp. 454
- Bonovich, M.T., and Young, R.** (1991). Dual start motif in two lambdoid S genes unrelated to lambda S. *J Bacteriol* **173**, 2897-2905.
- Bossi, L., Fuentes, J.A., Mora, G., and Figueroa-Bossi, N.** (2003). Prophage contribution to bacterial population dynamics. *J Bacteriol* **185**, 6467-6471.
- Brussow, H.** (2001). Phages of dairy bacteria. *Annu Rev Microbiol* **55**, 283-303.
- Brussow, H., and Desiere, F.** (2001). Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. *Mol Microbiol* **39**, 213-222.
- Brussow, H., and Hendrix, R.W.** (2002). Phage genomics: small is beautiful. *Cell* **108**, 13-16.
- Brussow, H., Canchaya, C., and Hardt, W.D.** (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**, 560-602.

-
- Bruttin, A., and Brussow, H.** (1996). Site-specific spontaneous deletions in three genome regions of a temperate *Streptococcus thermophilus* phage. *Virology* **219**, 96-104.
- Buchrieser, C., Rusniok, C., Kunst, F., Cossart, P., and Glaser, P.** (2003). Comparison of the genome sequences of *Listeria monocytogenes* and *Listeria innocua*: clues for evolution and pathogenicity. *FEMS Immunol Med Microbiol* **35**, 207-213.
- Campbell, A.** (1994). Comparative molecular biology of lambdoid phages. *Annu Rev Microbiol* **48**, 193-222.
- Campbell, A.** (2003a). Prophage insertion sites. *Res Microbiol* **154**, 277-282.
- Campbell, A.** (2003b). The future of bacteriophage biology. *Nat Rev Genet* **4**, 471-477.
- Carlton, R.M., Noordman, W.H., Biswas, B., de Meester, E.D., and Loessner, M.J.** (2005). Bacteriophage P100 for control of *Listeria monocytogenes* in foods: genome sequence, bioinformatic analyses, oral toxicity study, and application. *Regul Toxicol Pharmacol* **43**, 301-312.
- Casjens, S.** (1992). Evolution of dsDNA tailed-bacteriophage genomes. *Virology* **3**, 383-397.
- Casjens, S.R.** (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* **8**, 451-458.
- Chibani-Chennoufi, S., Dillmann, M.L., Marvin-Guy, L., Rami-Shojaei, S., and Brussow, H.** (2004). *Lactobacillus plantarum* bacteriophage LP65: a new member of the SPO1-like genus of the family Myoviridae. *J Bacteriol* **186**, 7069-7083.
- Christie, G.E., Temple, L.M., Bartlett, B.A., and Goodwin, T.S.** (2002). Programmed translational frameshift in the bacteriophage P2 FETUD tail gene operon. *J Bacteriol* **184**, 6522-6531.
- Cole, M.B., Jones, M.V., and Holyoak, C.** (1990). The effect of pH, salt concentration and temperature on the survival and growth of *Listeria monocytogenes*. *J Appl Bacteriol* **69**, 63-72.
- Cossart, P.** (2002). Molecular and cellular basis of the infection by *Listeria monocytogenes*: an overview. *Int J Med Microbiol* **291**, 401-409.
- Cossart, P., and Lecuit, M.** (1998). Interactions of *Listeria monocytogenes* with mammalian cells during entry and actin-based movement: bacterial factors, cellular ligands and signaling. *Embo J* **17**, 3797-3806.
- Cummins, A.J., Fielding, A.K., and McLauchlin, J.** (1994). *Listeria ivanovii* infection in a patient with AIDS. *J Infect* **28**, 89-91.
- Dalton, C.B., Austin, C.C., Sobel, J., Hayes, P.S., Bibb, W.F., Graves, L.M., Swaminathan, B., Proctor, M.E., and Griffin, P.M.** (1997). An outbreak of gastroenteritis and fever due to *Listeria monocytogenes* in milk. *N Engl J Med* **336**, 100-105.

- Desiere, F., Mahanivong, C., Hillier, A.J., Chandry, P.S., Davidson, B.E., and Brussow, H. (2001). Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology* **283**, 240-252.
- Dietrich, G., Bubert, A., Gentschev, I., Sokolovic, Z., Simm, A., Catic, A., Kaufmann, S.H., Hess, J., Szalay, A.A., and Goebel, W. (1998). Delivery of antigen-encoding plasmid DNA into the cytosol of macrophages by attenuated suicide *Listeria monocytogenes*. *Nat Biotechnol* **16**, 181-185.
- Dodd, I.B., Perkins, A.J., Tsemitsidis, D., and Egan, J.B. (2001). Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev* **15**, 3013-3022.
- Dodd, I.B., Shearwin, K.E., Perkins, A.J., Burr, T., Hochschild, A., and Egan, J.B. (2004). Cooperativity in long-range gene regulation by the lambda CI repressor. *Genes Dev* **18**, 344-354.
- Domann, E., Wehland, J., Rohde, M., Pistor, S., Hartl, M., Goebel, W., Leimeister-Wachter, M., Wuenscher, M., and Chakraborty, T. (1992). A novel bacterial virulence gene in *Listeria monocytogenes* required for host cell microfilament interaction with homology to the proline-rich region of vinculin. *Embo J* **11**, 1981-1990.
- Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127-6145.
- Duckworth, D.H. (1976). "Who discovered bacteriophage?" *Bacteriol Rev* **40**, 793-802.
- Dussurget, O., Cabanes, D., Dehoux, P., Lecuit, M., Buchrieser, C., Glaser, P., and Cossart, P. (2002). *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol Microbiol* **45**, 1095-1106.
- Farber, J.M., and Peterkin, P.I. (1991). *Listeria monocytogenes*, a food-borne pathogen. *Microbiol Rev* **55**, 476-511.
- Fujisawa, H., and Morita, M. (1997). Phage DNA packaging. *Genes Cells* **2**, 537-545.
- Gaeng, S., Scherer, S., Neve, H., and Loessner, M.J. (2000). Gene cloning and expression and secretion of *Listeria monocytogenes* bacteriophage-lytic enzymes in *Lactococcus lactis*. *Appl Environ Microbiol* **66**, 2951-2958.
- Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couve, E., de Daruvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Fsihi, H., Garcia-del Portillo, F., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueno, E., Maitournam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Perez-Diaz, J.C., Purcell, R., Remmel, B., Rose, M.,

-
- Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J., and Cossart, P.** (2001). Comparative genomics of *Listeria* species. *Science* **294**, 849-852.
- Groth, A.C., and Calos, M.P.** (2004). Phage integrases: biology and applications. *J Mol Biol* **335**, 667-678.
- Grundling, A., Blasi, U., and Young, R.** (2000). Biochemical and genetic evidence for three transmembrane domains in the class I holin, lambda S. *J Biol Chem* **275**, 769-776.
- Guhan, N., and Muniyappa, K.** (2003). Structural and functional characteristics of homing endonucleases. *Crit Rev Biochem Mol Biol* **38**, 199-248.
- Hendrix, R.W.** (2002). Bacteriophages: evolution of the majority. *Theor Popul Biol* **61**, 471-480.
- Hendrix, R.W.** (2003). Bacteriophage genomics. *Curr Opin Microbiol* **6**, 506-511.
- Hendrix, R.W., Hatfull, G.F., and Smith, M.C.** (2003). Bacteriophages with tails: chasing their origins and evolution. *Res Microbiol* **154**, 253-257.
- Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., and Casjens, S.** (2000). The origins and ongoing evolution of viruses. *Trends Microbiol* **8**, 504-508.
- Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F.** (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci USA* **96**, 2192-2197.
- Hodgson, D.A.** (2000). Generalized transduction of serotype 1/2 and serotype 4b strains of *Listeria monocytogenes*. *Molecular Microbiology* **35**, 312-323.
- Junttila, J.R., Niemela, S.I., and Hirn, J.** (1988). Minimum growth temperatures of *Listeria monocytogenes* and non-haemolytic *Listeria*. *J Appl Bacteriol* **65**, 321-327.
- Katsura, I.** (1990). Mechanism of length determination in bacteriophage lambda tails. *Adv Biophys* **26**, 1-18.
- Klaus, S., Krüger, D., and Meyer, J.** (1992). Bakterienviren. (Jena: Gustav Fischer Verlag), pp. 300.
- Knippers, R.** (2001). Molekulare Genetik. (Stuttgart: Georg Thieme Verlag), pp. 586
- Kutter, E., and Sulakvelidze, A.** (2005). Bacteriophages: biology and applications. (Boca Raton: CRC Press), pp. 510.
- Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J.** (2005). The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci USA* **102**, 5174-5179.

-
- Laemmli, U.K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Lauer, P., Chow, M.Y., Loessner, M.J., Portnoy, D.A., and Calendar, R.** (2002). Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J Bacteriol* **184**, 4177-4186.
- Lazarevic, V., Dusterhoft, A., Soldo, B., Hilbert, H., Mael, C., and Karamata, D.** (1999). Nucleotide sequence of the *Bacillus subtilis* temperate bacteriophage SPbetac2. *Microbiology* **145**, 1055-1067.
- Lebek, G., Teyssere, P., and Baumgartner, A.** (1993). A method for typing *Listeria monocytogenes* strains by classification of listeriocins and phage receptors. *Zentralbl Bakteriol* **278**, 58-68.
- Leverentz, B., Conway, W.S., Alavidze, Z., Janisiewicz, W.J., Fuchs, Y., Camp, M.J., Chighladze, E., and Sulakvelidze, A.** (2001). Examination of bacteriophage as a biocontrol method for *Salmonella* on fresh-cut fruit: a model study. *J Food Prot* **64**, 1116-1121.
- Leverentz, B., Conway, W.S., Camp, M.J., Janisiewicz, W.J., Abuladze, T., Yang, M., Saftner, R., and Sulakvelidze, A.** (2003). Biocontrol of *Listeria monocytogenes* on fresh-cut produce by treatment with lytic bacteriophages and a bacteriocin. *Appl Environ Microbiol* **69**, 4519-4526.
- Levin, M.E., Hendrix, R.W., and Casjens, S.R.** (1993). A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *J Mol Biol* **234**, 124-139.
- Li, Y., and Austin, S.** (2002). The P1 plasmid in action: time-lapse photomicroscopy reveals some unexpected aspects of plasmid partition. *Plasmid* **48**, 174-178.
- Loessner, M.J.** (1991). Improved procedure for bacteriophage typing of *Listeria* strains and evaluation of new phages. *Appl Environ Microbiol* **57**, 882-884.
- Loessner, M.J.** (2005). Bacteriophage endolysins - current state of research and applications. *Curr Opin Microbiol* **8**, 480-487.
- Loessner, M.J., and Busse, M.** (1990). Bacteriophage typing of *Listeria* species. *Appl Environ Microbiol* **56**, 1912-1918.
- Loessner, M.J., and Scherer, S.** (1995). Organization and transcriptional analysis of the *Listeria* phage A511 late gene region comprising the major capsid and tail sheath protein genes cps and tsh. *J Bacteriol* **177**, 6601-6609.
- Loessner, M.J., Goepl, S., and Busse, M.** (1991). Comparative inducibility of bacteriophage in naturally lysogenic and lysogenized strains of *Listeria* spp. by UV light and Mitomycin C. *Lett Appl Microbiol* **12**, 196-199.

-
- Loessner, M.J., Wendlinger, G., and Scherer, S.** (1995a). Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. *Mol Microbiol* **16**, 1231-1241.
- Loessner, M.J., Schneider, A., and Scherer, S.** (1995b). A new procedure for efficient recovery of DNA, RNA, and proteins from *Listeria* cells by rapid lysis with a recombinant bacteriophage endolysin. *Appl Environ Microbiol* **61**, 1150-1152.
- Loessner, M.J., Rudolf, M., and Scherer, S.** (1997). Evaluation of luciferase reporter bacteriophage A511::luxAB for detection of *Listeria monocytogenes* in contaminated foods. *Appl Environ Microbiol* **63**, 2961-2965.
- Loessner, M.J., Estela, L.A., Zink, R., and Scherer, S.** (1994a). Taxonomical classification of 20 newly isolated *Listeria* bacteriophages by electron microscopy and protein analysis. *Intervirology* **37**, 31-35.
- Loessner, M.J., Krause, I.B., Henle, T., and Scherer, S.** (1994b). Structural proteins and DNA characteristics of 14 *Listeria* typing bacteriophages. *J Gen Virol* **75**, 701-710.
- Loessner, M.J., Rees, C.E., Stewart, G.S., and Scherer, S.** (1996). Construction of luciferase reporter bacteriophage A511::luxAB for rapid and sensitive detection of viable *Listeria* cells. *Appl Environ Microbiol* **62**, 1133-1140.
- Loessner, M.J., Inman, R.B., Lauer, P., and Calendar, R.** (2000). Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. *Mol Microbiol* **35**, 324-340.
- Loessner, M.J., Kramer, K., Ebel, F., and Scherer, S.** (2002). C-terminal domains of *Listeria monocytogenes* bacteriophage murein hydrolases determine specific recognition and high-affinity binding to bacterial cell wall carbohydrates. *Mol Microbiol* **44**, 335-349.
- Lorber, B.** (1997). Listeriosis. *Clin Infect Dis* **24**, 1-9.
- Lowe, T.M., and Eddy, S.R.** (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955-964.
- Lucchini, S., Desiere, F., and Brussow, H.** (1999). Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J Virol* **73**, 8647-8656.
- Lynch, T.W., Read, E.K., Mattis, A.N., Gardner, J.F., and Rice, P.A.** (2003). Integration host factor: putting a twist on protein-DNA recognition. *J Mol Biol* **330**, 493-502.
- Madigen, T.M., Martinko, J.M., and Parker, J.** (1997). *Biology of Microorganisms*. (Upper Saddle River, New Jersey: Prentice-Hall, Inc), pp. 1038.

- Mahanivong, C., Boyce, J.D., Davidson, B.E., and Hillier, A.J.** (2001). Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. *Appl Environ Microbiol* **67**, 3564-3576.
- Marchler-Bauer, A., and Bryant, S.H.** (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, W327-331.
- Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., and Bryant, S.H.** (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33**, D192-196.
- McGrath, S., Fitzgerald, G.F., and van Sinderen, D.** (2004). The impact of bacteriophage genomics. *Curr Opin Biotechnol* **15**, 94-99.
- McLauchlin, J.** (1990a). Human listeriosis in Britain, 1967-85, a summary of 722 cases. 2. Listeriosis in non-pregnant individuals, a changing pattern of infection and seasonal incidence. *Epidemiol Infect* **104**, 191-201.
- McLauchlin, J.** (1990b). Human listeriosis in Britain, 1967-85, a summary of 722 cases. 1. Listeriosis during pregnancy and in the newborn. *Epidemiol Infect* **104**, 181-189.
- McLauchlin, J., Mitchell, R.T., Smerdon, W.J., and Jewell, K.** (2004). *Listeria monocytogenes* and listeriosis: a review of hazard characterisation for use in microbiological risk assessment of foods. *Int J Food Microbiol* **92**, 15-33.
- Mikkonen, M., Dupont, L., Alatosava, T., and Ritzenthaler, P.** (1996). Defective site-specific integration elements are present in the genome of virulent bacteriophage LL-H of *Lactobacillus delbrueckii*. *Appl Environ Microbiol* **62**, 1847-1851.
- Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W.** (2003a). Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156.
- Miller, E.S., Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Durkin, A.S., Ciecko, A., Feldblyum, T.V., White, O., Paulsen, I.T., Nierman, W.C., Lee, J., Szczypinski, B., and Fraser, C.M.** (2003b). Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol* **185**, 5220-5233.
- Nelson, D.** (2004). Phage taxonomy: we agree to disagree. *J Bacteriol* **186**, 7029-7031.
- Nelson, K.E., Fouts, D.E., Mongodin, E.F., Ravel, J., DeBoy, R.T., Kolonay, J.F., Rasko, D.A., Angiuoli, S.V., Gill, S.R., Paulsen, I.T., Peterson, J., White, O., Nelson, W.C., Nierman, W., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., Dodson, R.J., Durkin, A.S., Madupu, R., Haft, D.H., Selengut, J., Van Aken, S., Khouri, H., Fedorova, N., Forberger, H., Tran, B., Kathariou, S., Wonderling, L.D., Uhlich, G.A., Bayles, D.O., Luchansky, J.B., and Fraser, C.M.** (2004). Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria*

- monocytogenes* reveal new insights into the core genome components of this species. *Nucleic Acids Res* **32**, 2386-2395.
- Obregon, V., Garcia, J.L., Garcia, E., Lopez, R., and Garcia, P.** (2003). Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. *J Bacteriol* **185**, 2362-2368.
- Ochman, H., Gerber, A.S., and Hartl, D.L.** (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.
- O'Flaherty, S., Coffey, A., Edwards, R., Meaney, W., Fitzgerald, G.F., and Ross, R.P.** (2004). Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content. *J Bacteriol* **186**, 2862-2871.
- Paulsen, I.T., Banerjee, L., Myers, G.S., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., and Fraser, C.M.** (2003). Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* **299**, 2071-2074.
- Pedre, X., Weise, F., Chai, S., Luder, G., and Alonso, J.C.** (1994). Analysis of cis and trans acting elements required for the initiation of DNA replication in the *Bacillus subtilis* bacteriophage SPP1. *J Mol Biol* **236**, 1324-1340.
- Pedula, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs, W.R., Jr., Hendrix, R.W., and Hatfull, G.F.** (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**, 171-182.
- Peel, M., Donachie, W., and Shaw, A.** (1988). Temperature-dependent expression of flagella of *Listeria monocytogenes* studied by electron microscopy, SDS-PAGE and western blotting. *J Gen Microbiol* **134**, 2171-2178.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S.** (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.
- Pini, P.N., and Gilbert, R.J.** (1988). The occurrence in the U.K. of *Listeria* species in raw chickens and soft cheeses. *Int J Food Microbiol* **6**, 317-326.
- Rocourt, J.** (1986). [Bacteriophages and bacteriocins of the genus *Listeria*]. *Zentralbl Bakteriol Mikrobiol Hyg [A]* **261**, 12-28.
- Rohwer, F.** (2003). Global phage diversity. *Cell* **113**, 141.
- Rohwer, F., and Edwards, R.** (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* **184**, 4529-4535.

-
- Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., and Leiman, P.G.** (2004). The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol* **14**, 171-180.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B.** (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.
- Rybchin, V.N., and Svarchevsky, A.N.** (1999). The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol Microbiol* **33**, 895-903.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A.** (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.
- Sambrook, J., and Russell, D.W.** (2001). *Molecular Cloning - a Laboratory Manual*. (Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press), pp. 999.
- Schuchat, A., Swaminathan, B., and Broome, C.V.** (1991). Epidemiology of human listeriosis. *Clin Microbiol Rev* **4**, 169-183.
- Schultz, E.W.** (1945). *Listerella* infections: a review. *Stanford Medical Bulletin* **3**, 135-151.
- Shearwin, K.E., Dodd, I.B., and Egan, J.B.** (2002). The helix-turn-helix motif of the coliphage 186 immunity repressor binds to two distinct recognition sequences. *J Biol Chem* **277**, 3186-3194.
- Sitbon, E., and Pietrokovski, S.** (2003). New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem Sci* **28**, 473-477.
- Sleator, R.D., Wemekamp-Kamphuis, H.H., Gahan, C.G., Abee, T., and Hill, C.** (2005). A PrfA-regulated bile exclusion system (Bile) is a novel virulence factor in *Listeria monocytogenes*. *Mol Microbiol* **55**, 1183-1195.
- Southwick, F.S., and Purich, D.L.** (1996). Intracellular pathogenesis of listeriosis. *N Engl J Med* **334**, 770-776.
- Stoddart, B.L.** (2005). Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49-95.
- Svenningsen, S.L., Costantino, N., Court, D.L., and Adhya, S.** (2005). On the role of Cro in lambda prophage induction. *Proc Natl Acad Sci USA* **102**, 4465-4469.
- Thompson, J.D., Higgins, D.G., and 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.
- Tran, H.L., Fiedler, F., Hodgson, D.A., and Kathariou, S.** (1999). Transposon-induced mutations in two loci of *Listeria monocytogenes* serotype 1/2a result in phage

- resistance and lack of N-acetylglucosamine in the teichoic acid of the cell wall. *Appl Environ Microbiol* **65**, 4793-4798.
- Trussel, M.** (1989). [The occurrence of *Listeria* in the production of processed meat, salami and mettwurst]. *Schweiz Arch Tierheilkd* **131**, 409-412,417-421.
- van der Mee-Marquet, N., Loessner, M., and Audurier, A.** (1997). Evaluation of seven experimental phages for inclusion in the international phage set for the epidemiological typing of *Listeria monocytogenes*. *Appl Environ Microbiol* **63**, 3374-3377.
- Vazquez-Boland, J.A., Dominguez-Bernal, G., Gonzalez-Zorn, B., Kreft, J., and Goebel, W.** (2001a). Pathogenicity islands and virulence evolution in *Listeria*. *Microbes Infect* **3**, 571-584.
- Vazquez-Boland, J.A., Kuhn, M., Berche, P., Chakraborty, T., Dominguez-Bernal, G., Goebel, W., Gonzalez-Zorn, B., Wehland, J., and Kreft, J.** (2001b). *Listeria pathogenesis* and molecular virulence determinants. *Clin Microbiol Rev* **14**, 584-640.
- Vukov, N., Scherer, S., Hibbert, E., and Loessner, M.J.** (2000). Functional analysis of heterologous holin proteins in a lambdaDeltaS genetic background. *FEMS Microbiol Lett* **184**, 179-186.
- Vukov, N., Moll, I., Blasi, U., Scherer, S., and Loessner, M.J.** (2003). Functional regulation of the *Listeria monocytogenes* bacteriophage A118 holin by an intragenic inhibitor lacking the first transmembrane domain. *Mol Microbiol* **48**, 173-186.
- Wang, H., Yang, C.H., Lee, G., Chang, F., Wilson, H., del Campillo-Campbell, A., and Campbell, A.** (1997). Integration specificities of two lambdoid phages (21 and e14) that insert at the same attB site. *J Bacteriol* **179**, 5705-5711.
- Wang, I.N., Smith, D.L., and Young, R.** (2000). Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54**, 799-825.
- Weis, J., and Seeliger, H.P.** (1975). Incidence of *Listeria monocytogenes* in nature. *Appl Microbiol* **30**, 29-32.
- Weisberg, R.A., Gottesmann, M.E., Hendrix, R.W., and Little, J.W.** (1999). Family values in the age of genomics: comparative analyses of temperate bacteriophage HK022. *Annu Rev Genet* **33**, 565-602.
- Wendlinger, G., Loessner, M.J., and Scherer, S.** (1996). Bacteriophage receptors on *Listeria monocytogenes* cells are the N-acetylglucosamine and rhamnose substituents of teichoic acids or the peptidoglycan itself. *Microbiology* **142**, 985-992.
- Wommack, K.E., and Colwell, R.R.** (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**, 69-114.

-
- Yoichi, M., Abe, M., Miyanaga, K., Unno, H., and Tanji, Y.** (2005). Alteration of tail fiber protein gp38 enables T2 phage to infect *Escherichia coli* O157:H7. *J Biotechnol* **115**, 101-107.
- Young, I., Wang, I., and Roof, W.D.** (2000). Phages will out: strategies of host cell lysis. *Trends Microbiol* **8**, 120-128.
- Zdobnov, E.M., and Apweiler, R.** (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848.
- Zdobnov, E.M., Lopez, R., Apweiler, R., and Etzold, T.** (2002). The EBI SRS server--recent developments. *Bioinformatics* **18**, 368-373.
- Zimmer, M.** (2002). Complete genome sequence and characterization of the lysis system of the temperate *Clostridium perfringens* bacteriophage Φ 3626. In Institut für Mikrobiologie, Forschungszentrum für Milch und Lebensmittel Weihenstephan (Technische Universität München), pp. 117.
- Zimmer, M., Scherer, S., and Loessner, M.J.** (2002). Genomic analysis of *Clostridium perfringens* bacteriophage phi3626, which integrates into *guaA* and possibly affects sporulation. *J Bacteriol* **184**, 4359-4368.
- Zimmer, M., Sattelberger, E., Inman, R.B., Calendar, R., and Loessner, M.J.** (2003). Genome and proteome of *Listeria monocytogenes* phage PSA: an unusual case for programmed +1 translational frameshifting in structural protein synthesis. *Mol Microbiol* **50**, 303-317.
- Zink, R., and Loessner, M.J.** (1992). Classification of virulent and temperate bacteriophages of *Listeria* spp. on the basis of morphology and protein analysis. *Appl Environ Microbiol* **58**, 296-302.
- Zink, R., Loessner, M.J., and Scherer, S.** (1995). Characterization of cryptic prophages (monocins) in *Listeria* and sequence analysis of a holin/endolysin gene. *Microbiology* **141**, 2577-2584.
- Zink, R., Loessner, M.J., Glas, I., and Scherer, S.** (1994). Supplementary *Listeria*-typing with defective *Listeria* phage particles (monocins). *Lett Appl Microbiol* **19**, 99-101.

VI. PUBLICATIONS

- Dorscht, J., Klumpp, J., and Loessner, M.J.** (2006) *Listeria* bacteriophage genomics. Int J Med Microbiol (submitted for publication).
- Dorscht, J., Schmelcher, M., Calendar, R., Zimmer, M. and Loessner, M.J.** (2004) Comparative genomics of *Listeria* phages. Presented at the XV International Symposium on Problems of Listeriosis (ISOPOL), Sep 12-15, Uppsala, Sweden
- Dorscht, J., Zimmer, M., Schmelcher, M., Biemann, R., Calendar, R., and Loessner, M.J.** Comparative genomics and proteomics of *Listeria* bacteriophages: Extensive sequence mosaicism and programmed translational frameshifting as common elements (in preparation).
- Klumpp, J., Dorscht, J., Zimmer, M., Lurz, R., Calendar, R., and Loessner, M.J.** The large broad host range virulent *Listeria* bacteriophage A511 reveals an unusual genome organization (in preparation).

VII. APPENDIX

Bioinformatical results and protein similarities of *Listeria* bacteriophages A006, A500, B025, B054, P35, and A511

The annotation tables listed below comprise general features of the predicted gene products including the localisations of start and stop codons of the reading frames, the isoelectric point (pI), the length of the predicted peptide sequence represented by the number of amino acids (AA) and the molecular weight (MW). Protein similarities of the predicted gene products were identified using BlastP. Gene products that were annotated on the complementary strand are marked with an asterisk*.

Table VI.1. General features and database matches of the predicted gene products of bacteriophage A006.

A006	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp1	28	768	9.3	246	27.7	<i>L. innocua</i> lin2395 terminase small subunit, <i>L. monocytogenes</i> F6854 and H7858
gp2	761	2080	5.7	439	50.7	<i>L. innocua</i> lin2394 terminase large subunit, <i>L. monocytogenes</i> H7858 and F6854
gp3	2095	3651	5.5	518	58.7	<i>L. monocytogenes</i> F6854, <i>L. innocua</i> lin2393, <i>L. monocytogenes</i> H7858
gp4	3656	4696	6.2	346	40.7	<i>L. innocua</i> lin2392, <i>L. monocytogenes</i> H7858 and F6854; A118 gp4 minor capsid protein
gp5	4792	5346	4.6	184	21.0	<i>L. monocytogenes</i> F6854, H7858, <i>L. innocua</i> lin2391
gp6	5369	6241	5.4	290	32.2	<i>L. innocua</i> lin2390 major capsid protein, <i>L. monocytogenes</i> H7858, F6854
gp7	6255	6392	4.2	45	5.4	<i>L. monocytogenes</i> H7858, F6854, <i>L. innocua</i> lin2389
gp8	6424	6777	4.2	117	12.7	<i>L. monocytogenes</i> F6854, <i>L. innocua</i> lin2388, <i>L. monocytogenes</i> H7858
gp9	6777	7142	4.4	121	13.6	<i>L. monocytogenes</i> H7858, <i>L. innocua</i> lin2387
gp10	7132	7449	10.0	105	12.0	<i>L. monocytogenes</i> F6854 LMO2667
gp11	7446	7817	5.7	123	14.1	<i>L. monocytogenes</i> H7858 and F6854, <i>L. innocua</i> lin2386
gp12	7822	8508	4.4	228	23.5	<i>L. innocua</i> lin2385, <i>L. monocytogenes</i> H7858 and F6854; A118 gp13 major tail protein (Tsh)
gp13	8564	8995	4.3	143	16.1	<i>L. monocytogenes</i> H7858, <i>L. innocua</i> lin2384, <i>L. monocytogenes</i> F6854
gp14	9028	9303	10.1	91	10.7	<i>L. monocytogenes</i> F6854 LMO2662, <i>L. monocytogenes</i> H7858 LMO2424
gp15	9308	14107	9.7	1599	174.1	<i>L. monocytogenes</i> H7858; Phage L54a; <i>L. innocua</i> lin2383, <i>L. monocytogenes</i> F6854 putative tape measure protein (Tmp)

A006	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp16	14104	15672	5.2	522	58.5	<i>L. monocytogenes</i> F6854, <i>B. cereus</i> phage putative tail component
gp17	15685	17847	4.6	720	78.8	<i>L. monocytogenes</i> F6854 phage minor structural protein
gp18	17898	18203	9.2	101	11.5	<i>L. innocua</i> lin2376, <i>L. monocytogenes</i> F6854
gp19	18203	18484	10.2	93	10.5	<i>L. monocytogenes</i> F6854 holin, <i>L. innocua</i> lin2375, lin1295
gp20	18484	19191	10.1	235	25.7	<i>L. monocytogenes</i> F6854 phage lysin, <i>L. innocua</i> , A500 L-Alanoyl-D-glutamate-peptidase
gp21*	20020	19232	8.7	262	31	<i>L. monocytogenes</i> F6854 LMO2339
gp22*	20777	20280	10.2	165	19.3	<i>L. monocytogenes</i> EGDe (lmo2276) and F6854
gp23*	21251	20802	5.4	149	17.3	<i>L. monocytogenes</i> F6854 gp28, <i>L. monocytogenes</i> EGDe, <i>L. innocua</i> and A118 gp28
gp24*	21503	21252	4.3	83	9.9	
gp25*	21768	21535	4.1	77	9.3	<i>L. monocytogenes</i> F6854 LMO2335
gp26	22067	22300	5.0	77	9.4	<i>L. innocua</i> lin2372; PSA Gp22, <i>L. monocytogenes</i> EGDe (lmo2271) and H7858
gp27	22297	22500	6.6	67	7.9	<i>L. innocua</i> lin1302, PSA Gp23; <i>L. monocytogenes</i> H7858
gp28*	23910	22780	9.4	376	43.6	<i>Bacillus halodurans</i> integrase, Bacteriophage phig1e, <i>Enterococcus faecium</i>
gp29*	24483	23974	4.9	169	19.1	
gp30*	25001	24510	5.3	163	19.3	<i>L. monocytogenes</i> H7858 gp35, A118 gp35, <i>L. innocua</i> lin0160, <i>L. monocytogenes</i> EGDe lmo0113, F2365
gp31*	25341	25033	4.6	102	12.1	<i>L. monocytogenes</i> H7858 transcriptional regulator, <i>L. innocua</i> lin2422, A118 repressor protein
gp32	25490	25741	9.8	83	9.5	<i>L. innocua</i> lin2421, <i>L. monocytogenes</i> H7858 transcriptional regulator
gp33	25745	25981	4.4	78	8.8	<i>L. innocua</i> gp40, lin2420
gp34	25978	26331	4.7	117	13.2	<i>L. monocytogenes</i> EGDe lmo2326, <i>L. innocua</i> lin0078, <i>L. monocytogenes</i> H7858 and F6854; A118 gp41
gp35*	26851	26309	4.4	180	21.3	
gp36	26915	27688	10.0	257	29.3	<i>L. innocua</i> lin0080 anti-repressor homolog A118, lin2418, <i>Streptococcus pyogenes</i> orf010, phage 77
gp37	27810	28334	5.3	174	19.3	<i>L. innocua</i> lin0081, <i>L. monocytogenes</i> F6854, EGDe and H7858; A118 gp43
gp38	28341	28577	6.8	78	9.3	<i>Bacillus halodurans</i>
gp39	28685	28873	5.7	62	7.2	<i>L. innocua</i> lin0083, A118 gp45, <i>L. monocytogenes</i> H7858, <i>L. innocua</i> lin2415, <i>L. monocytogenes</i> EGDe lmo2321, F6854
gp40	28873	28962	9.5	29	3.4	A118 gp46
gp41	29109	30068	6.3	319	36.7	<i>L. innocua</i> gp47 lin0084, <i>L. monocytogenes</i> H7858 and F6854; A118 gp47
gp42	30068	30883	5.1	271	30.7	<i>L. innocua</i> putative recombinase lin2413, <i>L. monocytogenes</i> F6854; A118 gp48; lin0085
gp43	30903	31835	9.3	310	36.8	A118 gp49, <i>L. innocua</i> lin2412, <i>L. monocytogenes</i> F6854 DnaD domain protein, EGDe lmo2317, F6854, <i>L. innocua</i> lin0086
gp44	31832	32296	9.2	154	18.1	A118 putative methyltransferase gp50; <i>Enterococcus</i>

A006	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
						<i>faecalis</i> V583, Bacteriophage phig1e
gp45	32293	32853	4.4	187	21.3	A118 gp51, <i>L. innocua</i> lin0089, <i>L. monocytogenes</i> EGDe lmo2315, F6854, and H7858; PSA Gp35
gp46	32850	32996	4.2	48	5.5	A118 gp52, <i>L. innocua</i>
gp47	33074	33475	4.5	133	15.7	A118 gp54
gp48	33472	33681	4.0	69	8.1	A118 gp55, <i>L. innocua</i> lin0094, <i>L. monocytogenes</i> F6854
gp49	33678	34076	9.5	132	15.1	A118 gp56
gp50	34082	34192	4.6	36	4.0	A118 gp57
gp51	34192	34284	7,3	30	3.3	A118 gp58
gp52	34520	34921	5.7	133	15.2	A118 gp59, <i>L. monocytogenes</i> H7858 and F6854, <i>L. innocua</i> lin0095
gp53	34918	35400	5.1	160	17.8	A118 putative SSB protein gp60, <i>L. monocytogenes</i> EGDe lmo2398, H7858, and F6854; <i>L. innocua</i> lin0079, lin2402
gp54	35432	35737	10.6	101	11.7	A118 gp61, <i>L. monocytogenes</i> H7858 putative response regulator, <i>L. innocua</i> lmaD, EGDe lmo0115
gp55	35734	35874	5.0	46	5.6	A118 gp62
gp56	35840	36244	8.4	134	15.5	A118 gp63, <i>L. monocytogenes</i> H7858, <i>Lactobacillus plantarum</i> prophage LP2 protein 24
gp57	36237	36362	10.6	41	4.6	A118 gp64
gp58	36373	36537	4.2	54	6.3	A118 gp65, <i>L. monocytogenes</i> EGDe
gp59	36556	36990	9.0	144	17.1	A118 gp66, <i>L. innocua</i> gp66, <i>L. monocytogenes</i> EGDe and F6854 transcriptional regulator; <i>L. innocua</i> lin1739, lin1259
gp60	37036	37185	6.1	49	5.6	A118 gp67
gp61	37172	37804	4.7	210	25.0	
gp62	37885	38112	5.7	75	8.4	<i>L. innocua</i> lin2396, <i>L. monocytogenes</i> H7858 and F6854

Table VI.2. General features and database matches of the predicted gene products of A500.

A500	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp1	23	65	5.1	180	20.0	<i>L. innocua</i> lin0104 terminase small subunit; A118 gp1
gp2	534	1865	5.7	443	51.4	<i>L. innocua</i> lin0105 terminase large subunit, <i>L. monocytogenes</i> EGDe lmo2300, A118 gp2; <i>L. monocytogenes</i> F6854
gp3	1878	3377	4.7	499	56.8	A118 portal protein (gp3), <i>L. innocua</i> lin0106, <i>L. monocytogenes</i> F6854 and EGDe lmo2299
gp4	3383	4522	9.6	379	43.1	A118 minor capsid protein (gp4), <i>L. innocua</i> lin0107, <i>L. monocytogenes</i> F6854, EGDe
gp5	4601	5200	4.5	199	21.8	<i>L. monocytogenes</i> F6854 scaffold protein; A118 gp5, <i>L. innocua</i> lin0108; <i>L. monocytogenes</i> EGDe lmo2297
gp6	5223	6059	4.9	278	29.7	<i>L. monocytogenes</i> F6854; <i>Lactococcus lactis</i> phage TP901-1; <i>Streptococcus pyogenes</i> phage 315.4 major head protein
gp7	6059	6217	4.2	52	5.6	<i>L. innocua</i> lin0110, <i>L. monocytogenes</i> F6854; A118 gp7
gp8	6219	6614	5.5	131	14.6	<i>L. innocua</i> lin0111, A118 gp8, <i>L. monocytogenes</i> F6854 and EGDe
gp9	6614	6976	5.6	120	13.8	<i>L. innocua</i> lin0112, <i>L. monocytogenes</i> F6854; A118 gp9, <i>L. monocytogenes</i> EGDe
gp10	6976	7314	9.8	112	12.7	<i>L. innocua</i> lin0113, <i>L. monocytogenes</i> EGDe; A118 minor capsid protein (gp10), <i>L. monocytogenes</i> F6854
gp11	7314	7721	4.2	135	15.1	<i>L. innocua</i> lin0114, <i>L. monocytogenes</i> EGDe and F6854; A118 gp11
gp12	7724	8161	4.6	145	15.9	<i>L. innocua</i> lin0115, <i>L. monocytogenes</i> F6854 and EGDe; A118 major tail protein (Tsh) gp12
gp13	8184	8423	4.5	79	8.0	<i>L. monocytogenes</i> EGDe gp13; A118 gp13, <i>L. monocytogenes</i> F6854
gp14	8476	8898	4.9	140	16.6	<i>L. innocua</i> lin0117, <i>L. monocytogenes</i> F6854 and EGDe; A118 gp14
gp15	8904	9506	4.2	200	23.2	<i>L. innocua</i> lin0118 <i>L. monocytogenes</i> F6854 and EGDe; A118 gp15 putative tail protein
gp16	9517	14883	10.1	1788	185.6	<i>L. innocua</i> lin0019, <i>L. monocytogenes</i> EGDe (lmo2287) and F6854; A118 gp16 tape measure protein (Tmp)
gp17	14880	15707	6.8	275	31.7	<i>L. innocua</i> lin0120 gp17, <i>L. monocytogenes</i> EGDe and F6854; A118 tail or base plate protein (gp17)
gp18	15722	16744	5.2	340	40.0	<i>L. innocua</i> lin0121; A118 tail or base plate protein (gp18), <i>L. monocytogenes</i> F6854, EGDe and F2365
gp19	16745	17770	4.7	341	37.3	<i>L. innocua</i> lin0122, lin2380, <i>L. monocytogenes</i> H7858, EGDe, and F6854; A118 tail or base plate protein (gp19)
gp20	16767	18834	4.8	355	38.9	<i>L. innocua</i> lin0123, <i>L. monocytogenes</i> H7858, <i>L. innocua</i> lin2379 and lin2565; PSA Gp15; A118 gp20 putative tail fibre
gp21	18831	19172	4.0	113	12.5	<i>L. innocua</i> lin0124, lin2378, and lin2564; <i>L. monocytogenes</i> H7858 gp16; PSA Gp16

A500	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp22	19173	19319	9.1	48	5.6	<i>L. innocua</i> lin0125, lin2377; PSA Gp17
gp23	19356	19721	7.7	121	14.4	<i>L. monocytogenes</i> H7858, <i>L. innocua</i> gp23; A118 gp23, <i>L. monocytogenes</i> F6854 and EGDe
gp24	19725	20015	9.8	96	10.2	<i>L. innocua</i> holin, <i>L. monocytogenes</i> H7858; A118 gp24; <i>L. monocytogenes</i> F6854 and EGDe
gp25	20015	20884	9.8	289	33.4	<i>L. monocytogenes</i> H7858 (L-Alanoyl-D-glutamate-peptidase) and F6854 cply2438, <i>L. innocua</i> lin2563, lin2374, and lin1700
gp26	21378	21929	9.8	183	21.1	<i>L. monocytogenes</i> EGDe lmo2277
gp27	22516	23148	9.7	210	22.9	<i>L. monocytogenes</i> EGDe lmo0638
gp28	23171	23353	10.3	60	7.3	
gp29	23596	23829	5.4	77	9.4	<i>L. innocua</i> lin237; PSA Gp22, <i>L. monocytogenes</i> EGDe (lmo2271) and H7858
gp30	23826	24032	9.7	68	8.1	PSA Gp23, <i>L. monocytogenes</i> H7858, <i>L. innocua</i> lin1302
gp31*	25483	24281	9.9	400	46.2	<i>L. innocua</i> lin0071 integrase, <i>L. monocytogenes</i> F6854 putative site specific recombinase
gp32*	26289	25549	4.4	246	27.0	A118 gp32
gp33*	26852	26424	4.6	142	16.2	<i>L. monocytogenes</i> EGDe lmo2329 repressor protein; F6854 transcriptional regulator; <i>L. innocua</i> lin0073 and lin2422
gp34	27015	27236	8.7	73	8.2	<i>Moorella theramoacetica</i> transcriptional regulator, <i>Lactobacillus</i> phage phig1e repressor protein
gp35	27250	27477	9.6	75	8.4	<i>L. innocua</i> lin2420; A118 gp40
gp36	27489	27773	4.2	94	10.9	<i>L. innocua</i> lin2411 and lin2603
gp37	27786	28565	8.7	259	30.0	<i>L. monocytogenes</i> EGDe lmo2324, F6854 anti-repressor; A118 gp42; <i>L. monocytogenes</i> H7858
gp38	28689	29213	5.9	174	19.4	<i>L. innocua</i> lin0081, <i>L. monocytogenes</i> F6854, EGDe, and H7858 gp43; A118 gp43
gp39	29220	29405	10.0	61	6.9	<i>L. innocua</i> lin0082
gp40	29422	29511	9.5	29	3.3	A118 gp46
gp41	29504	29632	3.7	42	5.1	
gp42	29727	29921	5.4	64	7.6	<i>L. monocytogenes</i> EGDe lmo2320
gp43	29918	30394	5.2	158	18.2	<i>L. monocytogenes</i> EGDe (lmo2319), H7858, F6854; lmo2180 EGDe; <i>L. innocua</i> lin2284; PSA Gp46
gp44	30400	31059	7.6	219	23.8	<i>L. monocytogenes</i> EGDe lmo2318; Phage 42e ORF020; <i>S. aureus</i> phiPV83 orf17
gp45	31076	32053	5.7	325	37.7	<i>L. monocytogenes</i> EGDe (lmo2317), F6854 putative DnaD domain protein; A118 gp49; <i>L. innocua</i> lin2412
gp46	32050	32217	9.5	55	6.4	<i>L. innocua</i> lin0087; <i>Streptococcus thermophilus</i> phage Sfi11 gp57
gp47	32223	32516	10.4	97	11.6	<i>L. monocytogenes</i> F6854 LMO2692; PSA Gp33
gp48	32513	32722	9.9	69	8.1	<i>L. monocytogenes</i> F6854 LMO2691
gp49	32719	33327	4.7	202	23.3	<i>L. monocytogenes</i> H7858, F6854, and EGDe (lmo2315); A118 gp51
gp50	33324	33692	5.9	122	14.2	
gp51	33689	33865	4.4	58	6.6	
gp52	33852	34313	4.9	153	17.3	<i>L. monocytogenes</i> EGDe (lmo2313), F6854; <i>L. innocua</i> lin0090; PSA Gp37
gp53	34310	34831	4.3	173	20.3	
gp54	34828	35268	4.7	146	17.2	<i>L. monocytogenes</i> EGDe lmo2312, F6854

A500	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp55	35265	35465	5.4	66	7.5	
gp56	35465	35947	4.9	160	17.7	A118 gp60 putative SSB protein; <i>L. monocytogenes</i> EGDe (lmo2308), F6854, H7858; <i>L. innocua</i> lin0097, lin2402
gp57	35968	36159	4.2	63	7.1	<i>L. monocytogenes</i> F6854; <i>L. innocua</i> lin0098, <i>L. monocytogenes</i> EGDe (lmo2307); lin2401
gp58	36104	36508	10.2	134	15.9	<i>L. innocua</i> lin0099; <i>L. monocytogenes</i> F6854, EGDe lmo2306; <i>L. innocua</i> lin2400
gp59	36512	36895	9.6	127	15.5	<i>L. monocytogenes</i> EGDe (lmo2305) and F6854 <i>L. innocua</i> lin0100 and lin2399
gp60	36888	37013	10.0	41	4.6	A118 gp64
gp61	37025	37189	4.3	54	6.3	<i>L. monocytogenes</i> EGDe lmo2304; A118 gp65
gp62	37208	37642	9.3	144	16.9	<i>L. monocytogenes</i> EGDe (lmo2303) and F6854, <i>L. innocua</i> transcriptional regulator; A118 gp66
gp63	38006	38566	5.2	186	21.5	<i>Enterococcus faecalis</i> V583 EF0126
gp64	38663	38848	5.1	61	6.8	<i>L. innocua</i> lin1734, lin1265

Table VI. 3. General features and database matches of the predicted gene products of B025.

B025	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp1	62	361	5.4	99	11.5	<i>Streptococcus pyogenes</i> SSI-1; phage 2638A
gp2	358	2001	5.2	547	63.4	<i>Staphylococcus aureus</i> terminase large subunit; Phage 77, PSA Gp2
gp3	2013	3143	4.8	376	43.6	Phage 77 ORF008, <i>S. aureus</i> SAR2063; PSA Gp3 putative portal protein
gp4	3140	3889	4.9	249	27.5	<i>Clostridium tetani</i> E88 putative scaffold protein; Phage 77, <i>S. pyogenes</i> SSI-1
gp5	3916	5067	4.9	383	42.4	Phage 77 Orf006, <i>S. aureus</i> SAR2061, PSA Gp5 major capsid
gp6	5074	5244	10.1	56	6.6	
gp7	5254	5553	4.4	99	11.5	Phage 77 ORF042, Phage 2638A ORF036; PSA Gp6
gp8	5537	5902	6.2	121	13.9	Phage 77 ORF033, Phage 2638A ORF027
gp9	5899	6300	10.0	133	14.8	Phage 3A ORF028, Phage 47 ORF028
gp10	6297	6680	4.8	127	14.7	Phage 3A ORF029, Phage 47 ORF029
gp11	6702	7289	4.6	195	21.1	Phage 3A ORF017, <i>S. aureus</i> Phage phiSLT major tail protein (Tsh)
gp12	7324	7692	4.8	122	14.6	Phage 47 ORF033, Phage 42e ORF041, <i>S. aureus</i> protein MW1391
gp13	7743	7892	4.8	49	5.9	
gp14	7908	12803	10.0	1640	174.5	<i>L. innocua</i> lin1716, <i>Leuconostoc mesenteroides</i> COG5283, <i>Lactococcus lactis</i> prophage pi3 tail protein
gp15	12830	13660	6.5	276	31.7	Phage 47 ORF015, Phage 42e ORF014, <i>L. monocytogenes</i> EGDe (gp17) and F6854, <i>L. innocua</i> lin0169
gp16	13670	15262	6.4	530	60.1	Phage 47 ORF006; <i>S. aureus</i> SAS0946; Phage 42e and 3A ORF004; <i>L. monocytogenes</i> EGDe lmo0123
gp17	15262	15486	4.3	74	8.4	
gp18	15486	17405	5.4	639	71.1	<i>Bacillus licheniformes</i> , <i>S. aureus</i> SAS1841
gp19	17417	18691	4.6	424	47.0	Phage 37 ORF005, Phage EW ORF004, <i>L. monocytogenes</i> F2365 LMO0145
gp20	18708	18983	4.2	91	10.4	<i>L. monocytogenes</i> EGDe and F6854; A118 gp21 putative short tail fiber
gp21	18989	19153	4.9	54	6.4	A118 gp22, <i>L. monocytogenes</i> F6854 and EGDe, <i>L. innocua</i> lin1292 and lin1706
gp22	19179	19634	9.6	151	17.3	PSA Gp17-1, <i>L. innocua</i> lin1705 and lin1293
gp23	19601	19996	5.9	131	14.9	PSA Gp18, <i>L. innocua</i> lin1704 and lin1294
gp24	20009	20257	9.7	82	9.4	<i>L. innocua</i> lin1295 holin, Phage Tuc2009, PSA holin
gp25	20257	21087	10.0	276	31.7	<i>L. monocytogenes</i> EGDe endolysin (L-alanoyl-D-glutamate peptidase), A118 gp25, <i>L. monocytogenes</i> F6854
gp26	21099	21533	9.7	144	16.0	<i>L. innocua</i> lin0129
gp27	21942	22445	4.5	167	19.8	
gp28	22442	22834	8.9	130	15.2	
gp29	22859	23017	10.2	52	5.8	
gp30	23023	23175	10.2	50	6.1	
gp31	23229	23463	11.0	54	6.3	

B025	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp32	23531	23764	5.1	77	9.3	PSA Gp22, <i>L. monocytogenes</i> H7858 and EGDe (lmo2271), <i>L. innocua</i> lin2372
gp33	23761	23952	10.0	63	7.7	PSA Gp23, <i>L. monocytogenes</i> H7858, <i>L. innocua</i> lin1302
gp34*	25301	24147	9.8	384	44.6	PSA integrase Gp24, <i>L. innocua</i> lin2610
gp35*	26050	25436	3.8	204	22.3	<i>L. lactis</i> prophage pi3; <i>L. innocua</i> lin1471; PSA Gp25
gp36*	26553	26101	5.3	150	17.9	<i>L. innocua</i> lin2608, PSA Gp26; <i>L. monocytogenes</i> F2365, EGDe (lmo0113), F6854
gp37*	26890	26570	5.2	106	12.2	PSA Gp27 repressor protein, <i>L. innocua</i> lin2607; <i>Enterococcus faecalis</i> transcriptional regulator
gp38	27157	27360	9.6	67	7.7	PSA Gp28, <i>L. innocua</i> lin2605
gp39	27362	27604	9.5	80	9.6	PSA Gp29, <i>L. innocua</i> lin2604
gp40	27607	27792	10.0	61	7.5	PSA Gp30
gp41	28027	28179	4.8	50	5.5	PSA Gp31
gp42	28316	29029	4.6	237	27.6	<i>L. innocua</i> lin2602 and lin2410, <i>Streptococcus pyogenes</i> phage SSI-1
gp43	29040	29984	9.5	314	36.2	<i>L. innocua</i> integrase lin2409, lin2601, lin1743, lin1254, and lin0524
gp44	29997	30677	9.7	226	27.0	<i>L. innocua</i> lin2600 lin1255, lin1742, and lin2408
gp45	30674	31237	4.4	187	21.8	<i>L. monocytogenes</i> H7858; A118 gp51, <i>L. monocytogenes</i> EGDe lmo2315; PSA Gp35
gp46	31458	31859	4.5	133	15.7	A118 gp54
gp47	31856	32065	4.0	69	8.1	A118 gp55, <i>L. monocytogenes</i> F6854, <i>L. innocua</i> lin0094
gp48	32066	32395	4.0	109	12.5	
gp49	32392	32664	4.5	90	10.3	
gp50	32664	32843	9.2	59	6.8	<i>L. innocua</i> lin2403
gp51	32840	33379	4.4	179	20.1	<i>L. monocytogenes</i> F6854 and EGDe (lmo2312)
gp52	33379	33813	4.7	144	16.9	PSA Gp40
gp53	33810	34037	9.9	75	9.1	PSA Gp41, <i>L. innocua</i>
gp54	34076	34249	9.6	57	6.6	PSA Gp44, <i>L. innocua</i> lin2592
gp55	34246	34629	4.6	127	14.2	PSA Gp45
gp56	34631	35110	5.7	159	18.3	<i>L. innocua</i> lin2591, PSA Gp46, <i>L. monocytogenes</i> EGDe (lmo2319) and F2365, <i>L. innocua</i> lin2284
gp57	35123	35815	5.1	230	25.9	<i>S. pyogenes</i> SSI-1 COG0593, <i>S. suis</i> ATPase, <i>Lactococcus</i> Phage phi31, <i>L. innocua</i> lin2590, PSA Gp47
gp58	35993	37135	6.7	380	43.1	<i>L. innocua</i> lin2589 putative helicase, PSA Gp48, <i>Lactococcus</i> phage phi31
gp59	37158	37640	5.8	160	18.3	<i>L. innocua</i> lin2588, PSA Gp49
gp60	37663	39936	5.6	757	87.2	<i>L. innocua</i> lin2587, PSA Gp50 primase, <i>L. casei</i> phage A2
gp61	40225	40539	9.3	104	11.7	<i>L. innocua</i> lin2586, PSA Gp51, <i>S. pyogenes</i> SSI-1
gp62	40542	41180	8.7	212	25.5	<i>L. innocua</i> lin2585, PSA Gp52
gp63	41185	41610	9.7	141	16.8	PSA Gp53 transcriptional activator, <i>L. innocua</i> lin2583
gp64	42047	42250	8.6	67	7.9	
gp65	42295	42609	9.5	104	12.8	<i>S. aureus</i> phiN315, Phage 77 ORF040, PSA Gp54

Table VI.4. General features and database matches of the predicted gene products of B054.

B054	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp1	26	910	9.7	294	33.5	<i>L. innocua</i> lin1733 and lin1266, <i>L. monocytogenes</i> F6854 small terminase subunit
gp2	888	2348	5.7	486	56.1	<i>L. innocua</i> lin1732; <i>Enterococcus faecalis</i> V583 large terminase subunit; Phage Aaphi23
gp3	2362	3747	5.2	461	52.3	<i>L. innocua</i> lin1731, <i>E. faecalis</i> V583
gp4	3740	5149	8.98	469	52.8	<i>L. innocua</i> lin1730, <i>E. faecalis</i> V583 minor head protein
gp5	5146	5316	6.8	56	6.8	<i>L. innocua</i> lin1729
gp6	5365	6474	5.6	369	40.5	<i>L. innocua</i> lin1728, <i>E. faecalis</i> V583 EF1461
gp7	6474	6923	4.6	149	15.9	<i>L. innocua</i> lin1727, <i>E. faecalis</i> V583 EF1462
gp8	6944	7849	5.6	301	33.0	<i>L. innocua</i> lin1726, <i>E. faecalis</i> V583 EF1463
gp9	7872	8258	4.6	128	14.5	<i>L. innocua</i> lin1725, <i>E. faecalis</i> V583 EF1464
gp10	8270	8605	5.4	111	12.3	<i>L. innocua</i> lin1724, <i>E. faecalis</i> V583 EF1465
gp11	8605	9204	9.6	199	22.5	<i>L. innocua</i> lin1723, <i>E. faecalis</i> V583 EF1466
gp12	9204	9578	4.9	124	14.1	<i>L. innocua</i> lin1722, <i>E. faecalis</i> V583 EF1467
gp13	9568	10056	4.5	162	18.0	<i>L. innocua</i> lin1721, <i>E. faecalis</i> V583 EF1468
gp14	10061	11056	5.2	331	36.3	<i>L. innocua</i> lin1720, <i>E. faecalis</i> V853 EF1469
gp15	11073	11471	5.6	132	14.2	<i>L. innocua</i> lin1719, <i>E. faecalis</i> V583 EF1470
gp16	11524	11880	5.9	118	13.4	<i>L. innocua</i> lin1718, <i>E. faecalis</i> V583 EF1471
gp17	11843	12055	5.1	70	8.1	<i>L. innocua</i> lin1717, <i>E. faecalis</i> V583 EF1472
gp18	12057	16772	9.6	1571	168.0	<i>L. innocua</i> lin1716, <i>E. faecalis</i> V583 EF1473, Phage phig1e minor capsid protein
gp19	16778	17341	10.3	187	21.4	<i>L. innocua</i> lin1715, <i>E. faecalis</i> V583 LysM domain protein
gp20	17341	17706	3.7	121	13.9	<i>L. innocua</i> lin1714, <i>E. faecalis</i> V583 EF1475
gp21	17706	18512	9.0	268	29.6	<i>L. innocua</i> lin1713, <i>E. faecalis</i> V583 EF1476
gp22	18513	18851	4.9	112	12.4	<i>L. innocua</i> lin1712, <i>E. faecalis</i> V583 EF1477
gp23	18848	19210	4.2	120	13.9	<i>L. innocua</i> lin1711, <i>E. faecalis</i> V583 EF1478,
gp24	19203	20354	4.6	383	41.5	<i>L. innocua</i> lin1710, <i>E. faecalis</i> V583 EF1479, <i>Salmonella enterica</i> serovar <i>typhi</i>
gp25	20344	20985	5.3	213	23.1	<i>L. innocua</i> lin1709, <i>E. faecalis</i> V583 EF1480
gp26	21006	21734	5.3	242	26.5	<i>L. innocua</i> lin1708 and lin1290
gp27	21740	22081	4.1	113	13.1	<i>L. innocua</i> lin1291 and lin1707
gp28	22113	22238	6.3	41	4.7	<i>L. innocua</i> lin1292 and lin1706; A118 gp22, <i>L. monocytogenes</i> EGDe and F6854
gp29	22268	22714	9.3	148	16.7	<i>L. innocua</i> lin1705, lin1293, PSA GP17-1
gp30	22690	23136	5.8	148	17.0	<i>L. innocua</i> lin1294, lin1704, PSA GP18
gp31	23155	23370	8.9	71	7.9	<i>L. innocua</i> lin1703
gp32	23381	23638	6.4	85	10.0	<i>L. innocua</i> holin lin1295, Phage Tuc2009 holin, <i>L. lactis</i> subsp. <i>cremoris</i> COG5546
gp33	23638	24603	9.4	321	36.3	PSA endolysin, <i>L. innocua</i> lin2374 and lin2563 N-Acetylmuramoyl-L-alanine amidase (n-terminal) and L-alanoyl-D-glutamate peptidase (c-terminal); <i>Clostridium tetani</i>
gp34*	24887	24666	6.8	73	8.6	<i>L. innocua</i> lin1297 (gp28 A118 homolog)
gp35*	25118	24915	4.4	67	7.9	<i>L. innocua</i> lin1298
gp36*	25506	25132	4.5	124	14.5	<i>L. innocua</i> lin1299
gp37*	25912	25646	10.4	88	10.6	

B054	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp38*	27463	26288	9.5	391	45.8	<i>L. innocua</i> lin1765 integrase, <i>S. pyogenes</i> phage 315.5
gp39*	27952	27560	6.8	130	13.8	
gp40*	28419	28000	6.2	139	16.5	<i>L. innocua</i> lin1233, <i>Bacillus anthracis</i> lambda ba01 putative repressor protein
gp41*	28864	28436	8.0	142	16.6	<i>L. innocua</i> lin1234, <i>Bacillus cereus</i> ZK putative transcription regulator
gp42	29025	29210	9.3	61	7.3	<i>L. innocua</i> lin1235 transcriptional regulator, <i>E. faecalis</i> V583, <i>L. monocytogenes</i> EGDe lmo2328
gp43	29287	29811	5.3	174	20.0	<i>L. innocua</i> lin1760 and lin1236; A118 gp43
gp44	29826	30119	5.9	97	11.7	
gp45	30124	30288	4.9	54	6.1	<i>L. innocua</i> lin1238
gp46	30438	30728	5.3	96	11.2	<i>L. innocua</i> lin1759 and lin1239
gp47	30906	31133	9.3	75	9.1	<i>L. innocua</i> lin1758
gp48	31151	31492	4.7	113	12.6	<i>L. innocua</i> lin1757 and lin1240
gp49	31479	33422	5.0	647	74.4	<i>L. innocua</i> lin1756 and lin1241, <i>S. aureus</i> MSSA476
gp50	33423	34208	5.0	261	29.7	<i>L. innocua</i> lin1755 and lin1242, <i>S. pyogenes</i> MIGAS
gp51	34168	34905	8.6	245	28.1	<i>L. innocua</i> lin1243 and lin1754, <i>E. faecalis</i> V583 EF1433
gp52	34917	35834	5.7	305	35.3	<i>L. innocua</i> lin1244 and lin1753, Phage bIL286 replication protein, <i>E. faecalis</i> V583 DnaD domain protein
gp53	35758	36573	6.3	271	31.3	<i>L. innocua</i> lin1752 and lin1245, Phage 55 ORF016
gp54	36570	36764	5.0	64	7.6	
gp55	36751	36957	10.5	68	8.2	<i>L. innocua</i> lin1751 and lin1246
gp56	36958	37167	6.8	69	8.3	<i>L. innocua</i> lin1750 and lin1247
gp57	37148	37585	9.8	145	16.7	<i>L. innocua</i> lin1749 and lin1248, Phage LLH orf139, <i>Pediococcus pentosaceus</i> holliday junction resolvase
gp58	37606	38226	4.2	206	23.8	<i>L. innocua</i> lin1748 and lin1249, <i>E. faecalis</i> V583 EF2835
gp59	38267	38647	7.9	126	15.1	<i>L. innocua</i> lin1250 and lin1747
gp60	38607	38783	5.0	58	6.9	<i>L. innocua</i> lin1251 and lin1746
gp61	38795	39217	9.1	140	15.9	<i>L. innocua</i> lin1252 and lin1745; <i>S. pyogenes</i> phage 315.3; <i>E. faecalis</i> V583
gp62	39214	40575	9.5	453	53.5	<i>L. innocua</i> lin1253 and lin1744; <i>S. pyogenes</i> phage 315.3; <i>E. faecalis</i> V583
gp63	40588	40812	7.8	74	9.2	
gp64	40812	41756	8.8	314	36.2	<i>L. innocua</i> lin1254, lin1743, lin2409, lin2609, and lin0524, <i>Lactobacillus delbrueckii</i> integrase
gp65	41769	42449	9.6	226	26.9	<i>L. innocua</i> lin1255, lin1742, lin2600, and lin2408; Phage EJ-1 transcriptional regulator
gp66	42446	42739	4.8	97	11.8	<i>L. innocua</i> lin1256 and lin2597
gp67	42757	42897	4.0	46	5.4	
gp68	42884	43036	10.3	50	6.0	<i>Bacillus clausii</i> KSM-K16 ABC1364
gp69	43033	43284	5.4	83	10.1	<i>L. innocua</i> lin1741, <i>E. faecalis</i> V583 EF0509
gp70	43296	43856	9.1	186	21.6	<i>L. innocua</i> lin1740 and lin1258
gp71	43937	44359	9.2	140	16.6	<i>L. innocua</i> lin1739 and lin1259, <i>L. monocytogenes</i> EGDe, F6854 transcriptional regulator, A118 gp66
gp72	44372	45139	7.2	255	29.3	<i>L. innocua</i> putative antirepressor lin1738, lin1260
gp73	45087	45257	10.6	56	7.0	

B054	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp74	45270	45386	9.7	38	4.6	
gp75	45576	45710	8.4	44	5.5	
gp76	45667	45960	9.4	97	11.6	
gp77	45957	47294	4.9	455	51.1	<i>L. innocua</i> lin1737 and lin1262, <i>Ralstonia eutropha</i> COG0863
gp78	47342	47539	10.1	65	7.8	<i>L. innocua</i> lin1736 and lin1263
gp79	47670	47900	5.0	76	9.1	<i>L. innocua</i> lin1735 and lin1264, <i>E. faecalis</i> V583 EF1453
gp80	47928	48152	4.2	74	8.7	<i>L. innocua</i> lin1734 and lin1265

Table VI.5. General features and database matches of the predicted gene products of P35.

P35	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp1	29	649	5.4	206	23.5	
gp2	637	1959	5.3	440	50.7	<i>Bacillus subtilis</i> phage SPP1 terminase large subunit, <i>Clostridium thermocellum</i> COG1783
gp3	2008	3621	4.3	537	60.4	<i>E. faecalis</i> V583 portal protein EF0334, phage phiETA
gp4	3872	4780	8.1	302	34.7	<i>E. faecalis</i> V583 minor head protein EF0335
gp5	4918	5499	4.6	193	21.6	<i>E. faecalis</i> V583 scaffold protein EF0338
gp6	5543	6451	4.9	302	32.9	<i>L. monocytogenes</i> H7858 phage main capsid (Gp34), <i>L. innocua</i> lin2390
gp7	6518	7054	4.1	178	20.4	<i>E. faecalis</i> V583 EF0341
gp8	7054	7407	5.6	117	13.1	<i>E. faecalis</i> V583 EF0342
gp9	7409	7834	9.5	141	16.1	<i>E. faecalis</i> V583 EF0343
gp10	7834	8298	4.5	154	18.2	<i>E. faecalis</i> V583 EF0344
gp11	8299	9279	4.3	326	35	<i>E. faecalis</i> V583 EF0345 <i>C. acetobutylicum</i> CAC0057
gp12	9363	9821	5.6	152	17.4	
gp13	9875	10147	4.5	90	10.3	<i>E. faecalis</i> V583 methionine sulfoxid reductase
gp14	10153	12039	10.0	628	67.9	<i>E. faecalis</i> V583 EF048 (phage tail protein), <i>C. thermocellum</i> tail protein
gp15	12036	13295	4.1	419	46.5	
gp16	13288	14451	8.7	387	43.7	<i>Leuconostoc mesenteroides</i> Lmes02000557
gp17	14459	15295	9.4	278	30.6	<i>L. mesenteroides</i> Lmes02000556
gp18	15427	15876	5.9	149	16.6	<i>Streptococcus</i> phage Cp-1 holin, <i>C. thermocellum</i> holin
gp19	16299	16544	7.1	81	9.5	
gp20	16660	16905	8.6	81	9.7	
gp21	16905	17117	4.6	70	8.5	
gp22	17119	17448	9.9	109	12.3	
gp23	17448	18221	4.5	257	28.3	
gp24	18218	18460	6.1	80	9.2	
gp25	18444	19067	8.5	207	24.1	Vibriophage VpV262 HNH DNase, Mimivirus endonuclease
gp26	19103	19261	10.2	52	6.0	
gp27	19272	19433	9.2	53	6.3	
gp28	19450	20325	10.2	291	33.1	<i>L. innocua</i> endolysin, A500 L-alanoyl-D-glutamate peptidase
gp29	20439	20834	4.8	131	15.8	
gp30	20884	21456	6.0	190	22.3	
gp31	21497	22243	5.1	248	28.9	<i>L. johnsonii</i> prophage Lj965 protein Ljo0298
gp32	22263	22958	4.9	231	27	
gp33	23285	25192	6.9	635	73.3	<i>C. thermocellum</i> DNA polymerase elongation subunit
gp34	25205	27109	5.5	634	72.5	<i>C. thermocellum</i> ATPase, <i>S. thermophilus</i> DNA primase
gp35	27102	27341	9.5	79	9.0	
gp36	27437	27652	5.0	71	8.1	
gp37	27649	28203	5.4	184	21.5	
gp38	28203	28787	9.1	194	22.6	<i>L. innocua</i> lin2585
gp39	28784	29215	7.8	143	16.9	
gp40	29290	29490	5.2	66	7.6	
gp41	29487	29957	7.6	156	17.6	(LP65 orf143 HNH homing endonuclease) [#]
gp42	29957	30172	5.0	71	8.2	

P35	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp43	30183	30644	4.8	153	17.7	
gp44	30647	31012	6.2	121	14.4	
gp45	31005	31319	4.7	104	11.8	<i>Staphylococcus aureus</i> Phage K ORF17, <i>Enterococcus hirae</i> ArpR
gp46	31316	31624	4.6	102	11.5	
gp47	31617	31799	4.0	60	7.1	
gp48	31908	32012	7.6	34	3.8	
gp49	32009	32320	6.2	103	11.9	
gp50	32298	32465	10.2	55	6.7	
gp51	32449	33357	6.9	302	34.8	
gp52	33357	33548	8.6	92	7.5	
gp53	33550	34179	5.2	209	25.0	
gp54	34270	34674	5.5	134	15.5	
gp55	34677	35090	9.5	137	16.2	
gp56	35063	35326	4.7	87	10.3	

identity to only c-terminal 73 amino acids of LP65 orf143

Table VI.6. General features, protein similarities and functional assignments of putative gene products of A511. The predicted gene products are listed consecutively according to their genomic localisation. tRNA genes were predicted using tRNAscan-SE.

A511	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp10	52	438	9.3	128	14.7	P100 gp1, Phage G1 ORF168, Twort ORF140
gp11	422	694	10.4	90	10.5	P100 gp2, phage G1 ORF161, Twort ORF133, LP65 orf118
gp12	700	1116	4.2	138	15.3	P100 gp3, Twort ORF092, Phage G1 ORF133, Phage K ORF34, LP65 orf117
gp13	1116	1397	10.1	93	10.6	P100 gp4, Phage K ORF35, Phage G1 ORF10 large terminase, <i>B. subtilis</i> phage 1102phi1-3, Twort ORF151, LP65 orf115
gp14	1747	3300	5.8	517	59.1	P100 gp5, Phage K ORF35, Phage G1 ORF10 large terminase, Phage 1102phi1-3, LP65 orf113
gp15	3369	4208	5.2	279	31.5	P100 gp6, Phage K ORF36, Phage G1 ORF40, Twort ORF39
gp16	4213	4410	5.2	65	7.7	P100 gp7
gp17	4400	5038	4.6	212	24.4	P100 gp8, Twort ORF57, Phage K ORF37
gp18	5028	5408	7.2	126	14.4	P100 gp9
gp19	5472	6497	10.1	341	36.5	P100 N-Acetyl-muramoyl-L-alanine amidase, <i>Enterococcus faecalis</i> V583
gp20	6670	7398	9.4	242	26.1	P100 gp11
gp21	7500	7820	10.1	106	12.1	P100 gp12
gp22	7822	8172	5.3	116	13.6	P100 gp13, Twort ORF121, Phage K ORF40, Phage G1 ORF115
gp23	8177	9832	6.9	551	61.6	P100 gp14, Phage K portal protein ORF41, Phage G1 ORF14, Twort ORF12, LP65 orf112
gp1	9931	10725	5.0	264	29.7	P100 gp15, Phage K ORF42, Phage G1 ORF48, Twort ORF40, LP65 orf111 capsid protein
gp2	10718	11611	4.3	297	33.5	P100 gp17, Phage K ORF43, LP65 orf110
cps	11781	13184	5.2	467	51.4	P100 Cps major capsid, Twort Cps, Phage K ORF44, LP65 orf109
gp24	13266	13661	9.4	131	15.4	P100 gp18
gp3	13668	14549	4.9	293	33.2	P100 gp19, Phage K ORF45, Phage G1 ORF30, Twort ORF32, LP65 orf108
gp4	14567	15385	7.2	272	31.2	P100 gp20, Twort ORF43, Phage K ORF46, Phage G1 ORF34, LP65 orf107
gp5	15385	16002	10.7	205	23.9	P100 gp21, Twort ORF55, Phage K ORF47, Phage G1 ORF62, LP65 orf106
gp6	16015	16854	4.5	279	31.5	P100 gp22, Phage K ORF48, Phage G1 ORF39, Twort ORF34, LP65 orf105
gp7	16854	17174	9.2	106	12.3	P100 gp23, Twort ORF185, Phage G1 ORF202
tsh	17178	18866	4.7	562	61.3	P100 Tsh tail sheath protein, Phage K ORF49, Phage G1 ORF11, Twort ORF11, LP65 orf103
gp8	18934	19356	6.2	140	15.6	P100 gp25, Twort ORF142, Phage K ORF50, Phage G1 ORF105, LP65 orf102

A511	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp9	19508	19951	4.7	147	17.4	P100 gp26, Phage K ORF53, Phage G1 ORF95, Twort ORF229, LP65 orf100
gp25	20019	20612	4.0	197	23.4	P100 gp27, Twort ORF81, Phage K ORF54, Phage G1 ORF74, LP65 orf99
gp26	20674	24402	9.4	1242	131.1	P100 gp28 tail tape measure protein, Phage K ORF55, Phage G1 ORF1, Twort ORF2, LP65 orf98
gp27	24448	26835	5.0	795	88.5	P100 gp29, Phage K ORF56, Phage G1 ORF5, Twort ORF7, LP65 orf134
gp28	26853	28385	4.7	510	56.7	P100 gp30, Twort ORF10, Phage G1 ORF4, Phage K ORF58, LP65 orf97 putative tail fiber
gp29	28423	29136	5.1	237	25.7	P100 gp31, Twort ORF41, Phage K ORF59, Phage G1 ORF43, LP65 orf129 and orf130
gp30	29141	29674	4.9	177	20.2	P100 gp32, Phage K ORF60, Phage G1 ORF78, Twort ORF72, LP65 orf131
gp31	29661	30371	4.6	236	26.4	P100 gp33, Phage K ORF61, Phage G1 ORF52, Twort ORF50, LP65 orf132 base plate protein
gp32	30385	31431	4.9	348	39.2	P100 gp34, Phage K ORF62, Phage G1 ORF27, Twort ORF26, LP65 orf95 tail protein
gp33	31467	35396	4.7	1309	145.8	P100 gp35, Phage K ORF63, Phage G1 ORF3, Twort ORF5, LP65 orf94
gp34	35513	36034	5.8	173	19.1	P100 gp36, Phage K ORF64, Phage G1 ORF79, Twort ORF73, LP65 orf91
gp35	36051	39506	4.9	1151	128.1	P100 gp37, Phage K ORF65, Phage G1 ORF2, Twort ORF3, LP65 orf90
gp36	39552	39773	5.1	73	8.6	P100 gp38
gp37	39776	41068	6.1	430	46.7	P100 gp39, Phage K ORF68, Phage G1 ORF17, Twort ORF17, <i>L. monocytogenes</i> EGD lmo1188
gp38	41100	41510	4.4	136	15.3	P100 gp40
gp39	41507	41644	5.0	45	5.3	P100 gp41, PSA GP17, <i>L. innocua</i> lin0125 and lin2377
gp40	41745	43490	6.5	581	66.6	P100 gp42, Phage K ORF69 helicase, Phage G1 ORF12, Twort ORF6, LP65 orf123
gp41	43505	45145	6.8	546	62.9	P100 gp43, Phage K ORF70, Phage G1 ORF13, Twort ORF13 Rep protein
gp42	45163	46626	5.8	487	55.6	P100 gp44, Phage K ORF71 primase, Phage G1 ORF15, Twort ORF14, LP65 orf76
gp43	46641	47693	4.8	350	39.7	P100 gp45, Phage K ORF72, Phage G1 ORF28, Twort ORF27, LP65 orf75 DNA repair exonuclease
gp44	47788	49671	5.3	627	70.7	P100 gp46 and gp47, Twort ORF9, Phage K ORF74, Phage G1 ORF9, LP65 orf70
gp45	49691	50281	7.2	196	22.9	P100 gp48, Twort ORF65, Phage K ORF75, Phage G1 ORF67, LP65 orf69
gp46	50281	51342	4.7	353	40.4	P100 gp49, Phage K ORF76, Phage G1 ORF26, Twort ORF23, LP65 orf68 primase
gp47	51389	52036	5.7	215	23.5	P100 gp50, Phage 2638A ORF19, Phage ROSA ORF23, dUTPase
gp48	52033	52257	7.1	74	8.1	P100 gp51

A511	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp49	52254	52577	4.4	107	12.2	P100 gp52
gp50	52570	52992	5.2	140	16.1	P100 gp53, Phage G1 ORF98, Phage K ORF77
gp51	52995	53615	5.5	206	23.6	P100 gp54, Phage K ORF78, Phage G1 ORF64, Twort ORF63, LP65 orf62
gp52	53679	55079	6.3	466	53.5	P100 gp55, <i>L. innocua</i> ribonucleosid diphosphate reductase α subunit (n-terminal); Phage K ORF80
gp53	55290	56321	5.7	343	38.8	P100 gp56, <i>L. innocua</i> ribonucleosid diphosphate reductase α subunit (c-terminal); Phage K ORF80
gp54	56527	57558	4.8	343	39.5	P100 gp57, <i>L. innocua</i> ribonucleosid diphosphate reductase β subunit; Twort ORF99, Phage K ORF79, Phage G1 ORF102
gp55	57555	58007	4.3	150	17.3	P100 gp58, <i>L. innocua</i> flavodoxin homolog lin2257, <i>L. monocytogenes</i> lmo2153
gp56	58010	58306	4.9	98	10.8	P100 gp59
gp57	58330	58641	4.9	103	11.6	P100 gp60 (n-terminal) <i>L. innocua</i> putative thioredoxin lin2256, <i>L. monocytogenes</i> lmo2152
gp58	58616	59023	9.3	135	15.2	P100 gp60 (c-terminal)
gp59	59026	59163	10.2	45	5.0	P100 gp61
gp60	59166	60353	7.3	395	44.3	P100 gp62
gp61	60350	61264	5.3	304	34.7	P100 gp63, Phage K ORF110, Phage G1 ORF31 phosphoribosyl pyrophosphate synthetase
gp62	61275	63059	5.4	594	67.5	P100 gp64, Phage K ORF111, Phage G1 ORF118, nicotinamid phosphoribosyltransferase
gp63	63155	65614	8.4	819	94.8	P100 gp65, Phage K ORF18, Phage G1 ORF7
gp64	65708	66496	9.8	262	30.8	P100 gp66, Twort ORF64, Phage K ORF84, Phage G1 ORF66
gp65	66489	66803	9.7	104	12.0	P100 gp67, Phage K ORF85, Phage G1 ORF147, Twort ORF129; IHF integration host factor
gp66	66886	67722	5.4	278	31.9	P100 gp68, Phage G1 ORF35, Phage K ORF86/88/90 DNA Polymerase, Twort ORF4, LP65 orf59
gp67	67973	70165	5.7	730	84.3	P100 gp69, Phage G1 ORF37, Phage K ORF86/88/90 DNA Polymerase, Twort ORF4, LP65 orf59
gp68	70260	70736	4.9	158	18.6	P100 gp70, Phage K ORF91, Phage G1 ORF89, Twort ORF77
gp69	70774	72033	4.8	419	46.7	P100 gp71, Twort ORF22, Phage K ORF92, Phage G1 ORF20
gp70	72103	73347	7.9	414	46.1	P100 gp72, Phage K ORF93, Phage G1 ORF21, Twort ORF19, LP65 Recombinase A
gp71	73409	73786	9.3	125	14.5	P100 gp73, Twort ORF112, Phage G1 ORF121
gp72	73786	74424	7.7	212	25.2	P100 gp74, Phage K ORF94 putative σ -factor, Phage G1 ORF56, Twort ORF58, Phage SPO1
gp73	74483	74848	4.4	121	13.8	P100 gp75, <i>Bacillus licheniformes</i> , Phage SPP1 holin
gp74	74866	75573	4.7	235	26.1	P100 gp77, Phage K ORF95, Phage G1 ORF59, Twort ORF56
gp75	75681	76067	5.1	128	15.0	P100 gp78
gp76	76064	76993	6.2	309	35.5	P100 gp79

A511	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp77	77052	78323	7.2	423	47.6	P100 gp80, Phage K ORF98, Phage G1 ORF22, Twort ORF20, LP65 orf63 putative DNA repair exonuclease
gp78	78343	78726	10.0	127	13.9	P100 gp81
gp79	78734	79279	9.2	181	20.5	P100 gp82, Twort ORF71, phage G1 ORF75
gp80	79334	79531	8.7	65	7.1	P100 gp83
gp81	79582	80289	9.9	235	26.9	P100 gp84, Phage G1 ORF45, Phage K ORF101, Twort ORF43, LP65 orf45
gp82	80300	80782	10.7	160	18.5	P100 gp85, Twort ORF93, Phage K ORF102, Phage G1 ORF99
gp83	80842	81726	5.2	294	33.2	P100 gp86, hage G1 ORF36, Twort ORF38
gp84	81815	82252	5.4	145	16.6	P100 gp87
gp85	82258	82704	4.4	148	17.5	P100 gp88
gp86	82805	83509	7.9	272	27.2	P100 gp89
gp87	83514	84530	5.0	338	38.0	P100 gp90, Phage G1 ORF24, Phage K ORF15; ATPase
gp88	84517	85362	9.1	281	32.6	P100 gp91
gp89	85424	85879	4.9	151	17.4	P100 gp92
gp90	85885	86463	5.3	192	22.0	
gp91	86502	87074	9.8	190	21.5	P100 gp93
gp92	87071	87649	10.2	192	21.6	P100 gp94
gp93	87642	87974	10.1	110	12.7	P100 gp95
gp94	88260	88988	5.2	242	27.4	P100 gp96, Phage K ORF103, Phage G1 ORF47, Twort ORF48, LP65 orf41
gp95	89003	89470	4.3	155	17.9	P100 gp97, Phage K ORF104, Phage G1 ORF94, Twort ORF79
gp96	89585	90655	6.1	356	40.5	P100 gp98, Phage K ORF105, Phage G1 ORF100, Twort ORF88, transcription factor
gp97	90703	91272	5.0	189	21.3	P100 gp99
gp98	91275	91808	7.1	177	19.8	P100 gp100
gp99	91823	92581	9.7	252	29.5	P100 gp101
gp100	92594	92884	9.0	96	11.6	P100 gp102
gp101	93908	94267	4.5	119	13.6	
gp102	94352	94501	3.8	49	5.8	
gp103	94610	94873	10.4	87	10.4	P100 gp106
gp104	94957	95232	10.4	91	10.4	P100 gp107
gp105	95308	95760	9.5	150	17.1	
gp106	95762	95857	9.5	31	3.6	
gp107	95915	96568	9.9	217	25.4	LP65 orf128, Phage Mu <i>mom</i> protein
gp108	96585	96812	10.1	75	8.7	
gp109	96886	97083	4.5	65	7.2	
gp110	97102	97308	4.0	68	8.1	<i>L. innocua</i> lin1298
gp111	97312	97587	10.1	91	10.4	
gp112	97714	98016	4.6	100	11.9	
gp113	98088	98330	4.3	80	9.2	
gp114	98413	98556	9.2	47	5.1	
gp115	98759	99100	5.1	113	13.1	
gp116	99275	99496	3.5	73	8.4	<i>S. thermophilus</i>
gp117	99502	99945	4.1	147	17.0	<i>L. monocytogenes</i> F6854 LMO2652
gp118	100078	100620	5.0	180	20.6	

A511	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp119	100645	100923	9.9	92	10.4	
gp120	101247	101534	7.1	95	10.8	
gp121	101700	101954	8.0	84	9.0	P100 gp109
gp122	101954	102211	5.1	84	9.5	P100 gp110
gp123	102235	102420	7.6	61	7.3	P100 gp111
gp124	102438	102611	10.4	57	6.2	P100 gp112
gp125	102753	103028	5.1	91	10.3	P100 gp113
gp126	103042	103329	5.7	95	10.9	P100 gp114
gp127	103482	103754	4.2	90	10.3	P100 gp115
gp128	103887	104015	10.4	42	4.8	P100 gp116
gp129	104086	104232	10.6	48	5.9	
gp130	104318	104728	4.8	136	15.4	P100 gp117, <i>S. aureus</i> phage phi13 possible sensor protein
gp131	104718	104903	4.4	61	6.7	
gp132	104905	105126	4.5	73	8.4	P100 gp119
gp133	105133	105246	10.2	37	4.3	P100 gp121 (n-terminal)
gp134	105425	105925	4.7	166	18.9	P100 gp122
gp135	105928	106434	4.7	168	19.1	P100 gp123
gp136	106444	107628	9.0	394	46.1	P100 gp124, PSA GP52
gp137	107781	108206	9.0	141	16.6	P100 gp125
gp138	108224	108514	3.9	96	11.2	P100 gp126, <i>L. innocua</i> lin0093, PSA GP37
gp139	108511	108696	4.6	61	7.2	P100 gp127, <i>L. innocua</i> lin0092
gp140	108696	109043	5.3	115	13.2	P100 gp128, <i>Actinoplanes</i> phage phiAsp2 Pas24,
gp141	109075	109356	4.0	93	10.7	P100 gp129
gp142	109435	109767	5.1	110	12.9	P100 gp130
gp143	109867	110163	9.0	98	11.2	P100 gp131
gp144	110228	110407	9.5	59	6.3	P100 gp132
gp145	110430	110792	5.3	120	14.2	P100 gp133
gp146	110792	111148	5.5	118	13.7	P100 gp134
gp147*	111965	111192	5.6	258	30.1	P100 gp135, <i>L. monocytogenes</i> lmo0655 putative phosphatase, <i>L. innocua</i> lin0658, Phage K ORF6, Twort ORF49
gp148*	112298	111978	4.4	106	12.2	P100 gp136
gp149*	112461	112291	9.2	56	6.8	P100 gp137
gp150*	113032	112505	6.4	175	21.1	<i>L. monocytogenes</i> H7858 gp51, A118 gp51, PSA GP35
gp151*	113310	113029	8.8	93	10.9	P100 gp141
gp152*	113594	113370	9.0	74	8.9	P100 gp142
gp153	114137	113595	9.2	180	21.0	P100 gp143
gp154	114361	114134	4.0	75	9.3	P100 gp144
gp155*	115614	114364	6.6	416	48.3	P100 gp145
gp156*	116038	115625	4.8	137	15.8	P100 gp146
gp157*	116592	116104	9.7	162	18.8	P100 gp147
gp158*	117230	116598	9.9	210	24.0	P100 gp148
gp159*	117430	117233	5.2	65	7.7	P100 gp149
gp160*	117839	117420	8.8	139	15.5	P100 gp150
gp161*	118021	117836	6.8	61	6.8	
gp162*	118629	118012	6.9	205	23.6	P100 gp151, <i>E. faecium</i> putative phosphoesterase, Phage KVP40

A511	Start	Stop	pI	AA	MW (kDa)	Protein similarities in the database
gp163*	118844	118626	9.7	72	8.1	P100 gp152
gp164*	119238	118861	4.5	125	14.4	P100 gp153, Phage SPO1 ORF1
gp165*	119597	119241	9.7	118	13.5	P100 gp154
gp166*	120615	119674	5.6	313	36.1	P100 gp155, Phage K ORF21, Phage G1 ORF32; DNA ligase
gp167*	121165	120629	4.8	178	20.3	P100 gp156
gp168*	121353	121162	10.3	63	7.8	P100 gp157, <i>L. innocua</i> lin2404
gp169*	121857	121354	9.9	167	19.3	P100 gp158, Phage K ORF4, Phage G1 ORF86
gp170*	122128	121859	9.7	89	10.3	P100 gp159
gp171*	123386	122175	6.1	403	45.0	P100 gp160, <i>Bacillus cereus</i> putative rtcB protein
gp172*	123695	123492	9.7	67	7.8	P100 gp161
gp173*	124119	123688	5.0	143	16.8	P100 gp162
gp174*	124357	124133	9.3	74	8.3	P100 gp163
gp175*	124613	124371	9.8	80	9.6	P100 gp164, Twort ORF184, Phage G1 ORF187, A118 repressor protein
gp176*	126236	124716	6.3	506	57.3	P100 gp165, Phage KVP40 protein KVP40.0300
gp177*	126926	126594	10.2	110	12.9	Twort ORF180, Phage G1 ORF200
gp178*	127125	126907	10.1	72	8.2	P100 gp166
tRNA*	127859	127785				tRNA-Proline (CCA)
gp179*	128604	128197	10.4	135	16.5	P100 gp167
gp180*	128827	128630	7.3	65	7.7	P100 gp168
tRNA*	129047	128977				tRNA-Arginine (AGA)
gp181	129022	129186	10.5	54	6.3	
tRNA*	129364	129294				tRNA-Glycine (GGA)
tRNA*	129506	129434				tRNA-Asparagine (AAC)
gp182	130028	130282	10.1	84	9.6	
tRNA*	130218	130127				tRNA-Serine (UCA)
tRNA*	130302	130231				tRNA-Phenylalanine (UUC)
tRNA*	130379	130308				tRNA-Lysine (AAA)
tRNA*	130576	130505				tRNA-Tryptophan (UGG)
tRNA*	130651	130580				tRNA-Glutamine (CAA)
tRNA*	130741	130670				tRNA-Threonine (ACA)
tRNA*	130905	130824				tRNA-Tyrosine (UAC)
gp183*	131237	130962	5.7	91	10.4	P100 gp169, Phage KVP40 protein KVP40.0, <i>E. faecalis</i> V583
tRNA*	131349	131267				tRNA-Leucine (CUA)
gp184*	131786	131361	10.1	141	16.4	P100 gp170
tRNA*	131888	131817				tRNA-Aspartate (GAC)
tRNA*	132064	131993				tRNA-Isoleucine (AUC)
tRNA*	132327	132241				tRNA-Serine (AGC)
tRNA*	132479	132409				tRNA-Cysteine (UGC)
gp185*	133385	133149	4.8	78	8.8	P100 gp171
gp186*	133776	133435	5.4	113	13.0	P100 gp172, Phage G1 ORF209
gp187*	134145	133801	5.0	114	13.8	P100 gp173, Twort ORF158
gp188*	134430	134161	6.0	89	10.0	P100 gp174

ACKNOWLEDGEMENTS

Thanks to all of you, who supported and encouraged me in either way to finish this work! Thanks to my professors for the encouragement and for their patience; thanks to my lab mates and the “girls office” for any kind of help and the diverting chats, and special thanks to my friends and family for cheering me up, when I was down.

Many thanks to Professor Richard Calendar, University of California, Berkeley, USA for his support in phage purification and DNA extraction.

Parts of this work comprising construction of genomic libraries of phages B025 and B054, and annotation of B054 were the subjects of a Bachelor and a Diploma thesis, performed by Maja Lichstein and Mathias Schmelcher, respectively. Thank you for your interest and valuable assistance!

The research resulting in this doctoral thesis was funded by Competence Network PathoGenoMik, Center University Würzburg (project number PTJ-BIO/031U213B).

CURRICULUM VITAE

Julia Dorscht

Persönliche Angaben

Geburtsname	Gruner
Geburtsdatum	21.07.1975
Geburtsort	Bamberg
Familienstand	verheiratet
Staatsangehörigkeit	deutsch

Ausbildung

09.1987 – 06.1995	Dientzenhofer-Gymnasium, Bamberg Abschluss: Allgemeine Hochschulreife
10.1995 – 02.2000	Julius-Maximilians-Universität Würzburg Studium der Biologie; Diplomprüfung in den Fächern Tierphysiologie, Mikrobiologie und Humangenetik
03.2000 – 02.2001	Merck KGaA, Darmstadt Diplomarbeit im Fach Mikrobiologie
03.2001 – 05.2001	Merck KGaA, Darmstadt Laborpraktikum in der Abteilung Forschung, spezielle Analytik
09.2001 – 02.2002	Eberhard-Karls-Universität, Tübingen Stipendium des Graduiertenkollegs „Infektionsbiologie“ am Lehrstuhl für Mikrobielle Genetik
03.2002 – 08.2005	Technische Universität, München Promotionsstudium in der Abteilung für Mikrobiologie, Zentralinstitut für Ernährungs- und Lebensmittelforschung ZIEL

Berufliche Tätigkeit

Seit 05.2006	BIOTRONIK GmbH & Co. KG, Berlin Clinical Project Manager, Center for Clinical Research and Scientific Studies, Erlangen
--------------	---