KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoraatsschool Wetenschap & Technologie
Faculteit Bio-ingenieurswetenschappen
Departement Biosystemen

Doctoraatsproefschrift nr. 879 aan de faculteit Bio-ingenieurswetenschappen van de K.U.Leuven

ISOLATION AND CHARACTERIZATION OF LYTIC BACTERIOPHAGES INFECTING *PSEUDOMONAS AERUGINOSA*

Pieter-Jan Ceyssens

Promotor:
Prof. G. Volckaert, K.U.Leuven

Leden van de examencommissie:
Prof. R. Geers, voorzitter
Prof. B. Goddeeris, K.U.Leuven
Prof. J. Robben, K.U.Leuven
Prof. R. Lavigne, K.U.Leuven
Prof. M. Vaneechoutte, Universiteit Gent
Prof. R. Leplae, Université Libre de Bruxelles
Prof. H.W. Krisch, Université Paul Sabatier, France

Proefschrift voorgedragen tot het behalen van de graad van Doctor in de Bio-ingenieurswetenschappen

December 2009

ISBN 978-90-8826-123-7
D/2009/11.109/41

Table of Contents

Dankwoord **I**

Samenvatting **II**

Summary **VI**

List of abbreviations **IX**

List of publications **X**

Chapter 1. Introduction and Background **1**

1.1 Introduction on Bacteriophages 1
- 1.1.1 Short history on phage research 1
- 1.1.2 Phage morphology forms the basis for classification 2
- 1.1.3 The origin of bacteriophages 4

1.2 The impact of bacteriophages on the biosphere 6
- 1.2.1 Variation in bacteriophage abundance 6
- 1.2.2 'Killing the Winner' shapes community structures 8
- 1.2.3 Phage impact on host fitness 9
- 1.2.4 Impact of phages and Gene Transfer Agents on bacterial genetic exchange 11

1.3 The diversity of bacteriophages 12
- 1.3.1 The limits of current sampling methods 13
- 1.3.2 Diversity based on single-locus studies 14
- 1.3.3 Diversity based on metagenomics 16

1.4 Bacteriophage genomics 18
- 1.4.1 Introduction 18
- 1.4.2 General phage genome build-up 19
- 1.4.3 Genetic mosaicism in temperate phages 21
- 1.4.4 Conserved core and hyperplastic regions of T4-like genomes 23
- 1.4.5 ORFans, the dark matter of the phage universe 27
- 1.4.6 Phage evolution 28

1.5 Bacteriophages of Pseudomonas aeruginosa 30
- 1.5.1 The host bacterium: *Pseudomonas aeruginosa* 31
- 1.5.2 Bacteriophages of *P. aeruginosa* 32
- 1.5.3 Phage Therapy against *P. aeruginosa* 34

Chapter 2. Materials and Methods **36**

2.1 Bacterial strains, plasmids and growth media 36

2.2 Phage manipulations 37
- 2.2.1 Bacteriophage isolation, amplification and purification 37
- 2.2.2 Host range analysis and stability assays 38
- 2.2.3 One-step growth and adsorption curves 38
- 2.2.4 Electron Microscopy 39

2.3 Phage genome analysis 39
- 2.3.1 Phage DNA isolation and restriction analysis 39
- 2.3.2 Phage genome sequencing 40
- 2.3.3 Experimental promoter identification 43
- 2.3.4 *In vivo* splicing assay 43

2.4 Phage proteome analysis 44
2.4.1 Phage protein profiles and zymogram analysis 44
2.4.2 Mass spectrometric identification of virion proteins 44
2.4.3 '*de novo*' MS as a novel phage typing method 46
2.4.4 Tandem affinity purification (TAP) 47

Chapter 3. Isolation of new *P. aeruginosa* phages: '*de novo*' peptide sequencing as a novel tool to assess viral diversity 50

3.1 Compilation of a P. aeruginosa phage library 50
3.1.1 Phage isolation and preliminary characterization 50
3.1.2 Morphological characterization 53
3.1.3 Host range analysis and receptor analysis 54

3.2 Phage typing by de novo sequencing of structural proteins 55
3.2.1 Theoretical background 57
3.2.2 *De novo* peptide sequencing 57

3.3 Discussion 60

Chapter 4. *Pseudomonas* phage ϕKMV represents a conserved genus of widely spread virulent phages 62

4.1 Phage characteristics 62

4.2 Genome analysis 64
4.2.1 Sanger sequencing identified single-stranded nicks 64
4.2.2 Comparative annotation of the ϕKMV-like genomes 66
4.2.3 Transcriptional regulation 69
4.2.4 Genome organization 71

4.3 Influence of LUZ19 infection on the host RNA Polymerase complex 76

4.4 Discussion 78
4.4.1 Genome architecture of T7-related phages 78
4.4.2 The role of the conserved intergenic motif 80

Chapter 5. The intron-containing genome of lytic *Pseudomonas* phage LUZ24 resembles temperate phage PaP3 82

5.1 Phage characteristics 82

5.2 Genome analysis 83
5.2.1 Genomic properties, gene prediction and regulation 83
5.2.2 Genes involved in host conversion and metabolism 87
5.2.3 Genes involved in particle formation and host lysis 88

5.3 Discussion 90

Chapter 6. Bacteriophages LUZ7 and LIT1 are the first characterized N4-like *Pseudomonas* phages 92

6.1 Phage characteristics 93

6.2 Genome analysis 94
6.2.1 Determination of genomic phage ends 94
6.2.2 Genome annotation 95
6.2.3 Transcriptional control in N4-like *Pseudomonas* phages 98

6.2.4 Particle formation and host lysis 99

6.3 Discussion 101

Chapter 7. *Comparative analysis of the widespread and conserved PB1-like Myoviridae* **102**

7.1 Phage characteristics 103

7.2 Genome analysis 104
7.2.1 Genome properties and annotation 104
7.2.2 Transcriptional regulation 108
7.2.3 Clusters of small proteins 109
7.2.4 DNA replication machinery 109
7.2.5 Phage particle proteins 110
7.2.6 Host lysis 113

7.3 Discussion 113

Chapter 8. Bacteriophage YuA represents a new genus of *Pseudomonas*-infecting *Siphoviridae* **115**

8.1 Phage characteristics 115

8.2 Genome analysis 116
8.2.1 Genome annotation and similarity to other phages 116
8.2.2 Transcriptional regulation 120
8.2.3 Genes involved in DNA metabolism and genome replication 122
8.2.4 Genes involved in host interaction 124
8.2.5 Genes involved in particle formation and host lysis 125

8.3 Discussion 126

Chapter 9. General conclusions and perspectives **128**

9.1 The global spread of bacteriophage species 128
9.1.1 Genome sequencing of *Pseudomonas* phages 128
9.1.2 Every phage everywhere 129
9.1.3 How many types remain to be discovered? 130

9.2 Future applications and prospects 131
9.2.1 Phage therapy and targeted phage modifications 131
9.2.2 The exploitation of phage genetic information 132

References 134

Dankwoord

‘The field of bacterial viruses is a fine playground for serious children who ask ambitious questions.’ Hoewel deze uitspraak van Max Delbrück (1906-1981) dateert van 60 jaar geleden, past ze perfect bij de mijmeringen over de voorbije jaren als doctorandus aan het LoGT. Geen seconde heb ik spijt gehad van de tijd die in doorbracht in de speeltuin van de kleine virussen en hun wonderbaarlijke genomen. De persoonlijke evolutie in deze jaren was enorm, en is beïnvloed door velen.

Professor Volckaert, mijn promoter, stond 35 jaar geleden mee aan de wieg van de faaggenomica, en legde hiermee een hoeksteen waarop momenteel wolkenkrabbers gebouwd worden. Ik dank hem voor alle steun en interesse, en wens hem welverdiende emeritaatsjaren toe. Zijn opvolger, Rob, staat te trappelen om het LoGT verder uit te bouwen tot een gerenommeerd faaglab, een ambitie waarin ik mij perfect kan vinden. Bedankt voor de kennismaking met de bacteriofagen en de daaropvolgende aanjagingen, wilde ideeën en correcties, Rob, en dat het beste nog moge komen. Bij uitbreiding dank ik alle leden van de examencommissie hartelijk voor het lezen en corrigeren van deze tekst, en het IWT voor de broodnodige financiële steun.

Mijn eerste stappen als doctorandus werden vakkundig begeleid door Kirsten en Marleen, de twee ‘oermoeders’ van het LoGT die me de kritische blik en de knepen van het vak aanleerden, de basis van al wat volgde. Van evengroot belang waren mijn ‘compagnons de route’, Yves en Bart, echte wetenschappers en nog fijne mensen ook. Hoewel we elk naar andere oorden zullen vliegen, blijven de wortels dezelfde en wie weet kruisen onze wegen nog wel ergens in de toekomst.

In de loop van de jaren begeleidde ik vier thesisstudenten, die zich belangeloos en vrijwillig ingezet hebben ingezet en wiens resultaten een prominente plaats innemen in dit werk. Wesley, Mekonnen, Simon en Liesbet, veel succes in jullie verdere carrière. Collega-doctorandi komen en gaan, net zoals de gespreksonderwerpen aan tafel. Boeken en docusoaps kunnen geschreven worden over de dagelijkse beslommeringen binnen het LoGT en de warme mensen die er werken. Ik haal diep adem, en dank Tim, Gert, Tine, Sigrid, Maarten, Stijn, Jeroen, Elkes, Katrien, Anneleen, Joleen, Wesley, Barbara, Ingrid en Andrew voor de fantastische werkomgeving. Op naar meer, zou ik zeggen.

Interessante onderzoeksepisodes speelden zich ook af buiten de muren van het LoGT. Mijn bijzondere appreciatie gaat uit naar Johan Robben, Jean-Paul Noben en Erik Royackers, voor de introductie in de wondere wereld van de proteomics, en naar Jean-Paul Pirnay en Daniël De Vos voor al het werk op en rond onze *Pseudomonas* bacteriën. I would also like to express my gratitude to prof. Konstantin Miroshnikov and prof. Andew Kropinski for their heart-warming hospitality and the time I could spend in their labs. Andrew Chibeu, I wish you all the best in your postdoctoral research in this Canadian lab! And mind the bears my friend.

Dichter bij huis en dichter bij de avondklok zijn uiteraard andere mensen belangrijk. Het voorbije decennium zou heel wat grijzer en stiller geweest zijn zonder de leden van de Posse uit Dormaal. Een mens wordt spontaan goed gezind door tijd door te brengen met deze bonte verzameling van lichtelijk geschifte mensen. En nu we toch over een grote familie bezig zijn: Mijn lieve grootouders, papa, Lieve, Frederik, Jasmine, Vincent en Birgit, wat een aangenaam thuiskomen is het toch steeds. Een baken van zekerheid, om steeds op te vallen. En dat ben jij ook, mijn lieve Liesje, zelden heb ik zoveel tederheid verpakt zien zitten in zo'n klein volume. Eén uit de duizend zou ik zeggen, maar daarmee zou ik u tekort schieten.

Helaas zal 2009 niet geboekstaafd staan als het jaar waarin dit doctoraat geschreven werd. Het zal altijd het jaar blijven waarin ik veel te vroeg en veel te plots afscheid heb moeten nemen van u, mijn lieve moeder. Bedankt voor het onvoorwaardelijke geloof, de liefde en de vele woorden van wijsheid, en voor alle waarden die ik zal verder dragen in mijn leven. Dit boekje draag ik dan ook volledig aan u op. Always on my side.

Pieter-Jan.

Samenvatting

In het laatste decennium is er een brede wetenschappelijke consensus ontstaan over de enorme invloed die bacteriële virussen of bacteriofagen uitoefenen op de biosfeer. Bacteriofagen doden 4 tot 50% van de dagelijks geproduceerde bacteriën, dragen in belangrijke mate bij tot globale geochemische cycli en vormen het reservoir van de grootste genetische diversiteit op aarde (Suttle, 2005). De hoofddoelstelling van dit doctoraat was het onderzoek naar de globale genetische diversiteit van één specifieke faaggroep, nl. deze van de *Pseudomonas aeruginosa*-infecterende fagen. Deze alomtegenwoordige en metabolisch veelzijdige opportunistische pathogeen vertoont zowel intrinsieke als verworven resistentie tegen tal van antibacteriële middelen.

In een eerste stap werden 28 *P. aeruginosa* fagen geïsoleerd uit afvalwaterstalen en natuurlijk omgevingswater, verzameld in 14 verschillende landen verspreid over 4 continenten. Uit deze collectie werden 15 virulente fagen geselecteerd op basis van heldere plaquemorfologie; deze *Caudovirales* vertoonden alle unieke DNA- en eiwitprofielen en clusterden in zes verschillende morfologische groepen. Een uitgebreide infectieanalyse van 122 AFLP-getypeerde *P. aeruginosa*-stammen toonde aan dat 87% van deze stammen door minstens één faag uit de selectie kon worden afgedood. Analyse van de gevoeligheid van buitenmembraanmutanten voor faaginfectie toonde aan dat tien fagen afhankelijk zijn van de aanwezigheid van Type IV pili voor succesvolle infectie. Gedetailleerde moleculaire verwantschappen tussen de verschillende partikels werden in kaart gebracht door de primaire eiwitsequentie van de majeure structurele faageiwitten massaspectrometrisch *de novo* te bepalen. De hieraan gekoppelde staalbereiding werd voor de eerste keer toegepast op virale eiwitten, en liet toe om zeven verschillende faaggroepen af te bakenen, zonder voorafgaand morfologisch of genetisch onderzoek. Fagen van vijf verschillende groepen werden geselecteerd en onderworpen aan verdere analysen.

Fagen LKD16, LUZ19, LKA1 en ϕKF77 hebben een korte staart (*Podoviridae*), en dragen in hun capside een dsDNA genoom (41,5-43,5 kb) dat begrensd wordt door lange terminale herhalingen en codeert voor 52 tot 56 genen. De meerderheid van de voorspelde structurele eiwitten van LKD16 and LKA1 werd geïdentificeerd met ESI-MS/MS. Fagen LKD16, LUZ19 en ϕKF77 zijn nauw verwant met de reeds gekende faag ϕKMV (Lavigne et al., 2003), hetgeen meer doordachte genannotaties toeliet op basis van vergelijkende genoomstudies. Deze genomen bevatten vijf enkelstrengige breuken op de niet-coderende streng, en dit op specifieke intergenische locaties die mogelijk verband houden met faagspecifieke transcriptie-initiatie. Faag LKA1 is veel minder verwant met ϕKMV; er is geen DNA similariteit aanwezig, en slechts 48% van de gecodeerde eiwitten hebben een tegenhanger in ϕKMV.

Voornamelijk de vroege genoomregio's en staartvezeleiwitten zijn verschillen sterk, wat gereflecteerd wordt in opvallende verschillen in infectiegerelateerde parameters. Desalniettemin is de algemene genoomopbouw van deze fagen duidelijk bewaard doorheen de evolutie, en onderscheiden deze fagen zich van andere *Autographiviridae* (Lavigne et al., 2008) door de localizatie van het gen coderend voor het RNA polymerase, stroomopwaarts van de structurele genoomregio.

Faag LUZ24 (45.625 bp) behoort eveneens tot de familie van de *Podoviridae*, maar draagt een bidirectioneel afgeschreven genoom dat begrensd wordt door twee directe herhalingen van 184 bp, en bevat geen eigen RNA polymerase. *Promoter trap* analyse identificeerde zeven σ^{70} promotoren *in vitro*, met een duidelijke stijging van de promotorsterkte naar de latere genoomregio toe. Met behulp van *reverse transcriptie* werd de *in vivo* splicing van een intronsegment (668 bp) segment uit het DNA polymerase aangetoond. De mantel van LUZ24 is opgebouwd uit negen verschillende eiwitten. Hoewel deze faag 71% identiek is aan de getemperde faag PaP3 (Tan et al., 2007), wijzen genoomopbouw en infectieparameters duidelijk op een virulente levenscyclus.

De lineaire genomen van de *Podoviridae* LIT1 (72.543 bp) en LUZ7 (74.901 bp) coderen voor resp. 90 en 117 proteïnen. Massaspectrometrische analyse identificeerde 14 verschillende eiwitten in het partikel van LIT1. In tegenstelling tot de onderling nauw verwante groep van de phiKMV-achtige fagen, zijn LIT1 en LUZ7 niet verwant op nucleotidenniveau. Beide fagen delen wel hun algemene genoomopbouw met de modelfaag N4, hoewel significante verschillen aanwezig zijn in structurele genen en transcriptiecontrole. In tegenstelling tot N4, vertonen LUZ7 en LIT1 geen lysisinhibitie onder standaard laboratoriumomstandigheden.

Tijdens de genoomanalyse van vier strikt virulente fagen (LBL3, LMA2, 14-1 en SN) werden sterke gelijkenissen vastgesteld met de historisch belangrijke fagen PB1 en F8 (Holloway et al., 1960). Deze fagen dragen een lineair, niet-gepermuteerd genoom tussen 64.427 en 66.530 bp in een zuur-resistent virion dat opgebouwd is uit 22 structurele eiwitten. Deze structurele eiwitten vormen samen een icosahedraal capside (diameter 74 nm) en een niet-flexibele, contractiele staart van ongeveer 140 nm (*Myoviridae*). De faaggenomen zijn georganiseerd in (tenminste) zeven transcriptionele blokken en coderen tussen 88 (LBL3) en 95 (LMA2) eiwitten. Massaspectrometrische analyse identificeerde posttranslationele modificaties in twee structurele eiwitten, en een mogelijke leesraamverschuiving in gen 42.

Faag YuA tenslotte heeft een langwerpig capside waarin een circulair genoom van 58.663 bp vervat zit. Aan dit faagcapside is een flexibele en niet-contractiele staart bevestigd, en dus wordt YuA ingedeeld in de familie van de *Siphoviridae*. Het genoom codeert voor 78 proteïnen, en is mozaïsch gerelateerd aan tal van andere fagen zoals ϕJL001, 73, B3, DMS3 en D3112. Achttien van deze

eiwitten werden teruggevonden in het mature faagpartikel. Van de 5 verschillende bacteriële promotoren die experimenteel geïdentificeerd werden, hebben drie promotoren die de transcriptie van de structurele genen reguleren een σ^{54}-specifieke bindplaats. De afhankelijkheid van deze promotoren voor de σ^{54} factor van de gastheer werd bevestigd door analyse van een *rpoN* deletiemutant van *P. aeruginosa* PAO1. Hoewel YuA 91% identiek is aan faag M6 (Lindberg en Latta, 1974), coderen beide fagen samen voor 15 proteïnen die niet in de andere faag voorkomen.

Op deze wijze werden vijf nieuwe faaggenera van *P. aeruginosa*-specifieke fagen gedefinieerd, alle met duidelijk verschillende en unieke genoomarchitectuur. Ondanks de sterk uiteenlopende staalnameplaatsen werd zeer hoge sequentiesimilariteit en zeer beperkte horizontale gentransfer vastgesteld tussen individuele virussen van eenzelfde genus, indicatief voor de globale verspreiding van deze genera. Met uitzondering van YuA, omvatten de beschreven genera sterk lytische fagen die rechtstreeks toepasbaar zijn in therapeutische settings. De ontrafeling van deze genoomsequenties leidde tot de identificatie van 1.028 nieuwe genen, die een rijke bron zullen zijn voor toekomstig onderzoek naar biotechnologisch en antibacterieel interessante eiwitten.

Summary

During the last decennium, it has become widely accepted that bacterial viruses or bacteriophages are extremely abundant and exert enormous influences on the biosphere. Phages kill between 4–50% of the daily produced bacteria, contribute to global geochemical cycles and a reservoir of the greatest genetic diversity on Earth (Suttle, 2005). The main goal of this PhD dissertation was to assess the global genomic diversity of a specific group of virulent bacteriophages, i.e. phages infecting *Pseudomonas aeruginosa.* This opportunistic pathogen is ubiquitous in nature, has wide metabolic versatility and high intrinsic and acquired resistance to antimicrobials.

A collection of 28 newly isolated *P. aeruginosa* phages was established, sampled in 14 different countries traversing four continents from both natural environments and hospital sewage. Fifteen of these phages were selected based on their ability to form clear plaques, a hallmark feature of virulent phages. All displayed unique DNA and protein profiles and clustered morphologically into six groups within the three major families of the *Caudovirales*. Extensive host range studies on a library of 122 AFLP-genotyped clinical *P. aeruginosa* strains showed that the phages lysed 87% of the strains. Infection analysis of outer membrane mutants identified ten phages as type IV pili-dependent. More detailed information about their molecular relatedness was gathered by *de novo* peptide sequencing of major virion proteins using MALDI-TOF/TOF technology. Applying this technique for the first time to viruses, seven groups of closely related phages were identified without the need of prior knowledge of genome content and/or electron microscopic imaging. Phages from five different groups were selected for further genomic and proteomic characterization.

Podoviruses LKD16, LUZ19, LKA1 and ϕKF77 encapsulate linear dsDNA genomes of 41.5-43.5 kb in length, delineated by long terminal repeats (~400 bp) and encoding between 52 and 56 genes. The majority of the predicted structural proteins were experimentally confirmed as part of the phage particle using mass spectrometry. Phages LKD16, LUZ19 and ϕKF77 are closely related (83-91% overall DNA homology) to the previously sequenced phage ϕKMV (Lavigne et al., 2003), allowing thoughtful gene annotations based on comparative genomics. At specific intergenic locations on the non-coding strand, these genomes contain single-stranded nicks which might be related to phage-specific transcription. In contrast, LKA1 is more distantly related to ϕKMV, lacking significant DNA homology and showing protein similarity in 48% of its gene products. Especially the early region of the LKA1 genome has diverged strongly, and intriguing differences in tail fiber genes likely reflect observed discrepancies in infection-related properties. Nonetheless, the general genome organization is clearly conserved among all phages, as they all encode a single-subunit RNA

polymerase gene adjacent to the structural genome region, a feature which distinguishes them from other members of the *Autographiviridae* (Lavigne et al., 2008).

Bacteriophage LUZ24 (45,625 bp) also belongs to the family of the *Podoviridae*, but carries a bidirectionally transcribed dsDNA genome delineated by two direct terminal repeats of 184 bp. *In vitro* transcriptional analysis identified seven σ^{70} promoters, revealing a bias towards stronger promoter strength in the late genomic region. Reverse transcription demonstrated *in vivo* splicing of a 668 bp Group I intron embedded within the DNA polymerase gene. Using mass spectrometry, nine structural proteins were identified as part of the phage particle. Although LUZ24 displays an overall 71% sequence similarity to the temperate phage PaP3 (Tan et al., 2007), its genome architecture and infection characteristics clearly indicate a virulent nature.

Podoviruses LUZ7 and LIT1 encapsulate linear genomes of 74,901 and 72,543 bp, which include direct repeats of 660 and 655 bp and encode 117 and 90 proteins, respectively. In contrast to the close-knitted group of the 'phiKMV-like viruses', LUZ7 and LIT1 share no conservation at the nucleotide level, and display protein similarity in only 60 gene products. Both phages share their genome architecture with the well-studied coliphage N4, which was long considered a genetic orphan. However, contrary to N4, LUZ7 and LIT1 do not display lysis inhibition under standard laboratory conditions, and subtle differences in transcriptional control and particle formation were detected.

One genus of Myoviruses is described which encompasses four strictly virulent phages (LBL3, LMA2, 14-1 and SN) which share strong DNA similarity with the historically important *P. aeruginosa* phage PB1 (Holloway et al., 1960). They all encapsulate linear, non-permuted genomes between 64,427 and 66,530 bp within a solid, acid-resistant isometric capsid (diameter: 74 nm) and carry non-flexible, contractile tails of approximately 140 nm. The genomes are organized into (at least) seven transcriptional blocks, alternating on both strands, and encode between 88 (LBL3) and 95 (LMA2) proteins. Their virion particles are composed of at least 22 different proteins, which were identified using mass spectrometry. Posttranslational modifications are suggested for two proteins, and a frameshift hotspot was identified within ORF42, encoding a structural protein.

Finally, *Pseudomonas aeruginosa* phage YuA has an elongated head, encapsulating a circularly permuted genome of 58,663 bp, and a flexible, non-contractile tail which is terminally and subterminally decorated with short fibers (*Siphoviridae*). The YuA genome encodes 78 gene products and is mosaically related to phages ϕJL001, 73, B3, DMS3 and D3112. Eighteen YuA proteins were identified as part of the phage particle by mass spectrometry analysis. Five different bacterial promoters were experimentally identified using a promoter trap assay, three of which have a σ^{54}-specific binding site and regulate transcription in the genome region involved in phage particle

formation and host lysis. The dependency of these promoters for the host σ^{54} factor was confirmed by analysis in an *rpoN* mutant strain of *P. aeruginosa* PAO1. At the DNA level, YuA is 91% identical to phage M6 of the Lindberg Typing set. Despite this level of DNA homology throughout the genome, both phages combined have 15 unique genes that do not occur in the other phage.

As such, this study defined five new genera of *P. aeruginosa*-infecting phages, all comprising phages with highly discrete genome architecture. Despite large spatial separations between phage isolations, very high sequence similarity and limited horizontal gene transfer was found between individual viruses of the described genera, indicative for the global spread of *P. aeruginosa* phage species. Except for the YuA, all genera comprise highly lytic phages devoid of any virulence-associated genes, and thus suitable for applications in phage therapy. A total of 1,028 novel phage genes are encoded in these newly characterized phage genomes, providing a rich source for future screenings for potentially interesting biotechnological and antibacterial proteins.

List of abbreviations

aa	amino acid
ACN	acetonitrile
AFLP	amplified fragment length polymorphism
bp	basepair(s)
CHCA	α-Cyano-4-hydroxycinnamic acid
(k)Da	(kilo)Dalton
DIEA	diisopropylethylamine
ds	double-stranded
DTR	direct terminal repeat
EM	electron microscopy
ESI-MS	electrospray ionization mass spectrometry
gp(s)	gene product(s)
GTA	generalized transducing agents
HGT	horizontal gene transfer
ICTV	International Committee on the Taxonomy of Viruses
ICU	Intensive care unit
kb	kilobases
LPS	lipopolysaccharide
MALDI-TOF	matrix-assisted laser desorption/ionisation-time of flight mass spectrometry
MDR	multi-drug resistant
MOI	multiplicity of infection
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
nt	nucleotide(s)
ORF(s)	open reading frame(s)
(f)RFLP	(fluorescent) restriction fragment length polymorphism
RNAP	RNA polymerase
SDS-PAGE	sodium dodecyl sulphate polyacrylamide electrophoresis
SAR	single-arrest and release
SPITC	4-sulphophenyl isothiocyanate
ss	single-stranded
ssb	single-stranded binding protein
TAP	tandem affinity purification
TEV	tobacco etch virus
TFA	trifluoroacetic acid
VBR	virus / bacteria ratio
WSA	whole shotgun approach

List of Publications

Ceyssens, P.J., Miroshnikov, K., Mattheus, W., Krylov, V., Robben, J., Noben, J.P., Vanderschraeghe, S., Sykilinda, N., Kropinski, A.M., Volckaert, G., Mesyanzhinov, V., and Lavigne, R. (2009) Comparative analysis of the widespread and conserved PB1-like viruses infecting *Pseudomonas aeruginosa*. *Environmental Microbiology* **11**: 2874-2883.

Kulakov, L.A., Ksenzenko, V.N., Shlyapnikov, M.G., Kochetkov, V.V., Del Casale, A., Allen, C.C., Larkin, M.J., Ceyssens, P.J., and Lavigne, R. (2009) Genomes of "phiKMV-like viruses" of *Pseudomonas aeruginosa* contain localized single-strand interruptions. *Virology* **391**: 1-4.

Lecoutere, E., Ceyssens, P.J., Miroshnikov, K.A., Mesyanzhinov, V.V., Krylov, V.N., Noben, J.P., Robben, J., Hertveldt, K., Volckaert, G., and Lavigne, R. (2009) Identification and comparative analysis of the structural proteomes of phiKZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages. *Proteomics* **9**: 3215-3219.

Chibeu, A., Ceyssens, P.J., Hertveldt, K., Volckaert, G., Cornelis, P., Matthijs, S., and Lavigne, R. (2009) The adsorption of *Pseudomonas aeruginosa* bacteriophage phiKMV is dependent on expression regulation of type IV pili genes. *FEMS Microbiological Letters* **296**: 210-218.

Ceyssens, P.J., Noben, J.P., Ackermann, H.W., Verhaegen, J., De Vos, D., Pirnay, J.P., Merabishvili, M., Vaneechoutte, M., Chibeu, A., Volckaert, G., and Lavigne, R. (2009) Survey of *Pseudomonas aeruginosa* and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. *Environmental Microbiology* **11**: 1303-1313.

Lammens, E., Ceyssens, P.J., Voet, M., Hertveldt, K., Lavigne, R., and Volckaert, G. (2009) Representational Difference Analysis (RDA) of bacteriophage genomes. *Journal of Microbiological Methods* **77**: 207-213.

Ceyssens, P.J., Mesyanzhinov, V., Sykilinda, N., Briers, Y., Roucourt, B., Lavigne, R., Robben, J., Domashin, A, Miroshnikov, K., Volckaert, G., and Hertveldt, K. (2008) The genome and structural proteome of YuA, a new *P. aeruginosa* phage resembling M6. *Journal of Bacteriology* **190**: 1429-1435.

Ceyssens, P.J., Hertveldt, K., Ackermann, H.W., Noben, J.P., Demeke, M., Volckaert, G., and Lavigne, R. (2008) The intron-containing genome of the lytic *Pseudomonas* phage LUZ24 resembles the temperate phage PaP3. *Virology* **377**: 233-238.

Ceyssens, P.J., Lavigne, R., Mattheus, W., Chibeu, A., Hertveldt, K., Mast, J., Robben, J., and Volckaert, G. (2006) Genomic analysis of *Pseudomonas aeruginosa* phages LKD16 and LKA1: establishment of the phiKMV subgroup within the T7 supergroup. *Journal of Bacteriology* **19**: 6924-6931.

Lavigne, R., Noben, J.P., Hertveldt, K., Ceyssens, P.J., Briers, Y., Dumont, D., Roucourt, B., Krylov, V.N., Mesyanzhinov, V.V., Robben, J., and Volckaert, G. (2006) The structural proteome of *Pseudomonas aeruginosa* bacteriophage phiKMV. *Microbiology* **152**: 529-534.

Chapter 1

Introduction & Background

1.1 Introduction on Bacteriophages

1.1.1 Short history of phage research

During the last decades, it has become widely accepted that bacterial viruses or bacteriophages are extremely abundant and exert enormous influences on the biosphere. Phages kill between 4–50% of the bacteria produced everyday, are a driver of global geochemical cycles and a reservoir of the greatest genetic diversity on Earth (Suttle, 2005). In this light, it is surprising that they were not recognized for almost 40 years after the beginning of serious bacteriological work in European and American laboratories around 1880. Two independent observations made by Frederick W. Twort (1915) and Felix d'Herelle (1917) who described glassy transformations of micrococci colonies and antagonistic bacterial microbes, respectively, are seen as the onset of modern phage research (Summers, 1999). It was d'Herelle who came up with the name *bacteriophages*, derived from 'bacteria' and the Greek 'phagein', which means 'to eat'. He also introduced the term *plaque* to describe the circular area of clearing caused by infection of a single phage on double-layered agar plates. Soon after their discovery, d'Herelle observed that phage titers increased in stool samples from dysentery patients. This finding encouraged scientists almost immediately to explore the therapeutic potential of these bacterial viruses. The first report of the therapeutic use of phage was a note by Bruynoghe and Maisin (1921) from Louvain, who noticed reduction in swelling and pain after injection of staphylococcal phage in the local region of cutaneous boils.

While many early phage therapy trials were reported to be successful and many of the major pharmaceutical companies sold phage preparations, there were also failures. Several factors, including the success of broad-spectrum antibiotics, an inadequate understanding of phage biology and imperfections in the diagnostic bacteriology techniques available at that time, contributed to the

failure of early phage therapy studies and to a decline of interest in phage therapy in the West. While phage therapy was continued to be offered in the Eliava Institute (Tbilisi, Georgia) and later also by others, like the Hirszfeld Institute of Immunology and Experimental Therapy in Poland (Deresinki, 2009), phage research in the West continued at a more fundamental level. The study of phages played a central role in some of the most significant discoveries in biological sciences, from the identification of DNA as the genetic material, over the deciphering of the genetic code and the phenomenon of restriction and modification, to the development of molecular recombinant technology. Phage-derived proteins are currently being used as diagnostic agents (Smith et al., 2001), therapeutic agents (Loeffler et al., 2001; Schuch et al., 2002) and for drug discovery (Liu et al., 2004).

1.1.2 Phage morphology forms the basis for classification

Bacteriophages can be defined as obligatory intracellular bacterial parasites which lack an independent metabolism. A century of phage research has revealed these viruses as extremely diversified and ubiquitously present in the biosphere, preying on eubacteria and archaea in a wide range of biological niches. Accordingly, the genome sizes of phages vary enormously, from a few thousand base pairs up to 498 kilobases in phage G, the largest phage sequenced to date (Guttman, 2005). Although the size of this genome approximates that of an average bacterium, phage G lacks genes for essential cellular machinery like ribosomes, emphasizing the purely parasitic nature of these organisms.

A bacteriophage particle or virion consists of a single or double stranded (ss of ds) DNA or RNA molecule, encapsulated inside a protein or lipoprotein coat. Historically, the International Committee on the Taxonomy of Viruses (ICTV) used virion morphology and nucleic acid composition as a basis for the classification of phages into 13 families. Over 95% of all phages described in the literature belong to the order of the *Caudovirales* or tailed dsDNA phages. The three main families comprising the *Caudovirales* are distinguished by their very distinct tail morphologies: 60% of the characterized phages are *Siphoviridae*, with long, flexible tails; 25% are *Myoviridae*, with double-layered, contractile tails; and 15% are *Podoviridae* with short, stubby tails (Figure 1.1). Polyhedral, filamentous, and pleomorphic phages represent only 3–4% of the studied phages and belong to 10 families, some of which are very small (Ackermann, 2007). Since all phages encountered in this work are *Caudovirales* (Chapter 4), this literature review will be limited to this order.

Tailed virions are typically composed of a 1:1 ratio of proteins and dsDNA. In all dsDNA phages, the chromosome is highly condensed within the virion head, and represents 20-50% of its mass (Earnshaw and Harrison, 1977). The mature capsid lattice must be sufficiently robust to maintain this chromosome

in its condensed form: unlike many animal viruses, dsDNA phages deliver their genetic material through an injection mechanism rather than an uncoating process. All *Caudovirales* have heads with icosahedral (20 sides/12 vertices) symmetry or elongated derivatives thereof with known triangulation numbers of T=4, 7, 13, 16 and 52, assembled from many copies of one or two major head proteins. The size of the phage heads is correlated to the genome size being packaged, and varies in diameter between 45 and 100 nm. The corners of the icosaheder are generally made up of pentamers of the capsid proteins, while the rest of each side is made up of hexamers of the same (or a similar) protein (Gutmann, 2005).

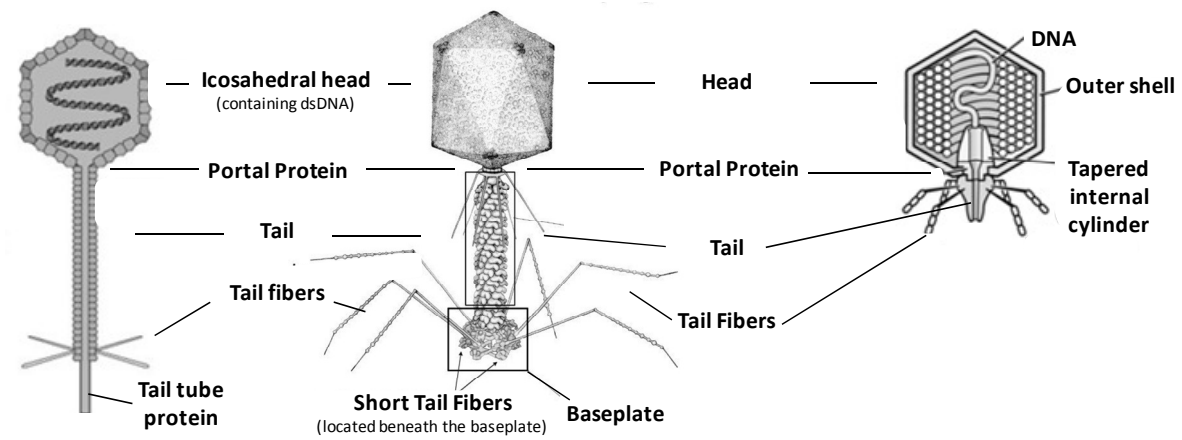


Figure 1.1. Basic morphology of the three families of the *Caudovirales*. From left to right, schematic drawings of a *Siphovirus* (long, non-contractile tail), *Myovirus* (long, contractile tail) and *Podovirus* (short tail) are shown. See text for more details (Figures adapted from Leiman, 2004 and Serwer et al., 2008).

The protein connecting the head and tail structures ('portal protein') is a key viral structure involved in several steps of the infection cycle. This connector is a homo-oligomeric structure found in a unique vertex of the phage head and controls the entry of the DNA and assembly of the tail to the immature head during morphogenesis of new particles. During infection, it changes its conformation and allows the DNA to exit the virion towards the bacterial cell (Valpuesta et al., 2000).

The tail attaches the phage to the host and serves as conduit for genome injection. Tail shafts have six-fold or three-fold symmetry, and are composed of helical (or stacks of disks of) subunits varying from 3 to 825 nm in length. They usually have base plates, spikes, or terminal fibers at the distal end. As stated before, long phage tails may be contractile, as found in the *Myoviridae* (e.g., phage T4), or noncontractile, as found in the *Siphoviridae* (e.g., phage λ). All long-tailed phages possess a large gene (usually, > 2 kbp) encoding a tape measure protein, which is responsible for precisely determining tail

length. Contractile tails are often very complex structures; for example, the tail of T4 is composed of 22 different proteins (Mesyanzhinov, 2004). The tube of noncontractile tails is composed primarily of multiple copies of one protein, known as the major tail protein. On infection, the overall shape of the noncontractile tail is maintained. In contrast, contractile tails include a central tail tube protein (TTP) that is surrounded by a tail sheath protein. On infection, the tail sheath contracts and the TTP penetrates the outer cell membrane and cell wall (Leiman et al., 2004). Short-tailed *Podoviridae* may have some key infection proteins enclosed inside the head that can form a sort of extensible tail upon contact with the host, as shown most clearly for coliphage T7 (Molineux, 2001).

The ICTV classification method as it is described above is currently being reevaluated, since it ignores the vast amount of available genome sequence data which occasionally causes contradictory aspects. For example, despite a complete lack of DNA sequence relatedness, the current taxonomic system lists both the *Salmonella* P22 phage and the coliphage T7 as belonging to the *Podoviridae* family based on the shared presence of short tails. However, it has been known for more than 35 years that P22 is genetically so related to the long-tailed phage lambda that recombination between two genomes forms functional hybrids (Botstein and Herskowitz, 1974). Cases like these encouraged scientists to look for alternative (molecular) classification methods, but this has been proven to be difficult because no universal gene, analogous to the 16S rRNA gene used for bacteria, exists throughout all phage families (Paul et al., 2002). Comparative approaches based on terminase sequences (Serwer et al., 2004), structural genes (Chibani-Chenouffi et al., 2004) and whole proteomes (Rohwer and Edwards, 2002) were proposed as novel taxonomic tools. In 2008, Lavigne and colleagues delineated new phage subfamilies and genera using a method which integrated BLAST-based tools with careful review of available literature. This work laid the basis for systematic and genome-based classification of sequenced phages, which is currently undergoing ratification by the ICTV.

1.1.3 The origin of bacteriophages

The mere question whether bacteriophages (and viruses in general) are living entities is the subject of a lively discussion among evolutionary biologists. The debate was fueled by the discovery of the Mimivirus, a giant amoebae virus which contains a 1.2 megabase chromosome encoding an unprecedented number of genes involved in transcription and translation (La Scola et al., 2003; Raoult et al., 2004). Inspired by the overlapping particle and genome sizes of viruses and cellular organisms

(Koonin, 2005) and the observation that the viral capsid protein is the only component that is unique to the viral world (Krupovic and Bamford, 2008), Raoult and Forterre (2008) divided all organisms into Ribosome-encoding and Capsid-encoding organisms. These authors suggest that all viruses descend from an ancient form of a capsid-containing virus which predates, or was a contemporary of, the Last universal Common Ancestor (Figure 1.2A). This view is supported by parallels drawn between the topology of capsid proteins of the bacterial tectivirus PRD1, the human adenovirus and an algal phycodnavirus, PBCV-1 (Figure 1.2B). Similar relationships were found between tailed dsDNA bacteriophages and herpesviruses (Baker et al., 2005), as well as dsRNA bacteriophages and reoviruses (Huiskonen et al., 2006). These observations gave rise to the viral lineage hypothesis, which predicts a common origin for viruses that share the same capsid architecture but infect hosts from different domains of life (Krupovic and Bamford, 2008).

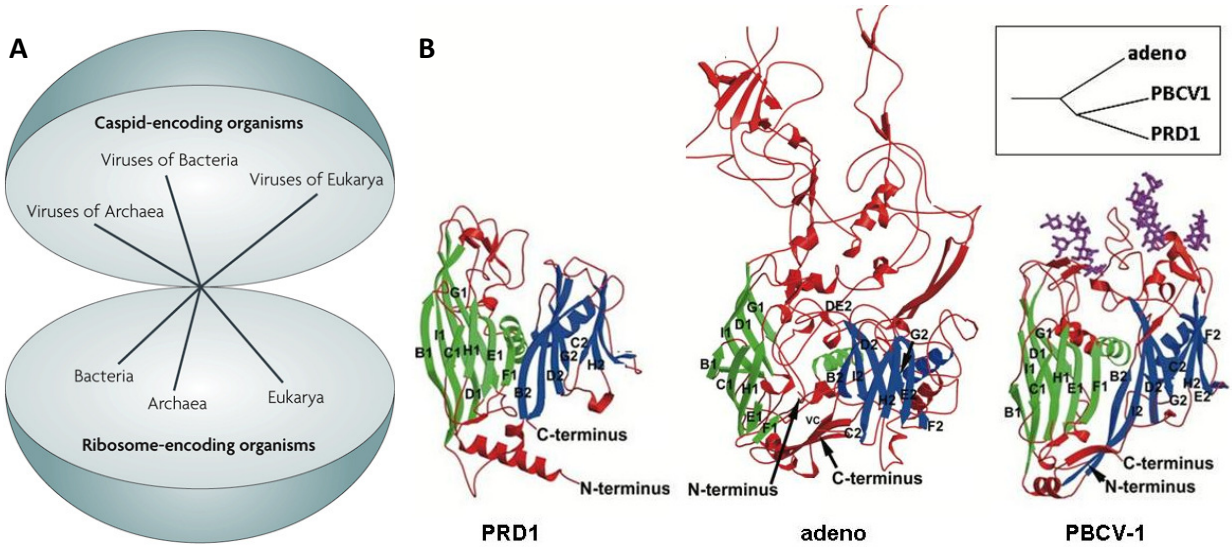


Figure 1.2. A. Representation of viruses with their capsids, and the three domains of life that have evolved from last universal common ancestor, as described by Raoult and Forterre (2008). **B.** Comparison of capsid proteins structures from viruses that can infect the three domains of life: the major capsid protein of bacteriophage PRD1, the human adenovirus and the *Paramecium bursaria* Chlorella virus 1 (PBCV1). These three capsid proteins contain the typical double-jelly-roll fold that is absent from cellular proteins. The eight β-strands and the flanking α-helix are displayed for the first (green) and second (blue) jelly rolls. Figure 1.2B was adapted from Benson et al., 2004.

However, the notion that viruses are living organisms that represent ancient lineages of the Tree of Life was recently strongly challenged. Moreira and Lopez-Garcia (2009) excluded viruses from the Tree of Life based on their lack of self-maintenance and their inability to evolve by themselves, as they are

'evolved by cells'. Based on the observation that no gene is shared by all viruses, they oppose the single, common origin of cellular life (Ranea, 2006) to a polyphyletic origin of viruses. These authors point at the cellular origin of viral metabolic and translational genes, and explain common capsid protein motifs by horizontal gene transfer and convergent evolution: the constraints to tertiary structures which are imposed by the simple geometric structures of viral capsids might have converged them into similar folds, as seen with other proteins under strong selection, like active site of enzymes (Gherardini et al., 2007). Based on a comparison of viral capsids with icosahedral shells of bacterial protein-based organelles such as carboxysomes, Moreira and Lopez-Garcia suggest that this type of molecular architecture could also be prone to convergence.

This view was again refuted by many evolutionary biologists, who emphasize the importance of viruses in the pre-cellular state of evolution (Koonin et al., 2009) and the limits of horizontal gene transfer (Ludmir and Enquist, 2009). Pointing at the unexplained origin of the majority of viral genes and their lack of resemblance to any cellular genes, Claverie and Ogata (2009) warn that today's dominant picture of evolutionary origin of life might be fundamentally incomplete and that any 'dogmatic' view on the role of viruses should be avoided.

1.2 The impact of bacteriophages on the biosphere

Whether or not phages should be included in the Tree of life or predate cellular evolution, their current abundance and their prominent role in biological evolution and global biochemical cycles are undeniable. It is now widely recognized that they are vehicles of horizontal gene transfer between cells, contribute to the acceleration of gene evolutionary rates and exert fundamental selection pressures on their host populations, thereby generating and regulating microbial diversity. Although pioneer work on the composition and influences of viral community structures was mainly focused on marine ecosystems, the impact of viruses on soil ecosystems is also more and more under investigation.

1.2.1 Variation in bacteriophage abundance

Although there was persuasive evidence in the late 1970s that viruses are abundant in the world's oceans (Torrella and Morita, 1979), it was not until a decade later that quantitative estimates using transmission electron microscopy (Bergh et al., 1989) and epifluorescence techniques (Hennes et al., 1995) revealed that each milliliter of seawater contains millions of these particles. In marine sediments, abundances are even higher, with 10^8-10^9 viruses cm^{-3} in nearshore surface sediments (Paul et al.,

1993). By extrapolation, the total population of phages is now estimated to be 10^8 species and 10^{31} particles in the biosphere, making them the most abundant organisms on this planet (Rohwer, 2003). Most viruses in seawater seem to be infective, as derived from measurements of DNA damage and photoreactivation (Wilhelm et al., 1998), and viruses infecting a specific host can be extremely abundant. For example, viruses infecting a single strain of the cyanobacterium *Synechococcus* (Suttle and Chan, 1994) or the photosynthetic flagellate *Micromonas pusilla* (Cottrell and Suttle, 1995) can occur in excess of 10^5 particles ml^{-1}. In absence of viral production, virions decay rapidly due to adsorption to particles, solar radiation and flagellate grazing, with an average half-life of 48 h in marine environments (Wommack and Colwell, 2000). However, data from some viral decay studies imply that virions display a rapid initial decay, followed by a period of slower decay since remaining particles seem more refractory to harmful environmental conditions (Heldal and Bratbark, 1991).

Subsequent application of these counting methods to soil samples has revealed that viruses are also extraordinarily abundant in a diverse range of soil types and locations. In a survey of Delaware soils which included wetlands and agricultural soils, viral abundance ranged from 8 to 39 $\times 10^8$ particles per gram dry weight of soils, while in cryptic Antarctic soils viral abundance was 23-64 x 10^7 (Williamson et al., 2003). Derived from these data, it was observed that despite the vast differences in composition and geographic location of these soil samples, there was only a 16-fold range in viral abundance. By comparison, across marine environments ranging from the deep sea to freshwater marshes, viral abundance changes over 2000-fold (Suttle, 2007). Interestingly, the ratio of virus particles to bacterial cells (VBR) does seem to vary enormously and is inversely correlated to the soil organic content, as illustrated by the VBRs of 10 and 8200 in Delaware and Antarctic soils, respectively (Williamson et al., 2003). This was explained by the relatively slow viral production and loss within soils as compared to highly variable host populations within these niches (Srinivasiah et al., 2008), a prediction supported by the high abundance of lysogenic bacterial hosts within soil environments (Williamson et al., 2007). Similar differences are also observed in VBRs from lakes (less than 5) and from deep waters of the Atlantic Ocean (more than 100). However, far-stretching conclusions cannot be drawn, since viral production occurs in microbial hot spots, illustrated by the order-of-magnitude variations in viral abundance and VBRs on spatial scales of centimeters that occur in aquatic environments (Seymour et al., 2006).

1.2.2 'Killing the Winner' shapes community structures

It was inferred from metagenomic data that at a specific location, the composition of the viral community follows a steeply declining rank-abundance curve (Figure 1.3), in which the most abundant viral genotypes comprise up to 5% of the total community and the vast majority of viral genotypes are extremely rare (< 0.01% of the community) (Breitbart et al., 2004). In this model, active virus-host pairs are behaving in a 'Kill-the-Winner' fashion, where the most abundant host population is reduced by its viral predators (Thingstad and Lignell, 1997). As a result, an open niche is created which enables a new host to become abundant. Supporting this model, culturing and genome size distribution studies provide evidence that specific viruses become abundant, decay to undetectable levels and then return on seasonal cycles (Wommack et al., 1999; Marston and Sallee, 2003). Moreover, study of phage-host systems isolated from seawater (Hennes et al., 1995) and sludge bioreactors (Kunin et al., 2008) showed that phages propagate as a function of host density. When the concentration of viruses was reduced to lower contact rates, taxa that were usually rare increased in abundance, whereas the taxa that were most abundant declined (Bouvier and del Giorgo, 2007). This indicates that the taxa that were initially numerically dominant were competitively inferior to the rare taxa, which were held in check because they were highly susceptible to viral infection.

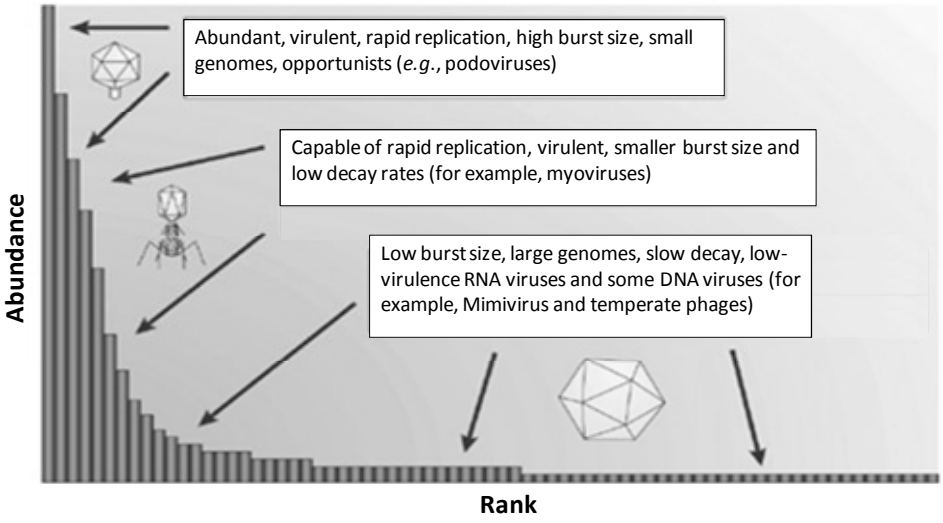


Figure 1.3. A rank-abundance curve showing the distribution of marine viruses. The population structure of these viruses is uneven, with many of the viruses at a given time being progeny from a limited number of lytic events. The most abundant viruses are virulent and are short-lived; the rarer phages decay slowly and can form stable relationships with their hosts (adapted from Suttle, 2007).

These interactions between hosts and parasites (also called 'antagonistic coevolution') have recently been recognized as a pervasive force in evolutionary ecology (Lindell et al., 2007). For example, Pal and colleagues (2007) investigated underlying selection pressures which drive bacteria with greatly elevated mutation rates (mutators). They showed that 25% of the *Pseudomonas fluorescens* populations which coevolved with phages had evolved to a 10- to 100-fold increase in mutation rates owing to mutations in mismatch-repair genes, giving them a significant higher probability to overcome phage infection. Interestingly, considerable variation in the evolution of resistance and counter-resistance exists, and the environment can strongly influence the level of parasite-mediated selection (Nuismer, 2006). In heterogeneous environments, where diversity is otherwise high, phages reduce diversity because phages cause a reduction in bacterial density, thus weakening negative frequency dependent selection (Buckling and Rainey, 2002). By contrast, in homogeneous environments where diversity is otherwise low, phages increase diversity as a result of fitness trade-offs between bacterial resistance and competitive ability. This allows coexistence of phage-sensitive and phage-resistant genotypes (Buckling et al., 2006).

An interesting light on the 'who is ahead?' question was shed by a recent survey of bacteria and their associated phages sampled from two 25-cm-by-25-cm grids on a grazed flood plain. This study showed that phages were on average 9% more infective to local than to foreign bacteria (Vos et al., 2009). This finding was particularly striking because the site is grazed and regularly flooded, which presumably would increase soil homogenization over such small spatial scales. The authors conclude that this local phage adaptation puts phages ahead of the bacteria in the coevolutionary arms race; first of all, selection acting on phages to overcome bacterial resistance is likely to be stronger than phage-imposed selection for bacterial resistance, simply because phages have to infect bacteria to replicate (Morgan et al., 2005). Second, the genetic diversity, and hence evolutionary potential, of the sampled phages (the whole community) is likely to be greater than that of the bacteria, which largely belong to a single taxonomic group (Vos et al., 2009).

1.2.3 Phage impact on host fitness

The host-specific nature of viral infection makes viruses powerful agents for controlling the composition of prokaryotic communities. However, many phages also have the capability to significantly augment their host's metabolism, immunity and evolution. A classic example involving virulent phages is the identification of photosynthesis genes in the genome of phages infecting *Synechococus* and *Prochlorococcus* species, which together account for about 25% of global

photosynthesis (Mann et al., 2005). These genes include *psbA* and *psbD*, which encode the most labile and hence rate-limiting protein in the photosystem II core reaction-centre (Lindell et al., 2005). These genes are expressed during the infection cycle, countering the virally induced decline in the host's transcription and translation (Sullivan et al., 2006). It is thought that phages prolong the photosynthesis during their infection cycle to generate the energy necessary for viral production, and to maintain the proton motive force necessary for host lysis (Rohwer and Thurber, 2009).

Many bacteriophages are also capable to form a stable symbiosis with their host bacterium, otherwise known as lysogeny. In this stage, the phage genomes are maintained in their hosts as prophages through integration into the bacterial chromosome or as an autonomous plasmid (Ackermann and Dubow, 1987). These residing prophages are very common: they can constitute as much as 10–20% of a bacterium's genome and are major contributors to differences between individuals within species (Casjens, 2005). Prophages can supply a key selective advantage to their bacterial host, ranging from protection against further phage infection to increasing the virulence of a pathogenic host (reviewed by Boyd and Brüssow, 2002). One of the best-known examples of this 'lysogenic conversion' occurs in *Vibrio cholerae*, a common near-shore bacterium that is normally harmless but becomes one of humanity's greatest scourges by incorporating phage cholera toxin (CTX) genes (Waldor and Mekalanos, 1996). Many other examples of phenotype-modifying genes are known, including genes that are involved in carbon metabolism, cellular invasion and pathogenesis (Table 1.1).

Table 1.1. Examples of phage-encoded proteins which modify the bacterial phenotype.

Protein	Function	Host genus	Phage	Reference
PsbA	Photosynthesis	*Prochlorococcus*	P-SSP7	Lindell et al., 2005
TalC	Carbon metabolism	*Prochlorococcus*	P-SSM2	Sullivan et al., 2006
SopE	Type III effector	*Salmonella enteric*	SopEϕ	Mirold et al., 1999
Tox	Pathogenesis	*Corynebacterium diphtheria*	β-phage	Freeman, 1951
CTX	Pathogenesis	*Vibrio cholerae*	CTXϕ	Waldor and Mekalanos, 1996

The total net effect of phage genomes (viromes) on the total microbial community metabolism cannot be readily determined. However, Dinsdale and colleagues (2008) compared 15 million microbial and viral metagenomic sequences from nine biomes to the SEED platform, which arranges metabolic pathways into a hierarchical structure by clustering genes required for a specific task (Overbeek et al., 2005). They observed that microbiomes and viromes carry almost equal frequencies of metabolic

genes involved in carbohydrate and protein metabolism. More unexpected was the finding that genes involved in vitamin and cofactor synthesis, stress-response genes such as those encoding chaperones, and genes associated with bacterial motility and chemotaxis were more common in viromes than in their corresponding microbiomes. This magnitude of microbial metabolic capabilities encoded by the viromes suggests that they serve as a repository for storing and sharing genes among their microbial hosts and influence global evolutionary and metabolic processes (Dinsdale et al., 2008).

1.2.4 Impact of phages and Gene Transfer Agents on bacterial genetic exchange

Unlike their cellular hosts, most phages carry a specific amount of DNA to correctly pack their capsids. If there is a selective advantage of extra DNA incorporated in the capsid (resulting in increased phage progeny), these genes are fixed in the phage genome; otherwise they will be lost. As shown by metagenomic analysis, there is little restriction to the types of genes carried by the viral community (Dinsdale et al., 2008). Because there are an estimated 10^{31} phages on the planet, the potential reservoir of genes that can be transferred by phages is enormous (Hendrix et al., 2002). The rate of this transduction (gene transfer by phage) was determined by measuring the horizontal spread of a gene encoding kanamycin resistance among marine bacterial isolates and concentrated natural bacterial communities. It was found that transduction frequencies ranged from 1.33×10^{-7} to 5.13×10^{-9} transductants/PFU, which can be extrapolated to 10^{24} genes being moved around by phages each year (Jiang and Paul, 1998).

However, the actual amount of transduction may be much greater than previously thought because of the action of generalized transducing agents (GTAs). These agents were first described in the bacterium *Rhodobacter capsulatus* (Marrs, 1974) but have now been found in various bacteria and archaeal organisms (Lang and Beatty, 2007). GTA particles are similar in morphology to phage, but they are smaller (with a head diameter of 30–50 nm) and contain a smaller amount (4-14 kb) of DNA (Figure 1.4A). What makes GTAs unique is that they are not able to lyse bacteria or to form plaques, and seem to function only in genomic transfer between cells. GTAs assemble into a virus-like particle which incorporates random fragments of host DNA and transfers these to other bacterial cells (Lang and Beatty, 2007). The genes encoding the GTA particle proteins have a distinct organization in the bacterial genome and are normally arranged in 15 ORFs (Figure 1.4B).

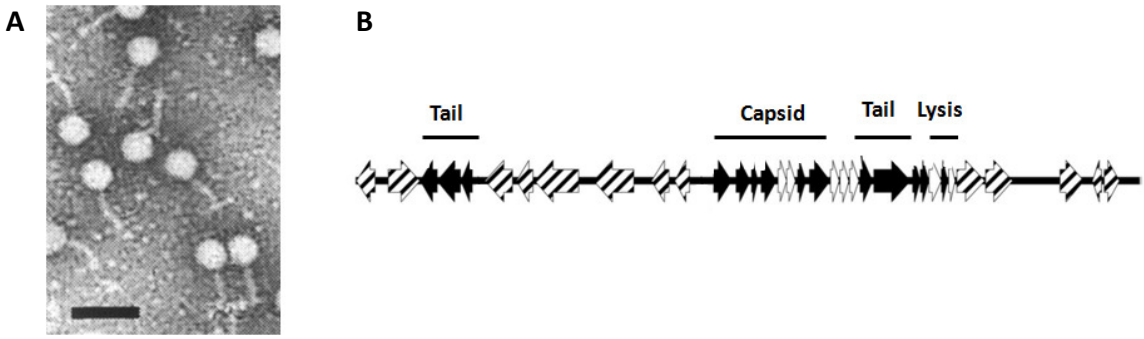


Figure 1.4. A. Transmission electron micrographs of VSH-1, a gene transfer agent of *Brachyspira hyodysenteriae* containing random 7.5 kb bacterial genomic fragments. The scale bar represents 100 nm (Matson et al., 2005). **B.** Genomic region encoding VSH-1 genes (black arrows), *B. hyodysenteriae* genes (striped arrows), and ORFs greater than 240 bp with no GenBank homologs (white arrows), oriented according to their direction of transcription. The figure was adapted from Stanton et al., 2009.

The regulation of production of GTAs has been linked to quorum sensing (Schaefer et al., 2002) and signaling proteins CckA and CtrA (Lang and Beatty, 2002), suggesting that gene transduction is induced under conditions where genetic recombination might provide a benefit for survival. A recent study has shown that genes encoding GTAs are found in most marine systems (Biers et al., 2008). The same study also showed that these GTAs are produced by marine bacteria and move genes between species of α-proteobacteria, suggesting that GTA-related gene swapping may contribute to the niche partitioning of closely related species.

1.3 The diversity of bacteriophages

Not until this 'Third Age of Phage' (Mann, 2005), researchers have begun to gain an idea of the biodiversity of phages in the environment, mainly through the metagenomic analysis of uncultured marine virus communities. Before (and still continuing in) this era, many research groups established large phage collections using a wide variety of prokaryotes as phage host (*e.g.*, Carey-Smith et al., 2006; Jamalludeen et al., 2007), among which representatives are sequenced and studied further. Although these phage collections provide insights into the diversity and evolution of specific phage groups (*e.g.*, Gill et al., 2003), culture-independent methods are required for the study of the diversity of natural phage populations.

1.3.1 The limits of current sampling methods

The typical procedure for isolating and propagating phages out of liquid or solid environmental samples is filtration, followed by a liquid enrichment culture or direct spotting on a bacterial lawn (Carlson, 2005). Two years ago, it was estimated that a total amount of 5,568 phages have already been visualized using electron microscopy, 96% of which belonged to the *Caudovirales* (Ackermann, 2007). However, it has become clear that the standard procedure appears to be incapable of isolating most bacteriophage species. For example, it was estimated that only a minor fraction (< 1%) of the bacteriophage species present in river samples could be detected by plaque assays (Chen et al., 2001). Ashelford and colleagues (2003) counted average bacteriophage concentrations of 10^8 per g in some soils, but were not able to extract any of these soil-borne phages into liquid medium. A method based on phage propagation in dilute agarose gels had some success in isolating poorly growing phages of *Pseudomonas chlororaphis* and *Bacillus subtilis* (Serwer et al., 2004). However, these and other methods suffer from the limitation that only 0.001–0.1% of the microorganisms in seawater, 0.25% in fresh water, 0.25% in sediments and 0.3% of soil microorganisms were found to be cultivable (Amann et al., 1995), a basic necessity for isolating and propagating phages.

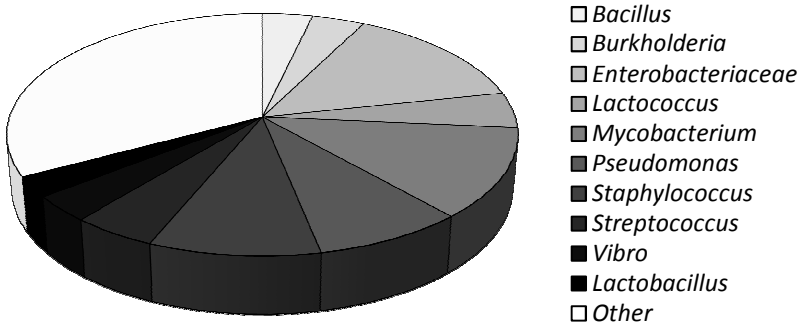


Figure 1.5. Division of 549 complete phage genomes deposited in GenBank on July 19th, 2009, arranged according to bacterial host organism. Bacteriophages of *Enterobacteriaceae*, *Pseudomonas*, *Staphylococcus* and *Mycobacteria* account together for >45% of all sequenced phages.

As a consequence, the characterized phages in current databases are clearly biased towards easily cultured phages which infect generally well known bacteria (Figure 1.5). Indicative of this problem is that only 8 (1.5%) sequenced phages have genomes as long as 200 kb, while statistical analysis reveals a significant under-sampling of long-genome bacteriophages (Claverie et al., 2006). The strong possibility exists that long-genome bacteriophages (> 200 kb genome) are more frequent and are major

contributors to microbial ecology, but are under-sampled because of the use of classical bacteriophage propagation procedures and possibly also classical processing of environmental samples for microscopy. For example, giant bacteriophage G was discovered by accident ~40 years ago through electron microscopy of a preparation of another bacteriophage (Donelli, 1968). This well-known plate count anomaly (Staley and Konopka, 1985) has forced microbial ecologists to rely on cultivation-independent approaches to study phage diversity in natural environments.

1.3.2 Diversity based on single-locus studies

Various research groups have employed some of the best conserved phage genes (polymerases, structural genes) as molecular markers for population studies. Many of these studies focused on the coliphage T4 portal-protein-encoding homologue, gene 20 (g20), as a diversity marker. The g20 homologue is ubiquitous among T4-like myoviruses with hosts ranging from proteobacteria to cyanobacteria (Mann et al., 2005). The evolution of g20 is likely to be constrained because its protein product initiates capsid assembly, a process involving geometric precision (Hsiao and Black, 1978). Early studies using non-degenerate PCR primers and DNA fingerprinting techniques provided evidence for an enormous genetic variation across gradients in space and time which was not represented in cultures (*e.g.*, Frederickson et al., 2003). Even more remarkable is that nearly identical g20 nucleotide sequences (> 99%) occurred in environments as far-reaching as the Southern Ocean, the Gulf of Mexico and a melt-water pond on an Arctic ice shelf (Short and Suttle, 2005). As more g20 sequences accumulated, it became clear that marine g20 sequences form nine phylogenetic clusters of which only three contained cultured representatives, and that g20 from freshwater environments formed a separate, tenth cluster (Figure 1.6).

Even when this analysis was recently expanded to all available g20 sequences, very few sequences fell into the clusters without cultured representatives, suggesting that the latter are rare, or sequencing artifacts (Wilhelm et al., 2006). In an attempt to identify more representatives of each cluster, Sullivan and colleagues (2008) sequenced g20 from 38 marine myophages isolated using a diversity of *Synechococcus* and *Procholorococcus* hosts. Interestingly, more than 90% of the sequences fell into clusters I, II and III, while the remaining sequences clustered in a monophyletic group indicated as PSSM9-cluster (Figure 1.6), strongly indicating that g20 sequences in other clusters represent myophages that infect other hosts. While some patterns of relationships between g20 sequence, host range and host of isolation emerged from this analysis, exceptions are so frequent that one must

conclude that phage portal proteins are not good predictors of a phage's host or the habitat in which a phage may thrive (Sullivan et al., 2008).

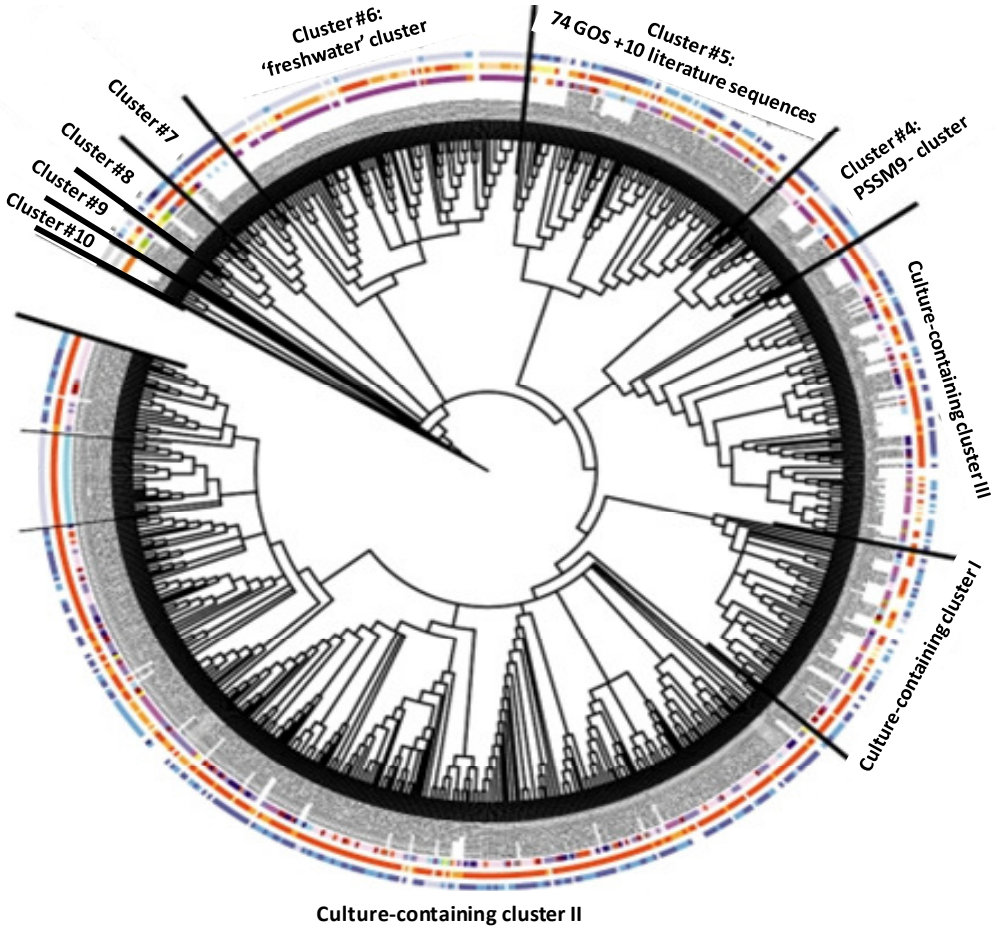


Figure 1.6. Evolutionary relationships determined using 554 base pairs of the portal protein gene (g20) from 769 available g20 sequences. These sequences include PCR amplicons from various environmental samples, non-PCR amplified sequences identified in the Global Ocean Survey (GOS) metagenomic database, as well as sequences from cultured phages. The tree shown is constructed using *Aeromonas* phage Aeh1 g20 sequence as an outgroup to root the tree (after Sullivan et al., 2008).

Many other single-locus studies have been performed, albeit to a lesser extent than the g20 analyses. For example, a study that targeted phage T7-like DNA polymerase genes (*pol*) found that these sequences were also ubiquitous in the environment (Breitbart et al., 2004). In addition, a recent study which targeted a larger fragment of this family A DNA *pol* gene showed that the diversity among marine podoviruses in a small geographic region is far greater than previously realized (Labonté et al., 2009). All these results indicate that closely related hosts and their phages are distributed widely and

that horizontal gene transfer occurs among phage communities from very different environments. Nevertheless, these single-locus studies still do not completely bypass the above-mentioned problems in assessing phage diversity, since environmental PCR primers are still based on sequences from cultured plates.

1.3.3 Diversity based on metagenomics

Metagenomics or whole-community shotgun sequencing can bypass this dilemma and provide quantitative as well as qualitative information of each genotype present in any microbial ecosystem. The online bioinformatic tool PHACCS makes it possible to use metagenomic data to mathematically model possible community structures. Using these models, estimates of viral community structures and of the richness, evenness and abundance of each viral genotype can be made (Angly et al., 2005). The majority of viral metagenomic studies to date have primarily focused on DNA isolated from material passing through 0.22 μm filters, the fraction that contains the bulk of virus-like particles, which is purified using 1.35-1.5 g ml^{-1} cesium chloride gradient centrifugation (Edward and Rohwer, 2005). Although the metagenomic approach holds drawbacks such as sampling bias towards "free-living" phages stable in CsCl, the technique gave valuable insights into the local and global distribution of phage genotypes.

Firstly, a study targeting RNA viruses failed to recover any RNA phage sequences, reinforcing the idea that phages with DNA genomes essentially dominate the marine environment (Culley et al., 2006). Secondly, numerous metagenomic studies point at local enrichment for specific phage types (Figure 1.7), such as ssDNA microphages in microbialites (Desnues et al., 2008), and prophages in potable water (Rosaio et al., 2009). In this respect, Angly et al. (2006) suggested that local viral community composition and nucleic acid type (i.e. dsDNA vs. ssDNA) is a function of the geographic location and regional selective pressure. Thirdly, extraordinary high phage diversity on both local and global scale was found. For example, at least 10^4 different viral genotypes were predicted in one kilogram marine sediment (Breitbart et al., 2004). In an analysis of marine viromes across four oceanic regions, no less than 129,000 genotypes were present in the viral metagenome of British Columbian coastal waters (Angly et al., 2006).

Interestingly, these studies also emphasize that regional diversity can be as high as global diversity due to viral migration. It has been shown that viruses taken from one ecosystem can replicate in a totally different environment, implying that phages can move between biomes (Sano et al., 2004). As an illustration, a phage like *Prochlorococcus* phage P-SSM4 was found at all sampling sites during the

Sorcerer II Global Ocean Sampling Expedition, suggesting a major role of these viruses in influencing the abundance, distribution and diversity of one of the most dominant components of picophytoplankton in oligotrophic oceans (Williamson et al., 2008).

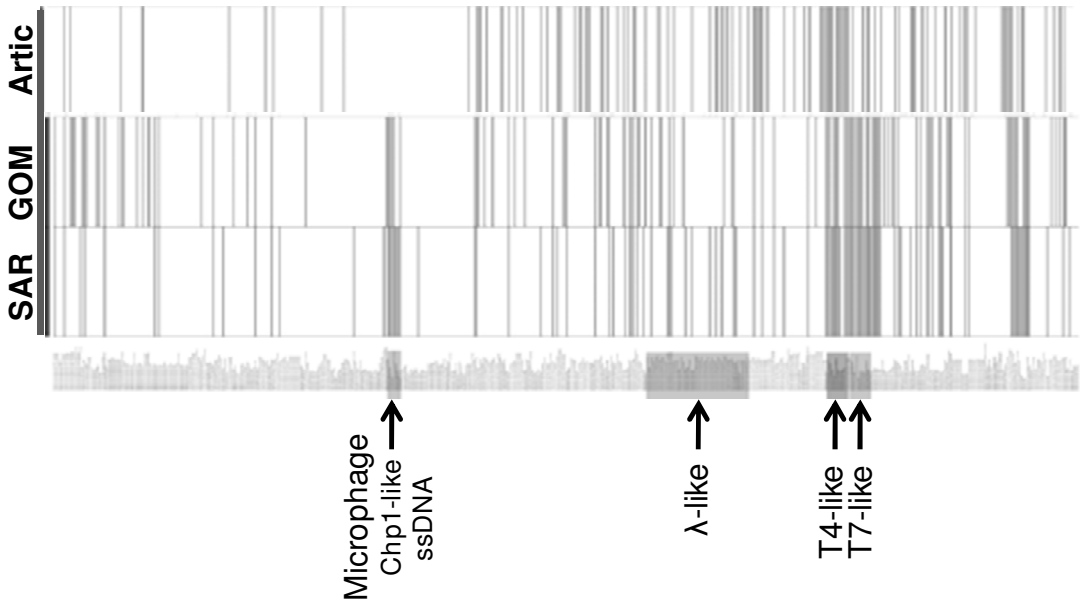


Figure 1.7. Types of phages in three metagenomes sampled at the Sargasso Sea (SAR), Gulf of Mexico (GOM) and the Artic Ocean. The metagenomic sequences were compared to 512 phages on the Phage Proteomic Tree (Rohwer and Edwards, 2002) using TBLASTx, with bars representing significant similarties (E<0.0001). Eighty-four phage species were specific to one marine region, whereas 45 were common for all three. Local enrichment for specific phage types can be derived from the figure (after Angly et al., 2006).

On average, 90% of metagenomic sequences do not display significant homology to anything in current databases. A partial explanation for this high percentage is almost certainly due to the short sequences (100-200 nt) that are generated by pyrosequencing, as the only metagenomic study that used traditional Sanger sequencing (650 bp fragments) found that only >60% of the sequences were unknown (Edwards and Rohwer, 2005). Nevertheless, these studies reveal massive amounts of uncharacterized gene diversity or database ORFans. Five years ago, it was calculated that 2 billion different phage-encoded ORFs remain to be discovered, and that less than 0.0002% of the global phage metagenome has been sampled (Rohwer, 2003). This reservoir of ORFans, still representing at least half of the total diversity, will be a rich area for future research (Section 1.4.5).

1.4 Bacteriophage genomics

1.4.1 Introduction

Phage research entered the genomic era with the completion of the 3,569 nt RNA genome of coliphage MS2 (Fiers et al., 1976) and the 5,386 nt DNA genome of phiX-174 (Sanger et al., 1977). The continuous introduction of new sequencing technologies in the last decade caused a dramatic increase in the number of completely sequenced phage genomes: from 40 in January 1997 to 552 in August 2009. Accordingly, a large number of peer-reviewed papers on phage genomics have been published lately, and this trend does not seem to decline (Figure 1.8). However, the estimates of thousands of sequenced and publicly available phage genomes within a few years (Hendrix, 2002) are probably exaggerated.

The accumulating data provide evidence that bacteriophage diversity is extraordinarily high. Phages isolated from different bacterial hosts typically have little or no identifiable nucleotide sequence similarity, and even phages capable of infecting a common host may exhibit great diversity (*e.g.*, Kwan et al., 2006). The origin of this diversity is obvious, as it was estimated that approx. 10^{25} phages initiate an infection every second, and in each of those infections phage encounter DNA (of bacterial or prophage origin) with which it can potentially recombine to generate new genomic arrangements (Pedulla et al., 2003). It is likely that this process has been ongoing for over 3 billion years, causing combinatorial chemistry on a grand scale (Hendrix et al., 1999). Since estimates of the number of existing bacterial species has ranged from ten million upwards, the number of phage species has been conservatively estimated at 100 million (Rohwer, 2003).

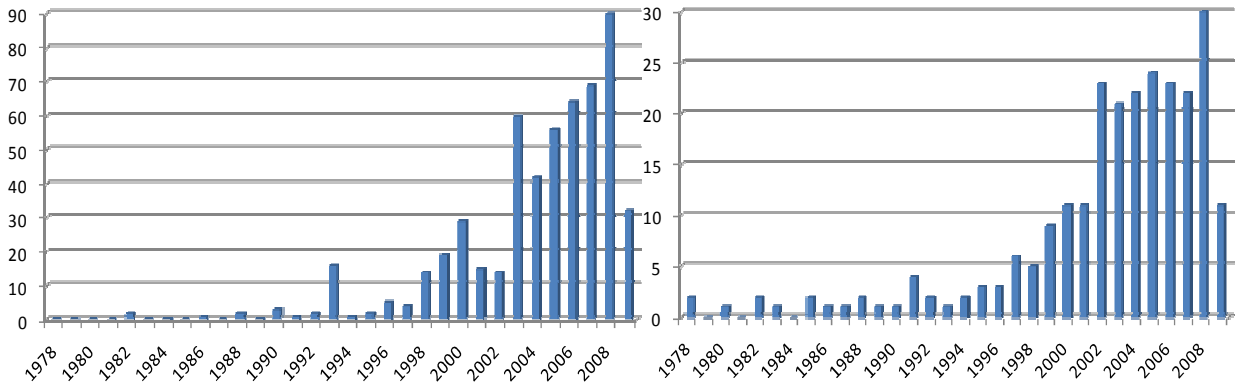


Figure 1.8. Overview of the amount of complete phage genomes yearly deposited in GenBank (left), compared to the number of published peer-reviewed papers describing phage genomes (right). The graphs describe the period from January 1978 until July 2009, and were constructed based on data derived from the NCBI databases on July 22th, 2009.

However, these enormous numbers do not make it impossible to understand phage diversity, since comparative genomics points at the existence of a limited number of phage 'types' or genera. For example, in *Escherichia coli* the well-studied tailed-phages λ, Mu, N4, Felix-O1, ε15, P1, P2, T1, T4, T5 and T7 each represent different phage groups in which each of the other 60 sequenced tailed *Enterobacteriaceae* phages can be grouped (Casjens, 2008). All individuals in each of these phage types have generally similar lifestyles, with a similar gene order and transcription pattern, and different members of a given type can infect a variety of different but closely related bacteria. Insights into genome architecture and the horizontal and vertical evolution which drove the establishment of the genetic diversity between and within these phage types were gathered by detailed examination of closely related phage genomes, and will be discussed in this chapter.

1.4.2 General phage genome build-up

The linear dsDNA genomes of the *Caudovirales* encode from 27 to over 600 genes that are typically tightly packed, highly clustered according to function and tend to be arranged in large operons. The minimal genome of tailed phages encodes DNA packaging, head, tail and tail fibers, DNA replication, transcription regulation and lysis genes. In addition, temperate phages like phage λ encode a cluster of genes responsible for integration into the host chromosome and maintenance of the lysogenic state. These genetic functions can be squeezed into a 20 kb genome, as demonstrated by the *Siphovirus* c2 and the *Podovirus* ϕ29 (Brüssow and Hendrix, 2002). As the genome size increases, the virion morphology and expression cascades become more complicated and the phage interferes more with cellular activities.

Bacteriophage gene expression is largely time-ordered and groups of genes are sequentially expressed. After internalization of the bacteriophage genome, the host-encoded RNA polymerase recognizes strong promoters located on the phage genome, leading to transcription of the phage *early genes*. Alternatively, internal virion proteins may be injected into the host cell together with the phage genome (Guttman, 2005). These early proteins are involved in the takeover of the host cell metabolism. Generally they provide protection from bacterial defense mechanisms or redirect the host-cell metabolism to establish optimal conditions for the synthesis of new virions. To establish these optimal conditions, various macromolecular processes like transcription, translation or replication, as well as single enzymatic functions may be redirected or inhibited (Miller et al., 2003). Early proteins are very phage-specific and generally poorly conserved, even between two closely related phages, and display typically no similarity to any known protein in the databases (Roucourt et al., 2009).

Once optimal metabolic conditions are established, replication of the bacteriophage genome starts. While most temperate phages depend on the host cell for their replication, virulent phages often encode their own replication machinery. Phages use a wide variety of different replication systems, extensively reviewed by Weigel & Seitz (2006). Before the replicated concatamerized phage genomes are packaged, the new phage heads or capsids have to be assembled by a highly regulated, multistage process encoded by the *late genes*. In this process, phages assemble a protein procapsid first, in which the DNA molecule is pumped by an ATP cleavage-powered DNA translocase. The two proteins that make up the motor of this DNA pump, the portal protein which forms the hole in the procapsid and the large terminase subunit which carries the ATPase site, are the only proteins that can be recognized by sequence similarity as being encoded by virtually all tailed phages (Casjens, 2008).

At the end of the infection cycle, phage terminases recognize DNA for packaging and have a nuclease activity that is responsible for creating the ends of the viral chromosome (Catalano et al., 1995). The chromosomes of the *Caudovirales* are known to have several different types of end structures, including (i) 5'- and (ii) 3'-protruding single-stranded cohesive ends with no terminal redundancy nor circular permutation (*e.g.*, 5' phages λ and P2, 3' HK97), (iii) direct terminal repeats (DTR) with no circular permutation (*e.g.*, phages T3 and T7), (iv) host DNA attached to both ends (*e.g.*, phage Mu), and (v) the headful packaging phages that have virion DNAs with both terminal redundancy and circular permutation (e.g., phages T4, SPP1, P1, and P22). Since these viral chromosome ends are generated by nucleolytic cleavage from replicating concatemeric or circular DNA molecules, the different types of ends reflect differing DNA replication strategies and especially differing terminase actions during DNA packaging. In addition, phage ϕ29 and its relatives are unique among the tailed dsDNA phages in that they do not cleave the DNA during packaging and have a protein covalently bound to each DNA end (Wang et al., 2004).

When constructing a neighbor-joining tree of large terminase subunits, Casjens and colleagues (2005) observed strong correlation between the structure of the virion DNA ends and clusters of related terminase genes. This might suggest that genomic end structures can be predicted based on terminase sequence; it is, however, important to realize that the current picture of terminase diversity is incomplete, and if a terminase sequence does not fall convincingly within a characterized group (currently including, for example, the long direct terminal repeat terminases of SPO1 and T5, headful terminases of P1, TP901-1, and Aaϕ23; *cos* phage VP16C, and short terminal repeat phage VpV260),

the predictive value of such a comparison declines abruptly. In general, directed experimental analysis is required to understand the nature of the virion chromosome ends (Casjens and Gilcrease, 2009).

1.4.3 Genetic mosaicism in temperate phages

In specific groups of temperate phages, genomes display a mosaic genetic structure, consisting of gene modules (individual genes or small groups of genes) which are exchanged using host- or phage-encoded recombination machinery. This phenomenon was first recognized in the "lambdoid" phages as a result of DNA-DNA heteroduplex experiments (Simon et al., 1971). From these data, Susskind and Botstein (1978) proposed a 'modular theory of phage evolution', in which recombination between different members of the family of phages took place at specialized 'linker' sequences at the mosaic boundaries. However, comparative analysis of sequenced phage genomes showed that this mosaicism most likely results from random, nonhomologous (illegitimate) recombination, resulting in a heterogeneous mix of recombinant types which are almost all defective. The few recombinants that survive are the ones in which the functions of essential genes are not disrupted and whose genomes still fit inside the virion capsid (Lawrence et al., 2002). In this way, genomes arise which share regions with obvious homology interspersed among regions that are unrelated.

One striking example of the phage capacity for mosaicism across initially considered distinct phage types was reported for *Pseudomonas* phage D3112. This phage resembles the *Myovirus* Mu in that it replicates its 40 kb-genome by transposition, and yet its tail genes are related to the classic *Siphovirus* lambda (Wang et al., 2004). Another example involves the lambdoid phages ST64T and P22, which have long patches of >90% sequence identity with interspersed unrelated sequences (Figure 1.9). For example, the endolysin gene (the *19* genes) of these two phages have no recognizable sequence similarity; the novel joint immediately to the left of the 3 genes, where very high sequence similarity ends abruptly at the gene boundary, is typical of the novel joints present when any two mosaically related tailed-phage genomes are compared (Figure 1.9).

Numerous other examples of this genetic mosaicism have been documented during the last decade, and this form of horizontal gene transfer (HGT) has been postulated repeatedly as being the major factor in bacteriophage evolution (*e.g.*, Hendrix et al., 2000; Hatfull et al., 2008). These observations formed the basis for a model for the genetic structure and dynamics of the global phage population,

which stated that all dsDNA phage genomes are mosaics with access, by horizontal exchange, to a large common gene pool (Hendrix et al., 1999).

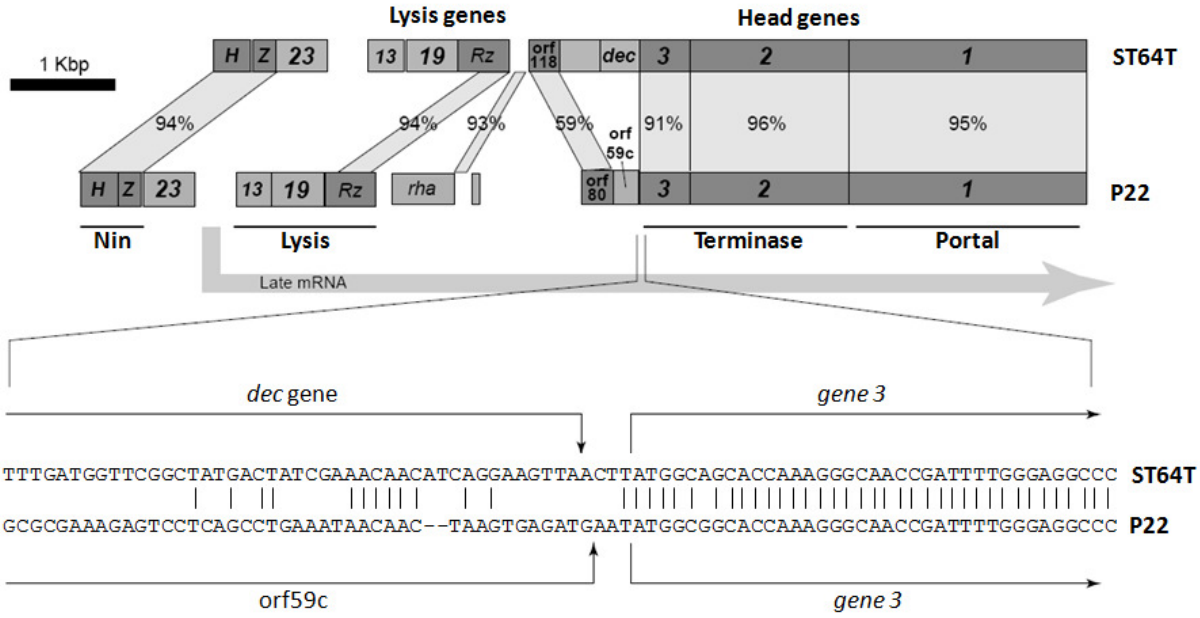


Figure 1.9. Classic example of phage genomic mosaicism, as exemplified by the late promoter regions of phages ST64T and P22. At the top of the figure, the two genomic regions are compared, with indication of the percentage nucleotide identity of each of the sections. Below, the nucleotide sequence of the novel sequence joint at the start of gene *3* is shown. See text for more details (after Casjens et al., 2005).

New genes in these temperate phage genomes are typically acquired by improper excision of prophages, recombination between ORFs generating chimera genes and acquisition of 'moron' sequences (Casjens, 2005). An example of non-homologous recombination is the N gene of phage HK97, which encodes a transcription factor. The encoded protein shows sequence similarity to the N gene homologue in P22, but immediately beyond the termination codon is a region with identity to the C-terminal part of phage 21 (Figure 1.10A). Morons (from 'more DNA') are genes which are inserted between two genes that are adjacent in related phages; these genes appear to be functionally unrelated to the operon function. They are typically preceded by a transcription promoter sequence and followed by a factor-independent (stem-loop) transcription terminator, an arrangement that allows the genes to be transcribed autonomously, even from a repressed prophage (Juhala et al., 2000). In fact, many virulence factors of pathogenic bacteria are actually encoded in the morons of integrated prophages (Figure 1.10B).

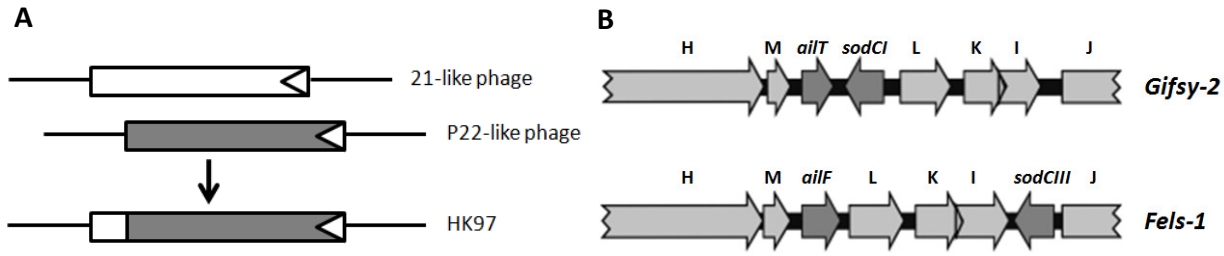


Figure 1.10. Examples of gene acquisition in temperate phages. A. Creation of the chimerical N gene of phage HK97. The N gene of phage HK97 has 17 codons matching the end of the phage 21 N gene, following the apparently complete P22-like N gene. **B. Morons.** The late operons of *Salmonella enterica* prophages Gifsy-2 and Fels-1 contain morons encoding bacterial virulence factors, which are not necessarily expressed together with the phage genes. *Sod* genes encode superoxide dismutases, *ail* genes encode bacterial outer membrane proteins that affect bacterial binding to mammalian cells (Figueroa-Bossi et al., 2001).

The nucleotide composition and codon usage of moron genes is typically substantially different from that of adjacent genes, arguing that the moron has recently entered the genome from an outside source (Lawrence, 1999). The original source of moron genes is unknown, but some (*e.g.*, *sodC, lom* and those encoding the putative DNA methylases) hae non-prophage-associated homologs in various extant bacterial genomes (Hendrix et al., 2000). In at least one case, homologous morons have been found in two very different phages (P22 and N4), each surrounded by different flanking sequences. This also argues for the view that morons are mobilizable elements capable of being inserted into different sequence contexts, and against the view that the presence of a moron in a single phage genome merely reflects deletion of the moron from phage lineages now lacking it (Hendrix, 2002).

1.4.4 Conserved core and hyperplastic regions of T4-like genomes

In contrast to phage groups which are strongly influenced by horizontal gene transfer, other groups are clearly shaped by vertical evolution. One of the most widespread and evolutionary successful of those phage groups are the T4-like phages, which is illustrated by the finding that five of the six proteins which were most enriched during the Global Ocean Sampling project of Craig Venter were T4-type phage proteins (Yooseph et al., 2007). Over 200 phages similar to T4 have been described (Ackermann and Krisch, 1997), and 19 completely sequenced phage genomes are currently (August 2009) listed as being T4-like in the NCBI-database. About 90% of the known T4-like phages grow on *Escherichia coli* or other enterobacteria, but the remaining 10% grow on phylogenetically more distant bacteria like

Aeromonas and *Vibrio* (*e.g.*, Kersters et al., 2003). The comparison between these genomes can provide insights into the evolutionary mechanisms that generate genomic diversity and the constraints that limit it.

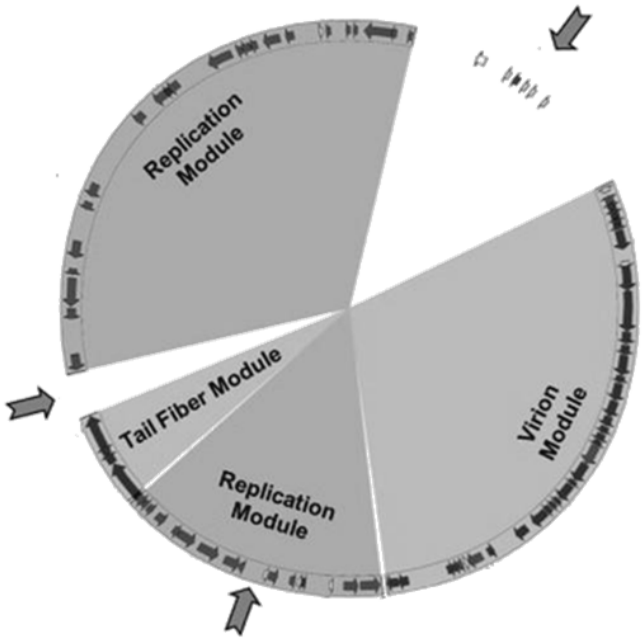


Figure 1.11. Schematic overview of a generic T4-type genome. The "core genes" and indicated by arrows, which are colored black, grey and white for structural, DNA manipulating, and unknown genes, respectively. The various hyperplastic regions are indicated with arrows (after Krisch and Comeau, 2008).

The phylogenetic analysis of more than a dozen T4-like genomes clearly pointed out that these genomes are bipartite. They consist of a conserved core that specifies the minimal essential genes for viral multiplication, interspersed with a highly variable set of facultative genes that allow the core genes to optimally fulfill their functions in a particular host cell and environment (Filée et al., 2006; Krisch and Comeau, 2008). Thus, this bipartite modular design of the T4-type genome can exploit a highly successful virus multiplication strategy and yet simultaneously provide the phage with the genetic flexibility required for adaptation to the environment. Depending on the precise subset of phages considered, the number of conserved genes is about 75, although very divergent cyanophages have less than 45 T4-like ORFs (Comeau et al., 2007). The most conserved segments of the T4-like genome appear to be the large module that encodes most of the structural components of the virion, and the genes encoding the DNA replication functions (Figure 1.11). Motif searches among T4-related phages identified putative promoter elements resembling early and late T4 promoters in all genomes,

indicative of a conserved transcriptional strategy (Nolan et al., 2006). Filée and colleagues (2006) analyzed the evolutionary history of the T4 core genes, and showed that they were derived from a common ancestor. The authors speculate that the numerous and complex protein-protein interactions during DNA replication (Karam and Konigsberg, 2000) and virion morphogenesis (Leiman et al., 2003) impose a barrier against lateral gene transfer. Comeau and colleagues (2007) expanded this idea, suggesting that the conserved core replication and virion super-modules were originally assembled from modular constituents, but as these various elements became better adapted to each other, their modular interchangeability was sacrificed.

The hypervariable regions in the T4-like genomes are mostly clustered in groups between these blocks of conserved genes (Figure 1.11). In contrast to the 'morons' and chimerical genes found in lambdoid phages (Section 1.4.3), novel T4-like ORFs do not differ significantly in codon bias from conserved genes and are subject to the same regulatory constraints as the rest of the phage genome, indicating that they are not recently acquired from host genes (Section 1.4.5; Nolan et al., 2006). In fact, these regions appear to be accreted from a large and diverse population of phages. As an example of a hyperplastic region, the tail fiber module is the most plastic structural region in the type T4-like phage genomes (Figure 1.12), as even in closely related T-even phages, the sequence and structure of this locus has diverged significantly (Montag et al., 1987). Only the first gene in the locus, *g34* (the proximal tail fiber subunit) is relatively well conserved; the downstream gene, *g35* encoding the tail fiber hinge joint, has diverged considerably. In Aeh1, for example, a sequence related to the N-terminal domain of the T4 gene product (gp) 35 has been fused to a large segment of novel protein sequence. Similarly, the small distal fiber (SDF; gp36) appears to be a collage of tail-fiber domains from a wide variety of phages with the homology to T4 gp36 only being retained in the N-terminal domain. In several phages, the homologues of the SDF (gp36) and the LDF (long distal fiber; gp37) are fused into a chimeric gene that has, in some cases, been duplicated (Krisch and Comeau, 2008).

It is obvious that lateral gene transfer must play a major role in this modular shuffling of the fibrous domains among the tail genes. Two important mechanisms to generate this variation seem to be the join-copy and join-cut-copy recombination pathways of T4 (Mosig, 1998). This recombination can contribute to evolution because it facilitates acquisition of any foreign DNA with limited sequence homology during horizontal gene transfer, without requiring transposition or site-specific recombination functions (Mosig et al., 2001).

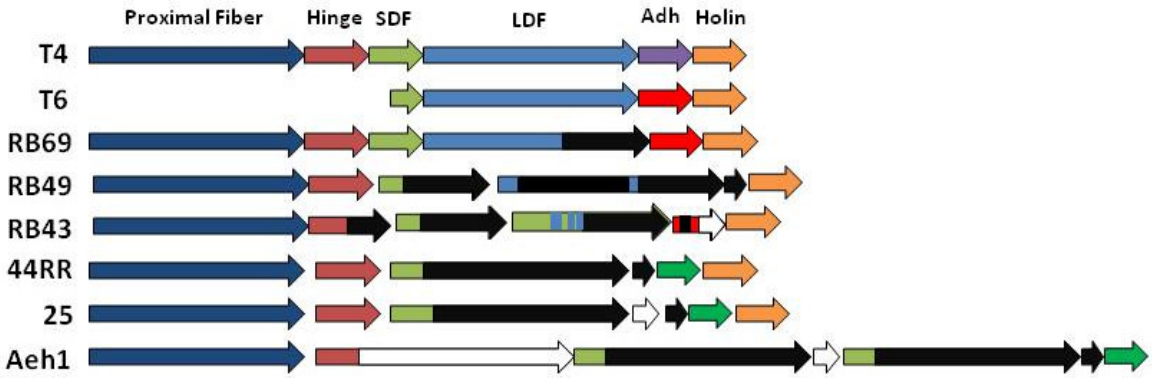


Figure 1.12. The hyperplastic tail fiber modules of six completed T4-related phages, compared to T4 and T6. For each gene, regions of T4 homology are indicated in the same color, regions of similarity to other phages are in black, and regions of unknown origin are in white. SDF, Short Distal Fiber; LDF, Long Distal Fiber; Adh, Adhesin (after Comeau et al., 2007).

It has been shown that T4 recombines most efficiently early in infection (Krisch et al., 1972) before foreign DNA is degraded, and that it requires only 50 bp of homology to recombine at reasonable frequencies (Singer et al., 1982). As an example, recombination in the highly diverged tail fiber locus apparently involve small conserved motifs that seem to be recombinational hotspots, like the His-boxes located within the C-terminal part of the gene encoding the long distal fiber of T4 (Tétart et al., 1996). Exchanges with such conserved motifs can replace one adhesion-recognizing domain by another one, and change the host range of the phage (Figure 1.12; Tétart et al., 1998). The important evolutionary advantage of this modular protein design for tail proteins is that it facilitates the swapping of recognition motifs among related phages. In other words, if one phage has somehow acquired a recognition domain for a novel bacterial receptor, then the entire population has potential access to it by recombination (Krisch and Comeau, 2008).

These exchanges of tail genes are not limited to T4-like phages, and are observed between many phages that infect enteric γ-Proteobacteria. For example, phage λ has side-tail fiber genes that are sufficiently similar to the tail fibers of T4 to be able to functionally replace them (Hashemolhosseini et al., 1996). Another yet completely unrelated example involves the T7-like phage SP6 and lambda-related phage P22, which have tail spike proteins (500 amino acids) that are 58% identical (Dobbins et al., 2004). The latter two phages have very different lifestyles and encode no other proteins that are closely related. It seems clear that one of these phages obtained the tail spike gene from the other's

phylogenetic group; it is simplest to imagine that an SP6 ancestor got it from a P22-like phage as no other known T7-like phage has this type of tail-spike gene and all known P22-like phages do (Casjens, 2005). It might be particularly advantageous for a phage to exchange tail genes, as a new tail tip might allow DNA delivery into a new host and by doing so might give access to new ecological niches.

1.4.5 ORFans, the dark matter of the phage universe

ORFans are defined as ORFs (Open Reading Frames) having no sequence homologues in other genomes (Fisher and Eisenberg, 1999). ORFan genes comprise approximately 30% of sequenced phage genomes compared with approximately 15% for the typical cellular genome (Frost et al., 2005). It has been suggested that ORFans are either misannotated genes, rapidly evolving sequences, newly formed genes, or genes recently transferred from not yet sequenced cellular or viral genomes (Siew and Fisher, 2003). Despite their abundance, very few ORFans have been experimentally characterized, and thus, most ORFans in the database are annotated as hypothetical proteins of unknown function. The source of these ORFans remains a major mystery of the post-genomic era since, contrary to previous expectations, their proportion remains stable despite the increasing number of complete genome sequences available (Figure 1.13).

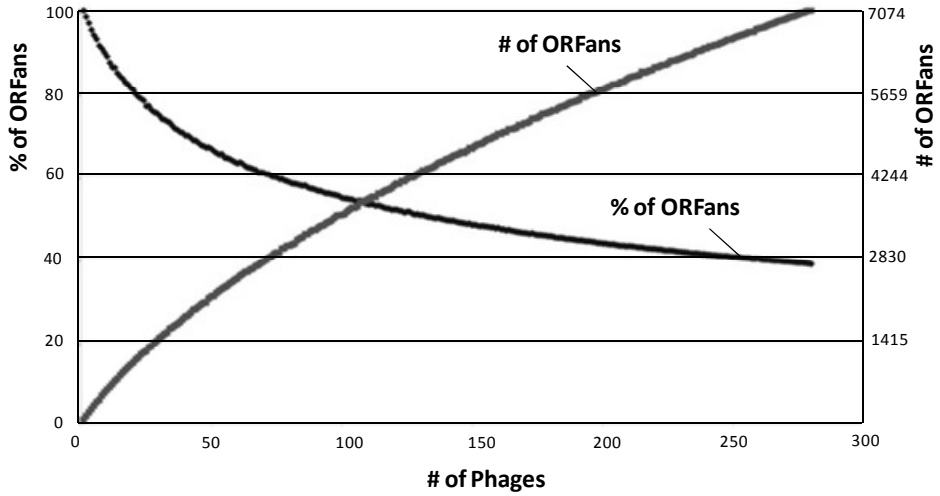


Figure 1.13. The number of phage ORFans is growing. The plot was computed by averaging the results of 1,000 random selections of subsets of the 280 phage genomes. The average number of ORFans in each subset size and the percentage of ORFans (computed as the number of ORFans divided by the total number of ORFs) are shown (after Yin and Fischer, 2008).

One of the most remarkable features of the ORFans is their enormous diversity, as even relatively closely related phages may have little or no overlap in their ORFan population. A recent interesting view on ORFans was provided by Comeau and colleagues (2008), who believe that many ORFans are a sort of "molecular splint" that binds to host proteins and either inhibits or modifies their function. This hypothesis fits well with the generally small size, extreme sequence diversity and lack of homologues in the databases. The authors state that in evolutionary terms, the construction of a small polypeptide that binds specifically to a host protein domain is not that difficult and can be accomplished in numerous ways, which in turn leads to highly diverse sequences. Many examples of this are present in literature, like Gp2 of T7 which binds to and inactivates the β' subunit of the host RNA polymerase (Nechaev and Severinov, 1999), or the 95 aa T4 internal protein IP1 which binds and inhibits a host restriction endonuclease that recognizes the hydroxymethylcytosine-modified DNA present in the T4 genome (Rifat et al., 2008). An interesting observation in this respect is the repeatedly reported underrepresentation of these novel ORFs in plasmid libraries constructed for shotgun sequencing, also pointing at putative functions of these ORFs in cell lysis or in commandeering host machinery for phage growth (Miller et al., 2003; Nolan et al., 2006).

Recent research on microbial ORFans pointed out that 56% of recently integrated foreign elements in bacteria and archaea (showing atypical sequence composition) are integrative elements, whereas only 7% likely originated via horizontal gene transfer from distant cellular sources (Cortez et al., 2009). This points to a major role for plasmids and viruses in microbial ORFan acquisition, which was consistent with the many common features observed between ORFans of γ-Proteobacteria and viral ORFans (Daubin and Ochman, 2004). These observations support the hypothesis that ORFan traffic is bidirectional, with ORFans being primarily created by phages to control host functions, but occasionally co-opted and eventually modified by bacteria for their own purposes (Comeau and Krisch, 2005). As this reservoir is largely unexplored, the study of the expression profiles, functions and structures of the ORFans should become one of the priorities of post-genomic studies (Cortez et al., 2009; Roucourt et al., 2008).

1.4.6 Phage evolution

Looking at the current insights in evolutionary mechanisms that shaped the tailed phages, different pictures emerge for different phage groups. Extensive genetic mosaicism is present between lambdoid phages, and these phages apparently easily replace essential genes with completely unrelated

sequences that are functionally equivalent (Section 1.4.3). In contrast, the genome of Roseophage SOI1 was found to be completely stable over ecologically significant time and distance (Angly et al., 2009), as were the genomes of three 936-like *lactococcal* phages (Mahony et al., 2006) and three Myoviruses infecting *Burkholderia cepacia* (Summer et al., 2006). Despite temporal and spatial separations between their isolation, the three latter phages (Bcep1, Bcep43 and Bcep781) exhibit 87% to 99% sequence identity, and most coding differences are due to synonymous nucleotide substitutions, a hallmark of neutral genetic drift (Summer et al., 2006). The T4 group seems to be a combination of both extremes since they encode both a core genome and display rampant lateral gene transfer in the hypervariable regions (Section 1.4.4).

It has been argued that these differences are correlated to the lifestyle of different phages, as virulent phages might have less opportunity to exchange genes among themselves and therefore are less diverse than temperate phages (Dobbins et al., 2004). It seems indeed reasonable that an infecting phage would encounter potential exchange partners more frequently in their latent prophage state compared to co-infecting phages from the environment. This was also observed by the construction of a reticulate classification system for phages, which showed indeed that mosaicism is less crucial in the diversity of virulent phages in comparison to temperate phages (Lima-Mendez et al., 2008). Exceptions exist however, and the presence of recently acquired temperate phage genes in virulent phage genomes (e.g. the SP6 tail-spike, Section 1.4.4) indicates that this might not be restricted to temperate phages, and that virulent phages might obtain genes from prophages (Casjens, 2005). Moreover, the recent identification of a prophage that is similar to the strictly virulent phage T7 in a *Pseudomonas putida* genome has further blurred the boundary between temperate and lytic phages (Casjens, 2003).

As more data sequence data accumulates, research groups began to develop ways to structure this information and to shape order and to reveal mutual relationships. The ACLAME database organizes phage proteins and prokaryotic mobile genetic elements into families based on sequence similarity, allowing crosslinks between various phages and more correct gene annotations (Leplae et al., 2004; Lima-Mendez et al., 2007). Similar procedures were used to define 'Phage Orthologous Groups' (POGs) as the set of orthologous genes that are shared by three or more phages each, and 'phageness' as the specific association of a gene with a (pro)phage genome as opposed to host genomes. Determination of the phageness of each POG indicated that about 80% of POGs are rarely or never exchanged with the host genomes, while 8% appear to be monophyletic within phages, most likely having been gained by

the phage kingdom in a single transfer event (Liu et al., 2006). By comparing the presence and absence of these POGs in 158 phages, 18 clusters of phages were delineated, with phages in most groups tending to have overlapping host ranges. By employing an algorithm for automated detection of horizontal gene transfer (HGT), is was shown that the majority of POGs have not been involved even in a single phage-phage HGT event. Moreover, while HGT is very common within each cluster, only 4 groups of phages with repeated between-group transfers were detected (Glazko et al. 2007). The same conclusion was drawn by Lavigne and colleagues (2008), who applied BLAST-based tools (CoreExtractor.vbs and CoreGenes) on all completely sequenced phage genomes. They were able to delineate specific genera of related phages, regardless patterns of frequent horizontal exchange. In other words, as frequent as HGT events may be in absolute terms, they occur mostly between related phages infecting related hosts, and are nevertheless not so frequent as to obscure the pattern of vertical, divergent evolution of phage genomes.

1.5 Bacteriophages of *Pseudomonas aeruginosa*

1.5.1 The host bacterium: *Pseudomonas aeruginosa*

1.5.1.1 Occurrence and virulence

Pseudomonas aeruginosa (Figure 1.14) is a ubiquitous Gram-negative soil organism which thrives in diverse humid environments. The bacterium is capable of utilizing a wide range of organic compounds as food sources, giving it an exceptional ability to colonize ecological niches where nutrients are limited. It has the ability to express a variety of virulence determinants which cause a wide range of infections in plants, nematodes, insects and animals. As a typical opportunist, it infects injured, burned and immunodeficient humans and causes persistent respiratory infections in individuals suffering from cystic fibrosis (Lyczak *et al*., 2000) and life-threatening infections in burn wound patients (Pruitt et al., 1998).

The virulence factors which are deployed by *P. aeruginosa* act at different stages of pathogenesis, starting from the bacterial adhesion and followed by colonization, invasion and dissemination (Lee et al., 2006). Of particular importance is a type III secretion system, which enables injection of effector proteins directly into the cytoplasm of host cells (Engel and Balachandran, 2009). Four effector proteins have been identified: ExoS and ExoT, which are bifunctional enzymes with GTPase activating protein

and ADP ribosylase activities, and ExoU (a cytotoxin) and ExoY (an adenylate cyclase). Different combinations of these proteins have profound and often devastating effects on epithelial barrier function and wound healing (Kerr and Snelling, 2009). During chronic adaptation in CF patients, numerous genetic changes occur that include conversion to mucoidy due to alginate production, mutations in regulatory genes, loss of the ability to add O-side chains to lipopolysaccharide core sugars and loss of motility (Smith *et al*., 2006), a process mainly catalyzed by hypermutation (Mena et al., 2008).

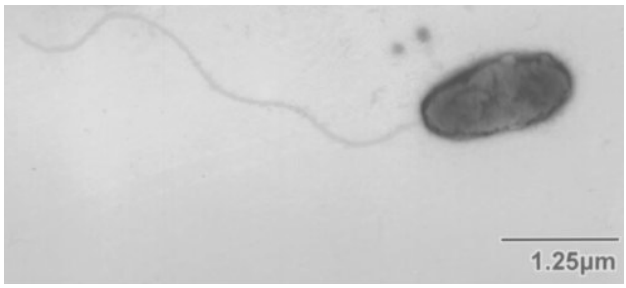


Figure 1.14. Electron micrograph of *Pseudomonas aeruginosa*. The bacterium measures 0.5 to 0.8 µm by 1.5 to 3.0 µm. The single polar flagellum can be distinguished (Dasgupta et al., 2000).

Infection with *P. aeruginosa* forms a major health problem mainly because of its inherent resistance to multiple drugs and the ability to gain resistance to all available treatments through mutation (Livermore, 2002). Multi-drug resistance (MDR) in *P. aeruginosa* is attributed to a combination of resistance mechanisms, mainly including multi-drug efflux pumps, aminoglycoside-modifying enzymes, β-lactamases and target site modifications. Especially the efflux system is considered to be a very effective mechanism of resistance due to the fact that it can mediate resistance to many classes of drugs (Tenover, 2006). Souli and colleagues (2008) recently published data from 23 countries on the European Antimicrobial Resistance Surveillance System, stating that 18% of all isolates were MDR *P. aeruginosa* strains.

1.5.1.2 Population diversity

To date, 16 complete genome sequences of *Pseudomonas* species have been determined. One of the widely studied strains of *P. aeruginosa* (PAO1) contains a genome of 6.3 million bp and shows a large proportion of regulatory genes for catabolism, transport, efflux of organic compounds and possible chemotaxis systems. These bacterial properties are indications for its environmental adaptability and ubiquitous nature (Stover *et al.,* 2000). However, the extreme adaptability of *P. aeruginosa* to different

habitats has its true origin in a mosaic bacterial genome of which the main part (90%) consists of a highly conserved core component, interrupted by strain-specific regions of genomic plasticity acquired by horizontal gene transfer (Wolfgang *et al*., 2003). Using this evolutionary mechanism, *P. aeruginosa* is able to shape its genomic composition to favor survival in the widest range of environmental reservoirs, with corresponding enhancement of its metabolic capacity (Mathee *et al*., 2008). This pattern of continually expanding the genomic repertoire is encountered in a number of environmental organisms, and is in contrast to the genomic decay observed during the adaptation of pathogens to a parasitic existence (Moran, 2002).

Besides this full-genome sequencing, the introduction of molecular typing methods has permitted improved characterization of outbreaks of *P. aeruginosa*. Several studies reported that individual patients were colonized by unique clonal lineages (e.g., Tramper-Stranders e*t al*., 2008), although polymicrobial infections are also observed in burn wound patients (De Vos *et al.,* 1997). Using a combination of DNA fingerprinting, sero- and pyoverdine typing, Pirnay and colleagues (2002) were able to elucidate the epidemic population structure of *P. aeruginosa*, comparable with that of *Neisseria meningitides*. According to this model, highly successful epidemic clones occasionally emerge through selection from a background of a large number of relatively rare and unrelated genotypes which are recombining at a high frequency.

Using the same methods, it was shown that a *P. aeruginosa* community sampled from a small river was almost as diverse as the global *P. aeruginosa* population, harboring members of nearly all successful clonal complexes but showing relatively low levels of antibiotic resistance (Pirnay et al., 2005). This study illustrates strongly the significance of river water as a reservoir and source of distribution of potentially pathogenic *P. aeruginosa* strains, which can infect immune-compromised patients with the traits acquired to survive in natural ecosystems. This confirms previous observations that clinical isolates of this bacterium are genotypically and functionally indistinguishable from environmental isolates (Alonso et al., 1999), and numerous outbreaks are indeed attributed to an environmental source (e.g., Trautmann et al., 2006).

1.5.2 Bacteriophages of *P. aeruginosa*

Phages of *Pseudomonas* have been known for a long time, and have been used in typing experiments (e.g., Bergan, 1972) and phage therapy treatments in Eastern Europe (Deresinski, 2009). Phage typing was long considered a useful method for the epidemiological tracing of *P. aeruginosa*, and at least 13

(mostly experimental) typing sets have been proposed since 1957. The most successful set was that of Lindberg, introduced in 1964 and containing first 15, and later 18 phages (Lindberg and Latta, 1974). Ackermann and colleagues (1988) identified 8 different morphotypes in this set (Figure 1.15), and some key representatives like ϕKZ, M6, F10, D3, F116 and gh-1 have been sequenced during the last decade. This paragraph briefly describes the landscape of fully sequenced dsDNA *P. aeruginosa* phages as it looks today, excluding the newly defined genera which are discussed in the results section.

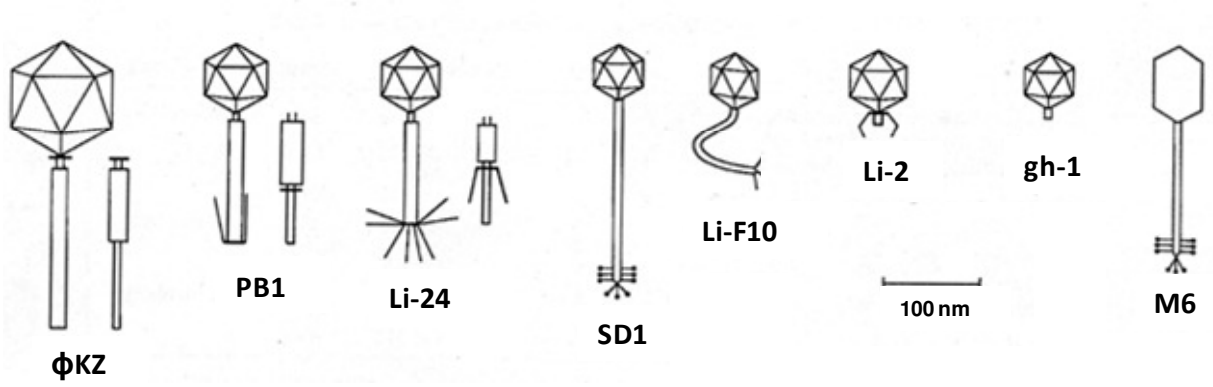


Figure 1.15. Scale drawing of the different morphotypes of the phages belonging the Lindberg typing set. No representatives of the PB-1, Li-24, Li-2, M6 and Li-F10 morphotype were sequenced at the start of this PhD (adapted from Ackermann et al., 1988).

Within the *Myoviridae* family, phage ϕKZ represents a genus of giant and broad-range *Pseudomonas* phages, containing unusually large heads and tails that have a conspicuous collar and are surrounded by a loose fibrous network (Ackermann et al., 1988). The 280 kb-genome of ϕKZ (Mesyanzhinov et al., 2002) is still among the largest phage genomes sequenced to date. Recently, phage SD1-M was reported to be up to 99% identical to ϕKZ (Kwan et al., 2006), and also phage 201ϕ2-1 contains a clearly conserved ϕKZ-like core genome (Thomas et al., 2008). The more distantly related phage EL (211 kb; Hertveldt et al., 2005) represents another group of giant *P. aeruginosa* phages. The only other fully sequenced *Pseudomonas*-infecting myovirus is the temperate, cytotoxin-converting phage ϕCTX, which is genetically and morphologically related to P2 (Nakayama et al., 1999).

All *Pseudomonas* phages belonging to the *Siphoviridae* family described to date have a temperate character, implying the possibility to undergo a lytic or lysogenic interaction with their host. Phages B3 and D3112 (SD1 morphotype, Figure 1.15) represent transposable phages which are commonly isolated

from natural populations and clinical isolates of *P. aeruginosa* (Heo et al., 2007; Zegans et al., 2009). As already stated in section 4.3, these phages share the transposable nature and propagation cycle with phage Mu, but resemble phage lambda morphologically (Braid et al., 2004). Homologs of gene products of these phages are extremely widespread, and can be found in - amongst others- strains of *E. coli*, *Salmonella enterica*, *Vibrio cholerae* and *Haemophilus influenzae*, indicative for the influence of these phages on bacterial evolution (Hertveldt and Lavigne, 2008). Phage DMS3 shares DNA similarity with phage D3112 and is able to transduce DNA between *P. aeruginosa* strains PA14 and PAO1 (Budzik et al., 2004). Another well-studied Siphovirus is phage D3, which causes host serotype conversion from O5 to O16 upon lysogenization (Kuzio and Kropinski, 1983). Although this phage is morphological identical to D3112 and lambda, the D3 genome is significantly larger (56 kb), it is not a transposable and integrates at two distinct loci in the host genome (Kropinski, 2000). Recently, the first D3-like phage was described by Uchiyama and colleagues (2009).

In current databases, five types of *Pseudomonas*-infecting *Podoviridae* can be distinguished based on genome organization and behavior during infection. While gh-1 is a *stricto sensu* T7-like lytic phage (Kovalyova and Kropinski, 2003), the 'ϕKMV-like viruses' form a genus of widespread and highly virulent phages with a genome architecture which is more distantly related to that of T7 (Lavigne et al., 2003). A third group includes uncharacterized temperate phages Pap2 (NC_005884) and 119X (NC_007807), which display 93% nucleotide sequence identity (Kwan et al., 2006). As a fourth type, the temperate and pilus-specific phage F116 (65,195 bp) replicates extrachromosomally as a plasmid during its lysogenic cycle and is considered as a genomic orphan (Byrne and Kropinski, 2005). Finally, phage PaP3 (NC_004466) is also described as a temperate phage and unrelated to any known phage. The PaP3 genome has 5'-protruding cohesive ends, and is able to integrate into the $tRNA^{lys}$ gene of its host (Tan et al., 2007).

1.5.3 Phage Therapy against *P. aeruginosa*

Although antibiotics have historically been very successful for the treatment of wound infections, the emergence of MDR bacteria and the failure of drug discovery programs over the last 10 years to provide new broad spectrum antibiotics with truly novel modes of action pose a major threat to public health worldwide. The reduction or closure of many antibacterial drug discovery programs is partly due to the huge economic cost of drug development, which is not very profitable when such agents rapidly

succumb to resistance (Thiel, 2004). In this light, the western critical skepticism towards phage therapy (Section 1.1.1) was again followed by renewed interest and reappraisal at the beginning of the 21st century (Stone, 2002). The innocuous nature of phages was demonstrated by adding high-titer T4 phage stock to drinking water of human volunteers (Bruttin and Brüssow, 2005). Several research groups performed controlled animal experiments, and reported the effectiveness of phage therapy for treatment of, amongst others, vancomycin-resistant *Enterococcus faecium* (Biswas et al., 2002) and methicillin-resistant *Staphylococcus aureus* infections (Matsuzaki et al., 2003).

Many research groups turned their attention to the phage treatment of *P. aeruginosa* infections in mice, guinea pigs and pet dogs (*e.g.*, Soothill et al., 2004). A report of the treatment of single cases of human burn wounds indicates that bacteriophage multiplication is associated with clinical improvement (Marza et al., 2006). A first randomized, double-blind and placebo-controlled Phase I/II clinical trial was performed on 24 patients suffering from chronic otitis caused by MDR *P. aeruginosa*, showing efficacy and safety in treatment of this infection (Wright et al., 2009). As many regulatory hurdles have to be taken in European legislation, however, we are still far from commercialization of these phage preparations (Verbeken et al., 2007). An important step was recently taken by the detailed description of a quality-controlled small-scale production of a bacteriophage preparation, leading to a safety trial in burn wound patients at the Burn Centre of the Queen Astrid Military Hospital in Brussels, Belgium (Merabishvili et al., 2009). Nevertheless, a crucial condition towards practical application of phage therapy will be the availability of a large library of well-characterized phages, subjected to in-depth genomic analysis to confirm the absence of undesired genes (Denou et al., 2009).

It is with these notions that this thesis work focused on the characterization of newly isolated, virulent bacteriophages infecting *P. aeruginosa.* The in-depth analysis of genomic differences between these phages not only helps in evaluating the suitability of these novel phages for use as antimicrobials, it also contributes to the understanding of phage population genetics and evolution.

Chapter 2

Materials and Methods

2.1 Bacterial strains, plasmids and growth media

P. aeruginosa PAO1 and 72 AFLP-genotyped clinical *P. aeruginosa* strains were provided by J.P. Pirnay (Pirnay *et al*., 2002). In December 2004 and April 2005, 49 additional *P. aeruginosa* isolates were isolated in ten different units within UZ Leuven from sputum (28), wounds (4), urine (9) and bronchus (8) samples by Prof. J. Verhaegen. Mutants of *P. aeruginosa* PAO1 defective in Type IV pili, flagella and AlgC production were kindly provided by Dr. Max Schobert (Technical University of Braunschweig, Germany). The *P. aeruginosa* PAO1 *rpoN* mutant strain (*rpoN::tet*) was kindly provided by H. Arai (Arai *et al*., 2005). All plasmids used during this study are listed in Table 2.1. *Eschericia coli* XL1-Blue MRF' cells (Stratagene, LA Jolla, CA, USA) were used for all standard cloning purposes.

Table 2.1. Plasmids used in this study. For each vector, sequencing primers are located 5' and 3' of the cloning site. Given the short average insert length (200 bp) in the promoter trap assays, one primer was sufficient to cover the whole insert.

Vector	Selection	Sequencing primers	Purpose	Section
pUC18	*bla*	TGTAAAACGACGGCCA CAGGAAACAGCTATGA	Shotgun sequencing library (blue/white selection)	2.3.2
pJET1.1	*bla*	GACTCACTATAGGGAGAGCGGC GAACATCGATTTTCCATGGCAG	Shotgun sequencing library (positive selection marker *eco47IR*)	2.3.2
pJET1.2	*bla*	GACTCACTATAGGGAGAGCGGC GAACATCGATTTTCCATGGCAG	Shotgun sequencing library (positive selection marker *eco47IR*)	2.3.2
pTZ110	*bla*	GCAAGCTTCTCGAGCCC	Promoter trap library; contains promoterless lacZ gene	2.3.3
pEX18Ap	*sacB, bla*	GTGCTGCAAGGCGATTAAGT GAGCGCAACGCAATTAATGT	Suicide plasmid; gene replacement in *Pseudomonas aeruginosa*	2.4.4

All chemicals were purchased from Acros Organics (Geel, Belgium), unless stated otherwise. Bacterial strains were grown in standard Lysogeny Broth (LB) consisting of 1% w/v/ bacto tryptone (Lab^M, Bury, UK), 1% w/v/ NaCl and 0.5% w/v/ yeast extract (Lab^M). To obtain soft agar or solid medium, 0.7% or 1.5% w/v of agar (Lab^M) was added, respectively. 100 µg/mL ampicillin (Ap^{100}, Roche Applied Science,

Basel, Switserland) or 200 µg/ml carbenicillin (Cb^{200}, Carl Roth GmbH & Co, Karlsruhe, Germany) was added to the medium to select for the *bla* resistance marker in *E. coli* and *P. aeruginosa* strains, respectively. The presence of 30 µg/ml gentamicin (Gm^{30}, Roche Applied Science) in the medium allowed the selection for the *aacC1* resistance marker, and counterselection against the *Bacillus subtilis SacB* gene was tested by adding 5% sucrose to the agar plates.

2.2 Phage manipulations

2.2.1 Bacteriophage isolation, amplification and purification

398 water samples (30 mL) were collected in sterile 50 mL falcons from environmental and hospital surroundings in 14 different countries spread over four continents (Table 2.2). Upon arrival, samples were cleared by filtration (0.45 µm pore size, Millipore, MA, USA) and incubated for 16 h. in the presence of exponentially growing *P. aeruginosa* cells at 37°C. Cultures which showed visible lysis were pelleted by centrifugation at 4°C and 8000 x *g* for 20 min. Bacterial cell debris was removed by filtration (0.45 µm filters), and the presence of phages was verified by plating 100 µl of the filtrate on lawns of *P. aeruginosa* using a standard soft agar overlay method (Carlson, 2005). For this agar overlay method, a hallmark procedure in phage research, 100 µl phage solution were added to 200 µl of an overnight culture of a susceptible *P. aeruginosa* strain and mixed with liquid soft agar (50 °C). This mixture was spread on solid LB medium, incubated overnight at 37 °C and controlled on the presence of visible zones of lysis or *plaques*.

Table 2.2. Number of water samples collected in each country.

Country	# Samples	Country	# Samples	Country	# Samples	Country	# Samples
Belgium	320	Sri Lanka	4	Kenya	3	Croatia	15
Netherlands	6	Mexico	6	South-Africa	6	Spain	8
Germany	4	Italy	8	Tenerife	2		
France	8	Australia	4	Greece	4		

Alternatively, the purified water samples were concentrated by ultracentrifugation (100,000 x *g*) and spotted directly on lawns of clinical *P. aeruginosa* strains. Single plaques were picked up and replated three times to ensure pure phage stocks. Phages were given a three-letter code based on the place of research and isolation (e.g. Leuven, Kasteelpark Arenberg), followed by the sample number.

Phages were typically amplified using soft-agar overlays, adding enough phages (approx. 10^4 pfu) to the indicator cells to produce confluent lysis on the plate. After overnight incubation, soft-agar layers were scraped off, supplemented with 5 mL phage buffer (10 mM Tris-HCl pH 7, 10 mM $MgSO_4$, 150

mM NaCl) and passed through a 0.45 µm pore-size filter. In contrast, highly lytic phages like LKD16 and LBL3 were amplified using liquid cultures (400 ml) of *P. aeruginosa* cells infected at an OD_{600} of 0.3 at a multiplicity of infection (MOI) of 1. Infected cells were incubated at 150 rpm until complete lysis was observed, and 2 ml of Chloroform (E. Merck, Darmstadt, Germany) was added to lyse residual bacterial cells. The mixture was left overnight at room temperature and cell debris was removed by centrifugation at 3,000 x *g* for 20 min at 4 °C. The supernatant was filtered through an 0.45 µm filter and stored at 4°C.

Phages were precipitated with polyethylene glycol 8000 (8%, wt/vol) -1 M NaCl for 4 hours, spun at 8000 x *g* for 20 min and resuspended in 15 ml phage buffer. This suspension was layered two consecutive times on a CsCl (MP Biomedicals, Illkirch, France) step gradient (1.33, 1.45, 1.50 and 1.70 g ml^{-1}), centrifuged at 140,000 x *g* for 3 hours and dialyzed three times for 30 min against 300 volumes of phage buffer using Slide-A-Lyzer dialysis cassettes (Pierce, Rockford, IL, USA). All analyzed phages were found stable in CsCl, and were stored at 4°C in the dark.

2.2.2 Host range analysis and stability assays

The stability of all analyzed phages in their storage conditions was tested each semester by titration of the stock solution. Only phage LUZ24 seems unstable in these conditions, as its titer dropped 80% in half a year. The ability of phages LKA1, LBL3, LUZ7, LUZ14, LUZ19, LUZ24 and LKR3 to retain their infectivity at different pH levels was assayed by incubating 10^7 -10^9 pfu for 1 h. at room temperature in 1 ml of different pH-buffers ranging from pH 1 to 13. The pH buffer was prepared with 150 mM KCl, 10 mM KH_2PO_4, 10 mM NaCitrate, 10 mM Boric acid and adjusting the pH using NaOH (10 M) or HCl. Each mixture was titrated using the standard double agar method, and compared to an untreated sample.

The host range of all newly isolated phages was examined by incubation of every phage (10^7 pfu) – host combination using standard soft-agar overlays in 6-well plates, in three independent assays. After overnight incubation at 37°C, plates were checked for confluent lysis or plaque formation against a negative, uninfected control. For phages LKA1 and LKD16, the result was compared (and found identical) to host range analysis by spot testing (Gill et al., 2003).

2.2.3 One-step growth and adsorption curves

For phages LKD16, LKA1, LIT1, LUZ7 and LUZ24, one-step growth curves were made using a procedure modified from Carlson (2005). Host bacteria were grown to OD_{600} of 0.3, and 1 ml of this bacterial solution was infected with phages at an MOI of 0.005. After 5 to 8 min incubation at 37 °C, the mixture was centrifuged at 13,000 x *g* for 1 min to remove unadsorbed phage particles and

synchronize phage infection. The pellet was resuspended and 10^5 times diluted with pre-warmed LB-media. Samples were taken at 2-20 min intervals and titrated immediately. The latent period was defined as the interval between adsorption of the phage to the host cell and host cell lysis. The burst size was determined as the number of progeny phage particles produced by a single host cell.

In adsorption assays, 1 ml of exponentially grown bacteria was infected with phage at an MOI of 0.001 and incubated immediately at 37 °C, the low MOI ensuring at least one receptor site for every phage. At regular 1 min.-intervals starting from 30 sec. after infection, 100 µl samples were taken and mixed with 850 µl LB-media, supplemented with 50 µl $CHCl_3$ to destroy the host cells. The supernatant was titrated to estimate the amount of non-adsorbed or reversibly adsorbed phages at the different time intervals.

2.2.4 Electron Microscopy

Purified virus samples were washed in 0.1 M ammonium acetate (pH 7.0) and pelleted by centrifugation for 1 h. at 25,000 x *g*. At the Laval University (Quebec, Canada), phage particles were deposited on carbon-coated copper grids, stained with 2% (w/v) potassium phosphotungstate (pH 7.0) or uranyl acetate (pH 4.5) and examined in a Philips EM 300 electron microscope. Magnification was controlled by means of T4 phage tails (Hans-W. Ackermann, personal communication).

2.3. Phage genome analysis

2.3.1 Phage DNA isolation and restriction analysis

Phage DNA was isolated using a modification of the standard protocol for phage λ DNA isolation (Sambrook and Russell, 2001). In short, purified phage particles were disrupted by incubation in 0.5% SDS (wt/v), 20 mM EDTA and 2 µg/ml proteinase K at 56°C for 1 h., followed by DNA extraction (phenol/chloroform) and precipitation (NaAc/ethanol). Typically, an additional treatment with RNase A (0.1mg/ml) for 1 hr at room temperature was required to remove residual RNA. Phage DNA was stored at -20°C in neutral Tris buffer.

For restriction analysis, 250 to 500 ng of phage DNA was digested with specific restriction enzymes, used in a total reaction volume of 20-50 µl according to the manufacturer's guidelines. Restriction enzymes were purchased from Roche Applied Sciences, New England Biolabs (Ipswich, MA, USA), or Fermentas (Burlington, Canada). To analyze DNA fragments, the molecules were separated by agarose gelelectophoresis (1 to 2% w/v agarose) in TAE electrophoresis buffer (40 mM Tris.HCl, 500 mM sodium acetate, 50 mM EDTA at a pH of 7.2). The DNA fragments were stained with ethidium bromide and visualized by UV illumination. To estimate the size and the amount of DNA fragments, a reference (λ/*Pst*I, a 100 or 1000 bp ladder) was ran along the fragments.

2.3.2 Phage genome sequencing

Recent technologies are rapidly replacing the traditional Sanger-based methods in full-genome sequencing, with a tremendous increase in throughput and reduction in costs (Morozova and Marra, 2008). However, all but one phage genomes described in this dissertation are still sequenced in-house using Sanger-based technology and capillary sequencing. The long sequence reads (up to 1000 nt) and the reduction of chemicals needed by the streamlining of sequencing procedures has proven to be very cost-efficient in sequencing small phage genomes (up to 60 kb). Moreover, the combination of shotgun sequencing and primer walking provided opportunities to observe sequence-specific properties like single-stranded genomic nicks (Chapter 4) and poly(G) stretches (Chapter 7). One phage, PB1, was submitted to the McGill University and Génome Québec Innovation Centre (Montréal, QC, Canada) and sequenced using 454 Technology.

2.3.2.1 Shotgun library construction

Although random fragmentation of the genomes of phages LKD16 and LKA1 was achieved by nebulisation and partial restriction digests, best results were obtained through shearing by sonication (Sonics, Vibra Cell™, USA). The generated genomic fragments between 800 and 1500 bp were extracted from ethidiumbromide-free gels using the QIAquick Gel extraction kit (QIAGEN GmbH, Germany). These fragments were end-repaired by 15 min incubation in the presence of 3 units T4 DNA polymerase and 0.1 mM dNTPs (Fermentas) at 16 °C, followed by the addition of 2 units of Klenow polymerase (New England Biolabs) and incubation at 22 °C for another 15 min The reaction was finally stopped by adding EDTA (final concentration of 10 mM) and heating for 30 min at 75 °C.

Prior to ligation in pUC18, the linearized vector was dephosphorylated with Antarctic phosphatase (New England Biolabs) and phage fragments were subjected to phosphorylation (Polynucleotide kinase, New England Biolabs) as described by the manufacturer. The ligation mixtures contained between 50 and 100 ng blunt-end vector DNA and repaired phage DNA in a 1:1 to 3:1 molar ration, in a reaction volume of 20 µl. While ligation in pUC18 was performed overnight at 16°C followed by heat-inactivation (10 min 70°C) and ethanol precipitation, ligation mixtures in pJET1 and pJET1.2 were incubated for 30 min at room temperature and immediately used for transformation. Electrocompetent *E. coli* XL1 Blue MRF' cells prepared as described by Tung and Chow (1995) were electroporated with the ligation mixtures in a 1 mm gap electroporation cuvettes (Eurogentec) using a Gene Pulser™ apparatus (Bio-Rad Laboratories, Hercules, CA, USA). Immediately afterwards, 1 ml of warm LB medium was added to the cells, followed by an incubation period of 45 min at 37 °C before plating the transformation mixtures on selective medium.

Plasmid DNA of individual clones was isolated by growing recombinant bacterial cells overnight in 96-well plates, and lysing them under alkaline conditions (NaOH-SDS) in the presence of RNaseA (0.1mg/l). The solution was subsequently neutralized with an acidic buffer and centrifuged (30 min, 3600 x *g*) to obtain clear cell lysate, which was precipitated using 60% (vol/vol) isopropanol and resuspended in 20 µl Tris-HCl (pH 8.0).

2.3.2.2 Sanger sequencing strategy

In a first round of Sanger sequencing, reactions were carried out in 96-well plates on a *Thermocycler UNO II* (Biometra, Göttingen, Germany), using only 0.2 µl of Big-Dye™ Terminator Mix 1.1 (Applied Biosystems, Foster City, CA, USA), 50 ng of vector DNA (Section 2.2.1) and 0.16 µM of vector-specific *forward* primer (Table 2.1) in the supplied buffer. After standard ethanol precipitation, the samples were separated and analyzed on an ABI 3100/3130 capillary sequencing device (Applied Biosystems). The sequences were analyzed and assembled into *contigs* of overlapping fragments using Sequencher 4.1 or 4.8 software (Genecodes, Ann Arbor, MI, USA). At a specific moment (after completing approx. 80% of the estimated genome length), new electropherograms add only limiting sequencing data due to overlap with existing contigs. At that point, individual plasmids carrying inserts located at the end of the contigs and in non-redundant regions were selected and subjected to sequencing reactions using the vector-specific *reverse* primer.

Table 2.3. Shotgun sequencing libraries of completely sequenced phage genomes described in this study.

Phage	Vector	Transf. eff.[b] (cfu/µg vector)	# Sequenced clones	# Walking primers	Total amount sequenced (bp)	Average Redundancy
LKD16	pUC18	9.2×10^4	354	42	344,235	8.1
LKA1	pUC18	1.6×10^5	401	35	297,957	7.1
YuA	pUC18	N.D.	282	94	314,875	5.3
LUZ24	pJET1	3.9×10^4	263	29	242,572	5.2
LUZ7	pJET1.2	N.D.	471	49	367,902	4.9
LIT1	pJET1.2	N.D.	402	52	344,235	4.7
LBL3	pJET1.2	7.3×10^4	464	34	412,982	6.4
LMA2	pJET1.2	4.6×10^4	420	24	414,315	6.2
14/1	pJET1	N.D.	440	30	330,242	4.9
ϕKF77[a]	-	-	-	21	125,129	2.9
LUZ19[a]	-	-	-	16	233,217	5.4

[a] From previous studies of the highly conserved phages ϕKMV and LKD16, a large set of 253 primers specific for ϕKMV-like phages was available. Sequencing reactions performed with these primers and genomic DNA of LUZ19 and ϕKF77 as template resulted in an 83% and 89% success rate, respectively. Remaining gaps and uncertainties were determined by phage-specific primers, thus completing both genomes using solely primer walking.

[b] N.D., Not Determined.

Remaining gaps, direct repeats and uncertainties in the sequence were determined by direct sequencing using phage-specific walking primers (Eurogentec, Seraing, Belgium), with the annealing temperature adapted to the primer used. In some cases, specific genomic regions had to be amplified using PCR in order to get sequenced; these reactions were performed with *Pfu* DNA polymerase (1.25 units, Fermentas), 0.5 µM of two adjacent walking primers, 0.2 mM of each dNTP (Fermentas) in the supplied buffer using a TRIO-Thermoblock (Biometra, Göttingen, Germany). In general, 10 ng of genomic phage DNA or PCR product is used in primer walking reactions. To ensure quality of the data, every position was sequenced at least twice, although total average redundancies up to 8.1 were reached (Table 2.3).

2.3.2.3 Determination of phage genomic ends

After assembly of all sequences into one circular contig, experimental restriction patterns were compared to theoretically generated ones (*e.g.*, http://www.restrictionmapper.org/). If both patterns did not match, the phage genome was reassembled in order to find the correct zero point. To determine this, the phage genomic ends are digested with BAL-31 exonuclease (40 µg genomic DNA; 0.5 units per µg) (New England Biolabs) at 30°C. Samples were removed at 0, 10, 20, and 40 min after the addition of the enzyme. All samples were purified by phenol-chloroform extraction and ethanol precipitation, and digested with a restriction enzyme which produces a clear pattern of fragments ranging from 0.5 to 15 kb. The two fragments at the physical ends of the genome will shorten over time, while the remaining fragments retain their original length. In a next step, outward-directed primers are specifically designed, leading to a direct stop of the sequencing reaction at the ends of the genome. In case of Direct Terminal Repeats (DTRs), primers that anneal within these repeats will anneal on both sides of the genome, causing the sequence reaction to continue after reaching the genome end. At that point however, a 50% drop in peak intensity is observed, which allows to exactly delineate the length of the DTR.

2.3.2.4 *In silico* analysis of phage genomes

All sequenced genomes were scanned for potential open reading frames (ORFs) by ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf) using the bacterial genetic code for gene prediction, and GeneMark.hmm algortorihm for prokaryotes (Lukashin and Borodovsky, 1998), with *P. aeruginosa* as reference genome and using a minimal ORF length of 120 bp. In addition, smaller, non-predicted ORFs which are conserved between two (or more) related phages were considered as functional ORFs, based on tBLASTx comparisons between the nexly sequenced genome and closely related reference genomes in the database. The annotation of phages PB1 and 14-1 was done using

Kodon (Applied Maths, Austin, TX). All annotated ORFs were manually inspected for the presence of convincing purine-rich ribosome binding sites, 5 to 13 nucleotides upstream of the start codon.
Translated ORF sequences were compared with known proteins using BLASTP (Altschul et al., 1990) with default settings, against the non-redundant database and considering an E-value threshold of 0.001. A search for transmembrane helices was done using the TMHMM algorithm (Möller et al., 2002), while prediction of structural motifs and signal peptides was performed by using the COILS (Lupas *et al*., 1991) and SignalP (Bendtsen *et al*., 2004) algorithms. Putative tRNA genes were searched by using the tRNAscan-SE program (Lowe and Eddy, 1997), and conserved domains were predicted by the Pfam algorithm (Finn et al., 2008).
In a next step, all intergenic regions were scanned for regulatory elements. Prokaryotic promoters were predicted using SAK (Gordon et al., 2003) and visual inspection for putative -35 (TTGACA) and -10 (TATAAT) elements. Potential conserved intergenic motifs were scanned with MEME/MAST (Bailey and Gribskov, 1998) and PHIRE (Lavigne et al., 2004) to identify phage-specific regulatory elements. Finally, putative factor-independent terminators were defined as palindromic sequences followed by a U-rich stretch and with a stable secondary structure (dG<-10 kcal/mol), as calculated by MFold (Zuker, 1989) and searched for using Transterm (Dalphin et al., 1999).

2.3.3 Experimental promoter identification

For phages YuA and LUZ24, promoter trap analyses were performed to evaluate the predicted promoter activity. In a first step, purified genomic phage DNA was randomly sheared by sonication. Fragments ranging from 200 to 400 bp were recovered from agarose gel, end-repaired, phosphorylated and ligated into the *Sma*I-digested and dephosphorylated vector pTZ110, a promoterless broad-host-range vector with a *lac*Z operon fusion (Schweizer and Chuanchuen, 2001). A threefold redundant promoter library was obtained after electroporation of the ligation mixture into freshly prepared electrocompetent *P. aeruginosa* PAO1 cells (Choi et al., 2006) and plating onto LB plates supplemented with 200 µg/ml carbenicillin and 40 µg/ml X-gal. Plasmid DNA from blue colonies was isolated using the QIAprep Spin miniprep Kit (Qiagen, Hilden, Germany) and inserts were sequenced using a vector-specific primer (Table 2.1). Promoter activity was determined quantitatively in *P. aeruginosa* cells by measuring the β-galactosidase activity in a Miller assay (Miller, 1992).

2.3.4 *In vivo* splicing assay

Using RFam (http://www.sanger.ac.uk/software/rfam), a putative group I intron was predicted residing in the DNA polymerase gene of LUZ24 (Section 5.2.2). To check whether this intron is indeed

spliced out during infection, *P. aeruginosa* strain Li010 was grown to an OD_{600} of 0.4 and infected with LUZ24 at a multiplicity of 10 per cell. Ten minutes after infection, 4 ml cells were harvested and quickly frozen to -80°C. Total RNA was isolated using an RNAeasy mini-prep kit (QIAGEN), removing contaminating DNA with RNase-free DNase. Two microgram of total RNA was incubated with 20 pmol of the sequence-specific primer P_R (5'-GCCAAAGGACAGTTAACCCCGAGGAGTTC-3'), followed by first-strand cDNA synthesis using the RevertAid™ H Minus kit (Fermentas). Resulting cDNA was purified, ten times diluted and used for PCR amplification and subsequent cycle sequencing with primers P_R, P_F (5'-ATCATCAGCCGCCAGA-GTAAACGTGGAG-3') and internal primers P_{f1} (5'-GGGTAAGTGGAAGACTGTAGGAAAG-3'), P_{f2} (5'-GGGCTATGGCTGATGAACGAG-3') and P_{ir} (5'-CAATGACTTGTGGATCATCACGTCGCC-3').

2.4. Phage proteome analysis

2.4.1 Phage protein profiles and zymogram analysis

Discontinuous sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis was performed according to Laemmli (1970). Phage proteins were extracted from approx. 10^{11} ultrapure virions, generated by two consecutive CsCl purifications, using a methanol/chloroform extraction (1:1:0.75, v/v/v). The phage protein pellet was resuspended in SDS-PAGE loading buffer (Moak and Molineux, 2004) and boiled for 5 min before loading onto a 12% polyacrylamide gel. Proteins gels were stained using SimplyBlue™ Safestain (Invitrogen) or Silver staining (Shevchenko et al., 1996).

For peptidoglycan hydrolase zymogram analysis, gels were cast containing 0.01% SDS and 350 mg dried and autoclaved *P. aeruginosa* cells. After electrophoresis, zymograms were washed for 30 min with water and then soaked for 3 days at room temperature in three buffers containing 150 mM sodium phosphate, pH 6.0-7.0, 1% - 0.001% Triton X-100 and 0-50 mM $MgCl_2$. Zymograms were stained for 3 h with 0.1% methylene blue in 0.001% KOH and destained with water. Peptidoglycan hydrolase activity was detected as a clear zone in a dark blue background.

2.4.2 Mass spectrometric identification of virion proteins

In recent years, the use of mass spectrometry (MS) for the identification of structural phage proteins has become increasingly popular and can be considered as a logical next step after phage genome sequencing. For MS-based identification, proteins are first digested with trypsin and analyzed, after ionisation, by either electrospray (ESI) or matrix-assisted laser desorption (MALDI). In peptide mass fingerprinting (PMF), the protein is identified by comparison of the peptide ion mass spectrum

obtained with the theoretical mass spectra generated from a computer database of protein sequences. In tandem mass spectrometry (MS/MS), peptide ions are individually mass-selected and further fragmented physically, and the resulting peptide fragmentation spectrum compared with computer-generated fragmentation spectra. In this way, MS/MS spectra allow amino acid sequence-based identification that is more confident than PMF.

2.4.2.1 Sample preparation

To prepare samples for MS analysis, the entire lane was cut into slices. Proteins in the gel slices were destained by addition of 1:1 (v/v) solution of 25 mM ammonium bicarbonate/ACN, reduced with 10 mM 1,4-dithiotreithol (Boehringer, Germany) at 56 °C for 1 hr followed by an alkylation of cysteine residues by a 45 min incubation with 55 mM iodoacetamide at room temperature in the dark. The gel pieces were twice washed with 100mM NH_4HCO_3 for 10 min and subsequently dehydrated with ACN for another 10 min. The gel pieces were dried using vacuum centrifuge (DNAplus, Belgium), submerged into digestion buffer [20 µg/ml trypsin (Promega), in 50mM NH_4HCO_3] and kept on ice for 45 min After removing the supernatant, the gel slices were incubated at 37 °C in 30 µl of 50mM NH_4HCO_3 for 16 hrs. The supernatant was saved and peptides were extracted from the gels by submerging them into 20 mM NH_4HCO_3 followed by sonication for 20 min. After saving the supernatant in the same tube, the gels were submerged in extraction buffer (5% formic acid, 50% ACN) and sonicated again for 20 min. Supernatants were pooled together and saved at -20 °C until analysis by Mass Spectrometry.

2.4.2.2 Identification by ESI-MS/MS

Trypsin digested samples were sent to Biomedical Research Institute of the Hasselt University, Belgium, and analysed by ESI-MS/MS on an LCQ Classic (ThermoFinnigan) equipped with a nano-LC column switching system. The trapped samples were separated over 55 min on an analytical column (Biosphere C18; NanoSeparations) using a linear gradient from 5 to 60% (v/v) ACN in water containing 100mM acetic acid. The eluate from the analytical column was introduced by a nanoelectrospray device and sprayed from a gold-coated fused silica emitter (NanoSeparations). The MS was operated in a data-dependent acquisition mode to automatically switch between MS (m/z 300-1500) and MS/MS acquisition on the three most intense precursor ions, controlled by Xcalibur 1.3 software. MS/MS data were analyzed using Sequest (Thermofinningan) against a local database of all possible phage proteins, considering minimal cross-correlation values of 1.8, 2.5 and 3.5 for single, double and triple charged peptide ions, respectively.

2.4.2.3 Whole Shotgun Approach (WSA)

For phages LKD16, LKA1, LUZ24, EL and YuA, an alternative/complementary approach for the identification of phage proteins was performed, being the digestion of whole phage particles instead of analysing individual proteins separated on an SDS-PAGE gel. To this aim, 10^{11} phages were destabilized by four successive rounds of freeze-thawing and sonication, heated for 10 min at 95°C and reduced in the presence of 10 mM DDT for 1 h at 56°C. Disulfide bounds were blocked by alkylation with 10 mM Iodoacetamide, followed by an overnight trypsinisation of the whole reaction mixture at 37°C. A more elaborate peptide separation is directly performed on the complex sample by combined reversed-phase HPLC and gas phase fractionation in the mass spectrometer prior to MS/MS analysis. Although whole phages generate a more complex peptide mixture, it is proven that nearly all predicted structural phage proteins can be identified, albeit with lower sequence coverage compared to the SDS-PAGE approach (Lavigne et al., 2006). An additional advantage is the need of only one major digest, implying a simplified sample preparation.

2.4.3 *'de novo'* MS as a novel phage typing method

2.4.3.1 Sample preparation

The 4 to 6 most intense protein bands on a phage protein profile (Section 3.2) were excised from the gel and destained, reduced, alkylated, washed and dried as described in section 2.4.2.1. Subsequently, an in-gel guanidination was performed as described in Samyn *et al.* (2006). The gel plugs were dried again, and each sample was rehydrated with approx. 10 µl of ice-cold trypsin solution (5 ng/µl in 50 mM NH_4HCO_3 pH 8.0), left on ice for 45 min and incubated overnight at 37°C after the addition of 20 µl 50 mM NH_4HCO_3. The supernatant was recovered and pooled with peptides extracted twice from the gel with 35 µl of 60% ACN/0.1% DIEA. The mixture was dried completely, redissolved in 6 µl 12.5 mM NH_4HCO_3 by vortexing and divided into two equivolumetric aliquots. The first aliquot (3 µl) was desalted using C18 ZipTip micropipette tips (Millipore). Before being desalted, the second aliquot was subjected to N-terminal derivatization using 4-sulphophenyl isothiocyanate (SPITC) by incubation with 3 µl of the reagent (20 mM in 12.5 mM NH_4HCO_3) at 37°C for 30 min. One-fifth volume of the eluate (0.5 µl) was spotted on an AnchorChip MALDI 384-target plate (Bruker Daltonics, Germany) by cocrystallization with 0.5 µl of CHCA matrix (10 mg/ml in 30% ACN and 0.3% (v/v) TFA), as described by Leon *et al.* (2007).

2.4.3.2 *'De novo'* MALDI-TOF/TOF analysis.

MALDI-MS was performed on an externally calibrated Ultraflex III TOF instrument. MS spectra were acquired in the positive ion reflector mode with 1250 laser shots per spot on both sulfonated and non-sulfonated samples. The most intense sulfonated ions were checked for the characteristic mass difference of 215 Da with their non-sulfonated counterpart, and subjected to laser induced dissociation with subsequent electrical potential lift, accumulating up to 1100 laser shots. This procedure allowed *de novo* determined peptide sequences to be deduced manually, followed by similarity searches using MS-BLAST (Shevchenko *et al*., 2001), MS-Homology (ProteinProspector, University of San Francisco, CA; http://prospector.ucsf.edu), and BLASTp (Altschul et al., 1990). BLASTp parameters were adjusted by using a word size of 2 instead of 3 and the PAM30 score matrix instead of BLOSUM62, and 'phage' [orign] was added as Entrez Query.

2.4.4 Tandem affinity purification (TAP)

2.4.4.1 Construction of the *P. aeruginosa rpoC::TAP* strain

A strain of *P. aeruginosa* PAO1 encoding the original TAP affinity tag appended to the 3′ end of the *rpoC* gene (which encodes the RNAP β′ subunit) was constructed using the suicide vector pEX18Ap (Table 2.1; Hoang et al., 1998). This TAP cassette consists of a fusion of two IgG-binding repeats of the Protein A (2PrA), a TEV protease recognition site, and a calmodulin-binding peptide (CBP) (Rigaut et al., 1999), and was selectively amplified from pEB304 using 5′ and 3′ primers containing *Nco*I and *Bam*HI restriction site, respectively (Table 2.4). The two flanking regions of the 3′ end of the *rpoC* gene of *P. aeruginosa* PAO1 were amplified by standard PCR reaction, with the 5′ tails of the forward and reverse primer for amplification of the 5′ fragment contain a *Hin*dIII and *Nco*I restriction site, respectively. Likewise, the primers for the 3′ fragment carried a *Bam*HI and *Eco*RI restriction site (Table 2.4). The fragments were ligated in *Hin*dIII and *Eco*RI digested and dephosphorylized pEX18Ap, together with the TAP cassette and a gentamicin resistance cassette. Since the cassette was derived from *Bam*HI digestion of pPS856 (Hoang et al., 1998) followed by gel extraction (Qiagen), all fragments are inserted in the correct order.

After verification, the construct was transformed to electrocompetent *P. aeruginosa* PAO1 cells (Choi et al., 2006). Double homologous recombination between the sequences cloned in the suicide vector and the *P. aeruginosa* genome replaces the targeted gene by the TAP and the gentamicin resistance cassette. In addition to the gene targeting construct, pEX18Ap carries a *bla* resistance marker and a *sacB* counter selectable marker. Both are not transferred to the genome in case of double homologous recombination. However, a single recombination event results in the integration of the whole suicide vector in the genome giving rise to merodiploids.

Table 2.4. Primers used for the construction of the *P. aeruginosa rpoC::TAP* strain. The restriction sites used for cloning are underlined. To ensure efficient restriction of the amplified fragments, extra nucleotides were added to the 5' end of the primers.

Primer	Sequence	Purpose
CTAP_F_Nco	AACCATGGTTCAGGCTGCCATTGATTAC	Amplification TAP cassette
CTAP_R_Bam	TAGGATCCAGCCAAGCTTCAGGTTGA	Amplification TAP cassette
PA4269_5F_Hin	TCAAGCTTGGCGACTCCAGCTTCATC	Amplification *rpoC* 5' flanking region
PA4269_5R_Nco	AACCATGGAGTTACCGCTCGAGTTCAG	Amplification *rpoC* 5' flanking region
PA4269_3R_Eco	CCGAATTCACCGATGTACGAGGAAAC	Amplification *rpoC* 3' flanking region
PA4269_3F_Bam	AAGGATCCGAGCGGTAACTAAGCATCAAC	Amplification *rpoC* 3' flanking region
PA4269_fl_F	GCGAACAGCGGATAGAGTTG	Verification
PA4269_fl_R	CTGTGCTCTTGCAGGTTGTGA	Verification
Gm_5F	GTGGGTCGATGTTTGATGTTATG	Verification
Gm_3R	CGGGATCGTCACCGTAATCT	Verification

The transformation mixtures were plated on LB + Gm^{30} selecting for insertion of the resistance expression cassette. A few colonies were patched on LB + Gm^{30} plates and LB + Cb^{200} plates to differentiate single from double cross-over events. Colonies only able to grow on LB + Gm^{30} but not on LB + Cb^{200} were selected. To ensure the strain originates from a single colony, this step was repeated once. As an addition control, the Gm^{R} colonies were struck for single colonies on LB + Gm^{30} plates containing 5% sucrose. Positive clones were confirmed by PCR clone analysis on *P. aeruginosa* cells using primers flanking the *rpoC* 3' gene region and the gentamicin resistance cassette (Table 2.4).

2.4.4.2 Purification of the RNA polymerase complex from phage-infected cells

To prepare phage-infected biomass, 500 ml of *P. aeruginosa rpoC::TAP* was grown to $OD_{600\ nm}$ of 0.3 and infected with phages at a multiplicity of infection of 10. Infection was halted 10-20 min post-infection by rapidly cooling the samples in icy water baths. Cells were harvested by centrifugation, washed once with ice-cold 10% (v/v) glycerol, and dissolved in 10 ml Protein A buffer [10 mM Tris.HCl pH 8, 150 mM NaCl, 0.1% (v/v) NP-40] supplemented with 100 µl Protease Inhibitor Mix (Amersham Biosciences), 10 mM DNase I (Sigma-Aldrich) 30 mM RNase A (Sigma-Aldrich) at room temperature for 15 min with gentle agitation. Cells were disrupted using a French press (Thermo Scientific), and the soluble fraction was isolated by two successive 10-min centrifugation steps at 27,200 x *g* at 4°C. Subsequently, this fraction was incubated with 200 µl pre-washed IgG Sepharose™ 6 Fast Flow beads (GE Healthcare) with gentle agitation for 2 h. at 4°C. The beads were collected in Bio-Rad Poly-Prep®

Chromatography column and washed three times with Protein A buffer. The beads containing the RNA polymerase complex were than incubated in ProTEV Protease (Promega) in the supplied buffer, and transferred to smaller Handee™ Spin Column (Pierce) for overnight incubation at 4°C. The eluates were collected, and the beads were washed with calmoduline binding buffer (CBB) [10 mM Tris.HCl pH 8, 150 mM NaCl, 0.1% (v/v) NP-40, 1 mM Mg acetate, 1 mM imidazole, 2 mM $CaCl_2$, 10 mM β-mercaptoethanol]. All eluates were pooled, and transferred to a new Bio-Rad Poly-Prep® column for incubation with pre-washed calmoduline-affinity beads (Stratagene, La Jolla, CA, USA) for 1 h. at 4°C. Finally, after three washing steps using CBB, the complex was eluted from the beads using calmoduline elution buffer [10 mM Tris.HCl pH 8, 150 mM NaCl, 0.1% (v/v) NP-40, 1 mM Mg acetate, 1 mM imidazole, 20 mM EGTA, 10 mM β-mercaptoethanol]. All fractions were analyzed by SDS-PAGE and subjected to mass spectrometric analysis as described in section 2.4.2.

Chapter 3

Isolation of new *P. aeruginosa* phages: *de novo* peptide sequencing as a novel tool to assess viral diversity [1]

As a consequence of the increased interest in phage research, numerous research groups established large phage collections using a wide variety of prokaryotes as phage host (*e.g.*, Stenholm et al., 2008; Carey-Smith et al., 2006; Jamalludeen et al., 2007). Newly isolated phages are typically characterized by restriction analysis, electron microscopy and host range analysis, classifying them into families according to the nature of the phage nucleic acid and their virion morphology (Section 1.1.2). These basic characterization techniques are also applied on a library of *P. aeruginosa* phages isolated from world-wide collected water samples, and are described in section 3.1.

Since genes defining structural/virion proteins are conserved between related phages (Proux et al., 2002), we argued that information about the primary sequence of structural phage proteins should be useful to pinpoint unsequenced phages to specific subfamilies or genera. In section 3.2, we present *de novo* sequence information of structural phage proteins gathered from an optimized MALDI post-source decay (PSD) analysis. This optimization resulted in easily interpretable MS/MS spectra and was already applied to proteomes of unsequenced bacteria (Samyn et al., 2006), nematodes (Leon et al., 2007) and primates (Tannu and Hamby, 2007). By applying this technique on (bacterial) viruses for the first time, 15 new isolates were clustered in seven groups of closely related phages.

3.1 Compilation of a *P. aeruginosa* phage library

3.1.1 Phage isolation and preliminary characterization

Between September 2004 and November 2006, a total of 28 *P. aeruginosa* phages were isolated from 398 water samples, worldwide collected from hospital sewage and natural environments in

[1] Data presented in this chapter are adapted from Ceyssens, P.J., *et al. (*2009) *Environmental Microbiology* **11**: 1303-1313.

Table 3.1. Overview of the isolation and preliminary characterization of phages described in this study.

Phage[a]	Isolation Details				Plaque	Genome	pH	Family	Morphotype[e]	Dimensions (nm)[f]	
	Water Source[b]	City/Country	Date	Host[c]	(mm)	(kb)[d]	stability			Head	Tail
LKA1	Dijle River	Leuven, Belgium	9-2004	PAO1	1	45	4-11	*Podoviridae*	T7-like	60	10
LKD16	Local pond	Holsbeek, Belgium	10-2004	PAO1	2-3	45	N.D.	*Podoviridae*	T7-like	65	12
LUZ24	Hospital sewage	Leuven, Belgium	10-2005	Li010	4-5	45	4-11	*Podoviridae*	T7-like	63	12
LUZ19	Hospital sewage	Leuven, Belgium	11-2004	PAO1	4-5	45	4-11	*Podoviridae*	T7-like	65	12
LUZ2	Hospital sewage	Leuven, Belgium	10-2004	PAO1	4-5	45	N.D.	*Podoviridae*	T7-like	63	12
LUZ14	Hospital sewage	Leuven, Belgium	11-2006	C1	1-2	45	4-10	*Podoviridae*	T7-like	62	12
LIT1	Local pond	Caligary, Italy	5-2006	Br257	2-3	65	N.D.	*Podoviridae*	Li2-like	74	30
LUZ7	Hospital sewage	Leuven, Belgium	11-2006	US449	3-4	75	4-11	*Podoviridae*	Li2-like	76	30
LAD2	Local pond	Adelade, Australia	9-2006	PAO1	2	65	N.D.	*Siphoviridae*	SD1-like	63	210
LPB1	Hospital sewage	Leuven, Belgium	6-2006	PAO1	1-2	45	N.D.	*Siphoviridae*	SD1-like	68	210
LUZ1	Hospital sewage	Leuven, Belgium	5-2006	PAO1	2	30	N.D.	*Siphoviridae*	SD1-like	67	210
LKR3	Local pond	Dubrovnik, Croatia	9-2006	LO053	1-2	50	4-12	*Siphoviridae*	Li-F10-like	63	120
LBL3	Local pond	Blanes, Spain	11-2006	Aa245	2	70	2-11	*Myoviridae*	PB1- like	73	148
LMA2	Maas river	Maastricht, Holland	6-2006	Bo559	2	70	N.D.	*Myoviridae*	PB1- like	74	138
LSL4	Sea water	Colombo,Sri Lanka	11-2006	Li012	3	65	N.D.	*Myoviridae*	Li24- like	75	100
LKA2*	Dijle river	Leuven, Belgium	9-2004	PAO1	1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LDA4*	Demer river	Aarschot, Belgium	10-2004	PAO1	1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LDB10*	Local pond	Heverlee, Belgium	10-2004	PAO1	2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LKD7*	Local pond	Holsbeek, Belgium	10-2004	PAO1	1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LIT2*	Local pond	Caligary, Italy	5-2006	Br257	1-2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LMX1*	Tap water	Mexico city, Mexico	6-2005	Br908	2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LTE1*	Local pond	Barrancho, Tenerife	7-2005	LO049	2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LKO1*	River source	Kos, Greece	7-2005	PAO1	1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LLO2*	Holy water	Lourdes, France	8-2005	PAO1	1-2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LSA1*	Local pond	South-Africa	9-2006	PAO1	1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LKR12*	Local pond	Dubrovnik, Croatia	9-2006	Li010	3	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LVA3*	Canal water	Leuven, Belgium	12-2005	US451	1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
LLE5*	River water	Lenne, Germany	9-2005	C	2-3	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

[a] Phages with an asterisk formed turbid plaques and were not selected for further characterization: no restriction analysis was performed on these phages, so there is also no indication whether these phages are all unique. [b] Hospital sewage samples taken at UZ Leuven, Belgium. [c] *P. aeruginosa* strain used for phage amplification. Clinical *P. aeruginosa* strains were named according to Pirnay and colleagues (2002). [d] Estimated from restriction fragments. [e] According to Figure 1.15 (Ackermann et al., 1988) [f] Dimensions of head diameter and tail length are approximations. N.D., Not Determined.

14 different countries (Table 3.1). During the multiple rounds of phage isolations, *P. aeruginosa* strains which were insensitive for all phages isolated at that time were specifically chosen as host for new phage enrichments. The observed ratio (28/398) is comparable to the isolation of phages infecting enterotoxigenic *E. coli* from soil and waste samples *(*Jamalludeen et al., 2007). There was a clearly positive correlation between the extent of pollution and/or organic waste in the water sample and the amount of *P. aeruginosa* phages present, an observation also seen for some bacterial surveys (Pirnay et al., 2005). As this research was initially focused on the isolation and characterization of potentially therapeutic phages, 15 phages which were able to form clear plaques on lawns of their bacterial host were selected for further analyses (Table 3.1, top 15 phages). Those phages were mutually distinguished by plaque morphology and restriction analysis, all of them showing unique restriction patterns which enabled us to estimate their dsDNA genome sizes between 30 and 75 kb (Figure 3.1C; Table 3.1).

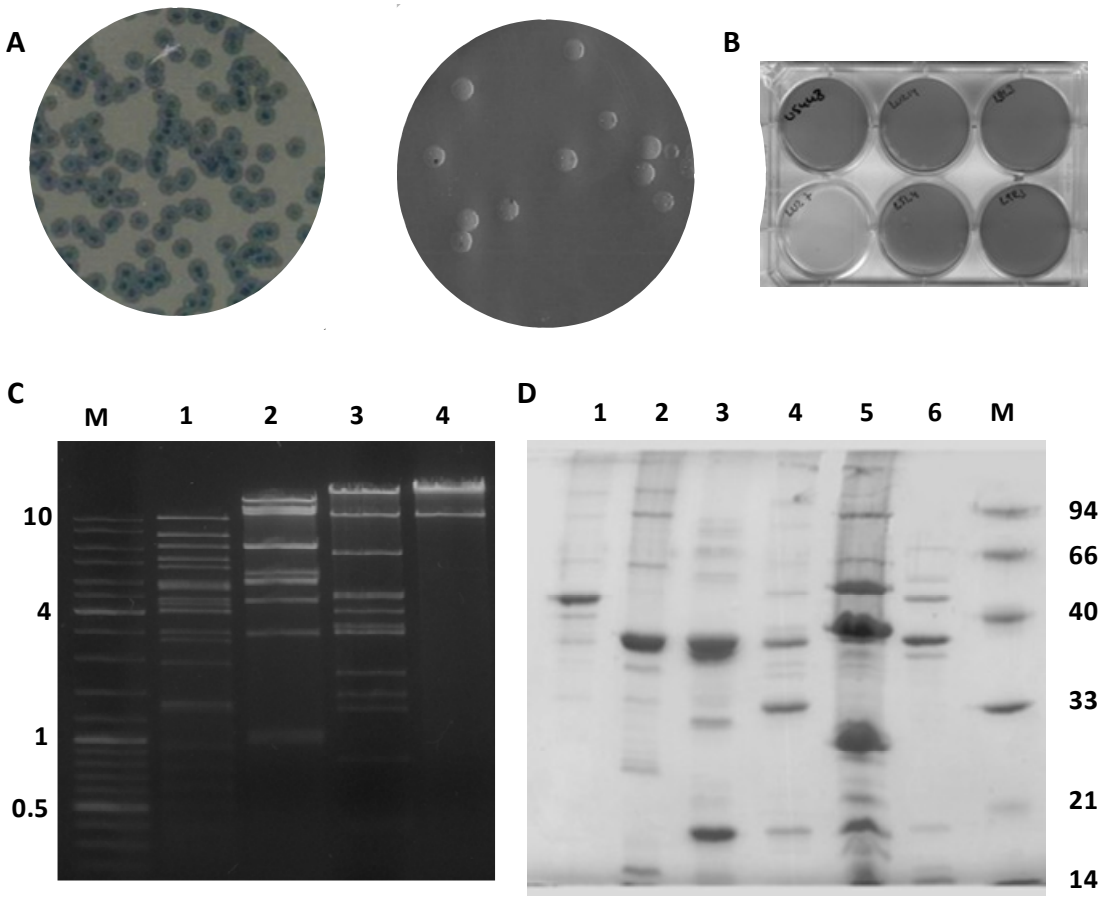


Figure 3.1. Illustration of preliminary phage characterization. A. Discrimination based on differences in plaque morphology between LUZ7 (left) and LKD16 (right). **B**. Host range analysis of five phages infecting *P. aeruginosa* US448, compared to a negative control. In this case, lysis was only observed with infection of LUZ7 (bottom left). **C.** Discrimination based on *Eco*RI restriction digest of phage genomes: lanes 1-4, phages LBL3, LIT1, LUZ7 and LSL4. M, marker (kb). **D.** Discrimitation based on 1D-SDS-PAGE profiles of structural phage proteins: Lanes 1-6: LUZ7, LUZ19, LAD2, LKR3, LBL3 and LSL4. Lane M, marker (kDa).

The ability of restriction enzymes to digest different phage genomes differed profoundly between phage groups, and will be discussed in more detail in later chapters. One-dimensional SDS-PAGE profiles of phage virion proteins clearly showed that the different morphotypes (section 3.1.2) display different protein patterns, with major virion (capsid) proteins varying in size between 35 and 45 kDa (Figure 3.1D). Phage particles from all virion morphologies were tested on their ability to remain infective after incubation at different pH levels; all phages were found to be stable between pH4 and pH11, except LUZ14 which lost its infectivity above pH10 (Table 3.1). Remarkably, 25% of the LBL3 particles still lysed their host after 1 hour of incubation at pH2 (Section 7.1), while 70% of LKR3 virions retained their infectivity at pH12.

3.1.2 Morphological characterization

Electron microscopic recordings of the 15 purified phage virions (performed by Hans-W. Ackermann at the Laval University of Quebec, Canada) revealed that all investigated phages are tailed, and have isometric heads. As summarized in Table 3.1, six different morphotypes can be distinguished in this collection. Six out of eight phages with a short tail (*Podoviridae*) are morphologically undistinguishable from phage T7, with a genome of 45 kb inside a head (diameter approx. 65 nm) and a tail of 12 x 8 nm (Table 3.1, Figure 3.2A). *Podoviridae* LUZ7 and LIT1 have slightly larger heads (diameter of 76 nm) incorporating a 70 kb genome, and very fine tails with an average length of 33 nm (Figure 3.2B). Both phages belong to the Li-2 morphotype (Figure 1.15), which also includes *Pseudomonas* phage 7 (Ackermann et al., 1988; Kwan et al., 2006).

Three phages (LUZ1, LPB1 and LAD2) are SD1-like *Siphoviridae* (Table 3.1), with a head size comparable to that of the "T7-like viruses", but a long and relatively rigid tail of approx. 210 x 10 nm decorated with terminal spikes (Figure 3.2C). These phages are morphologically identical to the widely-studied phage D3112 and have similar genome sizes (Wang et al., 2004). The fourth member of the *Siphoviridae* (LKR3, 50 kb) has a similar head size but a shorter (120 nm) and very flexible tail with terminal tail spikes (Figure 3.2D), typical for the Li-F10 morphotype (Ackermann et al., 1988).

Phages LBL3 and LMA2 belong to the PB-1 morphotype of the *Myoviridae*, carrying characteristic shallow capsomers which appear as cup-like depressions of about 8 nm on the phage heads (Figure 3.2E). Their heads are 73-74 nm in diameter and encapsulate 70 kb genomes, while their tails (148 x 17 nm) are contractile and rigid and carry tail fibers (40-50 nm) protruding from the baseplate. After contraction, the baseplate separates from the sheath (Figure 3.2F). The third myovirus, LSL4, has the same general morphology as the PB1-likes, but a slightly smaller genome (65 kb). It is distinguished

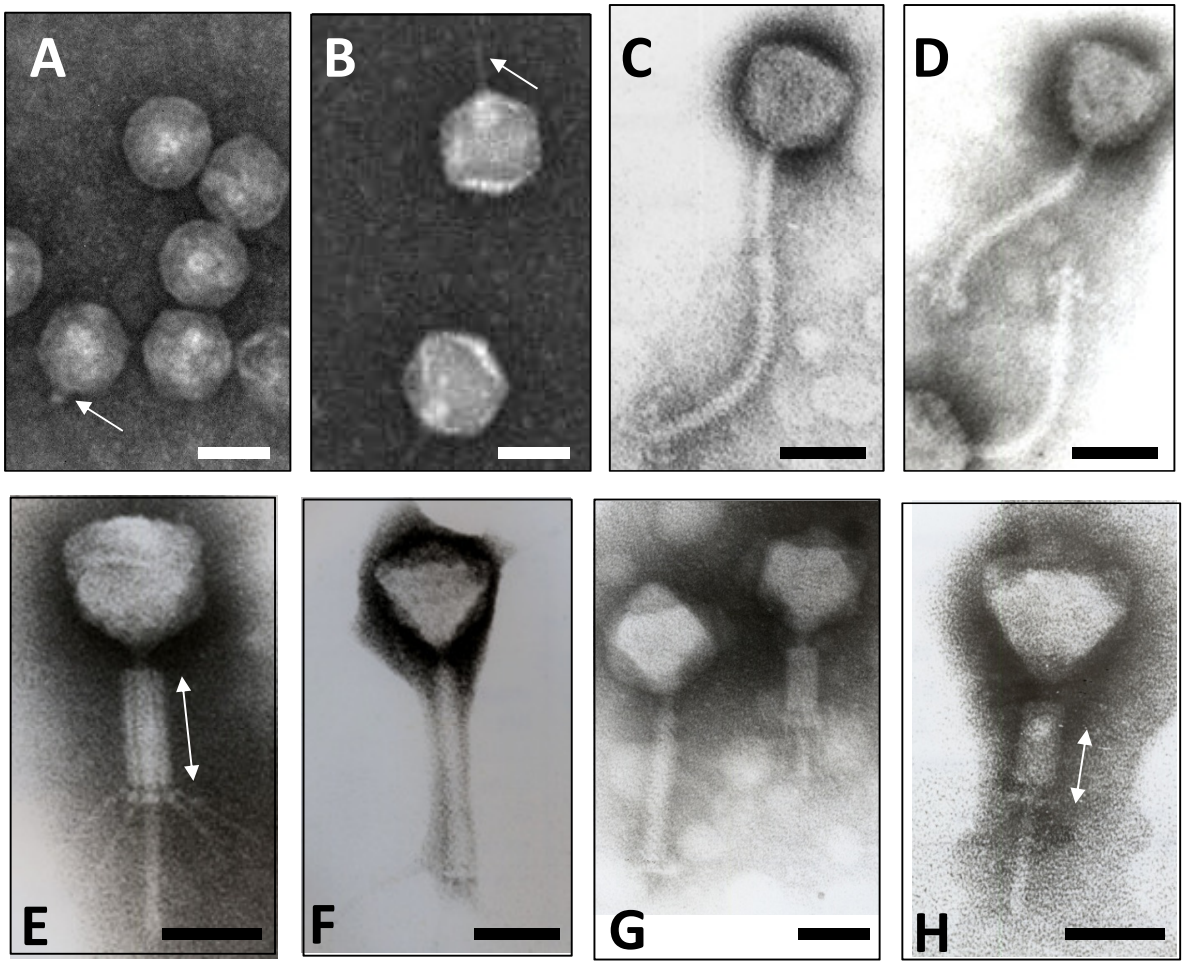


Figure 3.2. Electron microscopic images of negatively stained phage particles. Figs A-D: phages LKD16, LIT1, LUZ1, and LKR3, respectively. Figs. E and F, phage LMA2 with contracted (E) and extended (F) tail. Fig G and H, phage LSL4 with extended (G) and contracted (H) tail. Differences in tail morphology among *Podoviridae* and *Myoviridae* are indicated with white arrows. Scale bar represents 50 nm.

from PB1-like phages by the absence of visible capsomers and a shorter tail (126 nm, Figure 3.2G,F), which places it among the Li24 morphotype (Ackermann et al., 1988).

3.1.3 Host range analysis and receptor analysis

Each phage possesses a unique host range spectrum when tested on 122 AFLP-typed *P. aeruginosa* strains (Section 2.1), with the number of lysed strains ranging from 2% for LKA1 to 53% for LBL3 (Table 3.2). This variation was expected, given the vast diversity among both phage and *P. aeruginosa* isolates described above. All phages combined lyse 87% of the strains, while a mixture of LBL3, LUZ19 and LUZ24 lyses 72% of our *P. aeruginosa* collection (Table 3.3). However, no phage was able to lyse the predominant and multi-drug resistant clinical strains Br667 and Br735 (Pirnay et al., 2003). When comparing the host range data with AFLP typing results of the *P. aeruginosa* host bacteria, no

obvious correlation could be detected between AFLP clustering and infection patterns (data not shown). More detailed host spectra will be discussed in the following chapters.

Table 3.2. Host range and infection of *P. aeruginosa* PAO1 mutants.

Phage	Host range[a] (%)	wt PAO1	wt *pil*A Δ *fli*C	Δ *pil*A wt fliC	wt *alg*C Δ *fli*C[b]	Δ *alg*C Δ *fli*C[b]
LKD16	29	+	+	-	-	-
LUZ2	6	+	+	-	-	-
LUZ19	44	+	+	-	-	-
LKA1	2	+	+	+	+	-
LIT1	14	+	+	-	-	-
LUZ14	37	-	-	-	-	-
LUZ7	28	+	+	+	+	-
LUZ24	29	+	+	-	+	-
LUZ1	9	+	+	-	-	-
LPB1	10	+	+	-	-	-
LAD2	8	+	+	-	-	-
LKR3	16	+	+	-	-	-
LSL4	20	+	+	-	-	-
LMA2	22	-	-	-	-	-
LBL3	53	-	-	-	-	-

[a] Expressed as % lysis of a library of 122 clinical *P. aeruginosa* strains. [b] These mutants were derived from a *P. aeruginosa* PAO1 strain which is defective in *pil*A.

To obtain a more detailed understanding of the interaction between phages and host, infection tests were performed on *P. aeruginosa* PAO1 chromosomal *pil*A, *fli*C and *Alg*C mutants deficient in Type IV pili, flagella and AlgC production. Ten phages, including all isolated *Siphoviridae* and many of the *Podoviridae*, require Type IV pili for host infection (Table 3.2). Phages LKA1 and LUZ7 behave differently: their infection is dependent on the presence of the *Alg*C gene, strongly suggesting the involvement of the lipopolysaccharide (LPS) layer in phage adsorption. Phage LUZ24 is dependent on both *pil*A and *Alg*C, suggesting the use of secondary receptors. Three phages (LBL3, LMA2, and LUZ14) infected neither wild-type *P. aeruginosa* PAO1 nor the mutant strains, so no conclusions towards receptor use could be drawn for these phages.

3.2 Phage typing by *de novo* sequencing of structural proteins

In current phage research, mass spectrometry is mainly applied to experimentally identify structural proteins of sequenced phages (*e.g.*, Clokie et al., 2008; Labrie et al., 2008; Schwudke et al., 2008). Hereto, powerful search engines (*e.g.*, Mascot and Sequest) screen mass spectrometry data against primary sequence databases. A caveat of this database matching is that it can only be successful if the database being searched contains the identical amino acid sequence of the peptide being

Table 3.3. Host spectra of phage cocktails. Detailed overview of host ranges, expressed as % lysis of a library of 122 clinical *P. aeruginosa* strains, using a combination of two phages (right of the diagonal) or a mixture of LBL3 and two other phages (left of the diagonal, grey shading). From this table, it can (for example) be derived that all strains that can be infected by LKD16 are also susceptible to infection by phage LUZ19. The two combinations with the broadest host range are shaded in grey. Phages are grouped according to their family, which are separated by brokes lines.

	LKD16	**LUZ19**	**LKA1**	**LUZ2**	**LIT1**	**LUZ14**	**LUZ7**	**LUZ24**	**LUZ1**	**LPB1**	**LAD2**	**LKR3**	**LSL4**	**LMA2**	**LBL3**	
LKD16	**29**	44	30	29	38	48	46	49	32	32	33	41	43	42	61	
LUZ19	67	**44**	45	44	50	54	54	57	46	46	46	53	55	54	**67**	
LKA1	61	67	**2**	7	15	37	28	31	11	11	10	18	22	24	53	
LUZ2	61	67	54	**6**	20	39	31	32	12	13	11	20	23	24	54	
LIT1	63	67	55	56	**14**	44	36	37	21	20	21	27	28	28	55	
LUZ14	67	70	63	63	64	**37**	43	52	41	42	41	44	50	50	63	
LUZ7	67	71	59	60	62	66	**28**	48	33	33	33	37	45	43	59	
LUZ24	67	**72**	59	60	62	67	66	**29**	33	35	34	39	40	42	59	
LUZ1	62	67	54	55	57	63	61	61	**9**	15	14	21	24	26	54	
LPB1	63	68	56	57	58	65	63	63	57	**10**	15	24	26	28	56	
LAD2	63	68	55	56	58	65	62	62	57	58	**8**	20	24	28	55	
LKR3	65	71	58	59	60	66	63	64	59	61	59	**16**	28	30	58	
LSL4	64	70	56	57	59	65	63	62	58	59	59	59	**20**	28	56	
LMA2	64	70	57	57	59	65	63	63	59	60	59	59	59	**22**	57	
LBL3	**LBL 3 + two other phages** (% infection)														**53**	

Mixture of two phages (% infection)

investigated (Addona and Clauser, 2002), rendering it useless for proteomic analysis of novel and/or unsequenced organisms. Here we describe a novel phage typing method, gathering more detailed information about their evolutionary relatedness by *de novo* peptide sequencing of major virion proteins using MALDI-TOF/TOF technology.

3.2.1 Theoretical background

Mass spectrometry using matrix-assisted laser desorption/ionization (MALDI) is considered as a 'soft' peptide ionization technique which generates almost exclusively intact, mono-protonated ions used for protein identification (Kaufmann, 1995). In a reflector time-of-flight (TOF) mass spectrometer, these precursor ions are subject to metastable decay, and fall apart into different fragments which are separated by their difference in kinetic energy. In the past few years, many efforts have been undertaken to determine primary sequence information from this post-source decay (PSD) (Ackloo and Loboda, 2005). Interpretation of spectra generated by MALDI-PSD relies typically on measuring the mass differences between adjacent fragment ion peaks of one of the major ion (b or y-ion) series (Kaufmann et al., 1994). Unfortunately, some drawbacks impair the quality of regular PSD fragmentation spectra: peaks have often a weak intensity, different peptides show great variability in their ability to fragment and most importantly, two (or more) series of ions are present at the same time, complicating the interpretation of the spectra (Conrotto and Hellman, 2008).

This problem has been overcome by recent improvements in *de novo* peptide sequencing (Keough et al., 2003; Samyn et al., 2006). The interpretation of MS/MS spectra is greatly facilitated by the introduction of strongly charged groups at the C- or N-terminus (Keough et al., 1999; Marekov and Steinert, 2003), like the N-terminal derivatization with SPITC described in section 2.4.3. This sulfo group enhances fragmentation efficiency by lowering amide bound strength, and neutralizes the positive charge of the captured proton in the b_n fragment ions, rendering them neutral and thus undetectable by mass spectrometry. This allows *de novo* determined peptide sequences to be deduced manually. However, due to the reactivity of the sulfonic group, a preliminary step is required in which the ε-amino groups of lysines are blocked with O-methylisourea (Samyn et al., 2006).

3.2.2 *De novo* peptide sequencing

The most intense bands on 1D-SDS-PAGE profile gels (Figure 3.1C) were excised manually, subjected to derivatization and sequenced *de novo* using MALDI-TOF/TOF. The fragmentation yielded complete

y-ion series with lengths varying between 7 and 18 amino acids, which could be inferred by manual calculation of the differences between adjacent ions (Figure 3.3). These *de novo* peptides sequences were submitted to homology searches, showing that particle proteins of *Myoviridae* LBL3 and LMA2 are at least partially identical to major structural proteins of *Pseudomonas* phage F8. In contrast, identified peptides of the morphologically slightly distant phage LSL4 showed no significant similarity, suggesting the identification of a new subgroup of lytic phages (Table 3.4).

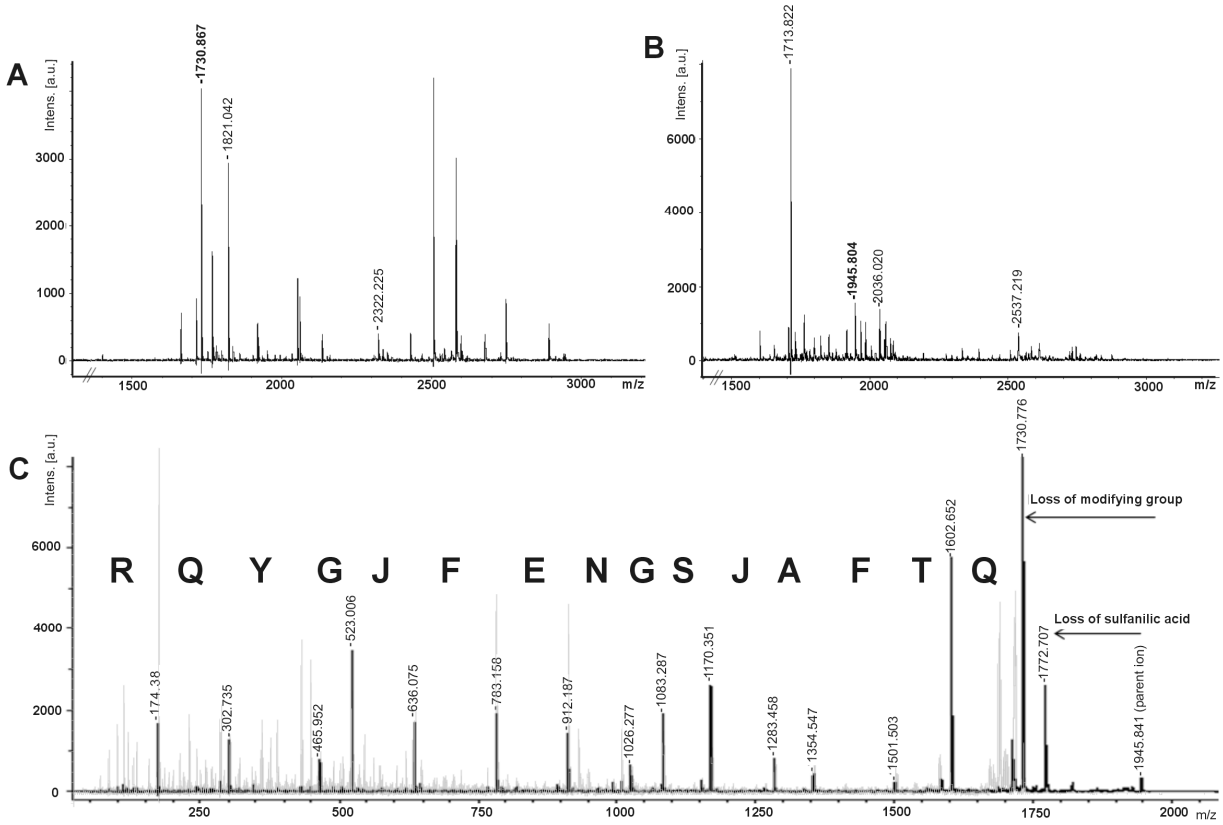


Figure 3.3. The *de novo* typing method, illustrated for phage LKR3. A. MALDI-TOF spectrum of tryptic peptides of phage LKR3. **B.** Spectrum after SPITC (215 Da) was used to introduce a fixed negative charge at the N-terminus of all peptides. The derivatized peptides were identified by a 215 Da increase in precursor ion mass. **C.** MALDI-TOF/TOF analysis of the parent ion with *m/z* 1945.841 (black peaks) compared to the MS/MS spectrum generated by fragmentation of its non-sulfonated counterpart (*m/z* 1730.867, grey peaks). Manual interpretation of this y-ion series is provided.

Analysis of the *Siphoviridae* confirmed the relationship between SD1-like phages isolated in this study (LAD2, LUZ1 and probably also LPB1) and phages D3112 (Wang *et al*., 2004), DMS3 (Budzik *et al*., 2004) and MP22 (Heo *et al*., 2007), which was suggested by electron microscopic recordings (section 3.1.2). Interestingly, virion proteins of phage LKR3 were nearly identical to gp8 and gp9 of the generalized transducing *Salmonella* phage ES18. This phage is related to lambdoid phages like coliphages HK97, λ and P22 (Casjens *et al*., 2005), but no related *Pseudomonas* phage has been described to date.

Table 3.4. *De novo* peptide sequencing of major virion proteins.

Phage (morphotype)	*de novo* peptide sequence[a]	Search algorithm[a]	Score/ e-value	Protein hit	Related bacteriophage
Myoviridae					
LBL3 (PB1-like)	JJSGVGAGAPVAGR	BLASTp	0.27	gp29	*Pseudomonas* phage F8
	GEJGMMVGTLEEGR	MS-Homology	58	gp23	*Pseudomonas* phage F8
	NAJGFYGWQSGIGNR	BLASTp	0.003	gp23	*Pseudomonas* phage F8
	QQAAJGJEJFR	MS-Homology	52	gp23	*Pseudomonas* phage F8
	NJAVJJDAZNR	BLASTp	3.6	gp30	*Pseudomonas* phage F8
LSL4 (Li24-like)	YTASDPDJJR			No identification	
	VPHFTYDDYJTPR			No identification	
	VEJFTDTPAPR			No identification	
	HAATMEFAR			No identification	
LMA2 (PB1-like)	JJSGVGAQAPVAG	BLASTp	0.069	gp29	*Pseudomonas* phage F8
	QQAAJGJQJFR	MS-Homology	49	gp23	*Pseudomonas* phage F8
	GEJGMMVGTJEEGR	BLASTp	0.0005	gp23	*Pseudomonas* phage F8
	NJAVJJDANR	BLASTp	0.039	gp30	*Pseudomonas* phage F8
Siphoviridae					
LKR3 (LiF10-like)	QTFAJSGNEFJGYQR	BLASTp	0.0002	gp9	*Enterobacteria* phage ES18
	DVVTPJVGEATGVVPJPR	BLASTp	0.0003	gp9	*Enterobacteria* phage ES18
	JYVJDC*AHHEGJNJR	BLASTp	0.028	gp8	*Enterobacteria* phage ES18
LAD2 (SD1-like)	QJDAEAAR	BLASTp	1.4	gp41	*Pseudomonas* phage DMS3
	JNGNVDE	BLASTp	4.1	orf45e	*Pseudomonas* phage MP22
	QVJ..TAPQPTJANR	BLASTp	1.8	gp39	*Pseudomonas* phage DMS3
LUZ1 (SD1-like)	YEYAATR	BLASTp	1.4	gp39	*Pseudomonas* phage DMS3
Podoviridae					
LUZ19 (T7-like)	AATVVAWSJR	BLASTp	0.012	gp30	*Pseudomonas* phage ϕKMV
	GSVDJYTHVQR	BLASTp	0.0006	gp30	*Pseudomonas* phage ϕKMV
	WNQSTJVGR	BLASTp	0.027	gp37	*Pseudomonas* phage ϕKMV
LUZ2 (T7-like)	VVESFJDR	BLASTp	2.0	gp32	*Pseudomonas* phage ϕKMV
	FATEAJAAH	BLASTp	0.002	gp32	*Pseudomonas* phage ϕKMV
LKD16 (T7-like)	AJSAHPJGR	MS-Homology	43	gp32	*Pseudomonas* phage ϕKMV
LUZ14 (T7-like)	NEJAYQJ	MSBLAST	193	major head	*Pseudomonas* phage PaP3
	AJTHEWQTD	BLASTp	0.12	major head	*Pseudomonas* phage PaP3
	QJJVEYTJR	BLASTp	0.38	major head	*Pseudomonas* phage PaP3
	FQNAJJVR			No identification	
LIT1 (Li2-like)	JHYJPJJDNR	BLASTp	0.6	Major coat	*Enterobacteria* phage N4
	MPTJTEJNR			No identification	
LUZ7 (Li2-like)	AMPTJTEVNR	BLASPp	3.6	gp36	*Enterobacteria* phage N4

[a] Residue J can be either Ile or Leu. Points in the sequence refer to gaps or ambiguous peaks in the spectra. The asterisk indicates the alkylation of Cys residus (+54 Da). [b] BLASTp parameters were adjusted for short input sequences, and similarity searches were limited to bacteriophages.

Within the *Podoviridae*, *de novo* sequencing identified three different subgroups. Phages LUZ19, LUZ2 and LKD16 belong to the ϕKMV-like phages, a group of highly virulent phages specific to *P. aeruginosa* (Lavigne et al., 2003), while LUZ14 and LUZ24 cluster with *Pseudomonas* phage PaP3 in a separate group (Tan *et al*., 2007). Most interestingly, phages LIT1 and LUZ7, which contain genomes of approx. 70 kb and have the Li-2 morphotype, seem to be distantly related to Enterobacteriophage N4. The type phage N4 is unique among dsDNA bacteriophages in that it packages a phage-encoded RNA polymerase into the capsid and injects the protein into the host cell upon infection (Choi *et al*., 2008). For a very long time, no N4-like phages have been described, although the genus of the "N4-like viruses" was already been recognized in the VIIIth ICTV report due to its unique characteristics (Hendrix and Casjens, 2005). All the relationships which are suggested in this analysis were confirmed later by partial (LKR3 and LSL4) or whole-genome sequencing, and are discussed in more detail in following chapters.

3.3 Discussion

This chapter describes the isolation of a set of 28 *P. aeruginosa* phages sampled in 14 different countries traversing four continents, from both natural environments and hospital sewage. This set was narrowed to 15 phages displaying clear-plaque morphology. This phage selection was subjected to preliminary characterization, all proving to be unique by differing clearly in plaque morphology, host range, restriction patterns and/or particle morphology. Based on virion morphology and the nature of the phage nucleic acid (dsDNA), they can be grouped into the three main families of tailed dsDNA phages: 8 *Podoviridae* (LKA1, LUZ2, LUZ7, LUZ14, LUZ19, LUZ24, LIT1, LKD16), 4 *Siphoviridae* (LAD2, LPB1, LUZ1, LKR3) and 3 *Myoviridae* (LBL3, LMA2, LSL4). Infection analysis of outer membrane mutants identified no less that ten of these phages as Type IV pili-dependent. This pili-dependence was also reported for many other *Pseudomonas* phages like ϕKMV, D3112, B3 and DMS3 (Roncero et al., 1990; Budzik et al., 2004; Chibeu et al., 2009).

Although a fairly broad spectrum (87%) of 122 clinical *P. aeruginosa* strains could be lysed by at least one phage, it is noteworthy that predominant and multi-drug resistant clinical strains like Br667 and Br735 (Pirnay et al., 2003) are not susceptible to any phage in the collection. This might suggest that co-evolution of phage and bacteria is disturbed in Intensive Care Units (ICU), leading to an imbalance between phages and host in this niche which, in turn, could allow epidemic strains to emerge more quickly. A link between epidemic outbreaks of cholera and prevalence of lytic *Vibrio cholerae* phages

was reported (Faruque et al., 2005; Jensen et al., 2006). However, this has to be examined further for outbreaks of *P. aeruginosa*.

To unravel more detailed relationships between the isolated phages, comparative digitized DNA fingerprinting using fRFLP can be used as first indication (Merabishvili *et al*., 2007), but classification in subfamilies and genera and elucidation of underlying evolutionary mechanisms can only be revealed by genome sequencing (Lawrence et al., 2002; Leplae et al., 2004; Lavigne et al., 2008). However, this is clearly not a standard practice, as genome sequences of only less than 10% of all isolated phages are available in public databases.

We implemented a new typing method based on recent advances in the field of proteomics, and were able to classify our phage collection into seven specific groupings without the need of prior DNA sequencing and/or EM studies. This clustering correlates to and even refines EM results (which only identified six morphotypes), and was confirmed by partial or whole-genome sequencing of unknown phages. As shown in section 3.1.2, EM studies also have high discriminatory power, but expertise is absolutely necessary for a correct interpretation of the recordings. For example, differences in head and tail morphology of the ϕKMV-like and LUZ24-like *Podoviridae* are very subtle (and both are identified as having T7-like morphotype, Table 3.1), while *de novo* sequencing allowed straightforward discrimination between these phage groups. As such, the described procedure can be seen as an additional tool for first-hand characterization and tentatively classification of unsequenced phages carrying protein similarity with phages in public databases. However, and in contrast to EM, it has to be stressed that this approach cannot be applied to uncharacterized phages with structural proteins that lack any similarity to database entries.

Chapter 4

Pseudomonas phage ϕKMV represents a conserved genus of widely spread virulent phages [2]

Six years ago, the genome of the virulent *P. aeruginosa* phage ϕKMV was published by Lavigne and colleagues (2003). Although this phage resembles the well-studied Podovirus T7 in overall genome architecture, it was the first known T7-like phage which encoded a single-subunit RNA polymerase gene downstream its DNA metabolism genes instead of in the early genomic region. This enzyme binds to phage-specific promoters and has an important role in both transcription and DNA replication, so this rearrangement implies profound differences in infection mechanism. Since then, several phages infecting *P. aeruginosa* (PT2, PT5), *Vibrio parahaemolyticus* (VP93) and *Ralstonia solanacearum* (RSB1) have been reported to be genetically related to ϕKMV (Kawasaki et al., 2009). In our phage collection, phages LKD16, LUZ19, LUZ2 and LKA1 were found to be relatives of ϕKMV. *Pseudomonas* phage ϕKF77, isolated in Russia in 1982, was added to this selection through research contacts with Dr. Leonid Kulakov (University of Belfast, Ireland). In this chapter, we present an in-depth characterization of these phages and a comparison of their genome structure to other *Podoviridae*.

4.1 Phage Characteristics

From the very first preliminary tests, it was clear that phage LKA1 behaved differently than the other three phages. While phages LUZ19, LKD16 and LUZ2 produced large clear plaques (diameter 3-5 mm) after infection of *P. aeruginosa* PAO1, LKA1 yielded small plaques (1 mm) surrounded by a halo. Both groups of phages also use different primary receptors, as observed during infection assays of outer membrane mutants of *P. aeruginosa*: while LKA1 requires LPS for phage adsorption, LUZ19, LKD16 and LUZ2 are clearly Type IV pili dependent (Table 3.2). Electron microscopic imaging of purified phage particles revealed these phages as typical members of the *Podoviridae*, with a head diameter

[2] Data presented in this chapter are partially derived from Ceyssens, P.J., *et al.* (2006) *Journal of Bacteriology* **19**: 6924-6931, and Kulakov, L., et al. (2009) *Virology* **391**: 1-4.

of approximately 60 nm and a stubby tail with a length of 8-10 nm (Figure 4.1). The characteristic portal vertexes are visible on the images of all phages, but possible tail fibers can only be distinguished in case of LKA1.

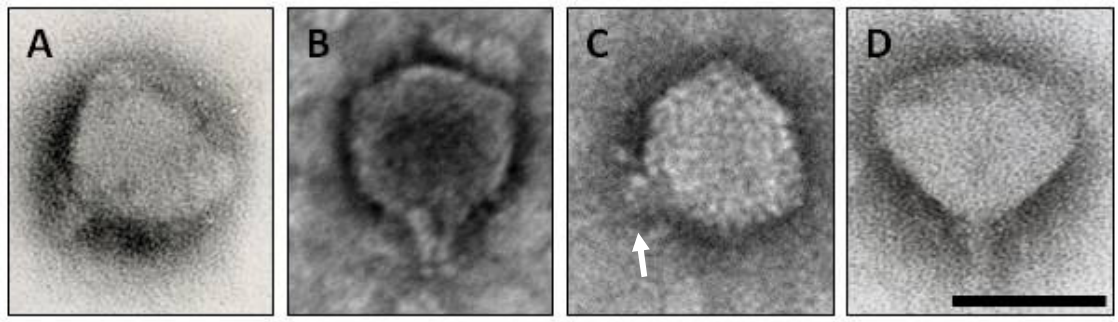


Figure 4.1. Electron micrographs of negatively stained LUZ19 (A), LKD16 (B), LKA1 (C) and LUZ2 (D) particles. Tail structures of LKA1 are indicated with an arrow. The scale bar represents 50 nm.

Host range studies on our *P. aeruginosa* library further aided mutual phage discrimination. LKD16 and ϕKMV show only minor differences in host spectrum, since 24% of the strains were infected by both phages while 2% and 5% were exclusively lysed by ϕKMV and LKD16, respectively. Phages LUZ2 and LKA1 display extremely narrow spectra, lysing only 6 and 1 strain besides *P. aeruginosa* PAO1, respectively. In contrast, LUZ19 is able to lyse a fairly broad spectrum of 44% of these strains, among which every strain sensitive for phage LKD16 (Table 3.3). The fact that LUZ19 is able to lyse more *P. aeruginosa* strains might have to do with the use of a different (more common) secondary receptor, or with post-infectious, intracellular reactions between phage and bacteria.

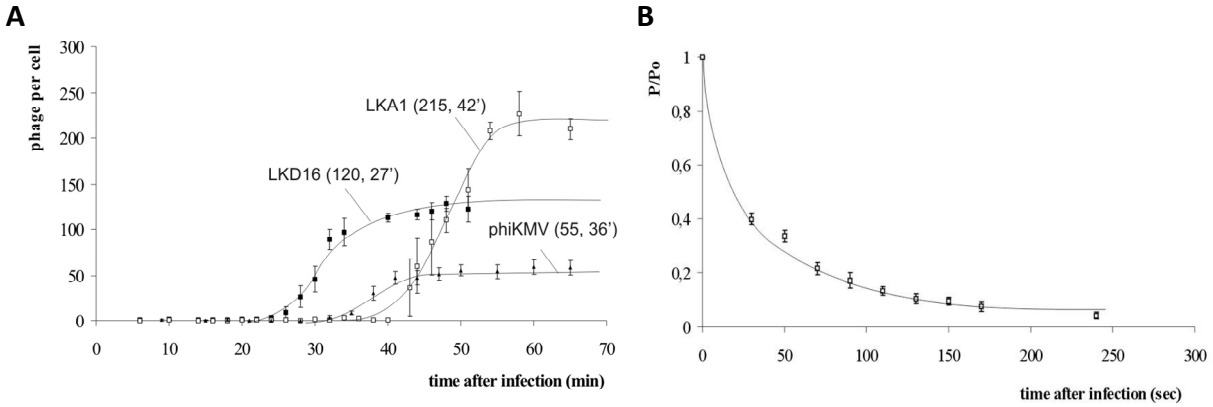


Figure 4.2. Infection analysis of ϕKMV-like viruses. A. One-step growth curves of phage LKD16, ϕKMV and LKA1, with indication of estimated burst size and latent period. **B.** Adsorption assay of phage LKA1 on *P. aeruginosa* PAO1 cells, expressed as the ratio of non-adsorbed phages (P) to the initial titer (P_0). All data are collected from at least three independent experiments, standard deviations are indicated.

The infection process towards *P. aeruginosa* cells was investigated by standard adsorption and one-step growth assays of LKD16 and LKA1, and compared to ϕKMV. Under laboratory conditions, the

infectivity properties of these phages differ considerably. Upon infection with LKD16, ϕKMV and LKA1, host cells lyse after 25, 35 and 40 minutes, respectively releasing approximately 120, 55 and 225 new phage particles per cell (Figure 4.2A). Adsorption assays show that LKA1 particles adsorb efficiently to *P. aeruginosa* cells (k_a = 3.9 10^{-9} ml min^{-1}), resulting in a 3-log reduction in non-adsorbed phage particles 4 minutes after infection (Figure 4.2B). This is in sharp contrast with LKD16 and ϕKMV which both failed to produce a clear adsorption curve, showing only minor adsorption to host cells under standard laboratory conditions (20-30% at MOI 0.001). This suggests a weak or an easily reversible attachment of both LKD16 and ϕKMV to the outer membrane of *P. aeruginosa*, as observed previously with phage phiS1 (Kelln and Warren, 1971). Under the same conditions, phage LUZ19 absorbs very efficiently to *P. aeruginosa* cells, and lyses its host after 30 minutes releasing 70 particles (E. Lecoutere, personal communication).

4.2 Genome analysis

The genomes of LKD16, LKA1, LUZ19 and ϕKF77 (L. Kulakov) were completely sequenced and annotated. In this section, the sequencing (4.2.1), annotation (4.2.2), transcriptional regulation (4.2.3) and genome architecture (4.2.4) of these phages are described.

4.2.1. Sanger sequencing identifies single-stranded nicks

While LKD16 and LKA1 were sequenced using a combination of shotgun sequencing and primer walking, the genomes of bacteriophages LUZ19 and ϕKF77 were sequenced solely using direct sequencing on genomic phage DNA (section 2.3.2.2). As summarized in Table 4.1, these four phages encapsulate linear, dsDNA genomes between 41.5 and 43.5 kb in length which are delineated by long Direct Terminal Repeats (DTRs), comparable in length to those of ϕKMV.

Table 4.1. Genome properties of the four ϕKMV-like viruses sequenced in this study. The genome of ϕKMV itself was described by Lavigne and colleagues (2003).

Phage	Genome (bp)	# ORFs	DTR length (bp)	GC (%)	# tRNAs	ϕKMV identity (%)[a]	Accession number
LKD16	43,200	54	428	62.6	-	83	NC_009935
LKA1	41,593	56	298	60.9	-	-	NC_009936
LUZ19	43,548	54	472	62.2	-	89	NC_010326
ϕKF77	43,152	53	454	62.8	-	90	NC_012418
ϕKMV	*42,519*	*52*	*414*	*62.3*	-	*100*	*NC_005045*

[a] The mutual DNA identity was calculated with Stretcher (EMBOSS package), using the default parameters.

In comparison to other *Pseudomonas* phages like ϕKZ (37%) and D3 (50%), the average G+C content of the ϕKMV-like phage genomes is quite high and comparable to the 66.6% of the host genome (Stover et al., 2000), suggesting a long history of interaction with its host.

At four specific points in the sequence (designed as α, β, γ and δ), Sanger sequencing of the non-coding strand of LUZ19 and ϕKF77 led to a full stop of the DNA polymerase (Figure 4.3A). While variation of reaction conditions had no effect, pretreatment of the phage DNA preparations with T4 DNA ligase was sufficient to completely restore sequencing reactions (Figure 4.3B).

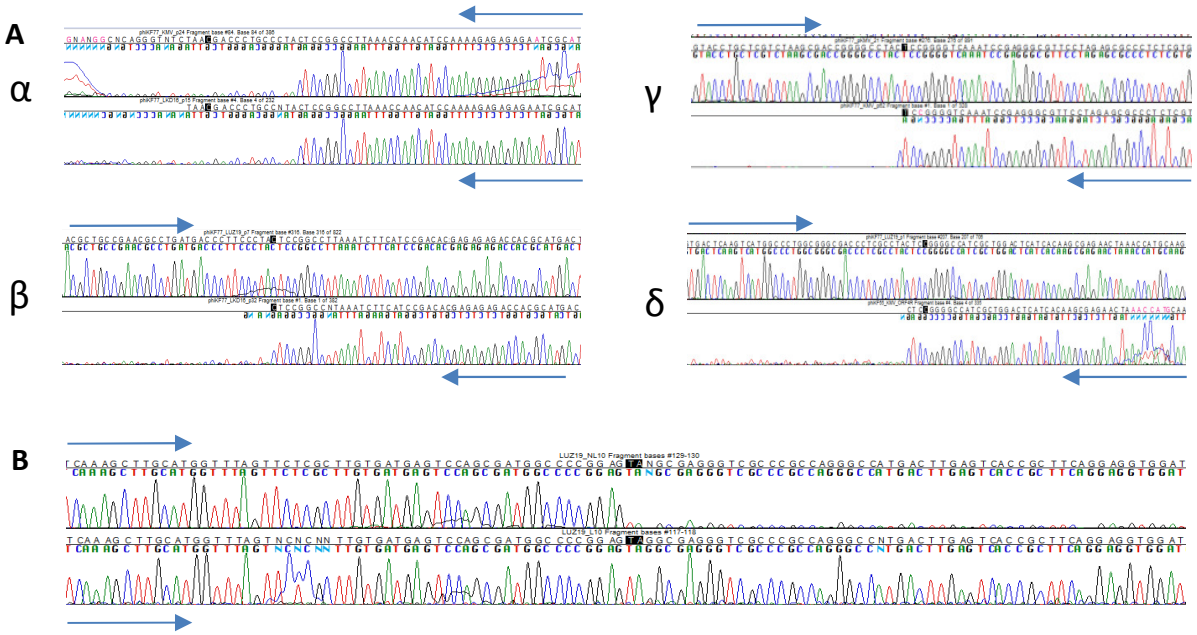


Figure 4.3. Full stop of Sanger sequencing reactions at distinct genomic locations. The arrows indicate the direction of sequencing. **A.** The DNA polymerase comes to a full stop at the four distinct locations on the minus (non-coding) strand. For nick α, no primer was available on the coding strand. **B.** Treatment of phage DNA with T4 DNA Ligase (lower chromatogram) is sufficient to restore DNA polymerase activity.

These observations indicated the presence of localized single-strand interruptions (nicks) in the non-coding strands of these phages, a distinguishing genomic trait previously thought to be unique for T5 related coliphages (Wang et al., 2005). The presence of these nicks was confirmed using alkali-denaturated DNA gels and visualization by electron microscopic recordings of ϕKF77 genomic DNA (Kulakov et al., 2009). Closer inspection of the sequences surrounding these nicks revealed that their 5'-ends have an identical six nucleotide sequence, -$_P$CTCCGG, while the common 3'-end sequence of each was determined as – CCTA$_{OH}$. Analysis of the genome sequences of ϕKMV and LKD16 showed that these bacteriophages also have the same 10 nucleotide sequence (5'-CCTACTCCGG-3') repeated several times throughout their genome. Sequences of two other ϕKMV-like viruses, PT2 (NC_011107) and PT5 (NC_011105), became available recently; although we were not able to analyze these

Table 4.2. Identification of the consensus sequences associated with nicks in genomes of the ϕKMV-like viruses. The position of the nick is indicated with an asterisk.

Nick	Phage	Location	Sequence
α	ϕKMV	24130-24146	CGACCCTGCCCTA*CTCCGGCCTTAAACT
	LUZ19	24767-24783	CGACCCTGCCCTA*CTCCGGCCTTAAACC
	LKD16	24723-24739	CGACCCTGCCCTA*CTCCGGCCTTAAACC
	ϕKF77	24652-24668	CGACCCTGCCCTA*CTCCGGCCTTAAACC
	PT2	24493-24509	CGACCCTGCCCTA*CTCCGGCCTTAAACT
	PT5	24495-24511	CGACCCTGCCCTA*CTCCGGCCTTAAACT
β	ϕKMV	13988-14004	CGACCCTTCCCTA*CTCCGGCCTTAAATC
	LUZ19	14415-14431	CGACCCTTCCCTA*CTCCGGCCTTAAATC
	LKD16	14568-14584	CGACCCTTCCCTA*CTCCGGCCTTAAATC
	ϕKF77	14320-14336	TGACCCTTCCCTA*CTCCGGCCTTAAATC
	PT2	14342-14358	CGACCCTTCCCTA*CTCCGGCCTTAAATC
	PT5	14353-14369	CGACCCTTCCCTA*CTCCGGCCTTAAATC
γ	ϕKMV	6316-6332	CGACTGGGGCCTA*CTCCGGGGTCAAATC
	LUZ19	6755-6771	CGACCAGGACCTA*CTCCGGGGTCAAATC
	LKD16	6658-6674	CGACCAGGGCCTA*CTCCGGGGTCAAATC
	ϕKF77	6379-6395	CGACCGGGGCCTA*CTCCGGGGTCAAATC
	PT2	6676-6692	CGACTGGGGCCTA*CTCCGGGGTCAAATC
	PT5	6681-6697	CGACTGGGGCCTA*CTCCGGGGTCAAATC
δ	ϕKMV	deletion	----------
	LUZ19	3132-3148	CGACCCTCGCCTA*CTCCGGGGCCATCGC
	LKD16	3259-3275	CGACCATCGCCTA*CTCCGGGGCCATCGC
	ϕKF77	2804-2820	CGACCCTCGCCTA*CTCCGGGGCCATCGC
	PT2	3112-3127	CGACCATCGCCTA*CTCCGGGGCCATCGC
	PT5	3112-3127	CGACCCTCGCCTA*CTCCGGGGCCATCGC

phages experimentally, the four putative canonical nick regions can be predicted in their genomes (Table 4.2). ClustalW-analysis of all nick regions identified an extra conserved stretch (GAC) at position -12 to -10, expanding the core sequence to 5' – GACxxxxxCCTA$_{OH\ P}$CTCCGG – 3'. The genome of ϕKMV lacks nick δ, showing that at least one of the nicks can be lost without noticeable loss of infectivity. The putative function of these nicks is discussed in section 4.4.1.

4.2.2 Comparative annotation of the ϕKMV-like genomes

Very strong sequence conservation between LKD16, LUZ19, ϕKF77 and ϕKMV is present throughout their entire genomes with exception of the early region and the tail fiber genes (Figure 4.4). These strong DNA similarities provided an additional tool in sequence analysis, resulting in more thoughtful ORF predictions in all phage genomes and a reannotation of the ϕKMV genome. Especially small putative ORFs (<150 bp) which previously fell below the threshold for gene annotation (*e.g.,* ORFs 13.1 and 25.1) and ORFs with partially overlapping sequences (*e.g.,* ORFs 12.1 and 46.1) are now

designated as putative ORFs based on their conserved presence in all phages. These small ORFs are usually ignored in the annotation of phage genomes and might be overlooked in later stages of phage research. The encoded gene products, however, may have important biological functions. A good example of such an active polypeptide is the phage T4-encoded Stp protein, which is only 26 amino acids long and activates the host PrrC lysyl-tRNA endonuclease (Penner et al., 1995). In addition, the 5' end of ORF 41 was extended with 198 bp. For reasons of consistency, the gene numbering of ϕKMV was maintained for these phages, and newly defined or phage-unique ORFs were numbered as the preceding ϕKMV-like ORF extended with the suffix 1.

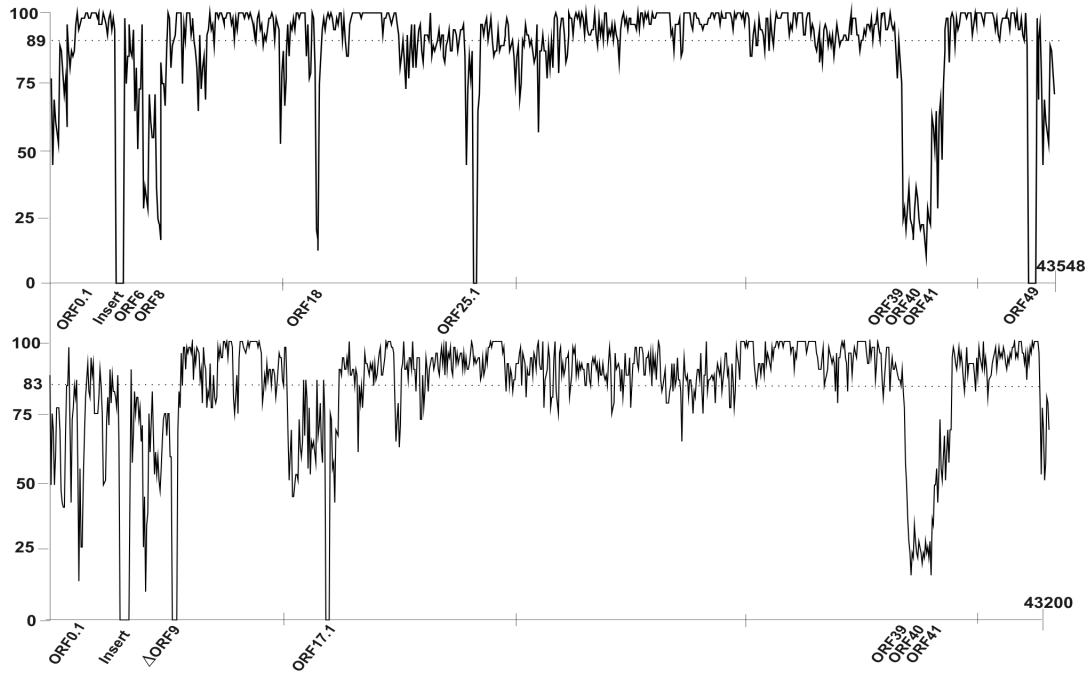


Figure 4.4. Pair-wise comparison of LUZ19 (top) and LKD16 (bottom) with ϕKMV at the nucleotide level, considering a sliding window of 60 bp. The average DNA identity is indicated by an intermitted line. ORFs displaying variation to the corresponding ϕKMV ORFs are marked. Mutual inserts/deletions (Δ) are shown.

Despite the major similarities at the genomic level, a limited number of discrepancies do exist between LKD16, LUZ19, ϕKF77 and ϕKMV. In all three genomes, a 0.5 kb-insert is present between ORFs 3 and 4, which is absent in ϕKMV. Two additional ORFs are defined in LKD16 (ORFs 0.1 and 17.1) and ORF9 is absent. In LUZ19, two additional insertion of approximately 0.3 kb are located downstream ORFs 25 and 48, encoding an extra ORF (ORF49) at the right end of the genome. In the early region of LUZ19, two major substitutions involving ORF6 and ORF8 were identified, but no putative function could be assigned to these new genes. However, the gene substituting ORF8 (ORF7.1) probably encodes a functional equivalent protein, as it contains two transmembrane domains like its counterpart in ϕKMV (Table 4.3).

In contrast, no identity at the nucleotide level could be detected between LKA1 and ϕKMV, and ORFs of LKA1 were given a unique gene number. Although only 27 gene products of LKA1 (48%) show protein similarity with ϕKMV counterparts, genome organization and gene order of ϕKMV is also largely conserved in LKA1 (Section 4.2.4). In total, between 52 and 56 potential ORFs were annotated in the ϕKMV-like viruses, which are summarized in Table 4.3.

Table 4.3. Annotation overview of the ϕKMV-like viruses described in this study.

ORF[a]	# AA	MW (kDa)	Absent in	Closest phage homologues beyond the ϕKMV genus [b] / Conserved domains – structural predictions	e-value [c]
0.1	42	5.9	All but LKD16 & LUZ19	none	
1	98	10.9	LKA1	none	
2	75	8.8	ϕKF77, LKA1	none	
3	179	19.5		Conserved hypothetical phage protein [PHA01782] hypothetical protein PaP3p70 [Pseudomonas phage PaP3]	9,00E-73 1,00E-24
3.1	34	4.5	All but ϕKF77	none	
3.2	41	5.5	All but ϕKF77	none	
4	123	13.1	LKA1	none	
5	76	8.4	LKA1	none	
5.1	78	8.2	All but LUZ19	none	
6	60	6.5	LUZ19, LKA1	none	
7	98	9.6	LKA1	none	
7.1	76	8.7	All but LUZ19	2 Transmembrane domains; Signal Peptide (P=0.67)	
8	83	8.8	LUZ19, LKA1	2 Transmembrane domains; Signal Peptide (P=0.83)	
9	98	11.1	LKA1	none	
10	139	14.2	LKA1	none	
11	120	13.3	LKA1	C-Terminal transmembrane domains	
12	270	27.5	LKA1	N-terminus related to binding domain of DNA polymerase A [cd06444] C-Terminal transmembrane domain	2,00E-04
12.1	140	15.7	LKA1	none	
13	181	20.3	LKA1	Predicted acetyltransferase [COG3153]	6,00E-03
13.1	45	4.4	LKA1, ϕKF77, ϕKMV	Transmembrane domain	
14	271	31.0		putative DnaG-like primase [PHA02031] Putative primase [Vibrio phage VP93]	1,00E-146 7,00E-18
15	398	44.5		putative DNA helicase [Vibrio phage VP93] DnaB Helicase, C-terminal domain [cd00984]	1,00E-70 2,00E-06
16	199	22.3	LKA1	Conserved hypothetical phage protein [PHA01806] hypothetical protein gh-1p13 [Pseudomonas phage gh-1]	2,00E-83 1,00E-03
17	314	34.3	LKA1	DNA Ligase A [pfam01068] DNA ligase [Vibriophage VP4]	2,00E-13 3,00E-24
17.1	87	10.0	ϕKMV, LUZ19	none	
18	114	13.5	LKA1	none	
19	808	91.9		putative DNA polymerase [Vibrio phage VP93] DNA Polymerase A [cd06444]	0 2,00E-20
20	103	11.2	LKA1	none	
21	351	37.1		Conserved hypothetical phage protein [PHA02030] hypothetical protein VPP93_gp17 [Vibrio phage VP93]	9,00E-135 4,00E-34
22	313	35.4		putative exonuclease [Vibrio phage VP93] 5'3' Exonuclease C [cd00008]	6,00E-61 5,00E-03
23	146	16.4		Endonuclease 7 [pfam02945] putative endonuclease [Vibrio phage VP93]	5,00E-10 7,00E-24
24	348	40.4		cd06126, DEDDy, DEDDy exonucleases hypothetical protein RSB1_gp22 [Ralstonia phage RSB1]	3,00E-04 2,00E-78
25	122	13.8		hypothetical protein PaP3p30 [Pseudomonas phage PaP3]	3,00E-03
25.1	48	5.5	ϕKMV	none	

Table 4.3. Continued

26	818	91.9		T3/T7-like RNA polymerase [PHA00452] putative RNA polymerase [Vibrio phage VP93]	0 3,00E-93
27	84	9.5		none	
28	157	18.0		Conserved hypothetical phage protein [PHA01807] hypothetical protein gh-1p24 [Pseudomonas phage gh-1]	4,00E-67 2,00E-07
29	98	10.5		Conserved structural phage protein [PHA01808]	1,00E-24
30	510	56.2		putative head-tail connector protein [Vibrio phage VP93]	2,00E-83
31	327	33.9		Conserved hypothetical phage protein [PHA01929] putative scaffolding protein [Vibrio phage VP93]	2,00E-81 3,00E-04
32	335	37.8		Conserved structural phage protein [PHA02004] Putative capsid protein [Vibrio phage VP93]	1,00E-137 3,00E-59
33	184	21.2		Conserved structural phage protein [PHA00428] Tail tuber protein A [Ralstonia phage RSB1]	8,00E-68 2,00E-09
34	826	91.8		Tubular tail protein B [Kluyvera phage Kvp1]	9,00E-41
35	181	18.8		Conserved hypothetical phage protein [PHA01547]	4,00E-60
36	898	98.2		Bacteriophage T4-like Lysozyme at C-terminus [cd00735]	4,00E-10
37	1337	144.1		Conserved structural phage protein [PHA02006] Phage internal core protein [Pseudomonas putida KT2440]	0 1,00E-12
38	251	28.3		Tail fiber [Cyanophage Syn5]	1,00E-07
39	152	16.2	LKA1	none	
40	302	33.1	LKA1	none	
41	201	22.7	LKA1	none	
42	101	11.7		Putative DNA maturase A [Vibrio phage VP93] Conserved hypothetical phage protein [PHA02046]	3,00E-04 9,00E-36
43	601	68.1		putative DNA maturase B [Vibrio phage VP93]	3,00E-145
44	66	7.1		Transmembrane domain	
45	160	17.3		Endolysin [cd00737]; Signal Peptide (P=0.998) putative endolysin [Enterobacteria phage RTP]	1,00E-19 3,00E-23
46	109	12.0		N-terminal transmembrane domain; Signal Peptide (P=0.994) Phage lambda Rz1-like protein [PHA02047]	 4,00E-32
46.1	66	6.8		Signal Peptide (P=1)	
47	104	10.5		none	
48	64	6.9	LKA1	none	
49	71	8.3	All but LUZ19	none	

[a] Gene numbering of LKA1 does not correspond to other ϕKMV-like phages, and is described in Ceyssens *et al*., 2006. [b] The ϕKMV genus was defined as comprising phages ϕKMV, LKA1, LUZ19, LKD16, ϕKF77, PT2 and PT5. [c] E-values are based on BLAST and pFAM searches using the corresponding protein of phage LKD16.

4.2.3 Transcriptional regulation

The intergenic genome regions of the ϕKMV-like phages were searched for transcriptional regulation elements. Analogous to ϕKMV (Lavigne et al., 2003), phages LUZ19 and ϕKF77 encode four potential host σ^{70} promoters in front of ORF 1. In contrast, LKA1 carries three strong host promoters before ORF 1, and a fourth one in front of ORF 2. Finally, five putative host promoters are identified within the first 1300 bp of the LKD16 genome (Table 4.4). The early factor-independent terminator for ϕKMV (bp 5087-5110) seems present in LKA1 and LUZ19, but absent in LKD16 and ϕKF77. This may suggest that transcription of the latter phage genomes by the host RNA polymerase continues at least until terminator behind the capsid gene is reached. Two factor-independent terminators located downstream of the major capsid protein and downstream the final gene are strongly conserved in all ϕKMV-like phages and are also found in other phages of the T7 group (Table 4.4).

Table 4.4. Strong early host σ^{70} promoters and factor-independent terminators in the genome of ϕKMV-like phages, illustrated for phages LKD16 and LKA1. The conserved -35 and -10 boxes, which clearly resemble the *P. aeruginosa* consensus sequence (TTGACa-N_{17}-TATaaT) (Dominguez-Cuevas and Sylvia, 2004), are indicated in grey, as are the stem-loop structures of the predicted terminators.

Phage	Start pos.	Upstream gene	Sequence
LKD16	779	1	GTTGACAGGGTAGGAGTGTGTCTGTAGAGTGCGCCCTGTCTCCAC
	857	1	GTTGACACACTGCCAGGGTATCGGTAGAGTGCGCAGCCAGAGGGA
	1009	1	ATTGACAAGGTAGTAAGATGCGG–TAAGATGTGCGGCGTCAAGTG
	1111	1	CTTGACACTGCAAGGCTCAGTCGGTAAGATGGGCACCCATAGCGG
	1260	1	GTTGACACGACAAGGCCAAGTCGATAACATGGCCACCACAAAGGG
	25782	33	CCCGCGCAGATTCCCTGCGTGGGTTTTT
	42628	DTR	AGTAGTTCAAGCCGAGCACCTGCATAGTCGGGTGCTCCACTGGAACTACTGAATTTTTATT
LKA1	475	1	GTTTGAATATAAGTACCACTCAGGTTATATAACGAATTGGTATTGA
	513	1	TATTGACGGGGCTTCAGCAGGTCTGTAGAGTTCGCAGCCATGGTTA
	685	1	GCTTGACAAGCCAA–GCGGATGCTGTAAGATGCGAGTCGTAACTCA
	1170	2	ACTTGACAAGCGAACGCCCAGTCTGTACGATGGGCACCAAGCAAGA
	6893	18	GGGGCTGGCCGCATAACCGGCCCCTTCT
	26102	44	GGACCTATCCTTCGGGGTAGGTCCTTTTTTTGGTT
	41296	DTR	CGCGCCCCCATGCCCCTGCGTGCGCGTTTCTT

The presence of a DNA-dependent RNA polymerase (gp26) in the genome of all ϕKMV-like phages implies the presence of phage-specific promoters to which this enzyme binds, leading to the transcription of the middle and late genomic region. All encoded RNA polymerases are definitely related to T7 (BlastP expected values between e^{-65} and e^{-68}) with broad conservation of the subdomains and of the essential catalytic residues Tyr639, Lys631, Asp537 and Asp812 (Cheetham and Steitz, 1999), strongly suggesting they are functional enzymes. However, no conserved T7-like promoter sequences were identified in the intergenic regions of the ϕKMV-like genomes. Accordingly, no homology can be detected in the recognition loop (AA 93-101) and specificity loop (AA 739-770), which interact with nucleotides -17 to -13 and -7 to -11 of the T7-like phage promoters, respectively (Cheetham and Steitz, 2000). These two regions are perfectly conserved between ϕKMV, LKD16, LUZ19 and ϕKF77, but not in LKA1 (Table 4.5), suggesting differences in phage promoter sequences between LKA1 and the other ϕKMV-like viruses.

Table 4.5. Alignment of two RNA polymerase subdomains of ϕKMV-like phages (gp26) and T7 (gp1).

Phage	Specificity loop	Recognition loop
ϕKMV	KTEEVRVRLRAEA---VEYVTLYEAKDEL	HQEAKAA--GPAAKL
ϕKF77	KTEEVRVRLRAEA---VEYVTLYEAKDEL	HQEAKAA--GPAAKL
LUZ19	KTEEVRVRLRAEA---VEYVTLYEAKDEL	HQEAKAA--GPAAKL
LKD16	KTEEVRVRLRAEA---VEYVTLYEAKDEL	HQEAKAA--GPAAKL
LKA1	DFVEHRVKIRSCG---ITDIVVREDLDQT	VVETKAR--GPGAAL
T7	KPIQTRLNLMFLGQFRLQPTINTNKDSEI	WFEEVKAKRGKRPTA

The only conserved motif which could be found in intergenic regions was the 5'-GACxxxxxCCTAC-TCCGG-3' string which was already associated with single-stranded nicks on the non-coding strand (Section 4.2.1). This sequence is not present in the LKA1 genome, but another conserved 11 nt sequence (5'-CSGCTGCACTC-3') in front of genes 5, 7, 18 and 30 (but not upstream the major capsid protein) was identified. The possible role of these sequences in the transcriptional regulation is discussed in section 4.4.1.

4.2.4 Genome organization

In various well-studies lytic *Podoviridae* like T7, T3 and gh-1, three major functional genome regions are distinguished which are involved in host conversion, DNA metabolism and particle formation, respectively (Pajunen et al., 2002; Kovalyova and Kropinski, 2003). Considering putative gene functions of the five ϕKMV-like phage genomes and taking into account the unusual localization of the RNA polymerase, we delineate two functional regions in these genomes. The first region ends downstream the RNA polymerase gene and encompasses genes for host conversion ('early' genes) and DNA replication, while the second ('late') region comprises genes coding for structural and lysis proteins (Figure 4.5).

4.2.4.1. Early genes and DNA replication genes

The phage proteins involved in interaction with host proteins are often produced immediately after infection. These 'early' proteins protect the bacteriophage from bacterial defence mechanisms or adapt the host-cell metabolism to establish an efficient infection cycle (Roucourt and Lavigne, 2009). Many of these genes can be deleted without seriously affecting phage production under standard laboratory conditions, suggesting that they are only necessary under specific environmental conditions, for infecting specific hosts, or have a redundancy in their function (Miller *et al*., 2003). The early genome regions of LKD16, LUZ19 and ϕKF77 are closely related to ϕKMV, although DNA homology is significantly lower in this region (50-60%, Figure 4.4). Most likely, the insertions and deletions within the early region account for the intriguing differences in burst size and latent period observed between LKD16 and ϕKMV (Section 4.1). In contrast, the corresponding region of LKA1 has strongly diverged and encodes 18 small proteins with no similarity to other gene products in the GenBank database. The only exception is gp4, which has a counterpart in all ϕKMV-like viruses and is also similar to gp70 of *Pseudomonas* phage PaP3 (Tan et al., 2007) and gp2 of *Ralstonia* phage RSB1 (Kawasaki et al., 2009), suggesting a strongly conserved function in phage infection or host conversion.

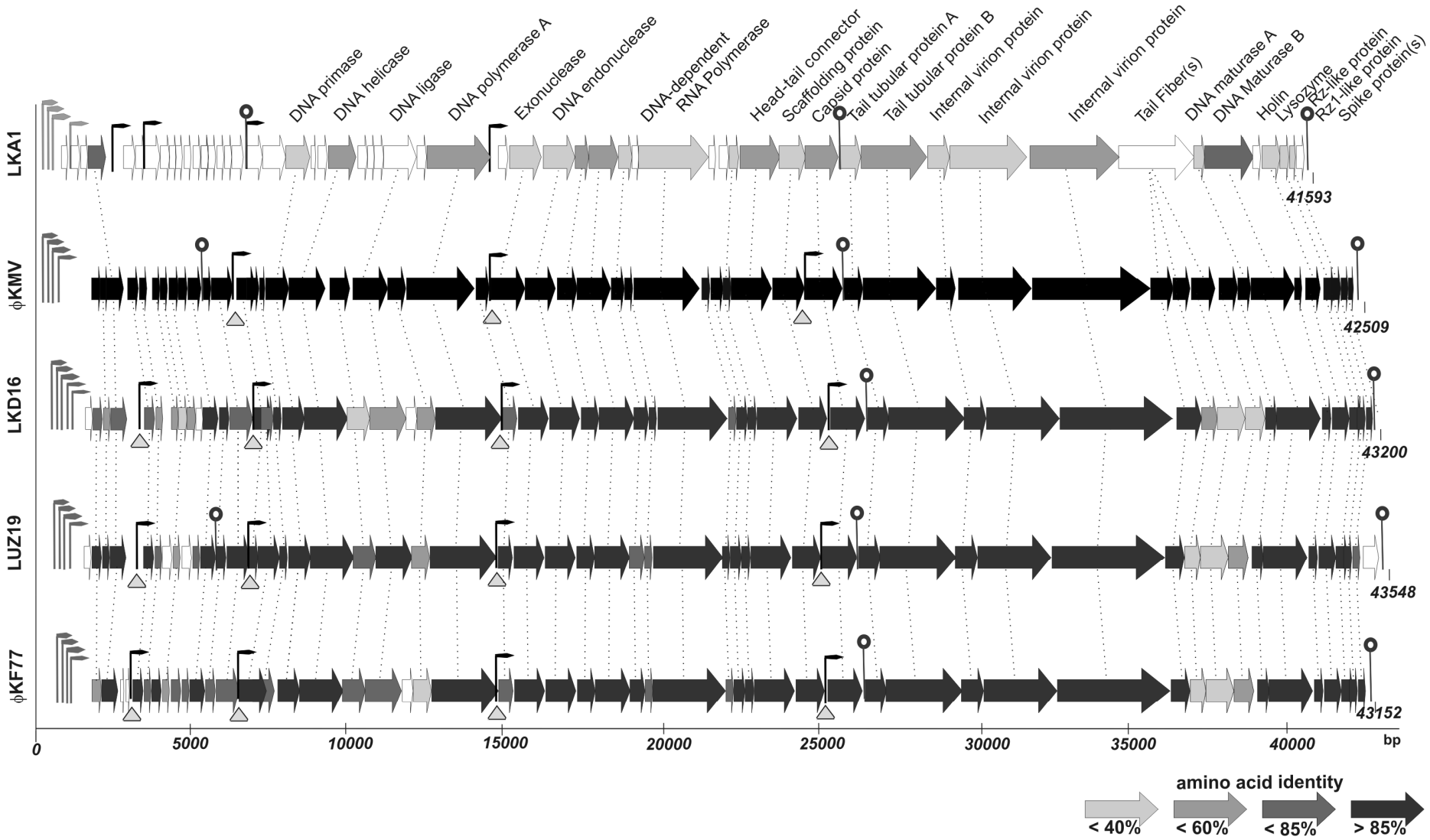


Figure 4.5. Comparative genomic analysis of the ϕKMV-like viruses. The predicted open reading frames of all ϕKMV-like phages and their amino acid identity to the corresponding ORF of ϕKMV, indicated in different shades of grey. ORFs unique to each phage are not colored, while predicted functions of the annotated genes are indicated on top. Corresponding proteins are connected with broken lines. Predicted factor-independent terminators and σ^{70} – and phage-specific promoters are shown as stem-loop structures and grey and black arrows, respectively. The position of the single-stranded nicks on the non-coding strand are indicated by grey triangles.

Gene order among the DNA replication proteins is conserved among all sequenced phages, except for gp16 (encoded between the helicase and ligase) which is replaced by three smaller gene products in LKA1 (Figure 4.5). This replication cluster corresponds to a T7-type replication module, which is basically composed of a Pol I-type DNA polymerase, a 5'-3' exonuclease, a $DnaG_{Eco}$-type primase and a $DnaB_{Eco}$-type helicase (Weigel and Seitz, 2006). It is unclear why ϕKMV-like phages encode a second 3'-5' exonuclease, because their DNA polymerases (gp19) already contain the typical 3'-5' exonuclease domains. The functional conservation of these DNA replication genes suggests that ϕKMV-like phages use the T7 replication mechanism, which is initiated by duplex melting by the phage RNA polymerase. In short, the RNA polymerase binds to a promoter on dsDNA and synthesizes a short (untranslated) transcript that remains attached to its template. The transcript provides the 3'-OH end used by DNA polymerase for displacement synthesis of a leading strand (Weigel and Seitz, 2006). This mechanism is widely spread, as exemplified by major similarities found with DNA replication genes of *Xanthomonas oryzae* Siphovirus Xp10 (Yuzenkova et al., 2003) and with a *Siphoviridae*-like prophage within the genome of *Burkholderia pseudomallei* 1710b (NC_007434).

4.2.4.2 Phage structural proteins

The ORFs encoding the phage virion proteins are clustered in the same relative order as in T7, although subtle differences are present. For example, all ϕKMV-like viruses carry a functional lysin domain at the C-terminus of the predicted internal protein gp36, while most members of the T7 group have this domain at the N-terminus of the protein encoded by the downstream gene (Lavigne et al., 2004). Twelve predicted ORFs were previously confirmed as part of the ϕKMV phage particle (Lavigne et al., 2006). Due to the major similarities in the structural gene region of LUZ19, LKD16, ϕKF77 (on average 96.5% AA identity to corresponding ϕKMV proteins), only LKD16 was selected for mass spectrometric (MS) analysis. In contrast to the ϕKMV particle analysis which was restricted to visual protein bands, the entire lane of a 1D-SDS-PAGE gel containing structural LKA1 and LKD16 proteins was cut into slices for subsequent trypsin digestion and ESI-MS/MS analysis. Consequently, less abundant proteins or peptides migrating between visible bands are not overlooked. Table 4.6 summarizes the experimentally confirmed structural proteins of both phages, displaying the number of significant sequence-identified non-sibling peptides and the resulting sequence coverage of the entire protein. As expected for LKD16, all the ORFs corresponding to the identified particle proteins of ϕKMV (Lavigne et al., 2006) were also found in LKD16 (Table 4.6). In addition, gp41 of LKD16 was identified as a structural protein. Analysis of the MS data pointed out that the two most significantly identified peptides of gp41 are located at the N-terminal of the gene product, which was previously not defined as part of the coding region of ORF 41 in ϕKMV (Section 4.2.2). This observation supports

the revised annotation of this ORF which expanded the 5' end of the gene with 198 bp. The smallest putative internal virion protein of LKD16 (gp35) was also confirmed as a structural protein. In this way, all predicted proteins in the structural genome region of LKD16 were experimentally identified, except for the scaffolding protein and gp39 which may not be part of the mature phage particle.

Table 4.6. Identification of the structural proteins of LKA1 and LKD16.

Putative function	LKA1				LKD16				ϕKMV[a]
	Gene	Mol. mass (kDa)	# peptides identified	Sequence coverage (%)	Gene	Mol. mass (kDa)	# peptides identified	Sequence coverage (%)	
Unknown	40	8.9	5	88.9	29	10.5	4	48.9	+
Connector	41	56.6	16	51.9	30	56.1	11	29.1	+
Capsid	43	36.7	23	83.2	32	37.7	21	64.6	+
Tail Protein A	44	23.4	2	21.4	33	21.2	2	18.2	+
Tail Protein B	45	84.4	17	31.2	34	92.0	12	21.6	+
Internal virion	46	21.6	2	16.1	35	18.8	5	30.9	-
Internal virion	47	93.9	50	69.3	36	98.2	19	37.2	+
Internal virion	48	108.0	32	50.1	37	143.8	17	18.7	+
Tail Fiber	49	80.4	16	33.6	38	28.5	6	32.2	+
Adsorption	-				40	31.8	5	27.5	+
Adsorption	-				41	22.7	4	35.8	-
Spike protein	56	10.7	2	42.4	47	10.5	7	86.5	+
Spike protein	-				48	6.8	2	42.2	+

[a] Proteins indicated with a '+' were previously reported as part of the ϕKMV phage particle (Lavigne et al., 2006).

One of the most intriguing differences between LKD16 and ϕKMV is located in the tail fiber section. In sharp contrast with the other structural proteins that are nearly perfectly conserved, gp39, 40 and 41 show less than 50% conservation (Figure 4.5). Similarity searches yielded no putative function for these genes, but gp40 and gp41 were shown to be part of the LKD16 phage particle, presumably playing a role in phage adhesion to the host. Interestingly, these three ORFs of LUZ19 are the only genes displaying significantly more similarity to LKD16 than to ϕKMV, which was already indicated by the overlapping host spectra observed for LUZ19 and LKD16. Even more remarkable is the perfect conservation of the three tail fiber genes of LUZ19 and ϕKF77. Unfortunately, the host range of the latter phage could not be checked due to unavailability of phage particles. Adsorption and infection of these phages might also be influenced by the two small proteins (gp47 and gp48) at the right end of the genome, which have only one counterpart (gp56) in LKA1 (Figure 4.5). These structural proteins encoded downstream the lysis cassette, are not found in T7 but are experimentally confirmed in phage SP6 (Scholl et al., 2004).

The structural region of LKA1 shares an overall 30-50% protein similarity to ϕKMV, encompassing proteins encoded between ORF 40 to ORF 49 and including ORF 56 (Table 4.6). Compared to ϕKMV, deletions are present in genes encoding the scaffolding protein (180 bp) and the largest internal virion protein (999 bp). Interestingly, ORFs 38, 39, 40 and 41 of ϕKMV are apparently replaced by a

single ORF in LKA1 (ORF 49). BLAST analyses on gp49 revealed similarity at the N-terminal to several T7-like tail fibers, probably attaching this N-terminal end of the protein to the phage virion. However, C-terminal of gp49 has further differentiated, and two iterative psiBLAST searches point at distant similarity with fungal *β1-3* glucanases (E-value of 7e-8). The combination of an N-terminal connection to the virion particle and the enzymatic activity at the C-terminal end would be consistent with tail structures found in other phages like K1F and PK1E (Scholl and Merril, 2005). The putative role of this protein in the binding and/or (partial) degradation of the host cell wall is currently under investigation.

4.2.4.3 Lysis mechanism

Most phages use a combination of a holin and an endolysin to release newly formed phage particles. These ORFs (44 and 45, respectively) are nearly perfectly conserved between LKD16, LUZ19, ϕKF77 and ϕKMV. Analysis of the N-terminal sequence of the endolysins (gp45) of both phages reveals the presence of an identical signal sequence (Table 4.7), suggesting a Signal-Arrest and Release (SAR) mechanism, which was recently demonstrated for ϕKMV (Briers, Y. et al., *in preparation*). In this phage release mechanism, first described in coliphage P1, the N-terminal signal sequence of the lysin mediates Sec-transport to the periplasm (Xu et al., 2004). In gp45, however, the C-terminal domain of the signal sequence, responsible for endopeptidase cleavage, is absent (Table 4.7), suggesting that lysin molecules are probably released upon destabilization of the inner membrane by a class II holin (Xu et al., 2005). Furthermore, many phages encode internally overlapping Rz/Rz1 proteins immediately downstream of the endolysin to improve bacterial lysis when the outer membrane is stabilised by divalent cations. These Rz/Rz1 gene products form a complex after translocation to the periplasm, link the inner and outer membrane and transform the mechanical stress caused by the holin lesion to the outer membrane, leading to its disintegration (Berry et al., 2008; Kuprovic et al., 2008). In-depth comparative analysis of ϕKMV-like genomes suggests that ORF46 contains a conserved out-of-frame internal ORF 46.1. Both ORFs have a clear ribosomal binding site and the corresponding proteins contain a true periplasmic signal sequence at the N-terminal, an essential feature for Rz/Rz-like proteins (Table 4.7).

The lysis cassette of LKA1 carries the same backbone structure. Despite only limited protein similarity, a signal peptide is predicted at the N-terminal ends of gp53, 54 and 55 which are presumably the endolysin and the Rz/Rz1-like proteins, respectively (Table 4.7).

Table 4.7. Features of the lysis-related genes of LKD16, ϕKMV and LKA1.

Phage	Protein	Function	N-terminal sequence[a]				Signal seq. probability[a]
LKD16/ϕKMV/	gp 45	Endolysin	VNKPLR	GAALAAALAGLVAL		EGSETTA	0.99
LUZ19/ϕKF77	gp 46	Rz protein	MPRT	IVAILVLAVVALG	ASYGFVQ	SYRALG	0.99
	gp 46.1	Rz1 protein	VPWTRT	ARGLTGLCLLLS	LTACAT	APAPA	1.00
LKA1	gp 53	Endolysin	MRKFK	ATLTAAALAL	GLGTGV	VYNAAEFI	0.99
	gp 54	Rz protein	MRLT	VALILALVVAV	GAGLYAA	QKAWQYHR	0.99
	gp 55	Rz1 protein	MKNLL	GGMLLCLLPFVAS	CASSSSA	EPVRC	1.00
				N H		C	

[a] Signal sequences were predicted using SignalP, showing for each protein the signal sequence probability. The N, H and C-domains in the signal sequence are indicated.

Moreover, two hydrophobic domains are present in gp52, suggesting its function as a class II holin. In contrast to ϕKMV and LKD16, a potential dual translational start motif is predicted in the coding sequence of gp52. In this motif (MTKM), two methionine residues are separated by a positively charged amino acid, leading to two proteins of slightly different size (68 and 65 AA). Analogous to the class II holin proteins of phage 21 (Barenboim et al., 1999), this could lead to a fine-tuned regulation of host lysis by a shorter holin-effector and a longer holin-inhibitor.

4.3 Influence of LUZ19 infection on the host RNA polymerase complex

The central component of bacterial transcription is the DNA-dependent RNA polymerase (RNAP), of which the core enzyme has a subunit composition of $\beta\beta'\alpha_2\omega$. The active site of the polymerase, which has determinants for the binding of both template DNA and the RNA product during transcription, is formed from the large β and β' subunits (Korzheva et al., 2000). Each of the two identical 329-residue α subunits consists of two independently folded domains that are joined by a flexible linker (Blatter et al., 1994). The larger amino-terminal domain (residues 1-235) dimerizes and is responsible for the assembly of the β and β' subunits. In contrast, the smaller carboxy-terminal domain (residues 250–329) is a DNA-binding module that has an important role at transcription of certain promoters, interacting not only with DNA but also with many activator and repressor proteins (Gourse et al., 2000). The small 91-residue ω subunit of the RNAP complex has no direct role in transcription, but seems to function as a chaperone to assist the folding of the β' subunit (Hampsy and Omega, 2001). For RNA polymerase to begin transcription at a particular promoter, it must first interact with a σ subunit to form the holoenzyme. The σ subunit has three main functions: to ensure the recognition of specific promoter sequences; to position the RNA polymerase holoenzyme at a target promoter; and to facilitate unwinding of the DNA duplex near the transcript start site (Wörsten, 1998).

Many phages encode proteins that directly bind to and regulate the activity of the bacterial host RNAP, and some phages encode enzymes that covalently modify the host transcriptional apparatus. These phage-induced modifications ensure coordinated temporal transcription of the phage genome and inhibition of host transcription (Nechaev and Severinov, 2003). Classic examples of this are gp0.7 and Alc of phage T7 and T4, respectively, which bind to the β' subunit and catalyze host transcription shutoff (Michalewicz and NLUicholson, 1992; Kashlev et al., 1993).

Bacteriophage ϕKMV and its relatives use the transcription apparatus of the host, at least until their own RNA polymerase (ORF26) is transcribed. We determined the effect of LUZ19 infection on the host RNAP using a tandem affinity purification (TAP) method, which allows rapid complex purification under native conditions, even when expressed at natural levels (Puig et al., 2001). To this end, we constructed a strain of *P. aeruginosa* PAO1 encoding the original TAP affinity tag appended to the 3' end of the *rpoC* gene, which encodes the RNAP β' subunit (section 2.4.4). The wild-type *P. aeruginosa* and *P. aeruginosa rpoC::TAP* strains exhibited almost identical growth curves in LB broth (data not shown), showing that the tag has no influence on the viability of the cells. Moreover, no difference in phage titer or plaque morphology was observed when the same phage preparation was plated on both strains. In a next step, both cell types were infected with LUZ19, RNAP was affinity-purified and its composition was compared to that of RNAP from uninfected cells. As such, uninfected *P. aeruginosa rpoC::TAP* cells yielded five distinct bands on a 1D-SDS-PAGE gel, which were identified using ESI-MS/MS as four subunits of the *P. aeruginosa* RNAP complex (Figure 4.6). In addition, the elongation factor Tu (PA4265) and the heat-shock protein PA4873 were co-purified.

After infection of *P. aeruginosa rpoC::TAP* with LUZ19, the purified complex contained an extra band with an apparent molecular mass of approx. 33 kDa, while the intensity of the band containing the α subunit significantly decreased (Figure 4.6). Mass spectrometry identified this protein also as the α subunit, but no peptides from the C-terminal part could be identified (Figure 4.6). Interestingly, *Pseudomonas* proteins PA1776 and PA4275 were co-purified with the RNAP complex, in contrast to any LUZ19-encoded proteins. PA1779 is a recently identified SigX sigma factor, which binds to σ^{70} promoters (Potvin et al., 2008). SigX shares 49% similarity to σ^{w} of *Bacillus subtilis,* which is induced by alkaline shock, salt, phage infection and antibiotics (Schobel *et al.*, 2004). PA4275 is a NusG antitermination protein which influences transcription termination and anti-termination (Greenblatt et al., 1993).

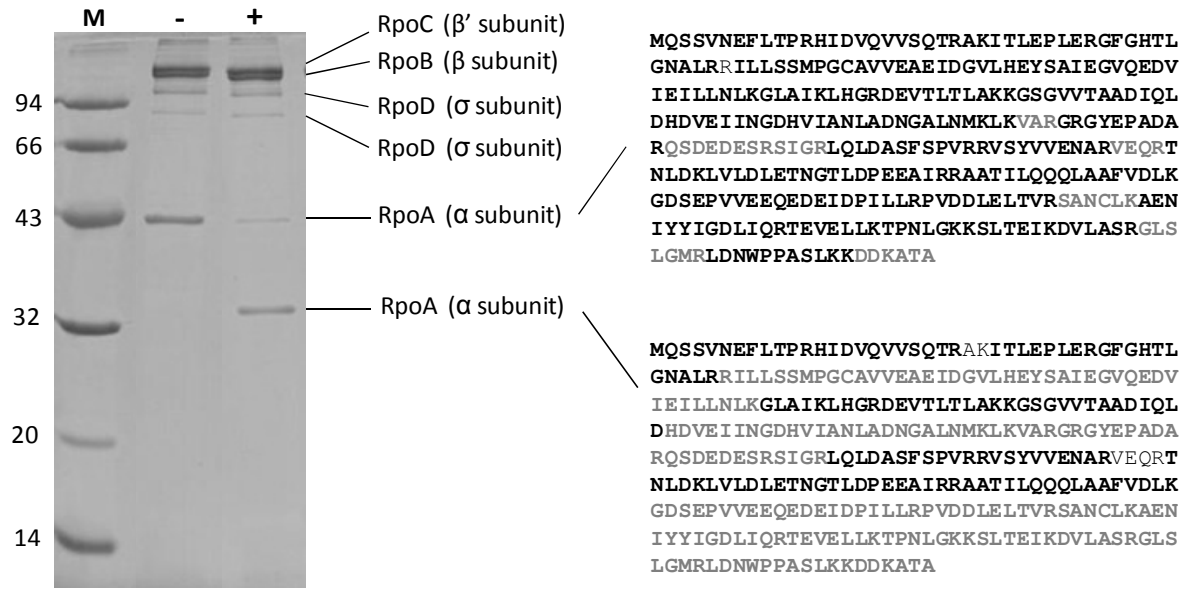


Figure 4.6. Tandem Affinity Purification of the *P. aeruginosa* RNA Polymerase complex, with (+) and without (-) LUZ19 infection. Visible proteins bands were excised and identified using ESI-MS-MS. The identified peptides of the RpoA protein, with and without phage infection, are indicated in black on the right.

These results suggest proteolytic cleavage of the α subunit of the RNAP core enzyme in the course of phage infection. To our best knowledge, this mechanism of transcriptional regulation has not been reported in literature, and will be subject of future research. A logical next step will entail mass spectrometric analysis of the whole gel lane, looking for low-abundant phage and host proteins, and analysis of the RNAP complex at different time points during infection.

4.4 Discussion

4.4.1. Genome architecture of T7-related phages

Two and a half decades after the completion of the genome sequence of coliphage T7 (Dunn and Studier, 1983), various phages sharing common morphological, biological and genomic characteristics have been sequenced (Figure 4.7). One of the major distinctive characteristics of the 'T7-like supergroup' is the presence of a single-subunit phage RNA polymerase encoded in the early genomic region, which recently formed the basis for the creation of the '*Autographivirinae*' subfamily (Lavigne et al., 2008). Phages closely related to T7 (the T7 *genus*) include, among others, coliphage T3 (Pajunen et al., 2002) *Yersinia* phages ϕA1122 (Garcia et al., 2003) and *Pseudomonas* phage gh-1 (Kovalyova and Kropinski, 2003). These phages were isolated over time, in different habitats and infecting different hosts and were shown to have up to 89% DNA identity with T7. A second genus in

this subfamily contains the slightly more diverged *Salmonella* phage SP6, coliphages K1-5 and K1E and *Erwinia* phage Era103. The genome organization of these phages resembles that of T7 (Figure 4.7), but numerous insertions and deletions throughout the genome are present, as shown by the presence of a P22-like spike protein (Dobbins et al., 2004).

This work led the basis for the creation of a third genus of the 'ϕKMV-like viruses' within the *Autographivirinae*, which was recently approved by the International Committee for the Taxonomy of Viruses (ICTV). The hallmark feature distinguishing this genus is the location of the RNA polymerase gene, adjacent to the structural genome region (Figure 4.7). Today, eight phages belonging to three different bacterial species are classified within this genus. Additional evidence of the wide environmental distribution of this type of lytic phages was provided by the isolation of several phages from various Russian environmental samples, which were annotated as ϕKMV-like phages based on Southern hybridization experiments (Burkalt'seva et al., 2006).

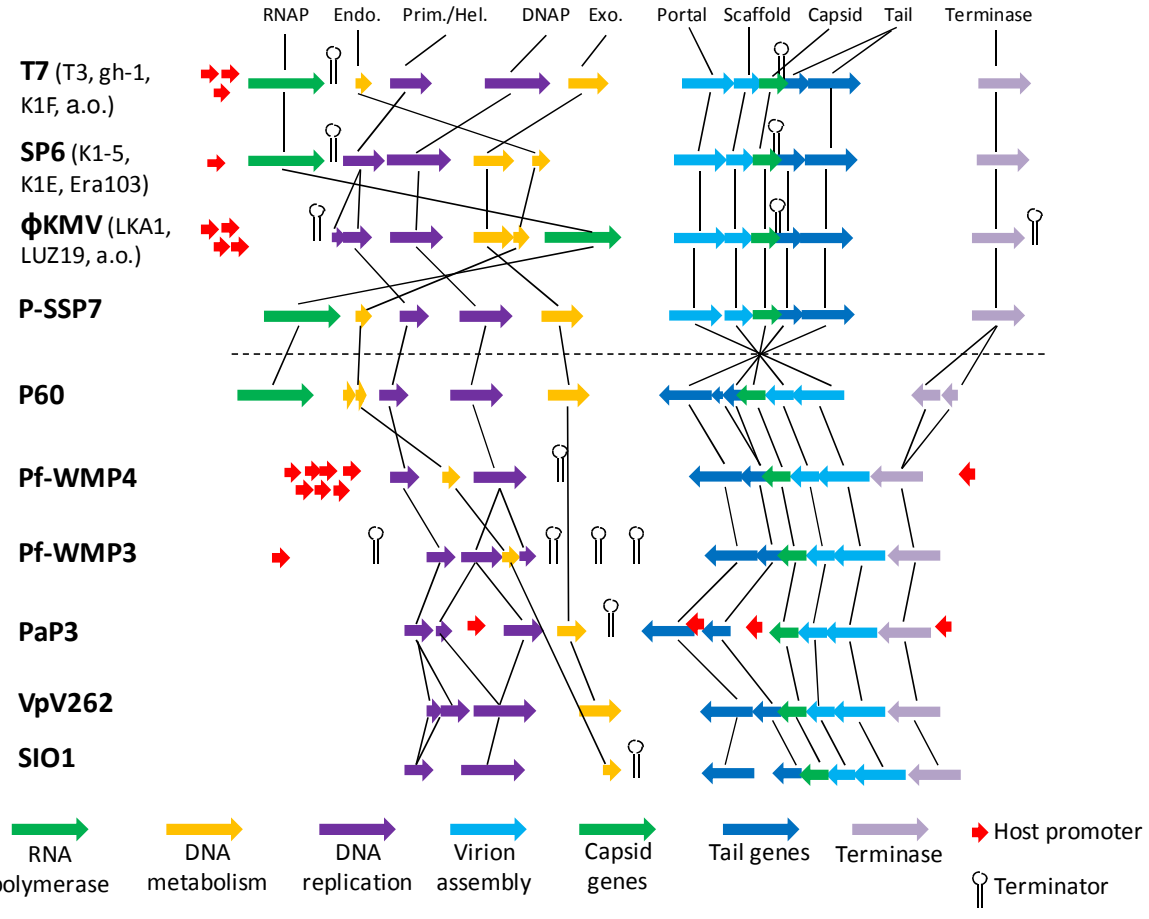


Figure 4.7. Comparative genomic arrangements of *Podoviridae* related to T7. The different colors indicate the functions of the ORFs. *RNAP* RNA polymerase, *DNAP* DNA polymerase, *Endo.* endonuclease, *Pri/hel.* primase/helicase, *Exo.* Exonucleas. This figures was constructed based on data presented by Pajunen et al., 2002, Chen and Lu, 2002, Lavigne et al., 2003, Hardies et al., 2003, Tan et al., 2007, Liu et al., 2007 and Liu et al., 2008.

Expanding this analysis to other virulent phages, many variations on the same theme are discovered, suggesting a common ancestry. For example, *Pseudomonas* phage PaP3 and marine phages Pf-WMP3, Pf-WMP4, VpV262 and SIO1 share extensive homology with T7 but lack the phage RNA polymerase, and probably constituting an ancient branch of the *Autographivirinae* (Hardies et al., 2003; Tan et al., 2007; Liu et al., 2008). Although the late modules encoded by these phages are oriented in different directions, their inner gene orders are not disrupted. In all these phages, the genes and gene order for the early region have strongly diverged, and the position of some conserved DNA replication genes, *e.g*., DNAP and DNA primase/helicase, is clearly not fixed (Figure 4.7). Cyanophage P60 presents an interesting chimera, containing both a phage-specific RNA polymerase and an inverted structural region (Chen and Lu, 2002). Future isolations will provide an even more comprehensive view on the diversity of this group of virulent phages.

4.4.2. The role of the conserved intergenic motif

The genomes of the *senso strictu* ϕKMV-like phages (LUZ19, LKD16, ϕKF77) have localized single-strand interruptions (nicks) on their non-coding strand, a distinguishing genomic trait previously thought to be unique for T5 related coliphages. No function has been attributed to these nicks, although viable T5 mutants have been obtained lacking major/minor nicks as well as mutants with increased frequency of canonical nicks (Rogers et al., 1979; Rhoades, 1984). Recent research on phage ϕKF77 DNA indicated that these nicks are repaired soon after infection, and that they are introduced into the phage DNA at late stages of a lytic cycle, probably during genome packaging (Kulakov et al., 2009). As these nicks remove the superhelicity of the genome, a function in facilitating phage genome packaging cannot be excluded.

Analysis of the genomic localization of the nicks indicates that they are all located in intergenic regions. Moreover, the conserved motif associated with these interruptions is the only conserved intergenic motif present in these phages. As such, it is very tempting to postulate these sequences as putative promoters, specifically binding the phage-encoded RNA polymerase. Compared to other T7-like phages, the number of putative promoters would be considerably lower and upstream AT-rich regions playing a role in modulating the efficiency of transcription initiation in T7-like phages (Tang et al., 2005) are absent. However, these differences might be correlated with the 'late' localization of the RNA polymerase gene and the higher dependency on the host RNA polymerase, illustrated by relatively high numbers of host promoters. It is interesting to note that LKA1, a more distant relative within the 'ϕKMV-like viruses' which lacks DNA homology, does not contain the conserved nicks' sequences and does not produce noticeable ssDNA fragments on agarose gels. However, this phage

carries another conserved 11 nt sequence (5'-CSGCTGCACTC-3') in front of genes 5, 7, 18 and 30 (Section 4.2.3). Research involving the ϕKMV-like *R. solanacearum* phage ϕRSB1 showed clear promoter activity associated with conserved intergenic regions, which also differ in the consensus sequence from the ϕkF77 nick sites (Kawasaki et al., 2009) but are situated at corresponding genomic locations (i.e. in front of the gene encoding a major capsid protein). This link between single-stranded nicks on the non-coding strand and putative promoter activity on the coding strand remains to be studied in more detail.

Chapter 5

The intron-containing genome of lytic *Pseudomonas* phage LUZ24 resembles temperate phage PaP3 [3]

It was inferred from the *de novo* peptide sequence analysis of *Podoviruses* LUZ14 and LUZ24 that these phages are at least structurally related to phage PaP3 (Section 3.2). *Pseudomonas* phage PaP3 (NC_004466) is described as a temperate phage with 5'-protruding cohesive ends, which is able to integrate into the $tRNA^{lys}$ gene of its host (Tan *et al*., 2007). Relatives of this phage seem widespread, as sequences related to this phage are commonly encountered in marine metagenomic projects (Bench et al., 2007; McDaniel et al., 2008). This chapter focuses on the genomic and proteomic characterization of phage LUZ24 (Leuven, Universitair Ziekenhuis, sample 24) as an example of this genus.

5.1 Phage characteristics

Phage LUZ24 was isolated from hospital sewage using the clinical *P. aeruginosa* strain Li010 (Pirnay *et al*., 2002) as enrichment host. Phage particle morphology clearly classifies LUZ24 within the *Podoviridae* family of short-tailed phages, consisting of an icosahedral head with a 63 nm diameter and a short and banded tail of 12 x 8 nm. Some decoration proteins can be distinguished, protruding from the phage capsid (Figure 5.1A). Standard stability assays show that LUZ24 retains its infectivity from pH 4 up to pH 11, but titer levels drop 80% after 9 months of storage at 4°C in standard phage buffer (pH 7.5). Extensive host range screenings on environmental and clinical *P. aeruginosa* strains (Pirnay *et al*., 2002) show that LUZ24 lyses 36 out of 123 (29%) strains, including multi-drug resistant *P. aeruginosa* strains Br642 and Br776. Surprisingly, only small and turbid plaques (1 mm) arise after infection of *P. aeruginosa* PAO1, although a slightly more efficient adsorption (91% in 5 min) and no significant difference in latent period or burst size could be noted (Figure 5.1B, C).

[3] Data presented in this chapter is adapted from Ceyssens, P.J., *et al.* (2008) *Virology* **377**: 233-238.

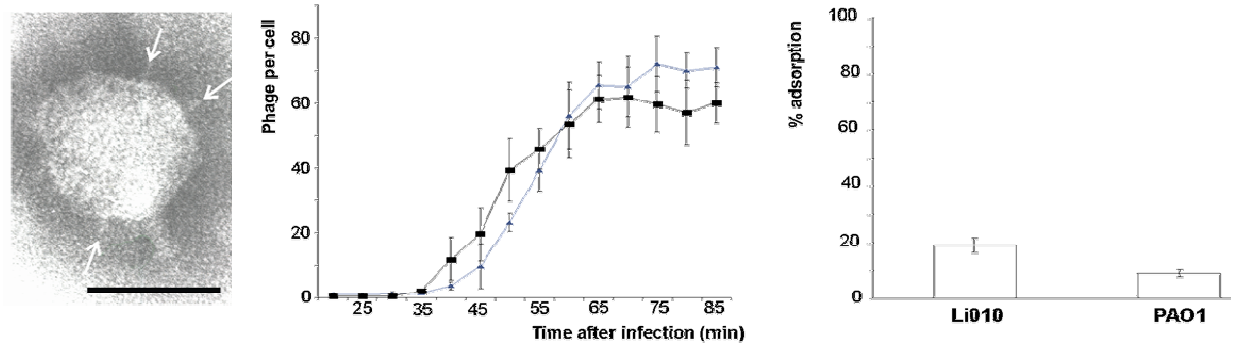


Figure 5.1. A. Electron microscopic image of a negatively stained LUZ24 particle. Putative head decoration proteins are indicated with white arrows, scale bar represents 50 nm. **B.** One-step growth curve of LUZ24, showing the amount of phages released per infected cell when either *P. aeruginosa* PAO1 (black) or *P. aeruginosa* strain Li010 (grey) are used as host. **C.** Adsorption assay, showing the amount of non-absorbed phage particles after 5 minutes of infection. In each case, the average and standard deviation of three independent experiments is given.

Despite the suspicion of lysogeny, all attempts to isolate stable lysogenic *P. aeruginosa* clones failed. DNA was extracted from phage-resistant *P. aeruginosa* clones isolated from turbid plaques and batch cultures. No phages could be induced, and no integrated LUZ24 DNA sequences could be demonstrated by PCR or restriction analysis.

5.2 Genome analysis

5.2.1 Genomic properties, gene prediction and regulation

The LUZ24 particle encapsulates a linear dsDNA molecule consisting of 45,625 bp (52% GC). The genome encodes 68 proteins, 47 of which are arranged in rightward orientation while 21 are encoded on the opposite strand (Figure 5.2). The genome of LUZ24 displays an overall nucleotide identity of 71% to phage PaP3. Eight small insertions/deletions are present and 88% of the encoded gene products are related between both phages (Table 5.1, Figure 5.2).

However, the *cos* site identified in PaP3 is not conserved in LUZ24, and no difference in restriction patterns could be observed if samples were heated to 80°C prior to separation on a 1% agarose gel (data not shown). In contrast, the genome of LUZ24 was found to be delineated by two 184 bp direct terminal repeats (DTRs) (Figure 5.3). Another mutual difference is that only two tRNA genes ($tRNA^{Asn}$ and $tRNA^{Pro}$) are present at the right end of the LUZ24 genome, compared to four in PaP3. The evolutionary advantage of these tRNA genes seems rather small, since no clear codon preference in LUZ24 compared to *P. aeruginosa* could be detected.

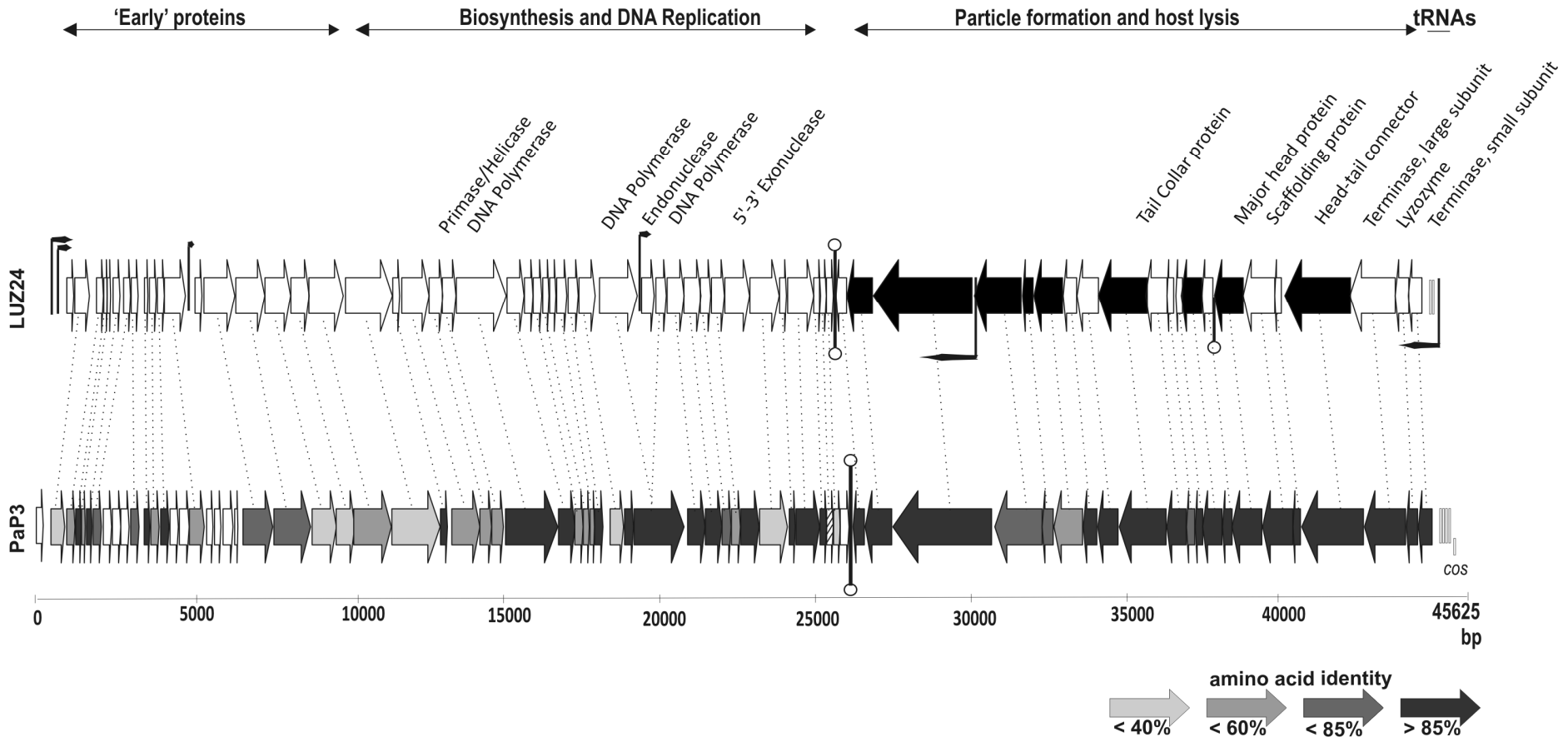


Figure 5.2. Side-by-side comparison of the LUZ24 and PaP3 genomes, marking the amino acid identities of the corresponding ORFs in different shades of blue. Genes with corresponding functions are connected with broken lines, and the experimentally identified σ^{70} promoters (arrows with length according to the promoter strength, Table 5.2) and rho-independent terminators (open circles) are indicated. The identified structural proteins of LUZ24 are colored black, and the newly annotated PaP3 gene is hatched.

Table 5.1. Summary of the *in silico* analysis of the LUZ24 genome.

ORF[a]	Location	Mw (kDa)	Strand	Best phage homolog	E-value	Comments
1	1020-1190	6.3	+	No similarity		
2	1208-1762	21.0	+	LKD16 gp3	3e-5	Also present in ϕKMV, LKA1 and PaP3
3	1974-2192	8.2	+	PaP3 p69	2e-6	
4	2189-2329	5.7	+	PaP3 p68	8e-14	
5	2326-2409	3.1	+	PaP3 p67	4e-5	
6	2509-2691	6.9	+	PaP3 p66	4e-6	
7	2823-3056	8.5	+	PaP3 p65	1e-10	
8	3058-3333	10.5	+	PaP3 p61	6e-30	
9	3487-3633	5.6	+	PaP3 p60	7e-21	
10	3642-3893	9.4	+	PaP3 p59	3e-25	
11	3883-4164	10.6	+	PaP3 p58	3e-47	
12	4164-4823	24.4	+	PaP3 p55	2e-27	C-terminal ATPase domain [Pfam domain PF02449, E= 0.75]
13	4979-5131	8.3	+	No similarity		
14	5443-6405	3.6	+	PaP3 p50	6e-39	
15	6424-7380	35.8	+	PaP3 p49	3e-50	
16	7377-8177	29.6	+	PaP3 p48	1e-94	Conserved in PA11 [E= 4e-30]
17	8170-8739	21.7	+	PaP3 p47	8e-47	
18	8715-9890	44.4	+	PaP3 p46	4e-115	Conserved in PA11 [E= 2e-22]
19	9902-11434	55.9	+	PaP3 p45	0.00	L-glutamine-D-fructose-6-phosphate amidotransferase GATase_2 domain [Pfam domain IPR000583, E= 2e-6]
20	11444-11665	8.6	+	PaP3 p44	5e-18	Conserved in PA11 [E= 1e-21]
21	11793-12677	32.2	+	PaP3 p43	9e-97	Gluthation synthase [Pfam domain RimK, E= 2e-3]
22	12677-13075	14.8	+	PaP3 p42	9e-43	Biosynthesis related [Pfam domain UPF0131, E= 2e-4]
23	13075-13452	14.2	+	PaP3 p41	2e-26	
24	13453-15162	64.4	+	PaP3 p40	0.00	Primase/helicase Conserved in PA11 [E= 8e-69]
25	15146-15655	19.7	+	PaP3 p39	4e-85	DNA polymerase (~T7 DNA pol. res. 2-186)
26	15668-15937	9.7	+	PaP3 p38	5e-20	Contains one transmembrane domain
27	15969-16202	8.8	+	No similarity		
28	16183-16374	7.2	+	PaP3 p37	2e-16	
29	16364-16525	5.8	+	PaP3 p36	3e-08	
30	16516-16935	15.8	+	Mx9 p02	5e-13	
31	16932-17216	10.7	+	PaP3 p35	8e-34	Putative holin; 3 transmembrane domains
32	17341-17796	17.9	+	PaP3 p34	1e-19	
33	17801-17998	7.5	+	PaP3 p33	2e-20	
34	17999-19180	45	+	PaP3 p32	0.00	DNA polymerase (~T7 DNA pol. res. 207-525)
35	19250-19666	16.1	+	phage Berlin	2e-21	Endonuclease_7 [Pfam family PF02945, E= 5e-15]
36	20024-20305	10.7	+	PaP3 p32	4e-11	DNA polymerase (~T7 DNA pol. res. 621-706)
37	20375-20929	20.7	+	PaP3 p31	8e-88	
38	20907-21410	18.3	+	PaP3 p30	9e-81	Also present in ϕKMV and LKD16
39	21436-21666	8.8	+	PaP3 p29	7e-20	
40	21666-22115	17.1	+	ES18 gp47	8e-20	Related to both gp2.8 and gp7.7 of T7
41	22200-23030	31.8	+	PaP3p28	4e-121	5'-3' Exonuclease [Pfam family PF01367, E= 4e-9] Conserved in PA11 [E= 5e-29]
42	23008-23952	35.5	+	PaP3 p27	1e-59	
43	24002-24184	7.2	+	PaP3 p26	4e-20	
44	24177-24938	29.3	+	PaP3 p25	1e-127	
45	24935-25153	8.1	+	PaP3 p24	2e-28	
46	25157-25366	7.8	+	PaP3 p23.1	6e-45	Newly defined in the PaP3 genome
47	25353-25559	7.7	+	PaP3 p23	2e-09	
48	26123-25767	12.6	-	PaP3 p21	2e-40	
49	27025-26138	31.6	-	PaP3 p20	1e-146	Structural protein
50	30204-27037	112.0	-	PaP3 p19	0.00	Structural protein
51	31872-30355	53.5	-	PaP3 p18	2e-169	Structural protein
52	32257-31877	12.6	-	PaP3 p17	1e-25	Structural protein
53	33198-32257	32.0	-	PaP3 p16	2e-113	Structural protein
54	33613-33179	16.6	-	PaP3 p15	2e-76	
55	34299-33610	25.2	-	PaP3 p14	3e-107	
56	35837-34296	57.5	-	PaP3 p13	0.00	Structural protein
57	36493-35846	22.0	-	PaP3 p12	1e-114	Phage Tail Collar Domain [Pfam family PF07484, E= 5.6e-19]
58	36731-36483	8.8	-	PaP3 p11	2e-17	
59	36927-36715	7.7	-	PaP3 p10	4e-17	
60	37543-36917	24.0	-	PaP3 p09	3e-111	Structural protein
61	37867-37547	12.0	-	PaP3 p08	3e-49	
62	38868-37915	35.0	-	PaP3 p07	2e-168	Major head protein

(Table 5.1. Continued)

63	39879-38887	37.2	-	PaP3 p06	2e-108	Scaffolding protein	
64	40094-39879	8.2	-	PaP3 p05	4e-21		
65	42244-40124	81.1	-	PaP3 p04	0.00	Portal protein	
66	43692-42244	54.6	-	PaP3 p03	0.00	Terminase, large subunit [Pfam family PF03237, 111]	E= 1.2e-
67	44090-43692	14.8	-	PaP3 p02	1e-55	T4-like Lysozyme	
68	44580-44122	16.7	-	PaP3 p01	1e-69	Terminase, small subunit (PsiBlast)	
	44680-44607		-			tRNA-Pro [tRNAscan-SE]	
	44852-44780		-			tRNA-Asn [tRNAscan-SE]	

[a] The genome of bacteriophage LUZ24 is deposited at Genbank under accession number NC_010325.

As found in similarly organized phages, the opposing transcriptional gene clusters are separated by a strong bidirectional terminator (Table 5.2). To experimentally identify bacterial promoters in LUZ24, a threefold redundant fragment (200-300 bp) library was generated in the broad-spectrum promoter trap vector pTZ110 (Schweizer and Chuanchuen, 2001). Promoter activity was determined quantitatively in *P. aeruginosa* cells by measuring the β-galactosidase activity in a Miller assay (Miller, 1992). Seven σ^{70}-specific promoter regions were identified, spread throughout the genome and displaying a clear *P. aeruginosa* consensus sequences (TTGACa-N_{17}-TATaaT, Table 5.2) (Dominguez-Cuevas and Sylvia, 2004). This indicates the dependency of LUZ24 on the transcriptional machinery of its host throughout the entire infection cycle. Promoters located in the right arm are clearly stronger, which might be explained by the presence of A-rich tracts upstream the -35 and -10 regions, a phenomenon also observed in phages T4 and T5 (Miller *et al*., 2003).

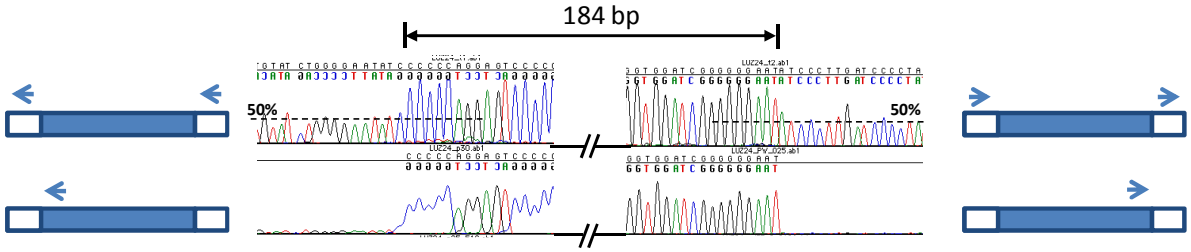


Figure 5.3. Determination of the exact length of the direct terminal repeats of phage LUZ24 by direct sequencing with outward-directed primers. Primers (arrows) annealing outside the DTR (white boxes) lead to a stop of the sequencing reaction at the ends of the genome (grey boxes), while primers annealing within these repeats lead to a 50% reduction in signal intensity.

All predicted proteins were subjected to an in-depth *in silico* analysis (Table 5.1), and seem to cluster into three genomic regions encoding processes in early, middle and late infection. These separate clusters are discussed in more detail in the following sections.

Table 5.2. Regulatory sequences of LUZ24. For the σ^{70} promoters, putative -35 and -10 boxes are indicated by grey boxes, as are the stem-loop regions of the putative factor-independent terminators.

Location	Promoter sequence	Activity [a] (Miller Units)
511-556	CGGGGGTTGACAAGGGATCGGTGAAGCGGTATAGTTCGCCCCGGTCAAG	357
615-660	CACCGCTTGACACTGTAAGGCTCAGTCGGTATGATGGGCACCACTCCCT	237
4828-4873	TAGGCCTTGACATTCTGCTGGAAGTGTGGTATAATAACCTTAAGGTGTC	120
20312-20357	TGGGGGTTGACTTTCAGCCCCCTCTGTGGTATAATACCTTCTTCCCTAC	289
45121-45076	TCAAAAAGTGCTTGACAAGGTATCGAAAATGTAGTATAATAGACCTATA	684
44923-44878	AGAAAAAAGACTTGACAAAATAGAAAAAGTGTGATATAATAGTATTATA	845
30353-30308	ATAAAAGGTACTTGACAAAGTATCCAAAATGTAGTATAATAGACCTATA	1084
	Terminator Sequences (ΔG, kcal/mol)	
25725-25768	AAAAAAAAGGCCCCAAGGGTATCATCCCAAGGGGCCTTATCTTT (bidirectional)	(-21.5)
37886-37907	ACGGGGCCCTTCGGGGCCCTTGTTT	(-17.8)

[a] Measurements were performed in threefold, activity units were calculated as described by Miller (1992).

5.2.2 Genes involved in host conversion and metabolism

Since LUZ24 (and PaP3) diverged strongly from other currently investigated phages, only a limited number of protein functions (26%) could be predicted by similarity searches. In particular, proteins assumed to be involved early in the infection process during host takeover (gp1-17) differ considerably, even among related phages, presumably because they are very host- or environmental specific. One exception is ORF2, which is conserved in PaP3 among all ϕKMV-like viruses (Chapter 4) and in phage PA11, suggesting a crucial role in infection and/or conversion of *P. aeruginosa*.

The middle genome region of LUZ24 encodes proteins involved in biosynthesis (gp18-22) followed by a phage T7-type genome replication module, with a primase-helicase (gp24) encoded directly upstream the DNA polymerase. Based on mutual sequence comparisons, we propose the annotation of a new ORF in PaP3 (ORF23.1, 21674-21465), which is perfectly conserved in LUZ24 (ORF46). While the PaP3 DNA polymerase is composed of two separate polypeptides (Tan *et al*., 2007), LUZ24 carries an endonuclease VII (ORF35) inserted within the 'core' polymerase, splitting the DNA polymerase gene in three separate parts (ORFs 25, 34 and 36, Figure 5.4). However, closer inspection revealed the presence of a group I intron-like sequence, which may have inserted into the phage genome through the process of homing (Sandegren and Sjöberg, 2007). To confirm the presence of this predicted intron, total RNA was isolated from infected *P. aeruginosa* cells. PCR performed on first-strand cDNA yielded smaller products compared to control reactions on LUZ24 DNA template (Figure 5.4). Subsequent cycle sequencing using ORF primer P_F and internal primers P_{f1}, P_{f2} and P_{ir} revealed the absence of the predicted 668 bp fragment (bp 19145-19812) in the cDNA (Figure 5.4). The self-splicing of this intron completely restores the reading frame of the large DNA polymerase subunit.

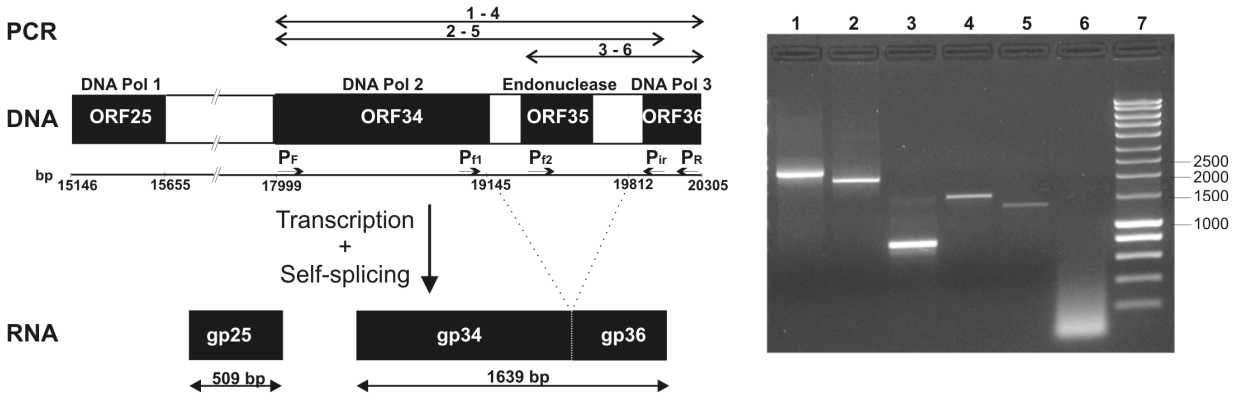


Figure 5.4. Schematic representation of the genomic region encoding the three DNA polymerase polypeptides, with demonstration of *in vivo* splicing of the intron DNA interrupting ORF34. Lanes 1, 2 and 3 of the 1% agarose gel contain PCR product obtained on LUZ24 DNA with primers P_F-P_R, P_F-P_{ir} and P_{f2}-P_R, respectively. Lanes 4, 5 and 6 contain PCR product obtained from a cDNA template, using the same primer combinations. Lane 7, Smartladder (Eurogentec).

Despite the lack of nucleotide similarity to other phage introns, the embedded endonuclease is related to the intron-associated endonuclease of *Synechococcus* phage S-PM2 (Mann *et al*., 2005) and resembles (among others) free-standing endonucleases of T7-like *Yersinia* phages Berlin and phiYeO3-12. A comparable 787 bp insertion containing a similar endonuclease is also present in the DNA polymerase gene of T7-like phage Pf-WMP3 (Liu *et al*., 2008), strongly suggesting the presence of a self-splicing intron in the genome of this cyanophages which was overlooked by these authors.

The presence of this intron within the DNA polymerase gene of LUZ24 confirms a common theme observed in many phage genomes, i.e., that phage introns target highly conserved regions within functionally important genes. It is speculated that these introns confer a selective advantage to the phage by offering the possibility of regulating the expression of the intron-containing genes by the regulation of splicing (Mann *et al*., 2005). By providing the first experimental evidence of a self-splicing intron in the genome of a *Pseudomonas* phage, we support the current hypothesis that there seems no barrier for intron maintenance within the phages of the Proteobacteria (Bonocora and Shub, 2004).

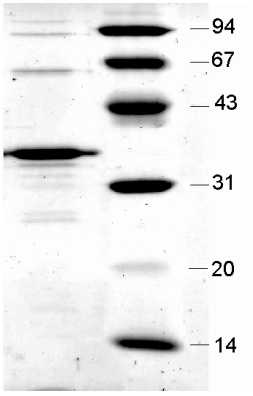
5.2.3 Genes involved in particle formation and host lysis

The genes transcribed from the right arm (Figure 5.2) are involved in virion assembly and host lysis, and follow a conserved gene order with the large subunit of the terminase (gp66) encoded downstream the portal (gp65), scaffolding (gp63) and the capsid protein (gp62). Following this

conserved genomic organization, upstream ORFs probably encode the internal core, tail tube and tail fibers, but only a tail collar domain (gp57) could be predicted through similarity searches. To obtain a more comprehensive view of the structural region, individual proteins within the mature phage particle were identified with ESI-MS/MS. Mass spectrometric analysis of denaturated phage particles, with and without prior fractionation on a 1D-SDS PAGE gel, identified nine structural gene products with a total protein coverage ranging from 13 to 88% (Table 5.3). This analysis delineates the structural region from gp49 to gp65, and confirms gp62 as the most abundant protein in the particle. Surprisingly, gp57 has a tail collar domain and displays similarity to tail fiber proteins of many *Yersinia* phages, but is not identified in this analysis. This is most likely due to the underrepresentation of arginine and lysine residues in its primary protein sequence, which precludes suitable trypsinization for peptide identification. All other non-identified predicted proteins in this region are rather small (<12 kDa) and may have fallen below the threshold for detection.

Table 5.3. ESI-MS/MS analysis of structural phage proteins. The protein of structural LUZ24 proteins extracted from purified virions (12% SDS-PAGE gel) next to a Low-Molecular Weight Ladder (kDa) is shown on the right.

Gene product	Mass (kDa)	Length (AA)	AA coverage in MS (%)	Predicted product function
gp49	31.6	296	181 (61.1)	unknown
gp50	112.0	1050	322 (30.5)	Structural protein
gp51	53.5	506	82 (16.2)	Unknown
gp52	12.6	127	77 (60.6)	Unknown
gp53	32.0	314	48 (15.3)	Unknown
gp56	57.5	514	73 (14.2)	Structural protein
gp60	24.0	209	27 (12.9)	Structural protein
gp62	35.0	318	281 (88.4)	Major head protein
gp65	81.1	707	254 (35.9)	Putative portal protein



The encoded terminases of LUZ24 (gp66) and PaP3 (gp3) are strongly related to the well-characterized terminase of phage P22 (Blastp e-value of 4E-69), which catalyses the sequential headful packaging of P22 DNA leading to both terminal redundancy and circular permutation (Wu *et al*., 2002). The lysozyme of LUZ24 (gp67) is encoded between both subunits of the terminase, and carries the conserved catalytic residues E32 and D41. It lacks a cell wall binding module, but is probably able to reach the peptidoglycan layer with the help of gp31. This short protein has three transmembrane domains and carries a positively charged intracellular C-terminal end (DRVFHKKA), typical for class I holins (Wang *et al*., 2000).

5.3 Discussion

It seems highly unlikely that the newly isolated phage LUZ24 has a temperate nature, despite being very homologous (71% sequence identity) to phage PaP3 which was reported to be temperate (Tan et al., 2007). LUZ24 shares its bidirectional genome organization with a vast number of virulent phages like cyanophages P60, Pf-WMP3 and Pf-WMP4, roseophage SIO1 and vibriophage VpV262 (Figure 4.7). Analogous to the latter phages, it carries DTRs at the genomic ends and lacks any gene or regulatory systems which suggest integration or lysogenic behaviour in its host bacterium. The *P. aeruginosa* $tRNA^{lys}$ gene is proposed as PaP3 integration site (Tan *et al*., 2007), but in contrast to other tRNA-integrating phages like mycobacteriophage Ms6 (Freitas-Vieira *et al*., 1998) and *Streptococcus* phage T12 (McShan and Ferretti, 1997), no site-specific recombinase is encoded upstream or downstream the PaP3 *attP* site. Although filamentous ssDNA phages like ctxϕ exist which recruit host-encoded recombinases, underlying mechanisms differ profoundly and integration-associated proteins like RstB are encoded in the ctxϕ genome (Davis and Waldor, 2000). Furthermore, neither immunity nor reactivation of the integrated PaP3 DNA was demonstrated by Tan and colleagues (2007). These arguments call into question the temperate nature of PaP3, although very closely related temperate and lytic phages have been reported before, i.e. phages L5 and D29 (Ford *et al*., 1998).

Another intriguing finding is that the end structures of virion chromosomes of both PaP3 (5'-protuding cohesive ends) and LUZ24 (184 bp DTRs, Figure 5.3) differ from those observed in P22, despite the fact that their terminases are strongly related. The same observation was made in the recently sequenced *E. coli* phage phiEco32, which carries 193-bp direct repeats and also encodes a P22-like terminase (Savalia *et al*., 2008). At first glance, this seems to contradict current theory that the structure of virion DNA ends can be accurately predicted based on amino acid similarity of the predicted terminase (Casjens *et al*., 2005). However, it is common among all the relevant terminases that the large terminase subunit does not determine the cut sites through sequence specificity. Since the four cuts in LUZ24 and PaP3 are made at different sequences, this might suggest that some other component of the packaging apparatus leads the terminase to its cut sites. This might explain in turn how the large terminase subunits can be 97% identical between LUZ24 and PaP3 and yet participate in a different end cutting strategy. This hypothesis implies, however, another (yet unspecified) function that must differ between LUZ24 and PaP3.

Another possibility is that the PaP3 ends are not correctly determined. Two reasons for this can be (a) Tan and colleagues (2007) performed a PCR with outwards facing primers and claimed to have amplified "across the ends", which would be indicative for the presence of cohesive genome ends.

However, if there was a terminal redundancy, they could have gotten amplification if one of the primers (whose sequences are not given in the paper) was within the terminal redundancy. (b) The experiment that these authors claim to have done to determine the PaP3 end structure seems to be perfectly impossible as described. They claim to have cloned the end fragment (a process which destroys any end structure that might be present) and used this clone to determine the end structure. Even if the experiment was done in fact with a purified virion DNA end fragment, the superfluous use of S1 nuclease (which efficiently digests dsDNA ends) without any controls may lead to erroneous interpretations.

Despite the differences stated above, LUZ24 (isolated in Belgium) and PaP3 (isolated in China) remain closely related phages and represent a new genus within the *Podoviridae* family, which was recently ratified by the ICTV. Because of the doubt which was raised on the temperate nature of PaP3, the proposed genus was termed after LUZ24.

Chapter 6

Bacteriophages LUZ7 and LIT1 are the first characterized N4-like *Pseudomonas* phages

Bacteriophage N4 is a lytic phage specific for *E. coli* K-12 originally isolated from the sewers of Genoa (Schito et al., 1967). The genome of N4 comprises 70,153 bp (NC_008720), encodes 72 proteins and is delineated by direct repeats varying in length between 390 and 440 bp (Ohmori et al., 1988). N4 is unique in using three different RNA Polymerases (RNAP) during its growth cycle, among which a giant (3500 aa) virion-encapsulated RNAP that is injected in the host cell upon infection. The central domain (mini-vRNAP) of this enzyme contains all RNAP functions and initiates RNA synthesis from three hairpin DNA promoters in the first 500 bp of the genome (Davydova et al., 2007). As this process pulls the next 10 to 40 kb out of the virus into the cell, N4 is the only bacteriophage known to date that does not require the activity of the host RNAP for transcription of its early genes (Choi et al., 2008). Injection of the remaining DNA requires transcription of the middle genes, carried out by a phage-encoded, heterodimeric T7-like RNAP which requires an additional N4 protein (gp2) for active transcription (Carter et al., 2003). One of the middle gene products, the N4 single-strand DNA binding protein, activates the E. coli σ^{70}-holoenzyme at late gene promoters (Beck et al., 2007).

Although the genus of the "N4-like viruses" was already recognized in the VIIIth ICTV report due to its unique characteristics (Hendrix and Casjens, 2005), N4 was a genetic orphan for over 40 years. Only very recently, phages DSS3ϕ2 and EE36ϕ1 infecting marine roseobacters *Silicibacter pomeroyi* DSS-3 and *Sulfitobacter* sp. EE-36, respectively, were reported to be related to N4 (Zhao et al., 2009). The *de novo* analysis of structural phage proteins identified (weak) similarity between virion proteins of N4 and two of our newly isolated Podoviruses, LIT1 and LUZ7 (Section 3.2). This chapter describes the characterization of both phages, in comparison with the well-studied phage N4.

6.1 Phage characteristics

Phages LUZ7 and LIT1 were visualized as *Podoviridae*, with a 70-nm capsid diameter, a small tail and host ranges of 28% and 14%, respectively (Figure 6.1; Section 3.1). Both phages are isolated and propagated using clinical *P. aeruginosa* strains (US449 and Br257, resp.) as host, on which they form clear plaques surrounded by a small halo (Figure 6.1). Host range studies of LUZ7 and LIT1 identified 28% and 14% of our *P. aeruginosa* collection to be susceptible to phage infection (Table 3.2).

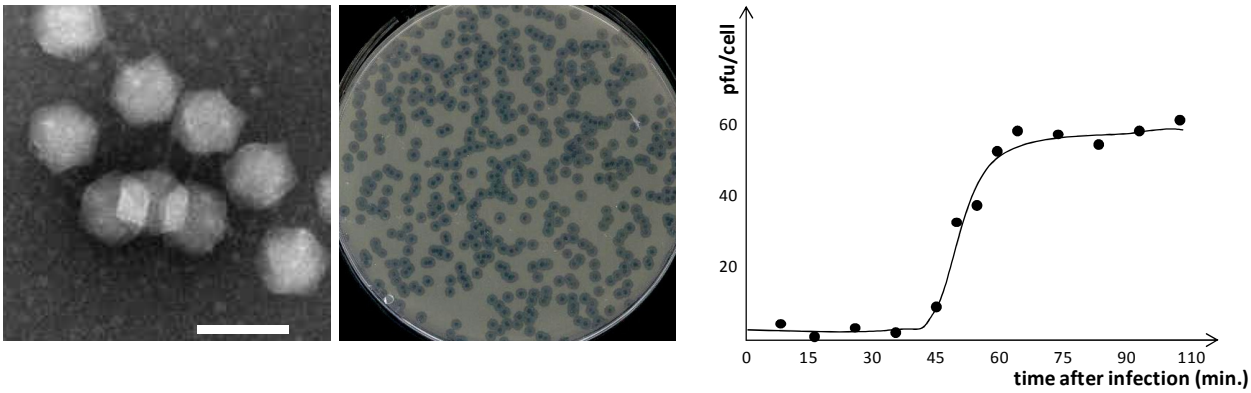


Figure 6.1. Particle analysis of the N4-like *Pseudomonas* phage LUZ7. From left to right, an electron microscopic image, plaque morphology and a one-step growth curve are shown (see text for more details). The scale bar represents 100 nm.

A distinct feature of N4 is its lysis-inhibited infection cycle, as it does not actively lyse the host cell and infected cells continue to grow while phage particles accumulate inside the cells. Eventually, phage burst starts asynchronously 180 min. postinfection, and releases over 3,000 phages per cell (Schito, 1974). This prolonged lysis period was also observed for N4-like phages DSS3ϕ2 and EE36ϕ1 (Zhao et al., 2009). However, when one-step growth curves of LIT1 and LUZ7 were constructed under standard laboratory conditions, *P. aeruginosa* cells were already lysed after 50 and 45 min., respectively releasing 60 and 55 newly formed phage particles per cell. In addition, no cell growth was observed when monitoring phage infection using a Leica DM Optical microscope (data not shown), suggesting profound differences in lysis mechanism (Section 6.2.4).

6.2 Genome analysis

6.2.1 Determination of genomic phage ends

Initially, the genome sequences of LU7 and LIT1 were assembled into circular contigs of 74.3 and 71.9 kb with G+C contents of 53.4 and 54.9 %, respectively, a significantly lower percentage than *P. aeruginosa*. However, we were surprised to find that experimentally obtained restriction enzyme fragment patterns were not in agreement with those predicted from the sequence (Figure 6.2). In most cases, one predicted fragment was missing or the fragment was significantly shorter than calculated. At the same time, one or two unexpected (additional) fragments appeared. For example, *Eco*RI digestion of LUZ7 DNA yielded additional fragments of approx. 6,500 bp and 8,000 bp, respectively, while the predicted 14,130-bp fragment was missing (Figure 6.2).

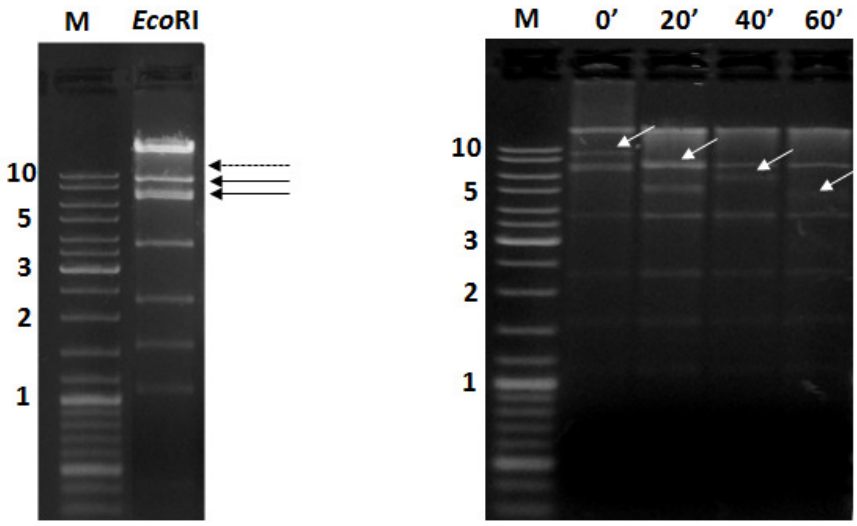


Figure 6.2. Enzymatic investigation of the genome ends of phage LUZ7. A. *Eco*RI digestion of LUZ7 DNA. *In silico* analysis of the preliminary LUZ7 genome predicted fragment sizes of 1.1, 1.5, 2.2, 3.8, 6.7, 14.1, 19.6 and 24.8 kb. Two additional bands were observed (black arrows), and the 14.1 kb fragment was missing (broken line arrow). **B.** *Eco*RI-digested LUZ7 DNA after treatment with Bal31 for the indicated time intervals. One restriction fragment disappearing over time is marked by white arrows. M, DNA marker (kb).

DNA treatment with BAL-31, an exonuclease that degrades double-stranded linear DNA from both ends simultaneously, and subsequent cleavage of the DNA by restriction enzymes revealed that the two non-predicted bands were simultaneously shortened over time (Figure 6.2 for LUZ7), and these fragments therefore must represent the physical genome ends. As a consequence, LUZ7 and LIT1 could not have a circularly permuted genome, because this would result in an even, simultaneous degradation of all restriction fragments (Loessner et al., 2000). Runoff Sanger sequencing of the left and right ends (as described for LUZ24, Figure 5.3) revealed the sequence of both ends and the extent of the terminal redundancy of 660 and 655 bp for LUZ7 and LIT1, respectively. When the

reassembled genomes were used for *in silico* restriction patterns prediction, results were then in perfect agreement with our experimental data.

6.2.2 Genome annotation

Phages LUZ7 and LIT1 encapsulate linear genomes of 74,901 and 72,543 bp and encode 117 and 90 proteins, respectively (Table 6.1). In contrast to, for example, the close-knitted group of the 'ϕKMV-like viruses' (Chapter 4), LUZ7 and LIT1 share no conservation at the nucleotide level, and display protein similarity in only 60 of their gene products. No tRNA genes were predicted in both phage genomes, which were deposited in GenBank under (confidential) accession numbers FN_422398 and FN_422399.

Table 6.1. *In silico* analysis of the LUZ7 and LIT1 genome. Only ORFs displaying similarity to database entries or encoding structural features are shown.

ORF LUZ7	ORF LIT1	Mut. ident.[a]	Mw (kDa)	Strand	Closest homologue	E-value	Comments
1	**1**	33	9.9	+	No similarity		Contains signal peptide (P=0.98)
2	**-**	-	12.3	+	No similarity		Two transmembrane domains
3	**5**	35	10.7	+	No similarity		
5	**6**	25	12.4	+	No similarity		
6	**6**	26	12.3	+	No similarity		
7	**7**	40	11.8	+	No similarity		
8	**2**	27	9.1	+	No similarity		
-	**8**	-	7.3	+	No similarity		Two transmembrane domains
-	**9**	-	7.9	+	No similarity		Two transmembrane domains; signal peptide (P=0.98)
-	**10**	-	8.5	+	No similarity		Two transmembrane domains
-	**11**	-	8.0	+	No similarity		Two transmembrane domains
10	**-**	-	3.6	+	No similarity		One transmembrane domains; signal peptide (P=0.9)
12	**-**	-	8.4	+	No similarity		C-terminal transmembrane domains
13	**-**	-	7.1	+	No similarity		Two transmembrane domains
14	**13**	37	11.1	+	No similarity		
16	**14**	19	8.7	+	No similarity		
-	**15**	-	14.6		gp14 [N4]	2e-23	
18	**16**	49	9.5	+	No similarity		
19	**17**	50	15.7	+	No similarity		
20	**19**	72	35.9	+	RNAP1 [N4]	6e-23	C-terminus: T3/T7-like RNA polymerase [PHA00452, 4e-22]
22	**23**	69	46.9	+	RNAP2 [N4]	2e-45	N-terminus: DNA-dependent RNA polymerase [pfam00940, 2e-14]
29	**24**	43	7.7	+	No similarity		Two transmembrane domains; signal peptide (P=0.99)
30	**-**	-	18.3	+	No similarity		TraB pilus assembly protein [pfam06447, 2e-3]
33	**30**	69	6.7	+	No similarity		
35	**-**	-	23.5	-	gp36 [phig1e]	7e-08	
36	**31**	46	30.7	+	Clp Protease [*P. uenosis* 60-3]	2e-15	ATPase family associated with various cellular activities [pfam07724, 3e-8]
-	**32**	-	9.1	+	hypothetical [*P. fluorescens*]	3e-5	
37	**33**	48	38.8	+	gp24 [N4]	1e-61	ATPase family associated with various cellular activities [pfam07728, 3e-5]
39	**35**	41	44.6	+	gp25 [N4]	3e-34	Predicted metallopeptidase [pfam09967, 3e-9]
40	**36**	51	46.6	+	Helicase [N4]	1e-31	
41	**37**	33	20.3	+	No similarity		
42	**38**	65	98.4	+	DNA Polym. [N4]	0.0	DNA polymerase I - 3'-5' exonuclease and polymerase domains [COG0749, 9e-38]
43	**39**	55	21.4	+	hypothetical [*T.*	2e-15	Deoxycytidylate deaminase [cd01286, 1e-19]

(Table 6.1. Continued)

					kodakarensis]		
45	**40**	38	7.2	+	No similarity		
-	**41**	-	13.2	+	gp22 [N4]	1e-31	
46	**42**	30	94.6	+	rIIA [N4]	2e-51	Only N-terminal part conserved
47	**43**	44	55.2	+	rIIB [N4]	8e-57	Central part not conserved
48	**44**	29	8.8	+	No similarity		Two transmembrane domains
49	**45**	62	11.9	+	No similarity		Three transmembrane domains
51	**47**	14	22.3	+	No similarity		
52	**48**	68	17.8	-	tail protein [E1]	1e-3	Protein of unknown function [pfam10721, 6e-8]
53	**49**	60	19.2	-	hypothetical [*D. acidovorans*]	1e-18	
54	**50**	43	9.2	-	PA0648 [*P.aeruginosa*]	2e-25	
55	**51**	27	10.7	-	PA0647 [*P. aeruginosa*]	1e-33	
56	**52**	**(C-term)**	115.8	-	PA0646 [*P. aeruginosa*]	6e-16	N-terminus similar to various tail fiber proteins
	53	**(N-term)**			Tail fiber[Berlin]	5e-10	C-terminus: SGNH_hydrolase [cd00229, 2e-5]
-	**54**	-	25.0	-	PA0642 [*P. aeruginosa*]	0.1	
59	**56**	63	46.3	-	tail tip fiber [phage 86]	7e-12	Related to C-terminal part of many phage tail fibers Phage-related tail component [COG4733, 3e-4]
60	**57**	70	15.5	+	No similarity		
61	**58**	47	14.9	+	No similarity		
62	**59**	63	36.9	+	gp42 [N4]	3e-78	
63	**60**	65	81.1	+	gp43 [N4]	0.0	
64	**61**	63	28.0	+	gp44 [N4]	8e-61	
65	**62**	42	26.7	+	gp45 [N4]	6e-24	Single-stranded DNA-binding protein
66	**63**	76	13.7	+	ORF144 [P. damselae]	1e-12	
-	**64**	-	17.1	+	gp48 [N4]	9e-12	Hypothetical protein [PHA01971, 3e-16]
67	**65**	61	20.0	+	gp46 [N4]	8e-38	
68	**66**	28	13.0	+	No similarity		
69	**67**	38	14.2	+	No similarity		Signal Peptide (P=0.98)
71	**69**	27	12.8	+	No similarity		
72	**70**	38	7.1	+	No similarity		Two transmembrane domains
73	**71**	65	370.1	-	vRNAP [N4]	1e-87	Only C-terminal and central part conserved N-terminus: TolA-like domain [PRK09510, 1e-3]
74	**72**	49	56.7	-	No similarity		
75	**73**	70	16.9	-	gp52 [N4]	0.089	One transmembrane domains
76	**74**	51	82.1	-	gp53 [N4]	1e-10	Three transmembrane domains
77	**75**	58	35.4	-	gp54 [N4]	1e-16	
78	**76**	49	24.1	-	gp55[N4]	2e-24	
79	**77**	77	44.2	-	gp56 [N4]	2e-121	Major coat protein
80	**78**	46	44.9	-	gp57[N4]	1e-36	
81	**79**	24	12.8	-	No similarity		
82	**80**	63	80.7	-	gp59[N4]	3e-180	
83	**81**	68	15.1	+	dUTPase [Pelobacter]	2e-21	Deoxyuridinetriphosphatase [pfam00692, 1e-21]
84	**83**	44	28.0	-	gp67[N4]	1e-4	
85	**84**	67	64.3	-	gp68[N4]	6e-153	Terminase, large subunit
86	**85**	56	26.9	-	gp69[N4]	2e-41	
-	**88**	-	10.3	+	No similarity		Three transmembrane domains

[a] Mutual similarity between corresponding LUZ7 and LIT1 proteins is expressed as % amino acid identity.

The main reason for the difference in gene number between both phages is a cluster of 29 small genes encoded at the right end of the LUZ7 genome, which are absent in LIT1 (Figure 6.3). These genes, which are likely to be expressed late in infection, lack any similarity to current database entries and any recognizable structural features, so their function remains unknown. Many early genes also diverged strongly, but are probably functionally equivalent to their counterpart in the other phage, since many of them encode membrane-spanning domains (Table 6.1). Bacteriophages LUZ7 and LIT1 clearly share their overall genomic architecture and have 25 proteins in common with

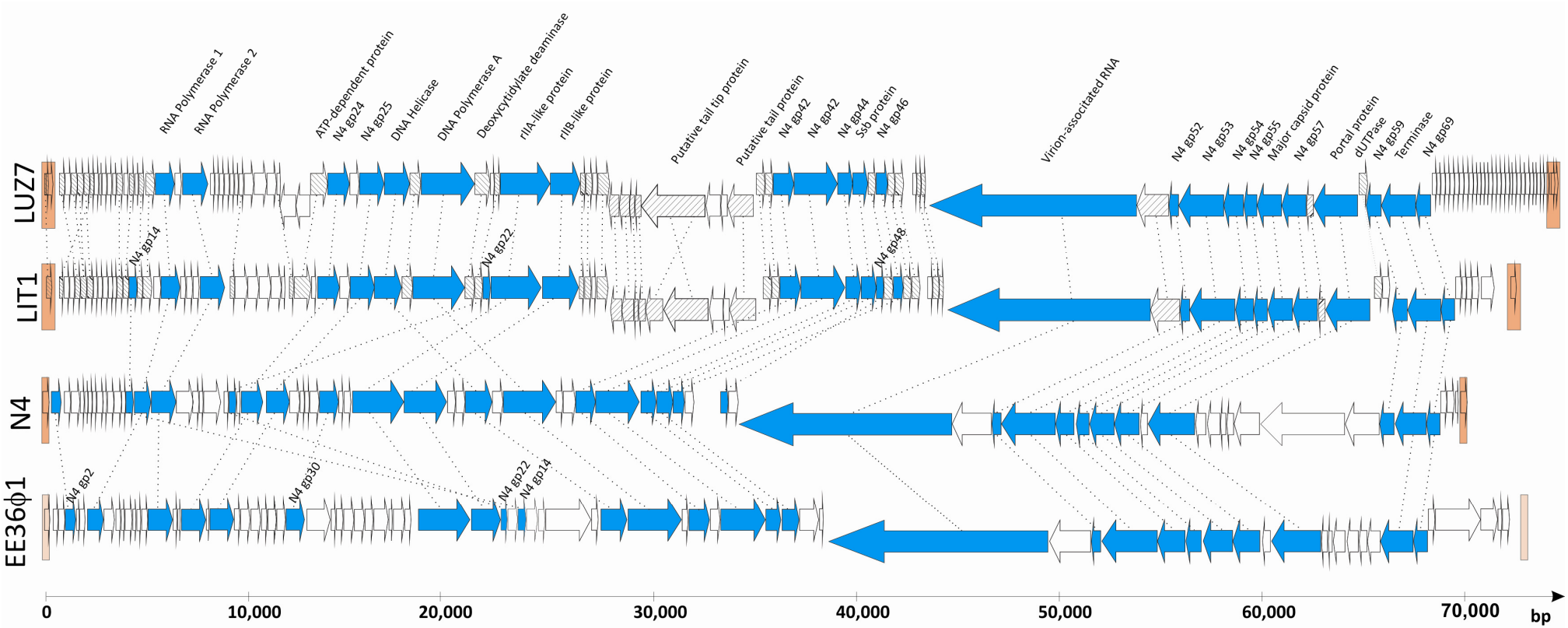


Figure 6.3. In silico analysis and comparison of four N4-like phage genomes. The arrows represent the predicted ORFs, and point at the direction of transcription. The direct repeats delineating the genomes (which were not reported by Zhao and colleagues (2009)) are indicated as orange boxes. Genes unique to *P. aeruginosa* infecting N4-like phages are hatched, and genes encoding proteins which share similarity (defined by BLASTp identification using default parameters) are connected by broken lines. Genes shared among N4-like phages are marked in blue, genes which are unique to each phage are white. As roseophage DSS3ϕ2 is very similar to EE36ϕ1 (Zhao et al., 2009), it is not included in this schematic drawing.

coliphage N4 and roseophages EE36ϕ1 and DSS3ϕ2, strongly suggesting a common origin. While gene order among the structural genes is strictly conserved, major rearrangements occurred among DNA replication genes (*e.g.*, N4-like ORFs 14 and 22, Figure 6.3). In general, genes involved in transcriptional control, DNA metabolism and capsid formation are conserved among the N4-like viruses, while major differences are evident in genes coding for host interaction, tail formation and host lysis.

6.2.3 Transcriptional control in N4-like *Pseudomonas* phages

An important gene in the LUZ7 and LIT1 genomes spans one-seventh of their entire genomes and encodes the virion-associated RNAP (vRNAP; 3,398 amino acids). Similar to the N4 RNAP, the two vRNAPs do not contain any cysteine residues. This could be an important requirement for the enzyme to pass through the host periplasm, which contains proteins that catalyze disulfide bond formation (Ritz and Beckwith, 2001). Interestingly, the C-terminal domain catalyzing the enzyme encapsulation (Kazmierczak and Rothman-Denes, 2006) and the middle domain containing all RNAP functions (Gleghorn et al., 2008) are conserved in the *Pseudomonas* phages, while the N-terminal domain required for injection of the genome into the host cell is not. In addition, although four T7-like RNAP motifs (motifs T/DxxGR, A, B and C) described in the N4 vRNAP are broadly conserved in LUZ7 and LIT1, reported catalytic residues (Kazmierczak et al., 2002) are absent in the latter phages, as is the conserved hairpin promoter.

Table 6.2. Putative early and middle promoter sequences, and factor-independent terminators in the genome of phage LUZ7. The conserved motifs and stem-loop structures of both elements are underlined.

Function	Name	Location[a]	Strand	Sequence
Early transcription				
	P_{e1}	5	+	TGCAAGGAGCAAAACCATGAAGACCATAGCT
	P_{e2}	8	+	ACCAAGGAACCAAACCATGAACCTGACTTTC
	P_{e3}	17	+	TCTCATGAGACAAACCATGAAGGGAAGGGAC
Middle transcription				
	P_{m1}	41	+	CATATGCTGCAAACCATGCTGAAAAGGAGTAAGCA
	P_{m2}	42	+	TCCCACTACGAAGCCATGCTGAAAAGGGTAACCTC
	P_{m3}	47	+	CTGCGTGAAGAAGCCATGGTGAAGAAGAATGCCAC
Factor-independent terminators				
	T_1	15	+/-	AAAACCATGTGCTAACCCATGGTTTT
	T_2	21	+	CCTCCCTTGGCTTCGGCTAAGGGAGGTTCCTTTT
	T_3	22	+	CTTCCCCTGAAAAGGGGAAGGCATCCTTTTTTT
	T_4	47	+	GGGGGAGGCAACTCCCCCACTCTTT
	T_5	49	+	AACCCCTCTTCGGAGGGGTTTTTTATT
	T_6	52	-	ATAAAAAAACCCCCTACAGGTAGCGACATGCCTGTAGGGGGTT
	T_7	63	+	GGGCCGCTTCCAGCGGCTCTTT
	T_8	72	+/-	AAAAAGAAAGGGGAGCATTAGCTCCCCTTCTTCTTT
	T_9	77	-	AAAAAGGGGGACCGAAGTCCCCCT
	T_{10}	79	-	AAAAAAGGGGAGGTTTCCCTCCCCT

[a] Upstream and downstream ORF are given for terminators and promoters, respectively.

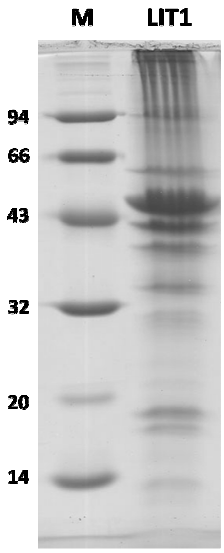
Based on these observations, we suggest subtle differences in infection mechanism and early transcription between N4 and the N4-like *Pseudomonas* phages. The only conserved motif in the early region is an 11 nt-sequence (5'-CAAACCATGAA-3', Table 6.2), which might serve as initiation site for vRNAP activity. This transcription leads to the formation of a heterodimeric RNAP (gps20 and 22 in LUZ7, gps19 and 23 in LIT1), which carries out middle transcription from three promoters located after approx. 20 kb in the genome (Table 6.2).

Late transcription in N4 is activated by the single-strand binding protein gp45, which interacts with a greatly conserved region at the C-terminus of the β' subunit of the *E. coli* RNAP (Miller et al., 1997). This protein does not bind to dsDNA, and seems to stimulate a (still unknown) step in transcription that occurs after initial binding of the polymerase to the promoter (Beck et al., 2007). As homologues of this protein are present in LUZ7 and LIT1 (gp65 and 62, resp.), these phages most probably also depend on the host RNAP for transcription of their late genes.

6.2.4 Particle formation and host lysis

The mature virion of N4 is composed of ten gene products, including the major capsid protein (gp56), the vRNAP (gp50), the tail sheath protein (gp65), the appendages (gp66) and the portal protein (gp59). The capsid is decorated by 175 copies of a gp17 (Choi et al., 2008). Using a combination of *in silico* and MS analysis, significant differences in particle composition were observed between N4 and the N4-like *Pseudomonas* phage LIT1.

Firstly, the decorating protein (gp17) and the proteins forming the tail appendages which are attached to the portal protein (gp66) are absent in LIT1, as is gp65, which attaches the phage to its primary receptor (McPartland and Rothman-Denes, 2009). In LIT1, genes 65 and 66 are apparently replaced by a cluster of genes located within the replication module on the opposite coding strand. These proteins display strong similarity to *P. aeruginosa* proteins and various tail proteins from other *Podoviridae* (Table 6.1), and were identified as part of the phage particle by ESI-MS/MS analysis (Figure 6.3). In contrast to its counterpart in N4, gp74 was identified as a structural protein. This protein plays a particularly intriguing role in particle formation is, as it contains three transmembrane domains at its C-terminus which might serve as a membrane anchor. However, this anchor mechanism may also attach this protein to *P. aeruginosa* membrane vesicles which are likely to be co-purified along with phages in CsCl gradients, leading to a false positive identification as a structural protein.



Gp	Mass (kDa)	No. Peptides	Coverage (%)	Homologue in N4	Proposed role in N4[a]	No. N4 copies[b]
52	29.8	2	16.4	Absent		
53	79.5	13	41.9	Absent		
54	25.0	3	30.4	Absent		
55	11.7	2	47.7	Absent		
56	46.4	7	31.0	Absent		
71	370.1	76	46.7	gp50	Virion RNAP	1 or 2
72	56.7	14	61.1	gp51	Structural protein	
73	16.9	2	24.5	gp52	Structural protein	
74	82.1	5	13.9	gp53	N.I.	
75	35.4	11	60.7	gp54	Structural protein	
77	44.2	16	82.7	gp56	Major Capsid	535
78	44.9	1	3.0	gp57	N.I.	
80	80.7	21	69.1	gp59	Portal protein	12
83	28.0	4	42.6	gp67	Structural protein	

[a] N.I. Homologues are present in N4, but were not identified by Choi et al., (2008). [b] As determined by Choi et al. (2008) by cryoelectron microscopy.

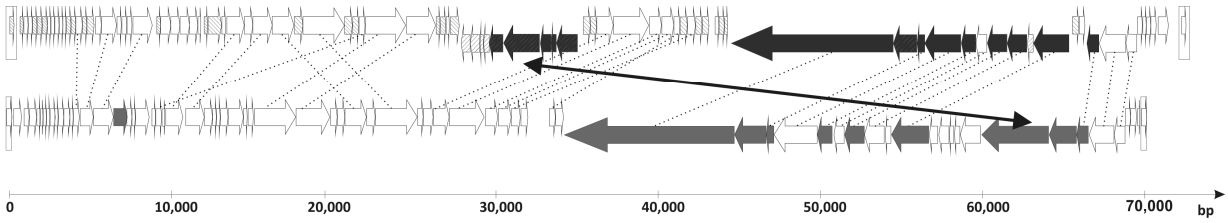


Figure 6.3. Analysis of the proteins composing the LIT1 particle. Structural phage proteins were extracted from purified virions, separated on a 12% SDS-PAGE gel (top left; M, marker in kDa) and identified using ESI-MS/MS. These identified proteins are marked in black in the figure below (upper genome is LIT1), and compared to the proteins identified as part of the N4 particle, which are indicated in grey. The rearranged position of the tail proteins is shown with an arrow.

Another interesting observation is that the gene encoding the N4 lysin, gp61, is deleted in the genomes of the N4-like *Pseudomonas* phages. In addition, no lysis-related protein can be predicted by similarity searches. Both LIT1 and LUZ7 do encode homologues of the RIIA and RIIB proteins of phage T4, which play a major role in protecting *E. coli* cells from superinfection. Normally, T4 infections are terminated by host lysis at 25 min. However, if T4-infected cells undergo secondary infections, lysis does not occur at the normal time, and the cells enter the lysis inhibition (LIN) state. In this state, the vegetative cycle can be prolonged up to several hours while virions continue to accumulate (Ramanculov and Young, 2001), as was noticed for N4 infections (Schito, 1974). While standard growth curves (which exclude superinfection) revealed rapid host lysis (Figure 6.1), LIN by superinfection is currently under investigation for phages LUZ7 and LIT1.

6.3 Discussion

The genomic characterization of LIT1 and LUZ7 opens a wide range of opportunities for future research. For example, as no lysis-related genes were identified despite rapid host lysis, these phages might encode a new type of murein hydrolase, as was already reported for N4 (Stojkovic and Rothman-Denes, 2007). The three-step transcriptional control is highly interesting and can be investigated further by rifampicin (which specifically inhibits bacterial RNAP) treatment of infected cells at various stages in phage infection. In addition, the *P. aeruginosa rpoC::TAP* strain (Section 4.3) can be infected by LIT1 and LUZ7, and proteins attached to the host RNAP can be co-purified. In this way, the functionality of the N4 ssb protein homologue which specifically binds to the β' subunit and directs transcription to the late promoters can be investigated. As a third example, the tails of LIT1 and LUZ7 are encoded by a cluster of genes which is totally absent in N4 and therefore has probably direct implications towards host receptor and tail morphology; Ultrastructural studies elucidating these differences would be highly interesting.

Chapter 7

Comparative analysis of the widespread and conserved PB1-like *Myoviridae* [4]

The strictly virulent and historically important *P. aeruginosa* phage PB1 was first described almost half a century ago (Holloway et al., 1960). This myovirus carries characteristic conspicuous capsomers that appear as 8 nm cup-like depressions on the phage heads. The tail exhibits no transverse striations but presents a criss-cross pattern (Ackermann et al., 1988), a rare feature that has only been observed with Salmonella phage Felix O1. It has long been known that PB1 and its relatives use the bacterial lipopolysaccharide layer as receptor (Jarrell and Kropinski, 1977). Over the years, no less than 42 phages were reported to be PB1-like, mainly based on cross-DNA hybridization and morphological studies (Ackermann and Dubow, 1987; Krylov et al., 1993; Pleteneva et al., 2008). One of the longest known relatives of PB1 is *Pseudomonas* phage F8, which was part of the Lindberg typing set (Lindberg and Latta, 1974; Section 2.5.3). Phages belonging to this group (e.g., phage 14-1) have been and are being used in human phage therapy trials (Merabishvili et al., 2009). In our collection, LBL3 and LMA2 were isolated from environmental water samples, and were suggested to be related to phage F8 by *de novo* mass spectrometric analysis of their main structural proteins (Chapter 3.2).

Despite being clearly a widespread and highly efficient virulent group of phages, no report on the genome composition of these phages was available. In this chapter, we describe and compare the genome sequences and particle composition of phage PB1 and four related phages (LBL3, LMA2, 14-1 and SN) to the originally unannotated *Pseudomonas* phage F8 sequenced in the same period (Kwan *et al*., 2006) and *Burkholderia* phage BcepF1, isolated from agricultural soils in New York. The genomes of PB1, SN and 14-1 were determined in collaboration with Andrew Kropinski (Laboratory for Foodborne Zoonoses, Guelph, Canada), Konstantin Miroshnikov (Institute of Bio-organic

[4] Data presented in this chapted are adapted from Ceyssens, P.J., *et al.* (2009) *Environmental Microbiology* **11**: 2874-2883.

Chemistry, Moscow, Russia) and Wesley Mattheus (Laboratory of Gene Technology, K.U.Leuven), respectively.

7.1 Phage characteristics

Pseudomonas phages SN, 14-1, LBL3 and LMA2 were isolated from water samples taken from very distinct natural environments, by different researchers, over a four-year period (Table 7.1). All phages displayed clear plaques (2-3 mm diameter), were easily cultivable using different strains of *P. aeruginosa* as host (Table 3.1) and are stable during long-term storage at 4°C in phage buffer. Transmission electron microscopy on uranyl acetate negatively stained particles revealed them to be morphologically indistinguishable from PB1, carrying non-flexible, contractile tails (ca. 140 nm) and isometric capsids with a diameter of 74 nm (Figure 7.1A). Based on these observations, these phages can be classified into the A1 morphological group of the *Myoviridae*. Virion particles of LBL3 were stable in acidic conditions: on average, 26% and 62% of the particles retain their infectivity after one hour incubation at pH2 and pH3, respectively (Figure 7.1B).

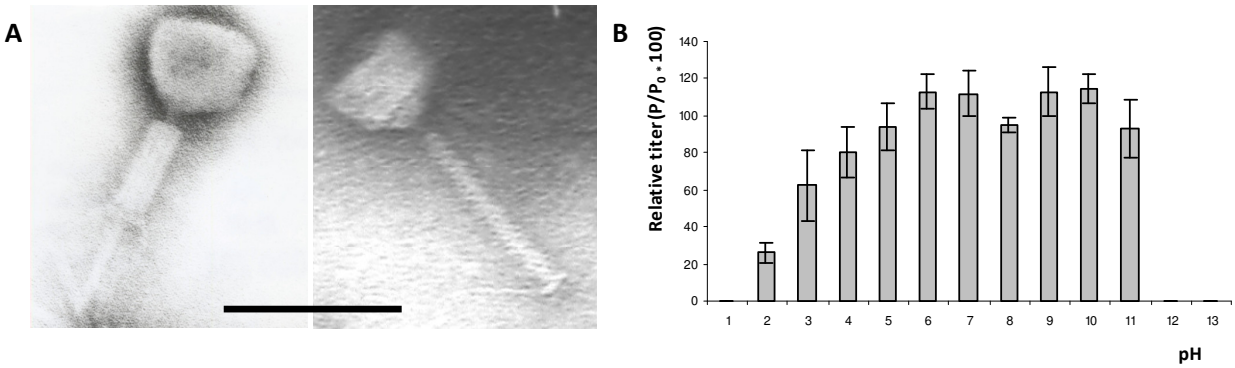


Figure 7.1.A. Electron microscopic images of negatively stained LBL3 (left) and SN (right) particles. Scale bar represents 100 nm. **B.** Relative amount of infectious LBL3 particles after one hour incubation at different pH levels. Three independent assays were performed, standard deviation is indicated.

Another remarkable feature of phage LBL3 is its broad host range, as it is able to lyse 53% of 122 *P. aeruginosa* strains (Table 3.2). Although a comparable broad host range was already reported for phages SN (Pleteneva et al., 2008) and 14-1 (Merabishvili et al., 2009), this does not apply on phage LMA2 which lyses only 22% of clinical *P. aeruginosa* strains (Table 3.2).

7.2. Genome analysis

7.2.1 Genome properties and annotation

After phage DNA isolation, DNA restriction fragments were subjected to a heat treatment (80°C) followed by rapid chilling prior to electrophoresis. This procedure did not alter restriction patterns (data not shown), excluding the possibility of cohesive genome ends. Restriction analysis after DNA treatment with BAL-31 (similar to the procedure described for LUZ7 in Section 6.2.1) revealed two specific fragments which shortened over time, while other fragments retained their length (Figure 7.2 for *Sal*I digest of phage 14-1). These bands represent the physical genome ends (Loessner et al., 2000), showing these phages carry a non-permuted, linear dsDNA genome.

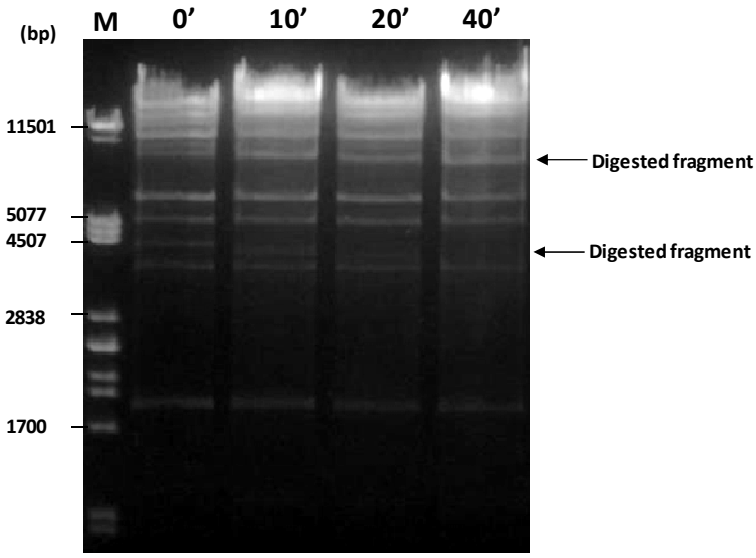


Figure 7.2. Enzymatic analysis of the physical phage genome ends. Agarose gelelectrophoresis (1%) of *Sal*I-digested 14-1 DNA after treatment with Bal31 for the indicated time intervals. Lanes: M, DNA marker (λ/*Pst*I); Restriction fragments reducing in size depending on the time of Bal31 treatment are indicated by arrows.

The complete genome sequences of phages PB1, LBL3, LMA2, SN and 14-1 were determined, and showed little variation in genome lengths (between 64,427 bp for LBL3 and 66,530 bp for LMA2). Remarkably, they display 87.2 to 93.5 % overall nucleotide identity to *Pseudomonas* phage F8 (Table 7.1) and carry very few insertions and deletions compared to this phage. In contrast to other lytic *Pseudomonas* phages (e.g. gh-1 and ϕKMV), the average GC content (55.5%, Table 7.1) in PB1-like phages is significantly lower than that of their host *P. aeruginosa* (66.6%). This GC content is fairly constant throughout the genome, and minor local decreases correlate to intergenic, regulatory regions (Figure 7.3).

Table 7.1. General properties of PB1-like phages described in this study.

Phage	Isolation details Source & Location	Year[a]	Genome size (%GC)	ORFs	Unique ORFs	DNA identity to F8 (%) [b]	Accession No.
PB1	Sewage, Edinburgh, Scotland	1960	65,764 (55.5)	93	-	93.5	NC_011810
SN	Lake, Pokrov, Russia	2004	66,390 (55.6)	92	2	87.2	NC_011756
14-1	Sewage, Regensburg, Germany	2004	66,238 (55.6)	90	-	87.4	NC_011703
LMA2	River, Maastricht, Holland	2007	66,530 (55.6)	95	2	87.6	NC_011166
LBL3	Pond, Blanes, Spain	2006	64,427 (55.5)	88	2	88.6	NC_011165
F8	Unknown	1972	66,015 (55.6)	93	1	100	NC_007810
BcepF1	Soil, New York, USA	2003	72,415 (55.9)	127	-	-	NC_009015

[a] Bergan (1972) tested phage F8 for suitability as a typing phage and suggested that it was isolated by Postic and Finland (1961), but there is nothing in this paper to support this. [b] Using Stretcher (EMBOSS package).

Despite this deviating GC content and the absence of phage-encoded tRNA genes, translational efficiency seems to be ensured by sharing the same dominant codon for each amino acid (except for glutamate) with its host (data not shown). The high degree of DNA sequence identity between the five sequenced phages and F8 allowed comparative ORF prediction for these phages. Based on mutual genome comparisons and aided by (t)BLASTx analysis, between 88 (LBL3) and 95 (LMA2) ORFs were predicted (Table 7.2, Figure 7.3), including smaller conserved ORFs (<150 bp, e.g. ORF 20), which are usually ignored in the annotation of phage genomes. For reasons of consistency and to ease comparison of the phage genomes, gene numbering of phage F8 was maintained for all genomes, and names of additional or unique genes are extended by a .1 suffix.

Table 7.2. Analysis of gene products encoded in the PB1-like bacteriophages.

ORF[a]	Strand	# AA	MW (kDa)	Absent in	Closest phage homologues beyond the PB1 genus [b] / Conserved domains – structural predictions	e-value
2	-	49	5.3	LBL3, SN	Coiled coils	
4	+	461	52.6		Terminase large subunit [Enterobacteria phage HK620] Terminase_3 [pfam 04466] with N-terminal motif present [pfam 07570]	3e-71 2e-39
7.1	-	75	8.7	All but LMA2	Coiled coils	
8	-	133	14.5		Signal peptide (P=0.94) and 2 transmembrane domains	
9	-	260	29.1		Coiled coils	
10	-	146	24.4		Signal peptide (P=0.81) and 3 transmembrane domains	
12	-	311	34		C-terminus similar to gp52 [Burkholderia phage BcepGomr], coiled coils	5e-20
12.1	-	166	18.8	All but SN and 14-1	Similar to gp65 [Mycobacteriophage Gumball] Possible methyltransferase [COG4627]	3e-17 1e-3
13	-	116	13.6		Coiled coils	
14.1	-	51	5.4	All but LMA2 and PB-1	Signal peptide (P=0.84)	
17	+	766	84.4		Phage-related protein [pfam 09714], coiled coils Uncharacterized protein conserved in bacteria [pfam 09978]	2e-15 1e-73
18	+	279	31.5		Phage Mu protein F like protein.[pfam 04233] – viral morphogenesis Coiled coils	3e-08
19	+	68	7.4		gp02 [Burkholderia phage Bcep43]	5e-09
20	+	46	5.2		One transmembrane domain	
21	+	478	50.4		Conserved N-terminal domain [COG 3566]	5e-16
22	+	212	21.6		gp24 [Burkholderia phage BcepB1A]	8e-07

(Table 7.2. Continued)

23	+	383	41.7		gp23 [Burkholderia phage BcepB1A]	1e-69
24	+	146	16.4		gp22 [Burkholderia phage BcepB1A]	4e-05
25	+	156	16.9		gp21 [Burkholderia phage BcepB1A]	6e-30
26	+	133	15.2		gp19 [Burkholderia phage BcepB1A]	7e-14
27	+	184	21.2		gp18 [Burkholderia phage BcepB1A]	1e-04
28	+	194	21.5		phiPLPE_42 [Iodobacteriophage phiPLPE]	5e-11
29	+	505	53.6		gp17 [Burkholderia phage BcepB1A]	1e-69
30	+	151	15.9		gp16 [Burkholderia phage BcepB1A]	6e-27
31	+	108	12.2		gp12 [Burkholderia phage BcepB1A]	0.053
34	+	168	17.7		gp16 [Burkholderia phage BcepB1A]	8e-10
35	+	180	19.8		gp13 [Burkholderia phage BcepB1A]	2.4
38	+	859	94.3		Transglycosylase SLT domain [pfam 01464], coiled coils	1e-17
					transglycosylase [Enterobacteria phage PRD1]	6e-24
39	+	288	32.4		gp11 [Burkholderia phage BcepB1A]	3e-14
40	+	178	19.6		gp15' E' [Burkholderia phage BcepB1A]	1e-04
41	+	222	24.2		gp10 V [Burkholderia phage BcepB1A]	9e-36
42	+	432	43.2		Baseplate_J Superfamily [COG 3299]	9e-03
					gp08 W [Burkholderia phage BcepB1A]	5e-56
43	+	505	54.6		gp07 [Burkholderia phage BcepB1A]	8e-17
44	+	965	102.9		phage tail fiber protein [Pseudomonas putida KT2440]	7e-14
45	+	143	16.4		Tail fiber assembly, ORF-191C [bacteriophage P-EibC]	3e-07
46	+	221	24.2		Lyzozyme: Glycoside hydrolase family 19 [cd 00325]	4e-10
					putative endolysin [Enterobacteria phage phiV10]	4e-26
48	-	304	34.9		ATP dependent DNA ligase domain [pfam01068]	2e-16
					PBCV-1 DNA ligase [Burkholderia pseudomallei 1710b]	6e-44
49	-	185	20.5		Uncharacterized bacterial protein family DUF2166 [pfam 09934]	1-03
					DNA-binding protein [Streptococcus phage EJ-1]	4e-11
50	-	202	21.1		Three transmembrane domain, signal peptide (P=0.99)	
					Bbp1 [Bordetella phage BPP-1]	2e-36
51	-	300	32.2		gp5 [Pseudomonas phage YuA]	1e-28
52	-	207	23.4		gp4 [Pseudomonas phage YuA]	3e-10
53	-	539	61.4		Conserved C-terminal helicase domain [pfam 00271]	3e-03
					gp2 [Pseudomonas phage YuA]	2e-55
54	-	139	15.8		gp31 [Bacteriophage Phi JL001]	0.029
55	-	1037	116.2		DNA polymerase III, alpha subunit [pfam 07733]	1e-97
					gp68 [Mycobacterium phage Gumball]	4e-97
56	-	195	21.3		DNA polymerase III, epsilon subunit [pfam 00929]	1e-12
57	-	339	37.5		PnkP, Polynucleotide kinase, Partial [Enterobacteria phage TLS]	3e-05
59	-	306	35.3		FAD-dependent thymidylate synthase [PRK 00847]	8e-21
					gp70 [Mycobacterium phage Gumball]	2e-13
67	-	396	45.1		ATP-dependent exonuclease V, alpha subunit [COG0507]	8e-09
70	+	423	47.5		Coiled coils	
73	+	62	6.6		Two transmembrane domains	
74	+	580	65.4		DNA primase [Burkholderia phage BcepNY3]	6e-13
					Virulence-associated protein E [pfam05272]	3e-20
86	-	67	7.2		One transmembrane domain	
91	-	63	7.3		hypothetical protein ORF041 [Pseudomonas phage 73]	0.88
92	-	241	27.6	LBL3, F8	Coiled coils	

[a] Only ORFs with predicted structural features or similarities to known proteins are shown. [b] Phages belonging to this genus are listed in Table 7.1.

As expected, most of the phage proteins (96%) are conserved among all sequenced phages, and only 6 true orphan genes were identified (Figure 7.3). Since recently acquired ORFs are expected to display atypical codon usage and compositional bias (Lawrence et al., 2001), we calculated individual N_c values (effective number of codons) for all ORFs using CodonW software. In extremely biased ORFs, N_c can approach 20, while in unbiased ORFs it will approach 61. Among the PB1 like phages, the average N_c is 48 ± 8.5. All but one single orphan genes (LBL3-ORFs 1.1 and 88.2, SN/ORFs 2 and 3 and LMA/ORF 3) exhibit extreme N_c values below 30, which might indicate that these genes were

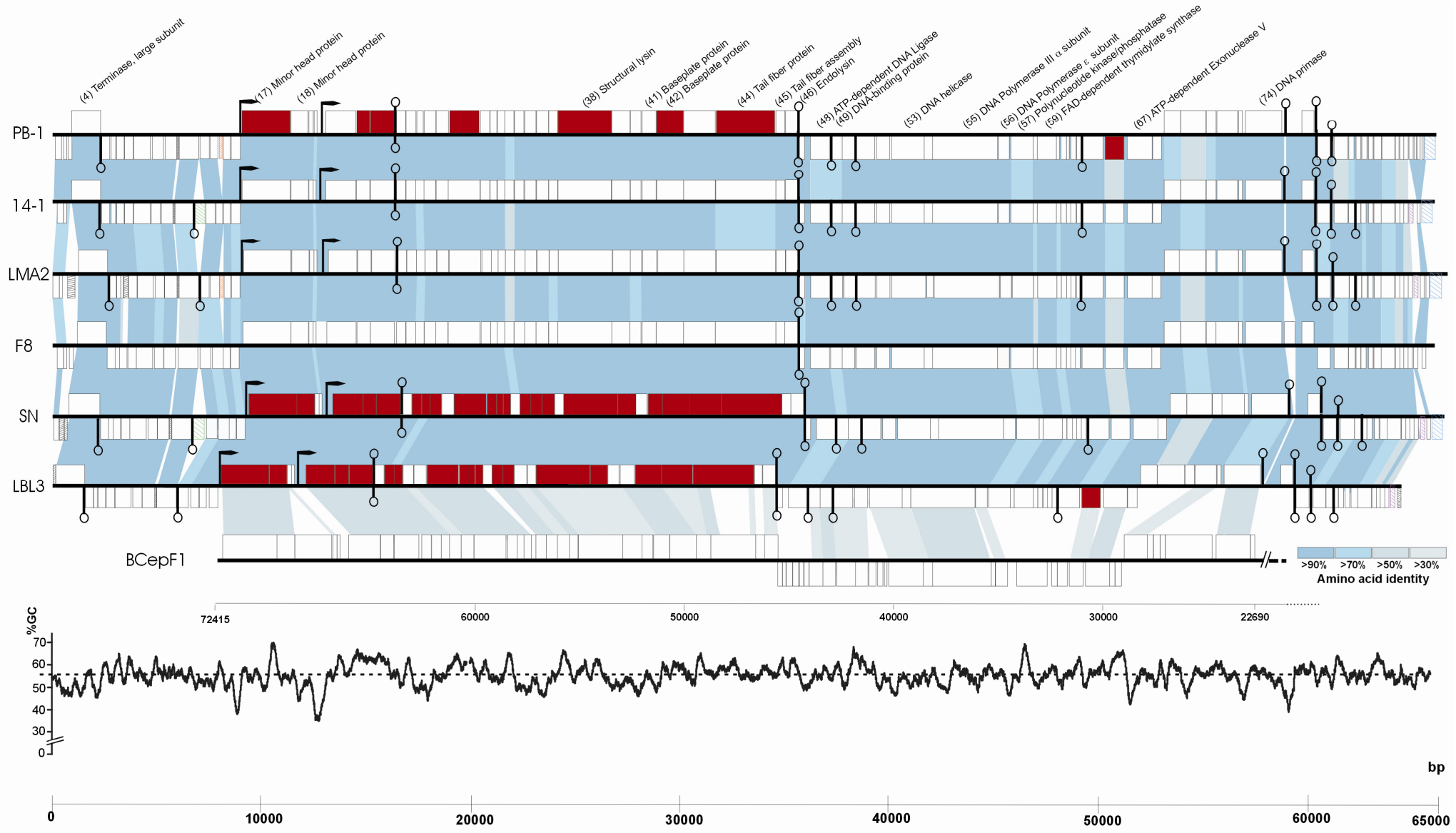


Figure 7.3. ***In silico*** **analysis of the PB1 genus. Upper part.** The predicted open reading frames of all PB1-like phages (white rectangles) and their amino acid identity to the corresponding ORF in phage F8, indicated in different shades of gray. Predicted terminators and σ^{70} – and phage-specific promoters are shown as stem-loop structures and black and open arrows, respectively. ORFs not present in F8 are hatched, and given unique colors when they appear in more than one phage. ORFs marked in red encode structural proteins which were identified as part of the phage particle using mass spectrometry. **Lower part.** Display of GC content throughout the PB1 genome. The plot was generated by the ISOCHORE tool on the EMBL-EBI website (http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html), with a window size of 200 bp.

recently acquired. ORFs present in more than one phage (e.g., ORF 14.1) have N_c values around 45, and may have lost their counterpart in phages PB1, LMA2, F8 and LBL3.

7.2.2 Transcriptional regulation

Genes are arranged in a typically compact manner with little intergenic space (on average 7.9%) and numerous overlapping start/stop codons. The genomes are organized into at least seven transcriptional blocks alternating on both strands, which are separated by (in some cases bidirectional) factor-independent terminators with perfectly conserved stem-loop (Table 7.3). Since PB1 phages lack a phage-encoded RNA polymerase, they depend entirely on the host transcriptional machinery. However, the only recognizable σ^{70}- specific promoters are located at the start of the structural region, upstream from ORFs 17 and 21 (Table 7.3). Additionally, conserved AT-rich boxes are found in intergenic regions spread throughout the genome; although no clear-cut consensus sequence could be derived, these motifs might be recognized by a putative phage sigma factor (Table 7.3).

Table 7.3. Regulatory elements in the PB1-like viruses. Conserved stem-loop sequences and -35 and -10 boxes are underlined.

Location[a]	Strand	Sequence		Algorithm[b]	Score
Factor-independent terminators			dG (kcal/mol)		
ORF5	-	GCCCGGAKCGMTCCGGGCTTTT	-14.8	TT	100
ORF13[c]	-	CGGTAGTGMGAAAKYGCTACCGTTTTT	-15.8		
ORF23	+	AAGGCCGGGCTTCCGGCCTTT	-15.1	TT	100
ORF46	bi	AAAAG(A)AAAGGCCGCTTATTCAGCGGCCTTTTT	-18.3	TT	100
ORF49	-	GCCCGGTTCGCCGGGCTTCTTT	-13.5		
ORF51	-	GAAAGGCGCTGTAA(T)GGCGCCTTTCTTTT	-16.1		
ORF66	-	GGGACTCATGAAAATGGGTCCCTTTTTT	-16.5		
ORF74	+	GGCCGCCTTCGGGCGGTCTTTT(T)	-16.8	TT	100
ORF76	bi	AAA(G)CCCCGGACTCTAGTTCAGAATCCGGGG(C)TTT	-22.7	TT	100
ORF78	bi	AAAACACCGCAGGA(R)CAGGCTGCGGTGTTT(T)	-17.9		
ORF80[c]	-	GCCTCTGC(S)GA(C)TCSGCAGAGGCTTT	-15.5		
σ^{70}-dependent promotors					
ORF17	+	TTGAACGTTTGATGTTTCCCCTATAAT		NN	0.97
				SAK	1.25
ORF21	+	TTGAAATTCAATTTTTAATYGGTAAATT		SAK	1.24
Putative phage-specific promoters					
ORF76	+	AATAGTAATTTGGTAATTTGGTATACTT		MEME	2.9E-15
ORF21	+	ATTGGTAAATTGGTAATTTGGATTAGTT		MEME	4.1E-14
ORF50	-	AAAGCTTAATATGTAATTTGCTTTACTT		MEME	5.6E-14
ORF65	-	TGAAGTAATTTACAAAAGTGCTTTACTT		MEME	1.2E-12
ORF92	-	ATAATTTTATTGTTAATACACTTTACTT		MEME	3.0E-12

[a] Upstream and downstream ORF are given for terminators and promoters, respectively. [b] TT, TransTerm (Ermolaeva *et al.*, 2000); NN, Neural Network (Reese, 2001); SAK, Sequence Alignment Kernel (Gordon *et al.*, 2003); MEME, Multiple Em for Motif Elicitation (Bailey and Elkan, 1994). [c] Not found in PB1.

7.2.3 Clusters of small proteins

A striking feature of the PB1 genomes is the clustering of a large number of small genes encoding hypothetical proteins. Three clusters of 8 kb (ORF1 through ORF16, except ORF4, Terminase), 4 kb (ORF77 through ORF91) and 1 kb (ORF60 through ORF64) are present, and encode small proteins of unknown functions (Table 7.2; Figure 7.3). A similar single cluster of 20 kb, encoding 62 genes unrelated to PB1, is found in *Burkholderia* phage BCepF1 (NC_009015). Since regions involved in BcepF1 DNA replication and particle formation are closely related to PB1 (Figure 7.3), this cluster largely accounts for the mutual differences in genome size and protein content.

Although no functional assignment could be made, seven proteins are predicted to have coiled coils, indicative for protein interactions with other phage or host proteins (see *e.g.*, Crucitti et al., 2003). Two proteins (gp*8* and gp*10*) carry both signal peptides and two and three transmembrane domains, respectively, and are clearly targeted to the outer membrane. Hypothetically and analogous to coliphage T4 (Miller et al., 2003), these proteins might have a role in exclusion of superinfection (Vallee and Cornett, 1972) or serve as an membrane anchor for DNA replication or particle assembly. Similar (but completely unrelated) large assemblies of small proteins with unknown functions are encountered in various phage families like for example the T4-like phages (Miller et al., 2003), the ϕKMV-like phages (Chapter 4) and in many cyanophages (e.g. Weigele et al., 2007). As stated in previous chapters, it has been speculated that they may have a crucial role as accessory factors that bind to and subtly modify the specificity of host proteins so that they function appropriately during phage infection.

7.2.4 DNA replication machinery

Based on sequence similarity searches, only two enzymes involved in nucleotide metabolism could be identified: a kinase/phosphatase (gp57) and a thymidylate synthase (gp59). In contrast, a wide assortment of conserved replication factors was found. Unlike other lytic *Pseudomonas* phages like gh-1, LUZ24 (Chapter 5) and ϕKMV (Chapter 4), PB1-like viruses use the DNA polymerase III holoenzyme for DNA replication, the main DNA replicating enzyme in bacteria (Kelman and O'Donnell, 1995), and supply the two main subunits of the catalytic core. The alpha subunit (gp55) catalyses the polymerization reaction and contains two conserved motifs ($_{342}$P**D**I**D**I**D**F and $_{492}$LL**K**I**D**A**LG**). The polIII epsilon subunit, encoded by gp56, is a DEDDh-type 3'-5' exonuclease responsible for the proofreading activity of the polymerase. The three conserved sequence motifs ExoI ($_{7}$**D**T**E**), ExoII ($_{95}$**N**LP**FD**) and ExoIII ($_{156}$**H**RALD**D**) (Bernad *et al*., 1989) are also perfectly conserved

in the phage protein. Despite being ubiquitous among prokaryotes, only two other phages (*Pseudomonas* phage F116 and *Staphylococcus* phage 47) encode this subunit. The presence of a phage-encoded DNA primase (gp74) and helicase (gp53) suggests that replication elongation is independent of the host machinery.

7.2.5 Phage particle proteins

Similarity studies delineate the genomic region involved in particle formation from ORF17 to ORF45 (Table 7.2), and suggest morphological relatedness to myophage BcepB1A, which itself is marginally related to the Bcep781 group of phages (Summer *et al*., 2006). One-dimensional 12% SDS-PAGE of CsCl purified particles of 14-1, LBL3 and LMA2 showed a highly conserved band pattern, with only slight variations in the 14-20 kDa region (Figure 7.4A). This conservation is reflected by the more than 95% nucleotide identity among PB1 phages in this region, with the only drop in sequence similarity is found in ORF*44*, encoding the presumed tail fiber.

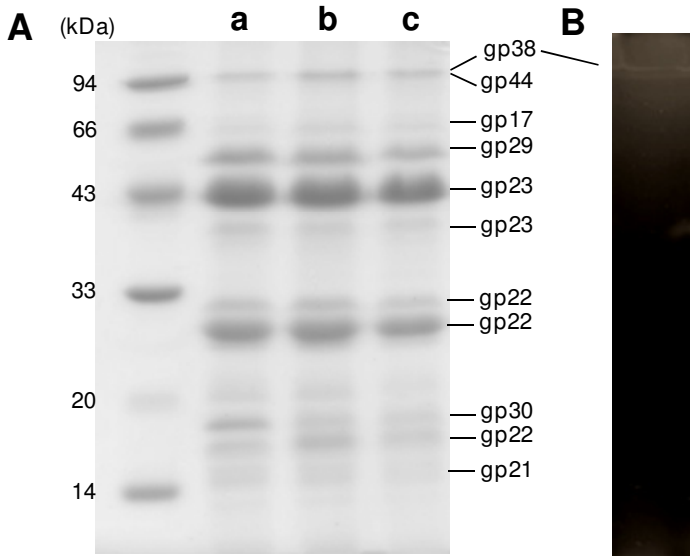


Figure 7.4. A. Phage particle proteins separated on a 12% SDS-PAGE gel, in parallel to a LMW-size ladder (kDa) and stained with Simply Blue^{TM} Safestain; a, 14-1; b, LBL3; c, LMA2. The positions of major structural proteins, as determined by ESI-MS/MS, are indicated. **B.** Zymogram analysis of phage 14-1. The clear zone associated with peptidoglycan degrading activity is indicated.

ESI-MS/MS analysis on PB1, SN and LBL3 led to the experimental identification of 22 predicted proteins, reaching sequence coverages up to 56 % (Table 7.4). The limited variation in SDS-PAGE patterns between the PB1-likes can be attributed to fluctuating amounts of low-molecular weight proteins gp30 (15.8 kDa) and gp22 (16.4 kDa). Gene products 22, 23, 29 and 30 were identified as major building blocks of the phage particle, and no N-terminal peptides of gp23 were found, hinting

at post-translational cleavage of this major head protein. This posttranslational processing can also be suggested for gp21, which migrated to an estimated mass of 15 kDa, much lower than predicted based on its DNA sequence (50 kDa). During MS analysis, no peptides corresponding to the carboxyl terminus of this protein were detected, supporting this hypothesis.

The high-MW structural protein gp*38* (94.3 kDa) carries a transglycosylase domain (residues 490-600), including an N-acetyl–D-glucosamine binding site and the catalytic residue E512. The muralytic activity of gp*38* of phages 14-1 and LBL3 was demonstrated using a zymogram assay, showing specific degradation of peptidoglycan of autoclaved *P. aeruginosa* cells upon renaturation of gp*38* (Figure 7.4B). These murein hydrolases are widespread in the virions of bacteriophages and assist efficient DNA transfer to the host cell, although they are often not required for phage infection (Rydman and Bamford, 2000). The combination of this hydrolase activity with a major structural component of the phage is a typical solution to prevent enzyme diffusion in order to constrain its activity and preserve the viability of the cell (Moak and Molineux, 2004). The exact location of gp38 in the phage particle cannot be derived from this functional analysis, since muralytic activity has been associated with the T4 baseplate (Kao and McClain, 1980), the T7 internal virion protein (Molineux, 2001), the TM4 tape measure protein (Piuri and Hatfull, 2006) and the T5 tail fiber protein (Boulanger et al., 2008).

Table 7.4 Summary of the mass spectrometric analysis of PB1, SN and LBL3. The analyses of PB1 and SN were performed at theLaboratory of Foodborne Zoonoses, Guelph, Canada (Prof. A. Kropinski) and the Biomedical Research Institute of the Hasselt University (Prof. J. Robben), respectively.

Gene Product	MW (kDa)	# Identified peptides			AA Coverage (%)			Comments
		SN	LBL3	PB1	SN	LBL3	PB1	
17	84.3	31	9	9	27	25	19	Minor head protein [pfam 09714, 2e-15]
18	31.5	5	1	-	28	4	-	Minor head protein [pfam 04233, 3e-08]
21	50.5	-	4	-	-	17	-	N-terminus conserved [COG 3566, 5e-16], no C-terminal peptides found
22	21.6	13	6	5	53	55	45	
23	41.6	27	9	4	50	45	21	No N-terminal peptides found
25	16.9	6	2	-	24	25	-	
26	15.2	10	2	-	56	20	-	Minor head protein [pfam 09714, 2e-15]
27	21.2	6	-	-	31	-	-	Minor head protein [pfam 04233, 3e-08]
29	53.7	12	11	5	40	43	22	
30	15.9	12	4	-	54	55	-	
31	12.2	3	1	-	42	14	-	
32	12.4	3	1	-	35	11	-	
34	17.8	5	4	-	32	40	-	
35	19.8	2	1	-	25	6	-	
36	21.6	1	-	-	11	-	-	
38	94.2	26	21	4	36	38	10	Structural lysin, contains Transglycosylase SLT domain [pfam 01464, 1e-17]
39	32.5	6	2	-	21	13	-	
41	24.2	1	-	-	11	-	-	Similar to various baseplate proteins
44	103.2	28	18	13	37	30	24	Member of Baseplate_J Superfamily [COG 3299, 9e-3]
66	33.2	-	1	4	-	4	23	

Curiously, gp*66* was also detected as a structural protein in both the PB1 and LBL3 particle. This 33 kDa protein is encoded in the DNA replication region and consists for 50% out of strongly basic (K,R) and strongly acidic (D,E) amino acid residues. This bizarre and apparently highly-charged protein can be linked to a bacterial cell-wall associated hydrolase, but its function in phage infection/ development remains unknown.

Another remarkable feature in the structural genome region of PB1-like phages is the observed variation in the amount of deoxyguanylate residues present in a G-rich stretch at the 5' of ORF42 (bp 32-42), which encodes a 44.8 kDa protein belonging to the Baseplate J-family (COG 3299). Sequencing of individual (shotgun) clones encompassing this region identified between 7 and 9 G's in LMA2, while this amount varied between 9 and 10 in both PB1 and LBL3. Primer walking on purified phage DNA produced mixed signals downstream this G-stretch (Figure 7.5).

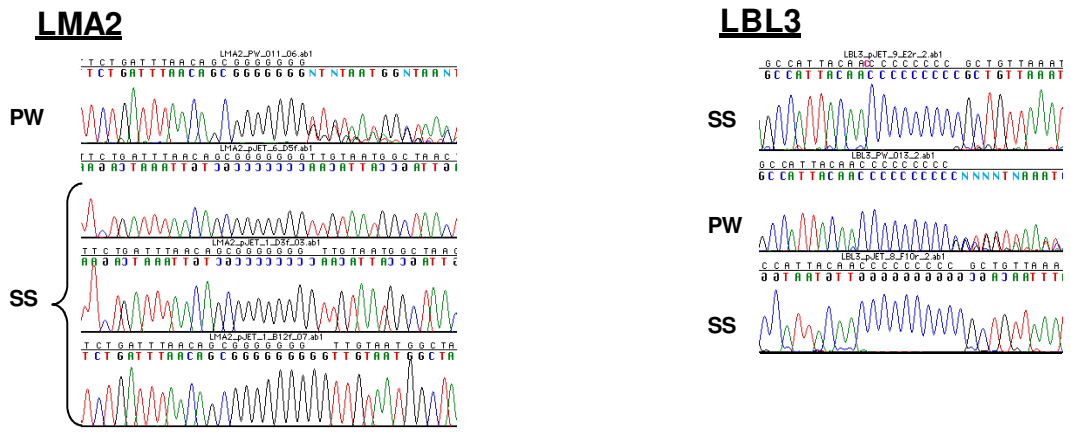


Figure 7.5. The construction of shotgun sequencing (SS) libraries resulted in clones of phages LMA2 (left) and LBL3 (right) which displayed a varying amount of G's in a homopolymer stretch. Primer walking (PW) on purified phage DNA produced mixed signals downstream this G-stretch, indicative for this mixed population. The same phenomenon was noticed in phage PB1 (data not shown).

As a consequence, a -1 or +1 frameshift is necessary for translation of the full-length baseplate protein for a part of the phage population. To our knowledge, this kind of variation between 8 and 10 G residues has only been reported for the *bpm* gene of the almost identical *Bordetella* phages BPP-1, BMP-1, and BIP-1 (Liu et al., 2004). Programmed translational frameshifting is well documented in dsDNA bacteriophages (e.g. Fortier et al., 2006), but occurs mostly at the C-terminus, to produce two overlapping proteins in appropriate ratio (Zimmer et al., 2003). In this case, a frameshifting event at the N-terminus might serve as a control point to regulate the levels of gp42 in the course of phage infection.

7.2.6 Host lysis

All analyzed phages use a classic two-component cell lysis system. The endolysin (gp46) is encoded at the end of the structural module, and carriers a single β-1,4-N-acetyl-D-glucosaminidase domain (Pfam pf00182, e = 4E-03) which presumably cleaves bonds in the aminosugar backbone of the peptidoglycan. The endolysins lack a secretory signal and a cell wall-binding domain, suggesting they accumulate in the cytoplasm, relying on their cognate holin to access the peptidoglycan layer (Young *et al*., 2000). Despite the lack of sequence similarity of gp50 to known proteins, its genomic location nearby the putative endolysin, its small size and three transmembrane domains with an N-terminal inside topology strongly suggest a function as a class I holin (Wang et al., 2000).

7.3 Discussion

The analysis of genome and particle content of five *P. aeruginosa* -infecting *Myoviridae*, originating from different European countries, revealed a tight evolutionary relationship between them. Together with the long-known phage F8 and BcepF1, isolated from agricultural soils in New York, they constitute the environmentally widespread genus of the 'PB1-like viruses'. This genus is marginally related to *Burkholderia* phage Bcep1A, and even more distantly to the Bcep781 viruses (Lavigne *et al*., *in press*). Although F8 has been known for a long time and was part of the Lindberg *P. aeruginosa* typing set (Lindberg and Latta, 1974), PB1 was chosen as reference phage, as it was first isolated (Table 7.1).

Considering only the PB1-like viruses infecting *P. aeruginosa*, the extent of conservation is truly remarkable: all phages exhibit >85% nucleotide identity and encode a completely conserved genome core region which extends from ORF4 to ORF88. When including the more distantly related phage BcepF1 in the comparison, genome conservation is limited to the modules involved in virion formation and DNA replication (ORF17 to 74). As stated in previous sections, this limited lateral gene transfer and genetic mosaicism in the phage genomic core is also found in the T4 subgenus, and was linked to the great complexity of these phages (Section 2.4.4). It would be very interesting to find out if more distantly related PB1-like phages, like T4 related viruses, are also capable of successfully infecting a broad range of hosts. The small variation in genome size (65-72 kb) and the lack of mobile elements (introns/HNH endonucleases) might suggest however that drastic expansion of the core genome is not possible.

The acid-resistant, broad-range PB1-like viruses are suitable for application in phage therapy experiments. However, the analysis described in this chapter illustrates the importance of genome

studies of phages which are applied in phage therapy settings. Similarity searches revealed a significant relationship between ORF74 and the Virulence-associated protein E [pfam05272, 3 e^{-20}] (Table 7.2). Phage-encoded VirE homologs have been suggested to be potential virulence factors, for example, *Vibrio* phage VP16T (Seguritan et al., 2003). VirE homologs, first identified on a genomic fragment present in virulent but not avirulent strains of *Dichelobacter nodosus*, the causative agent of bovine foot rot (Katz et al., 1991), are also encoded by *Burkholderia* phages Bcep1, Bcep43, Bcep781, and BcepF1 (Summer et al., 2007). The Bcep1 VirE homolog, however, was connected to the prim-pol primase superfamily of DNA polymerases (Iyer et al., 2005), strongly suggesting more conservative role for VirE in phage DNA replication, rather than in promoting host pathogenicity.

Chapter 8

Bacteriophage YuA represents a new genus of *Pseudomonas* – infecting *Siphoviridae* [5]

Phage YuA was isolated from a pond near Moscow in 2003, using *P. aeruginosa* PAO1 as host organism. Preliminary morphological data suggested relatedness to the Siphovirus M6 from the Lindberg typing set (Ackermann et al., 1988). The genome sequence of YuA was determined in the framework of a bilateral research project by combined effort of Vadim Mesyanzhinov and Konstantin Miroshnikov (Institute of Bio-organic Chemistry, Moscow, Russia) and Kirsten Hertveldt. This sequence was available at the start of this project, and comprised 58,663 bp arranged in a circularly permuted manner. Both YuA and M6 are significantly different from other *Pseudomonas*-infecting *Siphoviruses* deposited in the public databases, which are described in Section 1.5.2. Therefore, an in-depth analysis of the phage YuA genome sequence and the particle protein content was performed, and described in this chapter.

8.1 Phage characteristics

Bacteriophage YuA propagates on *P. aeruginosa* PAO1, producing small (1 mm diameter) and turbid plaques which suggests a temperate nature. The phage does not infect *P. putida* and *P. fluorescens* or other gram-negatives like *E. coli*, *Shigella* and *Salmonella* species, but is able to lyse *Burholderia solanacearum* at low efficiency (EOP 10^{-7}). Electron microscopic imaging revealed YuA as a typical member of the *Siphoviridae* family of dsDNA bacteriophages, having a flexible, non-contractile tail (Figure 8.1). In contrast to the well-known *Pseudomonas Siphoviridae* phages D3, B3 and D3112 that resemble phage λ, phage YuA has an elongated head (B2 morphotype) resembling *P. aeruginosa* phage M6 (Ackermann et al., 1988). The phage YuA head size is ~72 x 51 nm and the tail length is ~145 nm. Besides an elongated head, both YuA and M6 have striated tails which are terminally and subterminally decorated with short fibers. Phage M6 is reported to be morphologically identical to

[5] Data presented in this chapted are adapted from Ceyssens, P.J., *et al.* (2008) *Journal of Bacteriology* ***190****: 1429-1435.*

Xanthomonas oryzae phage XP12 (Ackermann et al., 1988), and shown to adsorb to non-retractile host pili (Bradley and Pitt, 1974).

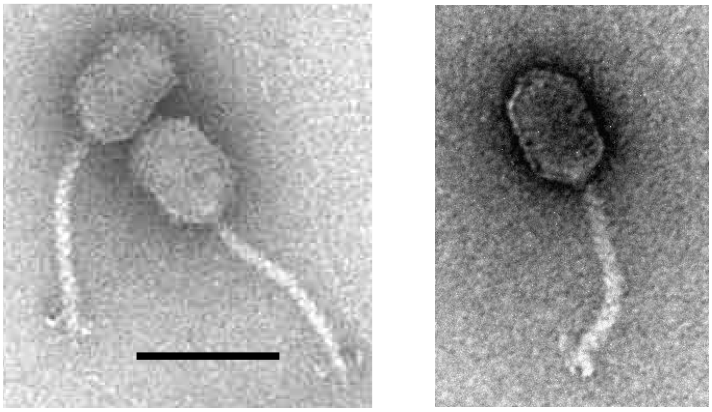


Figure 8.1. Electron microscopic images of phage YuA particles. Phage YuA has an elongated head and a flexible tail. The scale bar represents 100 nm.

YuA genomic DNA is insensitive to the activity of 10 out of 13 tested restriction enzymes, including many common enzymes like *Hind*III, *Bgl*II and *Eco*RI, and the methylation-dependent *Dpn*I (GA^{m}/TC). Given the sensitivity towards digestion with *Sau*3A (/GATC) and the methylation-sensitive *Sma*I ($C^{+}CC^{+}$/GGG), it can be concluded that YuA contains unmethylated adenine and cytosine residues. This is in contrast to the morphologically related phage XP12, which is known to contain a 5-methylcytosine instead of cytosine in its genome (Ehrlich et al., 1975). *In silico* analysis revealed the absence of 9 out of 10 recognition sites, although four *Eco*RI-sites are present despite the insensitivity towards the restriction enzyme. Furthermore, it became clear during genome sequencing that isolated YuA DNA is rather inaccessible to standard PCR amplification using various primers, annealing temperatures and commercially available DNA polymerases. These observations suggest the presence of another base substitution or modification, as is discussed in Section 8.2.3.

8.2 Genome Analysis

8.2.1 Genome annotation and similarity to other phages

As stated previously, the genome of YuA is comprised of 58,663 bp and has a G+C content of 64.3 %, strongly resembling the G+C rich host DNA (66%). In total, 78 open reading frames (ORFs 1 to 77 and ORF 60.1) were predicted on the circular genome map (Table 8.1; Figure 8.2), all oriented in the same direction and leaving only 4% of the YuA genome as non-coding. No tRNA genes were predicted. The genome of YuA is neither Mu nor λ-like and can be roughly divided in three functional regions,

containing gene products involved in (1) nucleotide metabolism and DNA replication, (2) host interaction and (3) particle structure, packaging and host lysis (Figure 8.2). The genome of YuA was deposited in GenBank under accession number NC_010116.

Table 8.1. Predicted gene products in the genome of phage YuA, and their similarity to database entries.

ORF[a]	Start	Stop	AZ	MW (kDa)	Closest phage homologue [b]	e-value	Structural Features / Putative Function [c]
1	68	412	114	12.5	gp04 [JL001]	3.0E-11	C-terminal transmembrane domain
2	840	2534	564	64.4	gp07 [JL001]	2.00-52	Helicase [HepA, COG0553, 7E-23]
3	2542	2895	117	12.9	gp08 [JL001]	5.0E-23	
4	2955	3500	181	20.1	gp10 [JL001]	1.0E-17	
5	3642	4481	279	30.4	gp11[JL001]	9.0E-32	
6	4583	5776	397	44.9	gp13 [JL001]	1.0E-98	RecD Exonuclease V, alpha subunit [COG0507, 7E-5]
7	5918	7660	580	63.7	gp14 [JL001]	0	Ribonucleotide reductase [CD002888, 9E-150]
8	7806	8489	227	24.4	none		
9	8500	8763	87	10.3	none		
10	8773	9648	291	33.5	hypothetical protein [phiW-14]	9.0E-11	
11	9650	10486	278	30.6	none		Threonine dehydratase [COG1171, 5E-06]
12	10556	11998	480	54.3	none		DNA Repair photolyase [COG1533, 3E-12]
13	12053	12616	187	20.4	none		
14	12616	13602	328	38.0	gp19 [JL001]	1.0E-36	
15	13629	14189	186	21.0	gp23 [JL001]	1.0E-17	
16	14202	14315	37	4.8	none		
17	14312	15367	351	40.1	gp24 [JL001]	2.0E-83	Thymidylate synthase and pyrimidine hydroxymethylase [cd00351, 9E-25]
18	15360	15956	198	22.3	gp25 [JL001]	4.0E-22	
19	15931	16158	75	9.1	none		
20	16148	16513	121	13.7	gp26 [JL001]	8.0E-5	
21	16513	18507	664	64.9	gp28 [JL001]	1.0E-111	DNA polymerase family A, 5'-3' polymerase domain [cd06444, 8E-48]
22*	18593	18790	65	7.0	gp68[BcepNazgul]	8.0E-09	T4 NrdA.1-like protein
23	18787	19248	153	17.1	gp27 [phage DSS3P2]	8.0E-18	Deoxycytidylate deaminase [cd01286, 8E-21]
24	19256	19639	127	14.8	gp31 [JL001]	1.0E-18	Transmembrane domain
25	19624	19950	108	12.8	gp32 [JL001]	2.0E-08	Putative repressor
26	19963	22374	803	91.9	gp33 [JL001]	7.0E-160	Putative Integrase; C-Terminal phage-associated DNA primase [COG3378, 2E-06]
27	22419	22634	71	8.1	none		
28	22746	23111	121	13.4	none		Transmembrane domain
29	23148	24326	392	27.3	none		
30	24323	24661	112	12.6	gp34 [JL001]	8.2E-02	
31	24685	24954	89	9.8	none		Signal Peptide (P = 1.0)
32	24957	25562	201	22.0	none		
33	25559	26035	158	17.6	none		
34	26032	26361	109	12.8	none		
35	26377	26676	99	10.9	none		
36	26679	26966	95	10.9	none		
37	26971	27240	89	9.8	none		
38	27237	27401	54	6.4	none		
39	27380	27997	205	19.8	none		
40*	27994	28275	93	10.8	gp78 [TLS]	2.0E-07	
41	28277	28849	190	20.3	gp34 [ε34]	3.0E-18	DEDDh 3'-5' exonucleases [cd06127, 3E-3]
42*	29122	29742	206	23.0	none		
43	29858	30610	250	28.2	ORF286c [RB43]	7.0E-06	
44*	30614	31084	156	17.2	none		Diguanylate-cyclase domain [pfam00990, 8E-14]

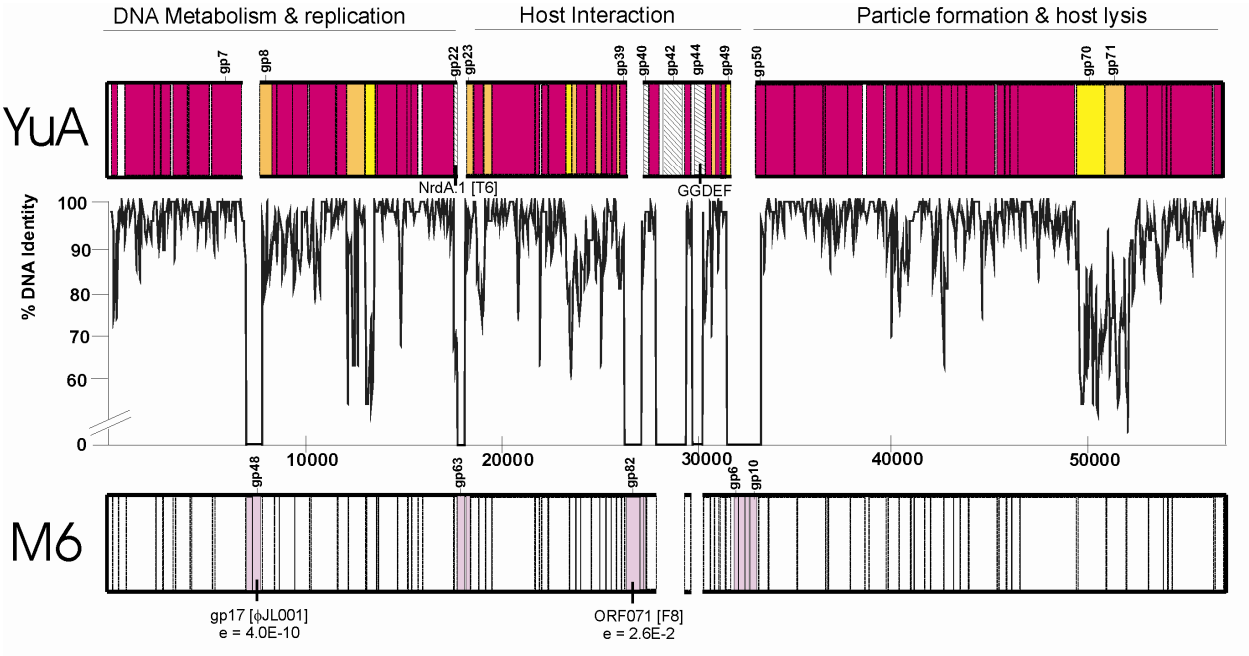
(Table 8.1. Continued)

45	31147	31572	141	16.3	none		Antirestriction protein [pfam03230, 1E-26]
46	31630	31965	111	12.7	none		
47	31988	32200	70	7.6	none		
48	32197	32601	134	14.3	none		
49	32591	32839	82	9.5	none		
50	33499	34026	175	19.3	gp58 [JL001]	2.0E-04	
51	34026	35573	515	58.9	gp59 [JL001]	5.0E-138	Terminase_6 [pfam03237, 7E-10]
52	35631	37154	507	56.1	gp5 [BcepGomr]	4.0E-20	Structural protein*
53	37151	37282	43	4.9	none		Transmembrane domain
54	37266	38474	402	45.0	gp62 [JL001]	1.0E-64	Putative minor head protein*
55	38543	39325	260	27.9	gp6 [BcepGomr]	1.0E-04	Structural protein*
56	39520	40452	310	32.2	[Xp15]	7.0E-86	Major structural protein*
57	40627	41136	169	19.2	none		Structural protein*
58	41164	41769	201	20.9	gp69 [JL001]	3.7E-01	
59	41797	41967	56	5.4	none		Structural protein*
60	41971	42540	189	20.0	p26 [phiE125]	1.0E-05	Rz protein; N-terminal Transmembrane domain, Signal Peptide (P = 0.99)
60.1	42269	42469	66	9.3	none		Rz1 protein, Signal Peptide (P = 1.0), predicted lipoprotein (LipoP = 1.0)
61	42537	42872	111	12.0	none		Holin; Contains 3 transmembrane domains
62	42857	43582	241	27.4	gp25 [B3]	4.0E-05	Endolysin, Signal Peptide (P = 1.0)
63	43582	44100	172	18.6	gp75 [JL001]	4.0E-09	Structural protein*
64	44104	44472	122	13.4	gp14 [73]	2.0 E-08	Structural protein*
65	44469	44921	150	16.3	gp11 [BcepGomr]	2.0E-27	Structural protein*
66	44955	46487	510	54.0	gp16[73]	8.0E-138	Major structural protein*
67	46550	47008	152	17.0	gp17 [73]	6.0E-20	
68	47282	47728	148	15.5	gp81 [JL001]	3.0E-33	
69	47731	50553	940	100.4	gp20 [73]	2.0E-70	Putative tail length tape measure protein*
70	50563	52086	507	56.2	ORF50 [B3]	7.0E-23	Structural protein*
71	52100	53095	331	35.7	p48 [D3112]	9.0E-10	Structural protein*
72	53097	54767	556	62.2	p49 [D3112]	3.0E-40	Structural protein*
73	54764	55573	269	30.1	ORF53 [B3]	2.0E-27	Structural protein*
74	55587	55820	77	8.1	p51 [D3112]	5.0E-04	Structural protein*, N-terminal Transmembrane domain
75	55817	56023	68	7.6	ORF55 [B3]	1.0E-08	
76	56106	58178	690	76.8	ORF56 [B3]	4.0E-124	Structural protein*
77	58178	58663	161	17.0	p54 [D3112]	6.0E-09	Structural protein*

[a] ORFs which are absent in M6 are marked with an asterisk. [b] With exclusion of similarity of the corresponding M6 protein. [c] Proteins marked with an asterisk are experimentally confirmed as structural proteins.

The YuA genome displays an overall 91% DNA identity to the completely sequenced, but initially poorly annotated phage M6 (Kwan et al., 2006), present throughout the entire genome and resulting in >80% amino acid identity in 92% of the predicted ORFs. The data obtained for YuA confirms the reannotation of M6, proposed in July 2007 by Andrew Kropinski (NC_007809). Only six genome regions contain unique YuA or M6 sequences, accounting in total for 15 differential gene products, four of which occur in YuA and eleven in M6. Further comparative analysis between YuA and M6 also revealed YuA gp70 and gp71 as most deviating in the structural region, hinting at a role of these proteins in host recognition and surface adhesion (Figure 8.2, top).

Comparison to other phages reveals significant similarity of 29 predicted YuA proteins to proteins encoded by ϕJL001 (63,469 bp), a phage that infects an uncharacterized marine α-proteobacterium


DNA Metabolism & replication
Host Interaction
Particle formation & host lysis
YuA
% DNA Identity
NrdA 1 [T6]
GGDEF
M6
gp17 [ϕJL001]
e = 4.0E-10
ORF071 [F8]
e = 2.6E-2

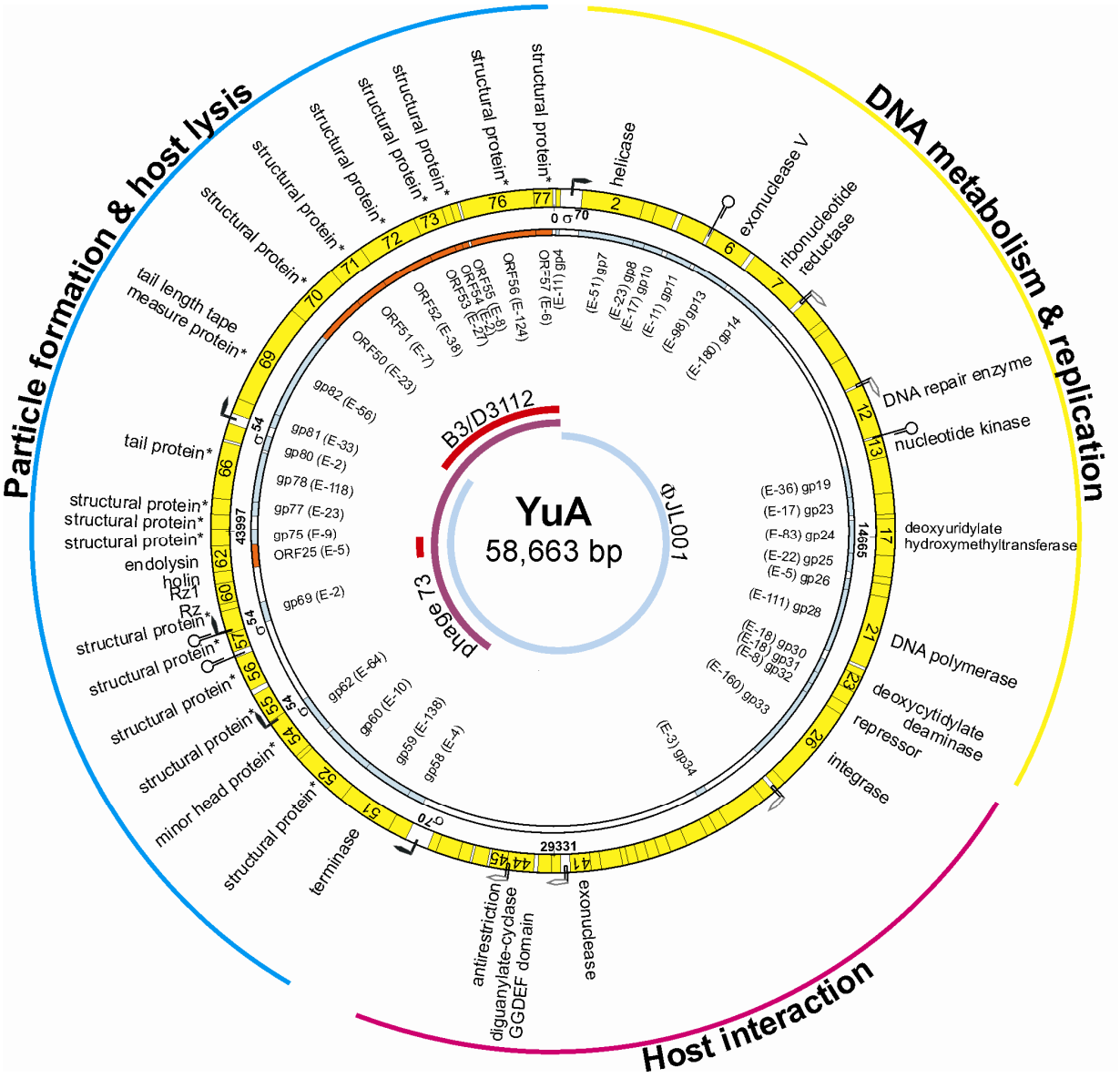

YuA
58,663 bp
ΦJL001
B3/D3112
phage 73
Particle formation & host lysis
DNA metabolism & replication
Host interaction
helicase
exonuclease V
ribonucleotide reductase
DNA repair enzyme
nucleotide kinase
deoxyuridylate hydroxymethyltransferase
DNA polymerase
deoxycytidylate deaminase repressor
integrase
exonuclease
antirestriction diguanylate-cyclase GGDEF domain
terminase
minor head protein*
structural protein*
tail length tape measure protein*
tail protein*
endolysin
holin
Rz1
Rz

Figure 8.2. Top: pair-wise comparison of bacteriophages YuA and M6. The predicted open reading frames and their mutual amino acid identity are indicated in red (>90%), orange (>80%) and yellow (>50%). ORFs unique to YuA (4) and M6 (11) are hatched and purple, respectively, and predicted functions are indicated. Nucleotide identity throughout both genomes is illustrated by the middle graph, comparing both phages ORF by ORF using a sliding window of 60 bp. **Bottom: Circular representation of the YuA genome.** The yellow outer circle represents the YuA ORFs, with indication of their predicted functions involved in DNA metabolism and replication, host interaction, particle formation and host lysis. Experimentally confirmed structural proteins are marked with an asterisk, confirmed phage promoters are indicated as black arrows. Predicted (non-confirmed) promoters and terminators are shown as open arrows and stem-loop structures, respectively. The inner circles represent similar ORF regions of phages 73 (purple), ϕJL001 (blue) and B3 and D3112 (red). Corresponding e-values are indicated for ϕJL001, B3 and D3112.

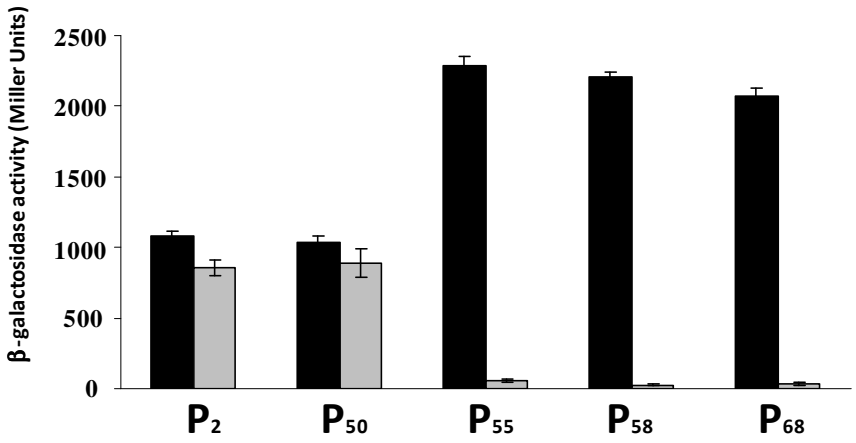
JL001. Phage ϕJL001 is reported as a temperate phage and shares roughly the YuA particle morphology (Lohr et al., 2005). In addition, eighteen particle structure-related YuA proteins share amino similarity to gene products of *P. aeruginosa* phage 73 from the Lindberg typing set (Lindberg and Latta, 1974). Despite major similarity in head morphogenesis genes, phage 73 is morphologically identical to phage D3112 (Ackermann et al., 1988), and does not show the elongated head morphology typical for YuA, M6 and ϕJL001 particles. This might be explained by the smaller genome content of phage 73 (42,999 bp) compared to the 60 kb-genomes of the latter phages. Finally, eight of the YuA proteins (gp70 to gp77) which are most probably involved in host attachment and interaction, share sequence similarity with the transposable and pilus-specific *P. aeruginosa* phages B3, DMS3 and D3112 (Figure 8.2, Table 8.1). The latter gene products also appear in several bacterial genome sequences, as prophage or cryptic phage elements in e.g. *Hahella chejuensis*, *Xylella fastidiosa*, *Burkholderia cepacia* and *Haemophilus ducreyi* and in *Burkholderia cepacia* phage BcepNazgul (NC_005091, 57,455 bp).

The choice of the YuA genome sequence zero point was based on genome comparisons with phages ϕJL001, D3112, DMS3, B3, predicted gene functions and promoter prediction/identification in phage YuA. The YuA zero point differs from the phage M6 zero point, which might be reconsidered for consistency among these related phages.

8.2.2. Transcriptional regulation

To experimentally identify host promoter sequences encoded in the YuA genome, promoter activity was determined quantitatively in *P. aeruginosa* cells by measuring the β-galactosidase activity of individual clones of the constructed promoter trap library (Section 2.3.3). Five YuA regions were identified from which transcription of the vector-encoded *lacZ* gene was initiated by the *P.*

aeruginosa transcriptional machinery (Figure 8.3). Two different promoter types were distinguished based on sequence information and promoter strength. The first type, found in front of genes 2 and 50, has a clear σ^{70}-like consensus sequence (TTAGGT-N_{17}-TtaAAT) and yields 1,021 Miller Units (MU) of β-galactosidase activity. The second promoter type is located in the genome region involved in particle formation and host lysis. It precedes genes 55, 58 and 68, and displays approximately a twofold activity (2,152 MU) compared to the first promoter. Conserved GG and GC elements in this second promoter type are separated by a DNA stretch corresponding to one helical turn, resembling σ^{54}-binding sites (Barrios et al., 1999). This finding was further investigated using an *rpoN* deletion mutant of *P. aeruginosa* PAO1, which is unable to produce the σ^{54} transcription factor (Arai et al., 2005). Infection studies (MOI 0.1-10^8) revealed the inability of YuA to replicate inside this mutant. In a second step, five pTZ110 constructs containing the identified promoter regions (Figure 8.3) were electroporated into this mutant and the transformants were assayed for β-galactosidase activity.



Promoter	YuA region	Core sequence
P_2	489-568	CTTAGGTTACTTCACTTTCTCCAGTTCAATCCTA
P_{50}	33014-33160	CTTAGGTTACTTGGGTTTCTTACCTGAAATAAAT
		-35 -10
P_{55}	38392-38570	GGCACAACACTGGCGGGTGATCTGCCGACAATGA
P_{58}	40888-41035	GACGATGCCCTGGCGGACCAGAAGCTGGAAGTGT
P_{68}	47152-47312	GATCAAGCCCTGGTGTCTGGGACGCCCGTGGACA
		-24 -12

Figure 8.3. Promoter assay of the YuA fragment library. Five promoter regions were identified and are named after the gene they precede. Measurements of β-galactosidase activity in wild-type *P. aeruginosa* PAO1 and the *rpoN* mutant are indicated in black and grey, respectively. Data represent the average and standard deviation of six samples per promoter construct for accurate quantification of the promoter activity by the Miller assay.

In contrast to the two identified σ^{70} promoters (P_2 and P_{50}) which largely maintained their activity, transcription of the reporter β-galactosidase gene downstream the putative σ^{54} promoters was absent (Figure 8.3). This proves that the three promoters preceding genes 55, 58 and 68 specifically require the σ^{54} binding factor to initiate transcription of the downstream genes. This is, to our knowledge, the first time that this is demonstrated for a bacteriophage, and implies the dependency of YuA on additional transcriptional factors (enhancer binding proteins) to initiate RNA synthesis of the late phage proteins (Morret and Segovia, 1993).

No promoter activity could be associated with a third, perfectly conserved intergenic motif (CTTTACTTACTTCGG-N9-TATACTT) preceding genes 8, 12, 28, 42 and 45 (Figure 8.2, open arrows), which might indicate the need for one or more phage-encoded protein(s) to control the expression of these YuA genes. Alternatively, another bacterial-encoded transcription factor that is only present under certain circumstances might be required to initiate transcription. In addition, four ρ-independent terminators were predicted, downregulating transcription beyond genes 5 (unknown function), 12 (predicted DNA repair enzyme), 55 (structural protein) and 56 (major capsid protein), respectively (Figure 8.2).

8.2.3 Genes involved in DNA metabolism and genome replication

A rather large region in the YuA genome (ORF2 to 23) encodes a cluster of proteins involved in nucleotide metabolism (gps 17, 22 and 23) and DNA replication. YuA encodes proteins required for the formation of the replisome, comprising a helicase (gp2), exonucleases (gp6 and gp41) and a type A DNA polymerase (gp21). Gene 22 encodes a putative ribonucleotide reductase which catalyses the committed step to DNA synthesis by reduction of ribonucleoside diphosphates (rNDPs) to the corresponding deoxyribonucleoside diphosphates (dNDPs). Deoxycytidylate (dCMP) deaminase (gp23) catalyzes the deamination of deoxycytidylate to deoxyuridylate (dUMP).

Interestingly, a dUMP hydroxymethylase (dUMP-HMase) function is predicted for gp17. In this enzymatic reaction, dUMP serves as substrate for the addition of a hydroxymethyl group, using CH_2H_4folate as a cofactor and generating the modified base hydroxymethyl-deoxyuridylate (Schellenberger et al., 1995). dUMP-HMases share similarity with the more widespread and highly conserved group of thymidylate synthases, which use the same cofactor in the reduction of dUMP to dTMP. Sequence alignment of YuA gp17 with dCMP hydroxymethylase of T4 and the corresponding gp17 homologous proteins in ϕJL001 and SPO1 clearly shows conservation of catalytically important

```
JL001          MKVIRARNVNDALPRGLAYLRESGLTRETR-NGGAVVAPEPVTTVYEQPRERVLFSKLRDANPFFHLFEALWMLAGRHDLAFVEKFA--------    86
YuA            MKVIKTRNVQQALPEALYQLSFEGVRRDSR-NGPVFMFPEPVTTVYLRPAERVLFWAERDANPFFHLMESLWMLGGRNDVEYVARFV--------    86
SP01           MYTFSGSKPTQLYMDILSTVIKEGDVLAPR-GKRIKEIRPVMIEFKNPIRRTTFLKGR-NINPFFQVAESLWILAGRSDVGFLLDYN--------    85
T4 TS          ---------MKQYQDLIKDIFENGYETDDRTGTGTIALFDTKLRWDLTKGFPAVTT--KKLAWKACIAELIWFLSGSTNVNDLRLIQHDSLIQGK    84
T4 CH          --MISDSMTVEEIRLHLGLALKEKDFVVDKTGVKTIEIIGASFVADEPFIF-------GALNDEYIQRELEWYKSKSLFVKDIPGET-------P    78

JL001          KNMRSFSDDG--------ETMWGAYGWRWR----------TFFDFDQLSWAVHRLRRLPNDRRVVVQMWAAGLDQERADAGGKDVPCNIAIHFRIN   164
YuA            DRMRSYSDDG--------LTFHGAYGFRWR----------QHFFEDQLPKIIAALKANRDDRRQVLSMWDADADLGRQ---GKDLPCNLQAIFQIA   161
SP01           KNMGQFSDDG--------VFFNAPYGERLRFWNRSDANNFIYNPLDQLRDVYEKIKADPDTRQAVAVIYNPLFDNINN--DTKDRPCNLLLSFKLR   173
T4 TS          TVWDENYENQAKDLGYHSGELGPIYGKQWRDFG----------GVDQIIEVIDRIKKLPNDRRQIVSAWNPAELKY-----MALPPCHMFYQFNVR   165
T4 CH          KIWQQVASSK--------GEINSNYGWAIWSEDN----------YAQYDMCLAELGQNPDSRRGIMIYTRPSMQFDYNKDGMSDFMCTNTVQYLIR   156

JL001          DRGQLDTVWNRSNDIIWGAYGANVVHMSILQEYVASAVG-------VEVGRYWQVSDDYHAYSDLFNKMLPLMS-----ASGVAPDDMYLAGVSEP   248
YuA            CDGRLDTVTNRSNDLIWGAYGANAVHFSYLHEYVARSVG-------VEQGIYRQVSANFHAYEEVLTKVAPLADLAANPMTGKETPDPYAAGIAEP   250
SP01           NGKLDLSVYNRSNDLHWGTFGANLCQFSTILEAMATWLG-------VEVGSYYQITDSLHVYLDDYGAKITEDIQK---VYGVNLETDEAPVVEDF   259
T4 TS          NGYLDLQWLQRSVDVFLG-LPFNIASYATLVHIVAKMCN-------LIPGDLIFSGGNTHIYMNHVEQCKEILRREPKELCELVISGLPYKFRYLS   253
T4 CH          DKKINAVVNMRSNDVVFG-FRNDYAWQKYVLDKLVSDLNAGDSTRQYKAGSIIWNVGSLHVYSRHFYLVDHWWKTGETHISKKDYVGKYA------   246

JL001          TPLMEQPESWQTWDQDLVMFLDE--------------------------PGAVGFRNAFFRKVCVPMMMAHKAYSQGRGEARYADAVEILSQMPPYN   319
YuA            YPLMSTDP—EEEWNQELMMFLSE--------------------------PDAVGFRDPFFRRVAIPMMKAHKAFKQTSNPSRFDAALAELDNVAAT-   320
SP01           EELLDPNPRMSYTMNELNQFLNEYFGIVDSLMRDDETYMHDGDCAQMLLNQIRHEPDEYIKVTLLLMVAKQALNRG-----MKDTVATAMNWIPLC   350
T4 TS          TEKQLKYVLKLRPKDFVLNNYVSHPPIKGKMAV----------------------------------------------------------   286
T4 CH          ----------------------------------------------------------------------------------------------   246

JL001          DWRIACMNWIQTRWDKYKKARDDGPNHDEV------------------ 350
YuA            DWKLAGVEWVERRRAAFEARKARAMDDGVAYE---------------- 351
SP01           EIKVSALRFLYKSYPEYINQYDLDEKLMKYIRREE------------- 383
T4 TS          ------------------------------------------------ 286
T4 CH          ------------------------------------------------ 246
```

Figure 8.4. Sequence alignment of the predicted dUMP hydroxymethylases of JL001, YuA, SPO1, and the T4 thymidylate synthase (TS) and dCMP hydroxymethylase (CH). The deoxyribose-binding, phosphate-binding and catalytically important residues (Song et al., 1999) are marked in blue, green and red, respectively. See text for more details.

residues (Glu60, Cys148 and Asp179), deoxyribose-binding residues (His216 and Tyr218) and phosphate-binding residues (Lys28, Arg123, Arg124, Arg168 and Ser169) (Figure 8.4). The presence of the modified base was shown in *Bacillus* phage SPO1 (Wilhelm and Rüger, 1992) and is also predicted for ϕJL001 and M6.

8.2.4 Genes involved in host interaction

The YuA genome region bearing genes with predicted functions in host interaction ranges from ORF25 to ORF45. Unique to YuA is the presence of a diguanylate-cyclase or GGDEF domain (ORF 44). This domain is widespread in bacterial proteins, functioning as a global second messenger controlling motility and adhesion in bacterial cells (Galperin et al., 2001). It was shown recently that GGDEF domains influence the transcriptional activity of genes encoding cell surface and membrane-bound proteins (Mendez-Ortiz *et al*., 2006) and that proteins containing these domains play a role in the formation and dissociation of biofilms of *P. putida* (Gjermansen *et al*., 2005). All 10 residues comprising the active site at the diguanylate cyclase dimerization interface are perfectly conserved in the N-terminal part of ORF 44, suggesting its functionality in influencing signal transduction pathways in *P. aeruginosa*. Remarkably, no similarity to ORF44 or to other GGDEF domains could be found using BLASTp, BLASTn or tBLASTn analysis against the current viral database.

The downstream gene product (gp45) displays high similarity with the ArdB antirestriction protein (pfam e-value of 10^{-30}), a plasmid-encoded protein which efficiently inhibits restriction by members of the three known families of type I restriction endonucleases (Belogurov *et al*., 1993). Although the presence of antirestriction proteins is common among T7-like *Podoviridae* (T7, T3, K1F, K1E, SP6), these enzymes are rarely found among temperate phages. The best-studied examples are the DarA and DarB proteins of the *E. coli* prophage P1, which are expressed late in the development cycle of P1 and are injected in the host cell along with the DNA, protecting the phage DNA upon infection (Lobocka *et al*., 2004). However, gp45 is not identified as part of the YuA phage particle (see below), suggesting a different role in the development cycle of this phage.

The putative repressor (gp25) contains helix-turn-helix motifs similar to those of the phage λ repressor, while the predicted integrase (gp26) differs significantly from the two major families of tyrosine and serine site-specific recombinases (Groth and Carlos, 2004). Thus far, we have not been able to isolate a stable lysogenic *P. aeruginosa* PAO1 strain. Our assays involved isolation of phage-resistant *P. aeruginosa* PAO1 clones from turbid plaques and analysis for the presence of phage sequences by PCR and DNA restriction analysis. Neither stably integrated, nor stably nonintegrated phage genome sequences could be detected. Similar results were reported for phage ϕJL001 (Lohr et

al., 2005) and *Vibrio parahaemolyticus* phages VP16T and VP16C (Seguritan et al., 2003). Strikingly, the YuA integrase exhibits 32% amino acid identity with the integrase of Vibriophages VP5 and VP2, which share similarity with VP16T and VP16C integrases. Apparently, these vibriophage-like integrase proteins are also unrelated to the well-studied tyrosine or serine recombinase families and exhibit distinct integrase behaviour.

8.2.5 Genes involved in particle formation and host lysis

ESI-MS/MS analysis of gel-separated phage particle proteins led to the experimental identification of 16 predicted proteins, reaching sequence coverages up to 66.1% (Table 8.2). Although no function could be assigned through homology searches, the two most abundant proteins were identified as gp56 and gp66, indicating a function as capsid and tail proteins. The abundance of these two proteins was confirmed by denaturation and subsequent MS analysis of whole phage particles, the so-called Whole Shotgun Approach (WSA, Section 2.4.2.3). The latter method allowed detection of additional peptides which were not found in the first analysis, thus increasing the overall coverage. Moreover, two additional YuA proteins (gp64 and gp75) were experimentally confirmed and identified as part of the virion particle (Table 8.2). These results delimit the YuA genome region involved in particle formation and host release to genes 52 up to 77, and it provides an experimental annotation of the particle proteins, accounting for 23% of the predicted ORFs.

Table 8.2. Identification of the structural proteins of YuA.

YuA gene	Mol. mass (kDa)	1-DE approach		WSA approach	
		# identified peptides	Sequence coverage (%)	# identified peptides	Sequence coverage (%)
52	56.1	13	28	-	-
54	45.0	14	32.3	2	10.4
55	27.9	4	14.6	-	-
56	32.2	20	66.1	22	88.2
57	19.1	2	22.9	2	23.4
59	5.5	2	35.7	3	46.4
63	18.6	8	36	3	29.7
64	13.4	-	-	3	63.1
65	16.4	3	22.7	1	10.4
66	54.0	33	63.3	21	66.1
69	100.4	24	26.3	-	-
70	56.2	6	13.2	-	-
71	35.7	9	40.6	4	24.5
72	62.2	9	17.6	3	15.1
73	30.1	6	24.5	-	-
74	8.2	-	-	1	37.5
76	76.9	16	27.0	2	3.4
77	17.1	3	36.6	-	-

Phage YuA encodes a delineated cluster of four overlapping lysis genes as typically observed in lambdoid and P22-like phages, and in the more closely related phage B3 (Young et al., 2000). The YuA lysis cassette is located within the genome region encoding phage particle proteins, presumably between the head and tail morphogenesis genes (Figure 8.2). Compared to the organization of the lysis cassette in lambdoid phages, the YuA genes within this cassette are rearranged since the putative endopeptidase *Rz* (gp60) and the embedded *Rz1* (gp60.1) (reading frame +1) genes precede the holin (gp61) and endolysin (gp62) gene (Bryl et al., 2000). YuA Rz and Rz1 proteins share similarity with the M6 gene products gp21 and gp21.1 (newly predicted based on homology to the YuA Rz1 protein), respectively, and serve presumable the same function as the corresponding proteins in ϕKMV-like viruses (Section 4.2.4.3). With the exception of phage M6 gp22, YuA gp61 does not show homology to known proteins. However, its position in the putative lysis cassette, its small size and three transmembrane domains with an N-terminal inside topology strongly suggest a function as a class I holin (Wang et al., 2000). Gp62 (the predicted endolysin) has a putative signal peptide sequence with a cleavage site at position 23 (ALA-QD). The presence of a signal peptide and lytic activity of gp62 is supported by the observation of cellular toxicity of recombinantly expressed gp62 in absence of sodium azide, an inhibitor of the Sec secretion complex. Expression in presence of sodium azide largely elevates toxicity of gp62 expression (Briers Y., unpublished results). The presence of a signal peptide sequence suggests the need for an inhibitory mechanism of the exported lysins to prevent early lysis of the host cells. Recent research showed that full activity of the previously exported fOg44 lysin is only achieved after sudden, ion-unspecific dissipation of the proton motive force, an event undertaken by the holin (Nascimento et al., 2008). In this way, holins still function as a time regulator of lysis.

8.3 Discussion

Despite producing small, turbid plaques, the temperate nature of YuA was not unambiguously confirmed by study of its genome architecture. Although YuA encodes proteins (gp25 and gp26) which are significantly similar to putative phage repressors and integrases, all attempts to isolate a stable lysogenic *P. aeruginosa* strain failed. The YuA integrase may require specific – not yet determined – physiological conditions or a different host strain for stable lysogenic establishment, or the phage may exhibit an extremely high level of spontaneous induction, as observed for phage phiHSIC (Williamson et al., 2001). Alternatively, the integrase gene may not be active anymore, and may reside as an artifact in the YuA genome, a hypothesis currently being evaluated by partial sequencing of 14 newly isolated YuA-like phages (K. Miroshnikov, personal communication).

The clear vertical evolution observed between YuA and M6, isolated in different continents over a 40 year period, contrasts strongly with the mosaic nature of their genome in respect to (among others) phages ϕJL001, 73, D3112 and B3 (Figure 8.2). Although this genome architecture implies a magnitude of phage-(pro)phage and phage-host interactions, no far-stretching conclusions towards infective nature can be drown from this. A possible clue might come from the observation that YuA seems to be able to replicate its genome independently from its host, which was postulated as a hallmark feature of virulent phages (Chopin et al., 2001; Chen and Lu, 2002): in addition, ORFs 2-6 in this genomic region display significant homology with the DNA replication genes of the strictly virulent PB1 genus (Chapter 7, Table 7.2). The relationship between phage and host needs to be clarified by future research, for example by monitoring host cell growth and bacteriophage production, by transcriptome analysis and/or by functional analysis of the integrase and repressor proteins.

Chapter 9

General Conclusions and Perspectives

"Now this is not the end.
It is not even the beginning of the end.
But it is, perhaps, the end of the beginning."

- Winston Churchill. London, 10 November 1942.

9.1 The global spread of bacteriophage species

9.1.1. Genome sequencing of *Pseudomonas* phages

The main goal of this PhD dissertation was to study the global genomic diversity of virulent bacteriophages infecting the opportunistic pathogen *Pseudomonas aeruginosa*. This bacterial species is ubiquitous in the biosphere, has wide metabolic versatility and high intrinsic and acquired resistance to antimicrobials. It can be found in a wide variety of ecological environments ranging from fresh and salt water to the rhizosphere in which they colonize the endemic fauna (*e.g.*, nematodes), flora and fungi (*e.g.*, *Pythium* spp.) (Goldberg, 2000). As a consequence, their natural predators are equally widespread and can be readily isolated from almost any rich organic soil or water sample. According to the most recent counts, 511 *P. aeruginosa* phages are already visualized using electron microscopy (Ackermann et al., 2007), but the great majority of that number is stored uncharacterized in various phage collections worldwide. At the start of this PhD research, six genera containing *P. aeruginosa* phages were defined (Section 1.5.2).

In the first two years of this research project, multiple rounds of phage isolation resulted in the in-depth characterization of a library of fifteen diverse and virulent dsDNA phages, varying in genome size between 30 and 75 kb and infecting a wide variety of *P. aeruginosa* strains. A newly developed typing method based on '*de novo*' mass spectrometric analysis of phage particle proteins, which could easily be applied to other uncharacterized phages, clustered these phages into seven distinct groupings (Chapter 3). Five of these groupings were studied in detail on a genomic and (structural) proteomic level, leading to the establishment of the newly defined genera of the ϕKMV-like (Chapter

4), LUZ24-like (Chapter 5), PB1-like (Chapter 7) and YuA-like (Chapter 8) viruses, and to the identification of the first N4-like *Pseudomonas* phages (Chapter 6).

One of most intriguing results which immediately emerged from this analysis was the level of sequence conservation among phages within specific genera. Apparently, closely related members of these ϕKMV-like and LUZ24-like *Podoviridae*, YuA-like *Siphoviridae* and PB-1-like *Myoviridae* can be readily isolated from different habitats across the globe. A striking example is the >90% DNA identity present between phages PB1 and 14-1, which were isolated in Scotland and Russia, respectively, with a 50 year interval between both isolations (Table 7.1). The main genetic variation within each genus is largely confined to genes encoding the early proteins and the tail fibers, indicative for local adaption necessary to infect specific hosts in specific environmental conditions. In any case, this finding implies that despite possible rapid phage evolution and short generation times, 'successful' viral genomes can be stably maintained over ecologically significant time and distance.

On a technical level, the occurrence of strongly related phages opens far-reaching opportunities for comparative genome annotations, providing deeper insights into their coding sequences and genome regulation. As automated annotation pipelines are developed to control the rapidly growing amount of DNA sequence data, small and/or overlapping potential ORFs (<150 bp) are usually ignored in the annotation of phage genomes and are consequently overlooked in later stages of phage research. However, embedded ORFs may encode important biological functions, like the phage T4-encoded Stp protein, which is only 26 amino acids long and activates the host PrrC lysyl-tRNA endonuclease (Penner et al., 1995). Based on sequence conservation, small and strongly conserved phage ORFs should be admitted in protein databases (*e.g.*, sections 4.2.2 and 7.2.1).

9.1.2. Every phage everywhere

At the moment, it remains elusive whether the observed genera actually exist, or whether they are an artifact of current sampling, and the 'phage universe' is a continuum where all phage types grade into each other (Casjens, 2005). For example, coliphage T7 was long considered to be the hallmark virulent phage and many closely related virulent phages have been isolated (section 4.4.2). However, its unique transcriptional system was also found in a prophage in the genome of *Pseudomonas putida* KT2440 (Nelson et al., 2002) and in the hybrid myovirus ϕEcoM-GJ1 (Jamalludeen et al., 2008), and a functional site-specific integrase is encoded by the marine T7-like phage P-SSP7 (Sullivan et al., 2005).

Reality no doubt lies somewhere in between, and comprehensive appreciation of phage diversity and evolution will, in any case, require the sequencing of additional phage genomes. However, the highly

discrete genome architectures observed among world-wide collected *P. aeruginosa* phages does suggest strongly the existing of distinct phage genera and that overall genetic phage diversity is large, but finite. It has been argued in literature that during the early phase of the evolution of an organism, horizontal genetic transfer may play a key role, but when it crosses the "Darwinian threshold", species borders are established and vertical genetic changes become more important (Woese, 2002; Kovalyova and Kropinski**,** 2003). It is reasonable to assume that this also occurred in the evolution of our virulent phage species, which developed a highly efficient life cycle which became globally spread. In the course of this research project, similar geographical distributions have been observed for many phage types infecting a wide variety of hosts like *E. coli*, *Xanthomonas oryzae* and *Lactococcus lactis* (Nolan et al., 2006; Lee et al., 2007; Labrie et al., 2008).

In conclusion, we speculate that all viral species infecting *P. aeruginosa* are globally widespread, reflecting the ubiquitous nature of its host. This apparent ambiguity between enormous local viral diversity observed in metagenomic projects (e.g., at least 10^4 viral genotypes in one kilogram of marine sediment) and the proposed wide global spread of viral species implies that phages are able to move between different habitats. Sano and colleagues (2004) incubated viral communities from soil, sediment and freshwater with marine microbial communities, and observed viral community growth within 48 hours, showing that phages can indeed find hosts in different biomes. This migration provides opportunities for global exchange of DNA among viral genomes, and is in agreement with the 'Kill-the-Winner' model described in section 1.2.2. What sets different phage communities apart is likely the change in abundance of its most abundant members, which supports to some extent the old microbiological tenet "everything is everywhere, but, the environment selects" (De Wit and Bouvier, 2006).

9.1.3. How many types remain to be discovered?

The question remains, how far will this diversity extend, and how many phage genera are yet to be discovered? With the global amount of phages being estimated at 10^{31} particles, at this time, any estimation regarding the size of the pool of unknown phage types remains extremely speculative. Considering well-studied model organisms like *E. coli*, the discovery of truly new phage types is quite rare, and most new isolates are simple variations of known phages with identical sets of genes and more than 90% identity at the DNA level (Savalia et al., 2008). Of the fifteen phages characterized in this study, only the N4-like phages LUZ7 and LIT1 (Chapter 6) and phages LKR3 and LSL4 (Table 3.4; unpublished data) represent new *Pseudomonas* phage types, as the majority of our isolates are at

least partially related to previously known (but not fully characterized, as was the case for PaP3, M6 and F8) *Pseudomonas* phages.

In contrast, the study of phages infecting non-model organisms is lagging far behind, as illustrated in Figure 1.5 and by the metagenomics-based (and again, extremely speculative) estimate of ten million different tailed-phage species in the environment (Rohwer, 2003). In this respect, it is clear that the point at which sequencing more phage genomes will fail to add significantly to our understanding of phage diversity and evolution is far from reached. However, a clear condition for this seems to be that the sampling and characterization of new phages moves away from standard isolation techniques and/or well-studied bacterial host organisms. Recent advances in field of culturing nonculturable bacteria (Lewis, 2007) can create a magnitude of new opportunities for phage research.

Another angle for future genomic diversity studies involves skipping the culturing step, and assembling whole phage genomes directly from metagenomic sequence data. While this is perfectly feasible on a technical level, major problems in phage classification emerge since no particle and host organism can be linked to these genomes, which seriously impedes any other analysis.

9.2 Future applications and prospects

9.2.1 Phage therapy and targeted phage modifications

The most straight-forward phage application is their direct use in phage therapy settings. Except for the 'YuA-like viruses' (Chapter 8), all genera comprise highly lytic phages devoid of any virulence-associated genes, and thus suitable for applications in phage therapy. Various phage cocktails are possible (Table 3.3), but a mixture of LBL3, LUZ19 and LUZ24 lyses 72% of a collection of 122 clinical *P. aeruginosa* strains and targets different primary receptors, which would strongly reduce the emergence of phage resistance (Tanji *et al*., 2004).

An interesting study by Vitiello and colleagues (2005) described spontaneous lambda phage mutants with 13,000-16,000-fold greater capacity to remain in the mouse circulatory system 24h after intraperitoneal injection, due to a single mutation in the phage capsid. However, in contrast to bacteria and temperate phages, no well-defined system for targeted genome modifications exists for virulent phages. Recently, a highly efficient system for directed mutagenesis of mycobacteriophages was described, based on recombineering of electroporated phage DNA (Marinelli et al., 2008). This system will be implemented for virulent *P. aeruginosa* phages, and has numerous possible

applications like directed capsid and tail fibers mutations, and the introduction of specific 'TAP' tags in phage genomes.

9.2.2 The exploitation of phage genetic information

Of the total of 1,028 novel phage genes which were newly identified in this study, a specific function in DNA replication, particle assembly or host lysis could be predicted for 361 genes, and 106 genes were experimentally confirmed as encoding structural proteins. As a consequence, 667 novel phage genes without any indication towards their possible role or function in phage infection have been deposited in the databases as "hypothetical proteins". While metagenomic studies suggest the presence of billions of unexplored phage genes (Rohwer, 2003), the gap between accumulating sequence data and designation of biological functions to each of these genes becomes almost impossible to bridge. For example, even despite decades of intensive study, 130 out of 300 potential T4 genes remain uncharacterized. Therefore, studies trying to unravel unknown gene functions (*e.g.*, Scaltriti et al., 2009; Roucourt et al., 2009) and ultrastructural studies elucidating the roles of individual particle proteins in phage virions (*e.g.*, Choi et al., 2008) should be greatly encouraged.

Nevertheless, the unraveling of each novel phage genome provides a direct supply of potentially interesting proteins, ready for further exploitation. Aiming at a high-throughput identification of novel antibacterial proteins among the 667 hypothetical gene products, a selection of 152 of these genes is currently being cloned and expressed in *P. aeruginosa* PAO1 to assess their antibacterial potential. *P. aeruginosa*-specific peptidoglycan degrading proteins (endolysins) which are identified in this study are currently being cloned and characterized, as are the putative polysaccharide depolymerases of LKA1 and LUZ7 which may allow these phages to degrade their way through host biofilms to gain access to the bacterial surfaces (Section 4.2.4.2; Table 6.1). Also of interest are the phage-encoded RNA Polymerases of the ϕKMV-like and N4-like phages (Sections 4.2.3 and 6.2.3, resp.) which bind to phage-specific promoters and can be potentially applied in *in vitro* transcriptional assays. As a final example, tail fiber proteins which efficiently bind specific bacterial outer membrane structures can be used in stead of antibodies in the detection of bacterial pathogens in food, feed, environmental samples and soil (*e.g.,* the VIDAS®UP *E. coli* O157 system, Profos AG, Regensburg, Germany).

Finally, an attractive target for development of new antibiotics is the bacterial transcriptional machinery, which is hijacked by all known phages at some point in infection. The bacterial RNAP, which catalyzes this transcription, is a large multisubunit enzyme which is expected to have multiple

sites that could be bound by small molecules (Chopra, 2007). Although the transcriptional machinery of prokaryotes and eukaryotes exhibits similarities in structure and mechanism, there is sufficient divergence in even the most conserved RNAP subunits to make specific targeting of bacterial or eukaryotic RNAPs possible (Sousa, 2008). Our Tandem Affinity Purification assay (section 4.3) allows the purification of phage proteins which specifically bind to and modify the bacterial RNAP complex in the course of infection. The analysis was already performed for phage LUZ19, and will be extended to all newly isolated phages. Identified RNAP-interacting proteins can be expressed individually and tested for antibacterial activity and/or their ability to abolish *in vitro* bacterial transcription.

Phage research is built on genetics. However, the genetic analyses are only the foundations of future research. So, already in 1942 and despite a slightly different context, Winston Churchill was right. This is not the end; we might just have reached the end of the beginning.

Reference list

Ackermann, H.W. (2007) 5500 Phages examined in the electron microscope. *Arch Virol.* **152**: 227-243.

Ackermann, H.W., and Krisch, H.M. (1997) A catalogue of T4-type bacteriophages. *Arch Virol.* **142**: 2329-2345.

Ackermann, H.W., and Dubow, M.S. (1987) Phages of *Pseudomonas* and related bacteria. *In: Viruses of Prokaryotes: Natural groups of Bacteriophages*. CRC Press, Florida, pp. 116-121.

Ackermann, H.W., Cartier, C., Slopek, S., and Vieu., J.F. (1988) Morphology of *Pseudomonas aeruginosa* typing phages of the Lindberg set. *Ann Inst Pasteur Virol.* **139**: 389-404.

Ackloo, S., and Loboda, A. (2005) Applications of a matrix-assisted laser desorption/ionization orthogonal time-of-flight mass spectrometer. l. Metastable decay and collision-induced dissociation for sequencing peptides. *Rapid Commun Mass Spectrom.* **19**: 213-220.

Addona, T., and Clauser, K. (2002) De novo peptide sequencing via manual interpretation of MS/MS spectra. *Curr Protoc Protein Sci.* **16**:16.11.

Alonso, A., Rojo, F., and Martínez, J.L. (1999) Environmental and clinical isolates of *Pseudomonas aeruginosa* show pathogenic and biodegradative properties irrespective of their origin. *Environ Microbiol.* **1**: 421-430.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol.* **215**: 403-410.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev.* **59**: 143-169.

Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., Felts, B., Nulton, J., Mahaffy, J., and Rohwer, F. (2005) PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *BMC Bioinformatics* **6**: 41.

Angly, F., Youle, M., Nosrat, B., Srinagesh, S., Rodriguez-Brito, B., McNairnie, P., Deyanat-Yazdi, G., Breitbart, M., and Rohwer, F. (2009) Genomic analysis of multiple Roseophage SIO1 strains. *Environm Microb., in press.*

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., et al. (2006) The marine viromes of four oceanic regions. *PLoS Biol.* **4**: e368.

Arai, H., Hayashi, M., Kuroi, A., Ishii, M., and Igarashi, Y. (2005) Transcriptional regulation of the flavohemoglobin gene for aerobic nitric oxide detoxification by the second nitric oxide-responsive regulator of *Pseudomonas aeruginosa*. *J Bacteriol.* **187**: 3960-3968.

Ashelford, K.E., Day, M.J., and Fry, J.C. (2003) Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol.* **69**: 285-289.

Bailey, T.L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers, *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, California, pp. 28-36.

Bailey, T.L., and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48-54.

Baker, M.L., Jiang, W., Rixon, F.J., Chiu, W. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol.* **79**: 14967-14970.

Barenboim, M., Chang, C.Y., dib Hajj, F., and Young, R. (1999). Characterization of the dual start motif of a class II holin gene. *Mol Microbiol.* **32**: 715-727.

Barrios, H., Valderrama, B., and Morett, E. (1999) Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Res.* **27**: 4305-4313.

Beck, L.L., Smith, T.G., and Hoover, T.R. (2007) Look, no hands! Unconventional transcriptional activators in bacteria. *Trends Microbiol.* **15**: 530-537.

Belogurov, A.A., Delver, E.P. and Rodzevich, O.V. (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J Bacteriol.* **175**: 4843-4850.

Bench, S.R., Hanson, T.E., Williamson, K.E., Ghosh, D., Radosovich, M., Wang, K., and Wommack, K.E. (2007) Metagenomic characterization of Chesapeake Bay virioplankton. *Appl Environ Microbiol.* **73**: 7629-7641.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* **340**: 783-795.

Benson, S.D., Bamford, J.K., Bamford, D.H., Burnett, R.M. (2004) Does common architecture reveal a viral lineage spanning all three domains of life? *Mol Cell* **16**: 673-685.

Bergan, T. (1972) A new bacteriophage typing set for *Pseudomonas aeruginosa.* I. Selection procedure. *Acta Pathol Microbiol Scand. B* **80**: 117-180.

Bergh, O., Børsheim, K.Y., Bratbak, G., and Heldal, M. (1989) High abundance of viruses found in aquatic environments. *Nature* **340**: 467-468.

Bernad, A., Blanco, L., Lázaro, J.M., Martín, G., and Salas, M. (1989) A conserved 3'-5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell* **59**: 219-228.

Berry, J., Summer, E.J., Struck, D.K., and Young, R. (2008) The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. *Mol Microbiol.* **70**: 341-351.

Biers, E.J., Wang, K., Pennington, C., Belas, R., Chen, F., and Moran, M.A. (2008) Occurrence and expression of gene transfer agent genes in marine bacterioplankton. *Appl Environ Microbiol.* **74**: 2933-2939.

Biswas, B., Adhya, S., Washart, P., Paul, B., Trostel, A.N., Powell, B., Carlton, R., and Merril, C.R. (2002) Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant *Enterococcus faecium*. *Infect Immun.* **70**: 204-210.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., Ebright, R.H. (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78**: 889-896.

Bonocora, R.P., and Shub, D.A. (2004). A Self-Splicing Group I Intron in DNA Polymerase Genes of T7-Like Bacteriophages. *J Bacteriol.* **163**: 8153-8155.

Botstein, D., and Herskowitz, I. (1974) Properties of hybrids between Salmonella phage P22 and coliphage lambda. *Nature* **251**: 585-589.

Boulanger, P., Jacquot, P., Plançon, L., Chami, M., Engel, A., Parquet, C., Herbeuval, C., and Letellier, L. (2008) Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. *J Biol Chem.* **283**: 13556-13564.

Bouvier, T., del Giorgio, P.A. (2007) Key role of selective viral-induced mortality in determining marine bacterial community composition. *Environ Microbiol.* **9**: 287-297.

Boyd, E.F., and Brussow, H. (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends Microbiol.* **10**: 521–529.

Bradley, D.E., and Pitt, T.L. (1974) Pilus-dependence of four *Pseudomonas aeruginosa* bacteriophages with non-contractile tails. *J Gen Virol.* **24**: 1-15.

Braid, M.D., Silhavy, J.L., Kitts, C.L., Cano, R.J., and Howe, M.M. (2004) Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa. J Bacteriol.* **186**: 6560-6574.

Breitbart, M., Miyake, J. H., and Rohwer, F. (2004) Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett.* **236**: 249-256.

Brüssow, H., and Hendrix, R.W. (2002) Phage genomics: small is beautiful. *Cell* **108**: 13-16.

Bruttin, A., and Brüssow, H. (2005) Human volunteers receiving *Escherichia coli* phage T4 orally: a safety test of phage therapy. *Antimicrob Agents Chemother.* **49**: 2874-2878.

Bruynoghe, R., and Maisin, J. (1921) Essais de thérapeutique au moyen du bacteriophage du Staphylocoque. *J Compt Rend Soc Biol.* **85**: 1120-1121.

Buckling, A., Rainey, P.B. (2002) Antagonistic coevolution between a bacterium and a bacteriophage. *Proc Biol Sci.* **269**: 931-936.

Buckling, A., Wei, Y., Massey, R.C., Brockhurst, M.A., and Hochberg, M.E. (2006) Antagonistic coevolution with parasites increases the cost of host deleterious mutations. *Proc Biol Sci.* **273**: 45-49.

Budzik, J.M., Rosche, W.A., Rietsch, A., and O'Toole, G.A. (2004) Isolation and characterization of a generalized transducing phage for *Pseudomonas aeruginosa* strains PAO1 and PA14. *J Bacteriol.* **186**: 3270-3273.

Burkal'tseva, M.V., Pleteneva, E.A., Shaburova, O.V., Kadykov, V.A., and Krylov, V.N. (2006) Genome conservatism of phiKMV-like bacteriophages (T7 supergroup) active against *Pseudomonas aeruginosa*. *Genetika* **42**: 33-38.

Byrne, M., and Kropinski, A.M. (2005) The genome of the *Pseudomonas aeruginosa* generalized transducing bacteriophage F116. *Gene* **346**: 187-194.

Carey-Smith, G.V., Billington, C., Cornelius, A.J., Hudson, J.A., and Heinemann, J.A. (2006) Isolation and characterization of bacteriophages infecting *Salmonella* spp. *FEMS Microbiol Lett*. **258**: 182-186.

Carlson, K. (2005) Working with bacteriophages: common techniques and methodological approaches. In: *Bacteriophages, Biology and Applications*. Kutter, E., and Sulakvelidze, A. (eds.). CRC Press, Florida, pp. 437-494.

Carter, R.H., Demidenko, A.A., Hattingh-Willis, S., and Rothman-Denes, L.B. (2003) Phage N4 RNA polymerase II recruitment to DNA by a single-stranded DNA-binding protein. *Genes Dev.* **17**: 2334-2345.

Casjens, S.R. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol.* **49**: 277-300.

Casjens, S.R. (2005) Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol.* **8**: 451-458.

Casjens, S.R. (2008) Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae. *Res Microbiol.* **159** : 340-348.

Casjens, S.R., and Gilcrease, E.B. (2009) Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol Biol.* **502**: 91-111.

Casjens, S.R., Gilcrease, E.B., Winn-Stapley, D.A., Schicklmaier, P., Schmieger, H., Pedulla, M.L., Ford, M.E., Houtz, J.M., Hatfull, G.F., and Hendrix, R.W. (2005) The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol.* **187**: 1091-1104.

Catalano, C.E., Cue, D., and Feiss, M. (1995) Virus DNA packaging: the strategy used by phage lambda. *Mol Microbiol.* **16**: 1075-1086.

Cheetham, G.M., and Steitz, T.A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**: 2305-2309.

Cheetham, G.M., and Steitz, T.A. (2000) Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr Opin Struct Biol.* **10**: 117-123.

Chen, F., and Lu, J. (2002) Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol*. **68**: 2589-2594.

Chibani-Chennoufi, S., Dillmann, M.L., Marvin-Guy, L., Rami-Shojaei, S., and Brüssow, H. (2004) Lactobacillus plantarum bacteriophage LP65: a new member of the SPO1-like genus of the family Myoviridae. *J Bacteriol*. **186**: 7069-7083.

Choi, K.H., Kumar, A., and Schweizer, H.P. (2006) A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J Microbiol Methods* **64**: 391-397.

Choi, K.H., McPartland, J., Kaganman, I., Bowman, V.D., Rothman-Denes, L.B., Rossmann, M.G. (2008) Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol.* **378**: 726-736.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D., and Chopin, M. (2001) Analysis of six prophages in *Lactococcus lactis*Lactococcus lactis IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* **29**: 644-651.

Chopra, I. (2007) Bacterial RNA polymerase: a promising target for the discovery of new antimicrobial agents. *Curr Opin Investig Drugs***. 8**: 600-607.

Claverie, J.M., Ogata, H. (2009) Ten good reasons not to exclude giruses from the evolutionary picture. *Nat Rev Microbiol*. **7**: 615.

Claverie, J.M., Ogata, H., Audic, S., Abergel, C., Suhre, K., and Fournier, P.E. (2006) Mimivirus and the emerging concept of "giant" virus. *Virus Res.* **117:** 133-144.

Clokie, M.R., Thalassinos, K., Boulanger, P., Slade, S.E., Stoilova-McPhie, S., Cane, M., *et al.* (2008) A proteomic approach to the identification of the major virion structural proteins of the marine cyanomyovirus S-PM2. Microbiology **154**: 1775-1782.

Comeau, A.M., and Krisch, H.M. (2005) War is peace – Dispatches from the bacterial and phage killing fields. *Curr Opin Microbiol.* **8**: 488-494.

Comeau, A.M., Bertrand, C., Letarov, A., Tétart, F., and Krisch, H.M. (2007) Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* **362**: 384-396.

Comeau, A.M., Hatfull, G.F., Krisch, H.M., Lindell, D., Mann, N.H., and Prangishvili, D. (2008) Exploring the prokaryotic virosphere. *Res Microbiol*. **159**: 306-313.

Conrotto, P., and Hellman, U. (2008) Lys Tag: an easy and robust chemical modification for improved de novo sequencing with a matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom*. **22**: 1823-1833.

Cortez, D., Forterre, P., and Gribaldo, S. (2009) A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol.* **10**: R65.

Cottrell, M. T. and Suttle, C. A. (1995) Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate, *Micromonas pusilla*. *Limnol Oceanogr.* **40**: 730-739.

Crucitti, P., Abril, A.M., Salas, M. (2003) Bacteriophage phi 29 early protein p17. Self-association and hetero-association with the viral histone-like protein p6. *J Biol Chem*. **278**: 4906-4911.

Culley, A.I., Lang, A.S., and Suttle, C.A. (2006) Metagenomic analysis of coastal RNA virus communities. *Science* **312**: 1795-1798.

Dalphin, M.E., Stockwell, P.A., Tate, W.P., and Brown, C.M. (1999) TransTerm, the translational signal database, extended to include full coding sequences and untranslated regions. *Nucleic Acids Res.* **27**: 293-294.

Dasgupta, N., Arora, S.K., and Ramphal, R. (2000) fleN, a gene that regulates flagellar number in *Pseudomonas aeruginosa*. *J Bacteriol.* **182**: 357-364.

Daubin, V., and Ochman, H. (2004) Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* **14**: 1036-1042.

Davis, B.M., and Waldor, M.K. (2000). CTXphi contains a hybrid genome derived from tandemly integrated elements. *PNAS USA* **97**: 8572-8577.

Davydova, E.K., Santangelo, T.J., and Rothman-Denes, L.B. (2007) Bacteriophage N4 virion RNA polymerase interaction with its promoter DNA hairpin. *PNAS USA* **104**: 7033-7038.

De Vos, D., Lim, A. Jr., Pirnay, J.P., Duinslaeger, L., Revets, H., Vanderkelen, A., *et al.* (1997) Analysis of epidemic *Pseudomonas aeruginosa* isolates by isoelectric focusing of pyoverdine and RAPD-PCR: modern tools for an integrated anti-nosocomial infection strategy in burn wound centres. *Burns* **23**: 379-386.

De Wit, R., and Bouvier, T. (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol.* **8**: 755-758.

Deresinski S. (2009) Bacteriophage therapy: exploiting smaller fleas. *Clin Infect Dis*. **48**: 1096-1101.

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., Liu, H., Furlan, M., Wegley, L. et al. (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452**: 340-343.

d'Herelle, F. (1917) Sur un microbe invisible antagoniste des bacilles dysentériques. *C. R. Acad. Sci. Paris* **165**: 373–375.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M.A., Nelson, K.E., Nilsson, C., et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* **452**: 629-632.

Dobbins, A.T., George, M. Jr., Basham, D.A., Ford, M.E., Houtz, J.M., Pedulla, M.L., Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2004) Complete genomic sequence of the virulent Salmonella bacteriophage SP6. *J Bacteriol.* **186**: 1933-1944.

Dominguez-Cuevas P. and Sylvia, S. (2004) Compiling sigma-70 dependent promoters. In: Ramos, J.L. (ed.), *Pseudomonas*: virulence and gene regulation, vol. II. Kluwer Academic/Plenum, Boston, MA, pp. 319-343.

Donelli, G. (1968) Isolamento di un batteriofago di eccezionali dimensioni attivo su *B. megatherium*. *Atti Accad Naz Lincei-Rend Clas Sci Fis Mat Nat.* **44**: 95-97.

Dunn, J. J., and Studier, F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol.* **166**: 477-535.

Earnshaw, W.C., and Harrison, S.C. (1977) DNA arrangement in isometric phage heads. *Nature* **268**: 598-602.

Edwards, R.A., and Rohwer, F. (2005) Viral metagenomics. *Nat Rev Microbiol.* **3**: 504-510.

Ehrlich, M., Ehrlich, K., and Mayo, J.A. (1975) Unusual properties of the DNA from Xanthomonas phage XP-12 in which 5-methylcytosine completely replaces cytosine. *Biochim Biophys Acta*. **395**: 109-119.

Engel, J., and Balachandran, P. (2009) Role of *Pseudomonas aeruginosa* type III effectors in disease. *Curr Opin Microbiol.* **12**: 61-66.

Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O., and Salzberg, S.L. (2000) Prediction of transcription terminators in bacterial genomes. *J Mol Biol*. **301**: 27-33.

Faruque, S.M., Islam, M.J., Ahmad, Q.S., Faruque, A.S., Sack, D.A., Nair, G.B., and Mekalanos, J.J. (2005) Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *PNAS USA* **102**: 6119-6124.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., and Ysebaert, M. (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**: 500-507.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D., and Bossi, L. (2001) Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. *Mol Microbiol.* **39**: 260-271.

Filée, J., Bapteste, E., Susko, E., and Krisch, H.M. (2006) A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. *Mol Biol Evol*. **23**: 1688-1696.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, J.C., Hotz, H.R., Ceric, G., Forslund, K., Eddy, K.R., Sonnhammer, E.L., and Bateman, A. (2008) *Nucleic Acids Res Database Issue* **36**: D281-D288.

Fischer, D., and Eisenberg, D. (1999) Finding families for genomic ORFans. *Bioinformatics* **15**: 759-762.

Ford, M.E., Sarkis, G.J., Belanger, A.E., Hendrix, R.W., and Hatfull, G.F. (1998). Genome structure of mycobacteriophage D29: implications for phage evolution. *J Mol Biol.* **279**: 143-164.

Frederickson, C.M., Short, S.M., and Suttle, C.A. (2003) The physical environment affects cyanophage communities in British Columbia Inlets. *Microbial Ecology* **46**: 348-357.

Freeman, V.J. (1951) Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *J Bacteriol.* **61**: 675-688.

Freitas-Vieira, A., Anes, E., and Moniz-Pereira, J. (1998). The site-specific recombination locus of mycobacteriophage Ms6 determines DNA integration at the tRNA(Ala) gene of Mycobacterium spp. *Microbiology* **144**: 3397-3406.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* **3**: 722-732.

Galperin, M.Y., Nikolskaya, A.N., and Koonin, E.V. (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol Lett.* **203**: 11-21.

Garcia, E., Elliott, L.M., Ramanculov, E., Chain, P.S., Chu, M.C., and Molineux, I.J. (2003) The genome sequence of *Yersinia pestis* bacteriophage phiA1122 reveals an intimate history with the coliphage T3 and T7 genomes. *J Bacteriol*. **185**: 5248-5262.

Gherardini, P.F., Wass, M.N., Helmer-Citterich, M., Sternberg, M.J. (2007) Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol*. **372**: 817-845.

Gill, J.J., Svircev, A.M., Smith, R., and Castle, A.J. (2003) Bacteriophages of *Erwinia amylovora*. *Appl Environ Microbiol.* **69**: 2133-2138.

Gjermansen, M., Ragas, P., Sternberg, C., Molin, S., and Tolker-Nielsen, T. (2005) Characterization of starvation-induced dispersion in *Pseudomonas putida* biofilms. *Environ Microbiol.* **7**: 894-906.

Glazko, G., Makarenkov, V., Liu, J., and Mushegian, A. (2007) Evolutionary history of bacteriophages with double-stranded DNA genomes. *Biol Direct.* **2**: 36.

Gleghorn, M.L., Davydova, E.K., Rothman-Denes, L.B., and Murakami, K.S. (2008) Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell.* **32**: 707-717.

Goldberg, J.B. (2000) Pseudomonas: global bacteria. *Trends Microbiol.* **8**: 55-57.

Gordon, L., Chervonenkis, A.Y., Gammerman, A.J., Shahmuradov, I.A., and Solovyev, V.V. (2003) Sequence alignment kernel for recognition of promoter regions. *Bioinformatics* **19**: 1964-1971.

Gordon, L., Chervonenkis, A.Y., Gammerman, A.J., Shahmuradov, I.A., and Solovyev, V. (2003) Sequence Alignment Kernel for recognition of promoter regions. *Bioinformatics* **19**: 1964-1971.

Gourse, R.L., Ross, W., and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol.* **37**: 687-695.

Greenblatt, J., Nodwell, J.R., and Mason, S.W. (1993) Transcriptional antitermination. *Nature* **364**: 401-406.

Groth, A.C., and Carlos, M.P. (2004) Phage integrases: biology and applications. *J Mol Biol.* **335**: 667-678.

Guttman, B., Raya, R., and Kutter, E. (2005) Basic Phage Biology. In: *Bacteriophages: Biology and Applications*, CRC Press, pp. 29-63.

Hampsey, M. (2001) Omega meets its match. *Trends in Genetics* **17:** 190-191.

Hardies, S.C., Comeau, A.M., Serwer, P., and Suttle, C.A. (2003) The complete sequence of marine bacteriophage VpV262 infecting *Vibrio Parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology* **310**: 359-371.

Hashemolhosseini, S., Stierhof, Y.D., Hindennach, I., and Henning, U. (1996) Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and lambda. *J Bacteriol.* **178**: 6258-6265.

Hatfull, G.F. (2008) Bacteriophage genomics. *Curr Opin Microbiol.* **11**: 447-453.

Heldal, M., Bratbak, G. (1991) Production and decay of viruses in aquatic environments. *Mar Ecol Prog Ser.* **72**: 205-212.

Hendrix, R., and Casjens, S.G. (2005) Caudovirales, *Podoviridae, Siphoviridae, Myoviridae*. In: *Virus Taxonomy. VIIIth Report of the International Committee on Taxonomy of Viruses.* Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., and Ball, A. (eds.). Elsevier Academic Press, Oxford, pp. 35-79.

Hendrix, R.W. (2002) Bacteriophages: evolution of the majority. *Theor Popul Biol.* **61**: 471-480.

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. *PNAS USA* **96**: 2192-2197.

Hennes, K.P., Suttle, C.A., and Chan, A.M. (1995) Fluorescently Labeled Virus Probes Show that Natural Virus Populations Can Control the Structure of Marine Microbial Communities. *Appl Environ Microbiol.* **61**: 3623-3627.

Heo, Y.J., Chung, I.Y., Choi, K.B., Lau, G.W., and Cho, Y.H. (2007) Genome sequence comparison and superinfection between two related *Pseudomonas aeruginosa* phages, D3112 and MP22. *Microbiology* **153**: 2885-2895.

Hertveldt, K., and Lavigne, R. (2008) Bacteriophages of *Pseudomonas*. In: *Pseudomonas, Model Organism, Pathogen, Cell Factory*. Rehm, B.A. (ed.) WILEY-VCH Verlag GmbH & Co, Weinheim, Germany, pp. 255-292.

Hertveldt, K., Lavigne, R., Pleteneva, E., Sernova, N., Kurochkina, L., Korchevskii, R., Robben, J., Mesyanzhinov, V., Krylov, V.N., and Volckaert, G. (2005) Genome comparison of *Pseudomonas aeruginosa* large phages. *J Mol Biol.* **354**: 536-545.

Hoang, T.T., Karkhoff-Schweizer, R.R., Kutchma, A.J., and Schweizer, H.P. (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* **212**: 77-86.

Holloway, B.W., Egan, J.B., and Monk, M. (1960) Lysogeny in *Pseudomonas aeruginosa*. *Austr J Exp Biol.* **38**: 321-330.

Hsiao, C.L., and Black, L.W. (1978) Head morphogenesis of bacteriophage T4. III. The role of g20 in DNA packaging. *Virology* **91**: 26-38.

Huiskonen, J.T., de Haas, F., Bubeck, D., Bamford, D.H., Fuller, S.D., Butcher, S.J. (2006) Structure of the bacteriophage phi6 nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* **14**: 1039-1048.

Iyer, L.M., Koonin, E.V., Leipe, D.D., and Aravind, L. (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.* **33**: 3875-3896.

Jamalludeen, N., Johnson, R.P., Friendship, R., Kropinski, A.M., Lingohr, E.J., and Gyles, C.L. (2007) Isolation and characterization of nine bacteriophages that lyse O149 enterotoxigenic *Escherichia coli*. *Vet Microbiol.* **124**:47-57.

Jamalludeen, N., Kropinski, A.M., Johnson, R.P., Lingohr, E., Harel, J., and Gyles, C.L. (2008) Complete genomic sequence of bacteriophage phiEcoM-GJ1, a novel phage that has myovirus morphology and a podovirus-like RNA polymerase. *Appl Environ Microbiol.* **74**: 516-525.

Jarrell, K., and Kropinski, A.M. (1977) Identification of the cell wall receptor for bacteriophage E79 in *Pseudomonas aeruginosa* strain PAO. *J Virol.* **23**: 461-466.

Jensen, M.A., Faruque, S.M., Mekalanos, J.J., and Levin, B.R. (2006) Modeling the role of bacteriophage in the control of cholera outbreaks. *PNAS USA* **103**: 4652-4657.

Jiang, S.C., and Paul, J.H. (1998) Gene transfer by transduction in the marine environment. *Appl Environ Microbiol.* **64**: 2780-2787.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W. (2000) Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol.* **299**: 27-51.

Kao, S.H., and McClain, W.H. (1980) Baseplate protein of bacteriophage T4 with both structural and lytic functions. *J Virol.* **34**: 95-103.

Karam, J.D., and Konigsberg, W.H. (2000) DNA polymerase of the T4-related bacteriophages. *Prog Nucleic Acid Res Mol Biol.* **64**: 65-96.

Kashlev, M., Nudler, E., Goldfarb, A., White, T., and Kutter, E. (1993) Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell* **75**: 147-154.

Katz, M.E., Howarth, P.M., Yong, W.K., Riffkin, G.G., Depiazzi, L.J., and Rood, J.I. (1991) Identification of three gene regions associated with virulence in *Dichelobacter nodosus*, the causative agent of ovine footrot, *J Gen Microbiol.* **137**: 2117-2124.

Kaufmann, R. (1995) Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: a novel analytical tool in molecular biology and biotechnology. *J Biotechnol*. **41**: 155-175.

Kaufmann, R., Kirsch, D., and Spengler, B. (1994) Sequencing of peptides in a time-of-flight mass spectrometer: evaluation of postsource decay following matrix-assisted laser desorption ionization (MALDI). *Int J Mass Spectrom and Ion Processes* **131**: 355.

Kawasaki, T., Shimizu, M., Satsuma, H., Fujiwara, A., Fujie, M., Usami, S., and Yamada, T. (2009) Genomic characterization of *Ralstonia solanacearum* phage phiRSB1, a T7-like wide-host-range phage. *J Bacteriol.* **191**: 422-427.

Kazmierczak, K.M., and Rothman-Denes, L.B. (2006) Bacteriophage N4. *In*: R. Calendar, Editor, *The Bacteriophages*, Oxford University Press, Inc., New York, pp. 302–314.

Kazmierczak, K.M., Davydova, E.K., Mustaev, A.A., and Rothman-Denes, L.B. (2002) The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *EMBO J.* **21**: 5815-5823.

Kelln, R.A., and Warren, R.A. (1971) Isolation and properties of a bacteriophage lytic for a wide range of pseudomonads. *Can J Microbiol*. **17**: 677-682.

Kelman, Z., and O'Donnell, M. (1995) DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem*. **64**: 171-200.

Keough, T., Youngquist, R.S., and Lacey M.P. (2003) Sulfonic acid derivatives for peptide sequencing by MALDI MS. *Anal Chem* **75**: 156-165.

Keough, T., Youngquist, R.S., and Lacey, M.P. (1999) A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. *Proc Natl Acad Sci USA* **96**: 7131-7136.

Kerr, K.G., and Snelling, A.M. (2009) *Pseudomonas aeruginosa*: a formidable and ever-present adversary. *J Hosp Infect.*, *in press*.

Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P., and Stackebrandt, E. (2003) Introduction to the *Proteobacteria*. In: M. Dworkin, Editor, *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*, Springer-Verlag, New York.

Koonin, E.V. (2005) Virology: Gulliver among the Lilliputians. *Curr Biol*. **15**: R167-169.

Koonin, E.V., Senkevich, T.G., Dolja, V.V. (2009) Compelling reasons why viruses are relevant for the origin of cells. *Nat Rev Microbiol* **7**: 615.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., Darst, S.A. (2000). A structural model of transcription elongation. *Science* **289**: 619-625.

Kovalyova, I. V., and Kropinski, A.M. (2003). The complete genomic sequence of lytic bacteriophage gh-1 infecting *Pseudomonas putida*-evidence for close relationship to the T7 group. *Virology* **311**: 305-315.

Krisch, H.M., and Comeau, A.M. (2008) The immense journey of bacteriophage T4--from d'Hérelle to Delbrück and then to Darwin and beyond. *Res Microbiol.* **159**: 314-324.

Krisch, H.M., Hamlett, N.V. and Berger, H. (1972) Polynucleotide ligase in bacteriophage T4D recombination. *Genetics* **72**: 187-203.

Kropinski, A.M. (2000) Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J Bacteriol.* **182**: 6066-6074.

Krupovic, M., and Bamford, D.H. (2008) Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat Rev Microbiol.* **6**: 941-948.

Krupovic, M., Cvirkaite-Krupovic, V., and Bamford, D.H. (2008) Identification and functional analysis of the Rz/Rz1-like accessory lysis genes in the membrane-containing bacteriophage PRD1. *Mol Microbiol.* **68**: 492-503.

Krylov, V.N., Tolmachova, T.O., and Akhverdian, V.Z. (1993) DNA homology in species of bacteriophages active on *Pseudomonas aeruginosa*. *Arch Virol.* **131**: 141-151.

Kulakov, L.A., Ksenzenko, V.N., Shlyapnikov, M.G., Kochetkov, V.V., Del Casale, A., Allen, C.C., Larkin, M.J., Ceyssens, P.J., and Lavigne, R. (2009) Genomes of "phiKMV-like viruses" of *Pseudomonas aeruginosa* contain localized single-strand interruptions. *Virology* **391**:1-4.

Kunin, V., He, S., Warnecke, F., Peterson, S.B., Garcia Martin, H., Haynes, M., Ivanova, N., Blackall, L.L., Breitbart, M., Rohwer, F., McMahon, K.D., Hugenholtz, P. (2008) A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res.* **18**: 293-297.

Kuzio, J., and Kropinski, A.M. (1983) O-antigen conversion in *Pseudomonas aeruginosa* PAO1 by bacteriophage D3. *J Bacteriol.* **155**: 203-212.

Kwan, T., Liu, J., Dubow, M., Gros, P. and Pelletier, J. (2006) Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J Bacteriol.* **188**: 1184-1187.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.M., Raoult, D. (2003) A giant virus in amoebae. *Science* **299**: 2033.

Labonté, J.M., Reid, K.E., and Suttle, C.A. (2009) Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. *Appl Environ Microbiol.* **75**: 3634-3640.

Labrie, S.J., Josephsen, J., Neve, H., Vogensen, F.K., and Moineau, S. (2008) Morphology, genome sequence, and structural proteome of type phage P335 from *Lactococcus lactis*. *Appl Environ Microbiol.* **74**: 4636-4644.

Lang, A.S., and Beatty, J.T. (2002) A bacterial signal transduction system controls genetic exchange and motility. *J Bacteriol*. **184**: 913-918.

Lang, A.S., and Beatty, J.T. (2007) Importance of widespread gene transfer agent genes in alpha-proteobacteria. *Trends Microbiol.* **15**: 54-62.

Lavigne, R., Briers, Y., Hertveldt, K., Robben, J., and Volckaert, G. (2004) Identification and characterization of a highly thermostable bacteriophage lysozyme. *Cell Mol Life Sci.* **61**: 2753-2759.

Lavigne, R., Burkal'tseva, M.V., Robben, J., Sykilinda, N.N., Kurochkina, L.P., Grymonprez, B., Jonckx, B., Krylov, V. N., Mesyanzhinov, V.V., and G. Volckaert (2003) The genome of bacteriophage phiKMV, a T7-like virus infecting *Pseudomonas aeruginosa*. *Virology* **312**: 49-59.

Lavigne, R., Noben, J.P., Hertveldt, K., Ceyssens, P.J., Briers, Y., Dumont, D., Roucourt, B., Krylov, V.N., Mesyanzhinov, V.V., Robben, J., and Volckaert, G. (2006). The structural proteome of *Pseudomonas aeruginosa* bacteriophage phiKMV. *Microbiology* **152**: 529-534.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W., and Kropinski, A.M. (2008) Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res Microbiol.* **159**: 406-414.

Lavigne, R., Sun, W.D., and Volckaert, G. (2004) PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics* **20**: 629-635.

Lawrence, J.G. (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol.* **2**: 519-523.

Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2002) Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol.* **184**: 4891-4905.

Lawrence, J.G., Hendrix, R.W., and Casjens, S. (2001) Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* **9**: 535-540.

Lee, C.N., Hu, R.M., Chow, T.Y., Lin, J.W., Chen, H.Y., Tseng, Y.H., and Weng, S.F. (2007) Comparison of genomes of three *Xanthomonas oryzae* bacteriophages. *BMC Genomics* **8**:442.

Lee, D.G., Urbach, J.M., Wu, G., Liberati, N.T., Feinbaum, R.L., Miyata, S., Diggins, L.T., He, J., Saucier, M., Déziel, E., et al. (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.* **7**: R90.

Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**: 419-429.

Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci.* **60**: 2356-2370.

León, I.R., Neves-Ferreira, A.G., Valente, R.H., Mota, E.M., Lenzi, H.L., and Perales, J. (2007) Improved protein identification efficiency by mass spectrometry using N-terminal chemical derivatization of peptides from *Angiostrongylus costaricensis*, a nematode with unknown genome. *J Mass Spectrom*. **42**: 781-792.

Leplae, R., Hebrant, A., Wodak, S.J., and Toussaint, A. (2004) ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res.* **32**: D45-49.

Lewis, K. (2007) Persister cells, dormancy and infectious disease. *Nat Rev Microbiol.* **5**: 48-56.

Lima-Mendez, G., Toussaint, A., and Leplae, R. (2007) Analysis of the phage sequence space: the benefit of structured information. *Virology* **365**: 241-249.

Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008) Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol Biol Evol*. **25**: 762-777.

Lindberg, R.B., and Latta, R.L. (1974) Phage typing of *Pseudomonas aeruginosa*: clinical and epidemiologic considerations. *J Infect Dis.* **130**: S33-42.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., Hess, W.R., Church, G.M., Chisholm, S.W. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83-86.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86-89.

Liu, J., Glazko, G., and Mushegian, A. (2006) Protein repertoire of double-stranded DNA bacteriophages. *Virus Res.* **117**: 68-80.

Liu, J., M. Dehbi, G. Moeck, F. Arhin, P. Bauda, D. Bergeron, M. Callejo, V. Ferretti, N. Ha, T. Kwan, J. McCarty, R. Srikumar, D. Williams, J. J. Wu, P. Gros, J. Pelletier, and DuBow, M. (2004) Antimicrobial drug discovery through bacteriophage genomics. *Nat Biotechnol.* **22**: 185-191.

Liu, M., Gingery, M., Doulatov, S.R., Liu, Y., Hodes, A., Baker, S., Davis, P., Simmonds, M., Churcher, C., Mungall, K., Quail, M.A., Preston, A., Harvill, E.T., Maskell, D.J., Eiserling, F.A., Parkhill, J., and Miller, J.F. (2004) Genomic and genetic analysis of *Bordetella* bacteriophages encoding reverse transcriptase-mediated tropism-switching cassettes. *J Bacteriol.* **186**: 1503-1517.

Liu, X., Kong, S., Shi, M., Fu, L., Gao, Y., and An, C. (2008) Genomic analysis of freshwater cyanophage Pf-WMP3 Infecting cyanobacterium *Phormidium foveolarum*: the conserved elements for a phage. *Microb Ecol*. **56**: 671-680.

Liu, X., Shi, M., Kong, S., Gao, Y., and An, C. (2007) Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium *Phormidium foveolarum*: complete genome sequence and DNA translocation. *Virology* **366**: 28-39.

Livermore, D.M. (2002) Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? *Clin Infect Dis.* **34**: 634-640.

Łobocka, M.B., Rose, D.J., Plunkett, G. 3rd., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M.B., Blattner, F.R. (2004) Genome of bacteriophage P1. *J Bacteriol.* **186**: 7032-7068.

Loeffler, J. M., Nelson, D., and Fischetti, V.A. (2001) Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science* **294**: 2170-2172.

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.

Lohr, J.E., Chen, F., and Hill, R.T. (2005) Genomic analysis of bacteriophage PhiJL001: insights into its interaction with a sponge-associated alpha-proteobacterium. *Appl Environ Microbiol*. **71**: 1598-1609.

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.

Ludmir, E.B., Enquist, L.W. (2009) Viral genomes are part of the phylogenetic tree of life. *Nat Rev Microbiol*. **7**: 615.

Lukashin, A.V., and Borodovsky, M. (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**: 1107-1115.

Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**: 1162-1164.

Lyczak, J.B., Cannon, C.L., and Pier, G.B. (2000) Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect.* **2**: 1051-1060.

Mahony, J., Deveau, H., Mc Grath, S., Ventura, M., Canchaya, C., Moineau, S., Fitzgerald, G.F., and van Sinderen, D. (2006) Sequence and comparative genomic analysis of lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage species. *FEMS Microbiol Lett.* **261**: 253-261.

Mann, N.H. (2005) The third age of phage. *PLoS Biol*. **3**: e182.

Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., et al. (2005) The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine *Synechococcus*. *J Bacteriol*. **187**: 3188-3200.

Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., and Krisch, H.M., (2005). The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. *J Bacteriol.* **187**: 3188-3200.

Marekov, L.N., and Steinert, P.M. (2003) Charge derivatization by 4-sulfophenyl isothiocyanate enhances peptide sequencing by post-source decay matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Mass Spectrom.* **38**: 373-377.

Marinelli, L.J., Piuri, M., Swigonová, Z., Balachandran, A., Oldfield, L.M., van Kessel, J.C., and Hatfull, G.F. (2008) BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. *PLoS One* **3**: e3957.

Marrs, B. (1974) Genetic recombination in *Rhodopseudomonas capsulata*. *PNAS USA* **71**: 971-973.

Marston, M.F., and Sallee, J.L. (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol*. **69**: 4639-4647.

Marza, J.A., Soothill, J.S., Boydell, P., and Collyns, T.A. (2006) Multiplication of therapeutically administered bacteriophages in *Pseudomonas aeruginosa* infected patients. *Burns* **32**: 644-646.

Mathee, K., Narasimhan, G., Valdes, C., Qiu, X., Matewish, J.M., Koehrsen, M., et al. (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *PNAS USA* **105**: 3100-3105.

Matson, E.G., Thompson, M.G., Humphrey, S.B., Zuerner, R.L., and Stanton, T.B. (2005) Identification of genes of VSH-1, a prophage-like gene transfer agent of *Brachyspira hyodysenteriae. J Bacteriol.* **187**: 5885-5892.

Matsuzaki, S., Yasuda, M., Nishikawa, H., Kuroda, M., Ujihara, T., Shuin, T., Shen, Y., Jin, Z., Fujimoto, S., Nasimuzzaman, M.D., et al. (2003) Experimental protection of mice against lethal *Staphylococcus aureus* infection by novel bacteriophage phi MR11. *J Infect Dis.* **187**: 613-624.

McDaniel, L., Breitbart, M., Mobberley, J., Long, A., Haynes, M., Rohwer, F., and Paul, J.H. (2008) Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. *PLoS One* **3**: e3263.

McPartland, J., and Rothman-Denes, L.B. (2009) The tail sheath of bacteriophage N4 interacts with the Escherichia coli receptor. *J Bacteriol.* **191**: 525-532.

McShan, W.M., and Ferretti, J.J. (1997). Genetic diversity in temperate bacteriophages of *Streptococcus pyogenes*: identification of a second attachment site for phages carrying the erythrogenic toxin A gene. *J Bacteriol.* **179**: 6509-6511.

Mena, A., Smith, E.E., Burns, J.L., Speert, D.P., Moskowitz, S.M., Perez, J.L., and Oliver, A. (2008) Genetic adaptation of *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients is catalyzed by hypermutation. *J Bacteriol.* **190**: 7910-7917.

Méndez-Ortiz, M.M., Hyodo, M., Hayakawa, Y., Membrillo-Hernández, J. (2006) Genome-wide transcriptional profile of *Escherichia coli i*n response to high levels of the second messenger 3',5'-cyclic diguanylic acid. *J Biol Chem.* **281**: 8090-8099.

Merabishvili, M., Pirnay, J.P., Verbeken, G., Chanishvili, N., Tediashvili, M., Lashkhi, N., Glonti, T., Krylov, V.N., Mast, J., Van Parys, L., Lavigne, R., Volckaert, G., Mattheus, W., Verween, G., De Corte, P., Rose, T., Jennes, S., Zizi, M., De Vos, D., and Vaneechoutte, M. (2009) Quality-controlled small-scale production of a well-defined bacteriophage cocktail for use in human clinical trials. *PLoS ONE* **4**: e4944.

Merabishvili, M., Verhelst, R., Glonti, T., Chanishvili, N., Krylov, V., Cuvelier, C., Tediashvili, M., and Vaneechoutte, M. (2007) Digitized fluorescent RFLP analysis (fRFLP) as a universal method for comparing genomes of culturable dsDNA viruses: application to bacteriophages. *Res Microbiol.* **158**: 572-581.

Mesyanzhinov, V.V. (2004) Bacteriophage T4: structure, assembly, and initiation infection studied in three dimensions. *Adv Virus Res.* **63**: 287-352.

Mesyanzhinov, V.V., Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M.V., Sykilinda, N.N., Krylov, V.N., and Volckaert, G. (2002) The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *J Mol Biol.* **317**: 1-19.

Michalewicz, J., and Nicholson, A.W. (1992) Molecular cloning and expression of the bacteriophage T7 0.7(protein kinase) gene. *Virology* **186**: 452-462.

Miller, A., Wood, D., Ebright, R.H., and Rothman-Denes, L.B. (1997) RNA polymerase beta' subunit: a target of DNA binding-independent activation. *Science* **275**: 1655-1657.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Rüger, W. (2003) Bacteriophage T4 genome. *Microbiol Mol Biol Rev.* **67**: 86-156.

Miller, J.H. (1992) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 431-435.

Mirold, S., Rabsch, W., Rohde, M., Stender, S., Tschäpe, H., Rüssmann, H., Igwe, E., and Hardt, W.D. (1999) Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. *PNAS USA* **17**: 9845-9850.

Moak, M., and Molineux, I.J. (2004) Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol Microbiol.* **51**: 1169-1183.

Molineux, I.J. (2001) No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. *Mol Microbiol*. **40**: 1-8.

Möller, S., Croning, M.D., and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**: 646-653.

Montag, D., Riede, I., Eschbach, M.L., Degen, M., and Henning, U. (1987) Receptor-recognizing proteins of T-even type bacteriophages. Constant and hypervariable regions and an unusual case of evolution. *J Mol Biol.* **196**: 165-174.

Moran, N.A. (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **108**: 583-586.

Moreira, D., López-García, P. (2009) Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol.* **7**: 306-311.

Morett, E. and Segovia, L. (1993) The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. *J Bacteriol.* **175**: 6067-6074.

Morgan, A.D., Gandon, S., and Buckling, A. (2005) The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* **437**: 253-256.

Morozova, O., and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**: 255-264.

Mosig, G. (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu Rev Genet.* **32**: 379-413.

Mosig, G., Gewin, J., Luder, A., Colowick, N., and Vo, D.(2001) Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *PNAS USA* **98**: 8306-8311.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y, and Hayashi, T. (1999) The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Mol Microbiol.* **31**: 399-419.

Nascimento, J.G., Guerreiro-Pereira, M.C., Costa, S.F., São-José, C., and Santos, M.A. (2008) Nisin-triggered activity of Lys44, the secreted endolysin from *Oenococcus oeni* phage fOg44. *J Bacteriol.* **190**: 457-461.

Nechaev, S., and Severinov, K. (1999) Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. *J Mol Biol.* **289**: 815-826.

Nechaev, S., and Severinov, K. (2003) Bacteriophage-induced modifications of host RNA polymerase. *Annu Rev Microbiol.* **57**: 301-322.

Nelson, K.E., Weinel, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A., Fouts, D.E., Gill, S.R., Pop, M., et al. (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol.* **4**: 799-808.

Newton, G.J., Daniels, C., Burrows, L.L., Kropinski, A.M., Clarke, A.J., and Lam, J.S. (2001). Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. *Mol Microbiol.* **39**: 1237-1247.

Nolan, J.M., Petrov, V., Bertrand, C., Krisch, H.M., and Karam, J.D. (2006) Genetic diversity among five T4-like bacteriophages. *Virol J.* **3**: 30.

Nuismer, S.L. (2006) Parasite local adaptation in a geographic mosaic. *Evolution* **60**: 24-30.

Ohmori, H., Haynes, L.L., and Rothman-Denes, L.B. (1988) Structure of the ends of the coliphage N4 genome. *J Mol Biol.* **202**: 1-10.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**: 5691-5702.

Pajunen, M. I., Elizondo, M.R., Skurnik, M., Kieleczawa, J., and Molineux, I.J. (2002) Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3. *J Mol Biol. 319*: 1115-1132.

Pal, C., Maciá, M.D., Oliver, A., Schachar, I., Buckling, A. (2007) Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**: 1079-1081.

Paul, J.H., Sullivan, M.B., Segall, A.M., Rohwer, F.(2002) Marine phage genomics. *Comp Biochem Physiol B Biochem Mol Biol.* **133**: 463-476.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., et al. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**: 171-182.

Penner, M., Morad, I., Snyder, L., and Kaufmann, G. (1995) Phage T4-coded Stp: double-edged effector of coupled DNA and tRNA-restriction systems. *J Mol Biol.* **249**: 857-868.

Pirnay, J.P., De Vos, D., Cochez, C., Bilocq, F., Pirson, J., Struelens, M., et al. (2003) Molecular epidemiology of *Pseudomonas aeruginosa* colonization in a burn unit: persistence of a multidrug-resistant clone and a silver sulfadiazine-resistant clone. *J Clin Microbiol.* **41**: 1192-1202.

Pirnay, J.P., De Vos, D., Cochez, C., Bilocq, F., Vanderkelen, A., Zizi, M., Ghysels, B., and Cornelis, P. (2002) *Pseudomonas aeruginosa* displays an epidemic population structure. *Environm Microbiol.* **4**: 898-911.

Pirnay, J.P., Matthijs, S., Colak, H., Chablain, P., Bilocq, F., Van Eldere, J., et al. (2005) Global *Pseudomonas aeruginosa* biodiversity as reflected in a Belgian river. *Environ Microbiol.* **7**: 969-980.

Piuri, M., and Hatfull, G.F. (2006) A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Mol Microbiol.* **62**: 1569-1585.

Pleteneva, E.A., Shaburova, O.V., Sykilinda, N.N., Miroshnikov, K.A., Krylov, S.V., Mesianzhinov, V.V., and Krylov, V.N. (2008) Study of the diversity in a group of phages of *Pseudomonas aeruginosa* species PB1 (Myoviridae) and their behavior in adsorption-resistant bacterial mutants. *Genetika* **44**: 185-194.

Postic, B., and Finland, M. (1961) Observations on bacteriophage typing of *Pseudomonas aeruginosa*. *J Clin Invest.* **40**: 2064-2075.

Potvin, E., Sanschagrin, F., and Levesque, R.C. (2008) Sigma factors in *Pseudomonas aeruginosa. FEMS Microbiol Rev.* **32**: 38-55.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., *et al.* (2002) The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. *J Bacteriol*. **184**: 6026-6036.

Pruitt, B.A. Jr., McManus, A.T., Kim, S.H., and Goodwin, C.W. (1998) Burn wound infections: current status. *World J Surg.* **22**: 135-145.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Séraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods.* **24**: 218-229.

Ramanculov, E., and Young, R. (2001) An ancient player unmasked: T4 rI encodes a t-specific antiholin. *Mol Microbiol.* **41**: 575-583.

Ranea, J.A., Sillero, A., Thornton, J.M., Orengo, C.A. (2006) Protein superfamily evolution and the last universal common ancestor (LUCA). *J Mol Evol.* **63**: 513-525.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306**: 1344-1350.

Raoult, D., Forterre, P. (2008) Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol.* **6**: 315-319.

Reese, M.G. (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput Chem.* **26**: 51-56.

Rhoades, M. (1984) A bacteriophage T5 mutants with an increased frequency of single-chain interruptions. *J Virol*. **51:** 533-555.

Rifat, D., Wright, N.T., Varney, K.M., Weber, D.J., and Black, L.W. (2008) Restriction endonuclease inhibitor IPI of bacteriophage T4: a novel structure for a dedicated target. *J Mol Biol.* **375**: 720-734.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol.* **17**: 1030-1032.

Ritz, D., and Beckwith, J. (2001) Roles of thiol-redox pathways in bacteria. *Annu Rev Microbiol.* **55**:21-48.

Rogers, S.G., Godwin, E.A., Shinosky, E.S., and Rhoades, M. (1979). Interruption-deficient mutants of bacteriophage T5. I. Isolation and general properties. *J Virol*. **29**: 716-725.

Rohwer F, Edwards R. (2002) The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol*. **184**: 4529-4535.

Rohwer, F. (2003) Global phage diversity. *Cell* **113**: 141.

Rohwer, F., and Thurber, R.V. (2009) Viruses manipulate the marine environment. *Nature* **459**: 207-212.

Roncero, C., Darzins, A., and Casadaban, M.J. (1990) *Pseudomonas aeruginosa* transposable bacteriophages D3112 and B3 require pili and surface growth for adsorption. *J Bacteriol* **172**: 1899-1904.

Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., and Breitbart, M. (2009) Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol.*, *in press.*

Roucourt, B. and Lavigne, R. (2009) The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environ Microbiol*., *in press.*

Roucourt, B., Lecoutere, E., Chibeu, A., Hertveldt, K., Volckaert, G., and Lavigne, R. (2009) A procedure for systematic identification of bacteriophage-host interactions of *P. aeruginosa* phages. *Virology* **387**: 50-58.

Rydman, P.S., and Bamford, D.H. (2000) Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Mol Microbiol.* **37**: 356-363.

Sambrook, J., and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual, 2nd ed., *In* Molecular cloning: a laboratory manual, 3rd ed., vol. 1. CSHL Press, New York, pp.1.32-1.34.

Samyn, B., Sergeant, K., Memmi, S., Debyser, G., Devreese, B., and Van Beeumen, J. (2006) MALDI-TOF/TOF de novo sequence analysis of 2-D PAGE-separated proteins from *Halorhodospira halophila*, a bacterium with unsequenced genome. *Electrophoresis* **27**: 2702-2711.

Sandegren, L. and Sjöberg, B.M. (2007). Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. *J Bacteriol.* **189**: 980-990.

Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchison, C.A. 3rd, Slocombe, P.M., and Smith, M. (1978) The nucleotide sequence of bacteriophage phiX174. *J Mol Biol.* **125**: 225-246.

Sano, E., Carlson, S., Wegley, L., and Rohwer, F. (2004) Movement of viruses between biomes. *Appl Environ Microbiol.* **70**: 5842-5846.

Savalia, D., Westblade, L.F., Goel, M., Florens, L., Kemp, P., Akulenko, N., Pavlova, O., Padovan, J.C., Chait, B.T., Washburn, M.P., Ackermann, H.W., Mushegian, A., Gabisonia, T., Molineux, I., and Severinov, K. (2008) Genomic and proteomic analysis of phiEco32, a novel *Escherichia coli* bacteriophage. *J Mol Biol.* **377**: 774-789.

Schaefer, A.L., Taylor, T.A., Beatty, J.T., and Greenberg, E.P. (2002) Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J Bacteriol.* **184**: 6515-6521.

Schellenberger, U., Livi, L.L., and Santi, D.V. (1995) Cloning, expression, purification, and characterization of 2'-deoxyuridylate hydroxymethylase from phage SPO1. *Protein Expr Purif.* **6**: 423-430.

Schito, G.C. (1974) Development of coliphage N4: ultrastructural studies. *J Virol.* **13**: 186-196.

Schöbel, S., Zellmeier, S., Schumann, W., and Wiegert, T. (2004) The Bacillus subtilis sigmaW anti-sigma factor RsiW is degraded by intramembrane proteolysis through YluC. *Mol Microbiol.* **52**: 1091-1105.

Scholl, D., and Merril, C.R. (2005). The genome of bacteriophage K1F, a T7-like phage that has acquired the ability to replicate on K1 strains of *Escherichia coli*. *J Bacteriol*. **187**: 8499-8503.

Scholl, D., Kieleczawa, J., Kemp, P., Rush, J., Richardson, C.C., Merril, C.R., Adhya, S., and Molineux, I.J. (2004). Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup. *J Mol Biol.* **335**:1151-1171.

Schuch, R., Nelson, D., and Fischetti, V.A. (2002) A bacteriolytic agent that detects and kills *Bacillus anthracis. Nature* **418**: 884-889.

Schweizer, H.P. and Chuanchuen, R. (2001). Small broad-host-range lacZ operon fusion vector with low background activity. *Biotechniques* **31**: 1258-1260.

Schwudke, D., Ergin, A., Michael, K., Volkmar, S., Appel, B., Knabner, D., *et al.* (2008) Broad-host-range *Yersinia* phage PY100: genome sequence, proteome analysis of virions, and DNA packaging strategy. *J Bacteriol*. **190**: 332-342.

Seguritan, V., Feng, I.W., Rohwer, F., Swift, M., and Segall, A.M. (2003) Genome sequences of two closely related *Vibrio parahaemolyticus* phages, VP16T and VP16C. *J Bacteriol.* **185**: 6434-6447.

Serwer, P., Hayes, S.J., Zaman, S., Lieman, K., Rolando, M., and Hardies, S.C. (2004) Improved isolation of undersampled bacteriophages: finding of distant terminase genes. *Virology* **329**: 412-424.

Serwer, P., Wright, E.T., Hakala, K.W., and Weintraub, S.T. (2008) Evidence for bacteriophage T7 tail extension during DNA injection. *BMC Res Notes* **26**: 36.

Seymour, J.R., Seuront, L., Doubell, M., Waters, R.L., Mitchell, J.G. (2006) Microscale patchiness of virioplankton. *J Mar Biol Assoc.* **86**: 551-561.

Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W., and Standing, K.G. (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem.* **73**: 1917-1926.

Short, C.M., and Suttle, C.A. (2005) Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. *Appl Environ Microbiol.* **71**: 480-486.

Siew, N., and Fischer, D. (2003) Analysis of singleton ORFans in fully sequenced microbial genomes. *Proteins* **53**: 241-251.

Simon, M.N., Davis, R.W., and Davidson, N. (1971) Heteroduplexes of DNA molecules of lambdoid phages: physical mapping of their base sequence relationships by electron microscopy. In A. D. Hershey (ed.), *The bacteriophage lambda.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New Tork, pp. 313-328.

Singer, B.S., Gold, L., Gauss, P., and Doherty, D.H. (1982) Determination of the amount of homology required for recombination in bacteriophage T4. *Cell* **31**: 25–33.

Smith, D. E., Tans, S. J. , Smith, S. B., Grimes, S., Anderson, D. L., and Bustamante, C. (2001) The bacteriophage ϕ29 portal motor can package DNA against a large internal force. *Nature* **413**: 748-752.

Smith, E.E., Buckley, D.G., Wu, Z., Saenphimmachak, C., Hoffman, L.R., D'Argenio, D.A., et al. (2006) Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *PNAS USA* **103**: 8487-8492.

Song, H.K., Sohn, S.H., and Suh, S.W. (1999) Crystal structure of deoxycytidylate hydroxymethylase from bacteriophage T4, a component of the deoxyribonucleoside triphosphate-synthesizing complex. *EMBO J.* **18**: 1104-1113.

Soothill, J., Hawkins, C., Anggård, E., and Harper, D. (2004) Therapeutic use of bacteriophages. *Lancet Infect Dis.* **4**: 544-545.

Souli, M., Galani, I., and Giamarellou, H. (2008) Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. *Euro Surveill.* **13**: 19045.

Sousa, R. (2008) Tie me up, tie me down: inhibiting RNA polymerase. *Cell* **135**: 205-207.

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K.E. (2008) Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res Microbiol.* **159**: 349-357.

Staley, J.T., and Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol.* **39**: 321-346.

Stanton, T.B., Humphrey, S.B., Bayles, D.O., and Zuerner, R.L. (2009) Identification of a divided genome for VSH-1, the prophage-like gene transfer agent of *Brachyspira hyodysenteriae*. *J Bacteriol.* **191**: 1719-1721.

Stenholm, A.R., Dalsgaard, I., and Middelboe, M. (2008) Isolation and characterization of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*. *Appl Environ Microbiol*. **74**: 4070-4078.

Stojković, E.A., and Rothman-Denes, L.B. (2007) Coliphage N4 N-acetylmuramidase defines a new family of murein hydrolases. *J Mol Biol.* **366**: 406-419.

Stone, R. (2002) Bacteriophage therapy. Stalin's forgotten cure. *Science* **298**: 728-731.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., et al. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959-964.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., et al. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959-964.

Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., Defrancesco, A.S., Tan, G., Fu, R., Lee, J.A., Waterbury, J.B., Bielawski, J.P., and Chisholm, S.W. (2008) Portal protein diversity and phage ecology. *Environ Microbiol.* **10**: 2810-2823.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., and Chisholm, S.W. (2005) Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3**: e144.

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol*. **4**: e234.

Summer, E.J., Gill, J.J., Upton, C., Gonzalez, C.F., and Young, R. (2007) Role of phages in the pathogenesis of Burkholderia, or 'Where are the toxin genes in Burkholderia phages?'. *Curr Opin Microbiol*. **10**: 410-417.

Summer, E.J., Gonzalez, C.F., Bomer, M., Carlile, T., Embry, A., Kucherka, A.M., Lee, J., Mebane, L., Morrison, W.C., Mark, L., King, M.D., LiPuma, J.J., Vidaver, A.K., and Young, R. (2006) Divergence and mosaicism among virulent soil phages of the *Burkholderia cepacia* complex. *J Bacteriol.* **188**: 255-268.

Summers, W.C. (1999) Bacteriophage discovered. In: *Felix d'Herelle and the Origins of Molecular Biology*, Yale University Press, New Haven, CT, pp. 47-59.

Susskind, M.M., and Botstein, D. (1978). Molecular genetics of bacteriophage P22. *Microbiol Rev.* **42:** 385-413.

Suttle, C. A., and Chan, A. M. (1994) Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl Environ Microbiol.* **60**: 3167-3174.

Suttle, C.A. (2005) Viruses in the sea. *Nature* **437**: 356-361.

Suttle, C.A. (2007) Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol.* **5**: 801-812.

Tan, Y., Zhang, K., Rao, X., Jin, X., Huang, J., Zhu, J., Chen, Z., Hu, X., Shen, X., Wang, L., and Hu, F., (2007). Whole genome sequencing of a novel temperate bacteriophage of *P. aeruginosa*: evidence of tRNA gene mediating integration of the phage genome into the host bacterial chromosome. *Cell Microbiol*. **9**: 479-491.

Tang, G.Q., Bandwar, R.P., and Patel, S.S. (2005). Extended upstream A-T sequence increases T7 promoter strength. *J Biol Chem*. **280**: 40707-40713.

Tanji, Y., Shimada, T., Yoichi, M., Miyanaga, K., Hori, K., and Unno, H. (2004) Toward rational control of Escherichia coli O157:H7 by a phage cocktail. *Appl Microbiol Biotechnol.* **64**: 270-274.

Tannu, N.S., and Hemby, S.E. (2007) De novo protein sequence analysis of *Macaca mulatta*. *BMC Genomics* **8**: 270.

Tenover, F.C. (2006) Mechanisms of antimicrobial resistance in bacteria. *Am J Infect Control.* **34**: S3-10.

Tétart, F. Repoila, F., Monod, C., and Krisch, H.M. (1996) Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesion. *J Mol Biol.* **258**: 726-731.

Tétart, F., Desplats, F., and Krisch, H.M. (1998) Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: Recombination between conserved motifs swaps adhesin specificity. *J Mol Biol.* **282**: 543–556.

Thiel, K. (2004) Old dogma, new tricks--21st Century phage therapy. *Nat Biotechnol.* **22**: 31-36.

Thingstad, T. F., Lignell, R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**: 19–27.

Thomas, J.A., Rolando, M.R., Carroll, C.A., Shen, P.S., Belnap, D.M., Weintraub, S.T., Serwer, P., and Hardies, S.C. (2008) Characterization of *Pseudomonas chlororaphis* myovirus 201varphi2-1 via genomic sequencing, mass spectrometry, and electron microscopy. *Virology* **376**: 330-338.

Torrella F, Morita R.Y. (1979) Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of Yaquina Bay, oregon: ecological and taxonomical implications. *Appl Environ Microbiol*. **37**: 774-778.

Tramper-Stranders, G.A., van der Ent, C.K., Wolfs, T.F., Kimpen, J.L., Fleer, A., Johansen, U., Johansen, H.K., and Høiby, N. (2008) *Pseudomonas aeruginosa* diversity in distinct paediatric patient groups. *Clin Microbiol Infect.* **14**: 935-941.

Trautmann, M., Bauer, C., Schumann, C., Hahn, P., Höher, M., Haller, M., and Lepper, P.M. (2006) Common RAPD pattern of *Pseudomonas aeruginosa* from patients and tap water in a medical intensive care unit. *Int J Hyg Environ Health* **209**: 325-331.

Twort, F.W. (1915) An investigation on the nature of ultra-microscopic viruses. *Lancet* **186**: 1241–1243.

Uchiyama, J., Rashel, M., Matsumoto, T., Sumiyama, Y., Wakiguchi, H., and Matsuzaki, S. (2009) Characteristics of a novel *Pseudomonas aeruginosa* bacteriophage, PAJU2, which is genetically related to bacteriophage D3. *Virus Res.* **139**: 131-134.

Vallée, M., and Cornett, J.B. (1972) A new gene of bacteriophage T4 determining immunity against superinfecting ghosts and phage in T4-infected Escherichia coli. *Virology* **48**: 777-784.

Valpuesta, J.M., Sousa, N., Barthelemy, I., Fernández, J.J., Fujisawa, H., Ibarra, B., and Carrascosa, J.L. (2000) Structural analysis of the bacteriophage T3 head-to-tail connector. *J Struct Biol*. **131**: 146-155.

Verbeken, G., De Vos, D., Vaneechoutte, M., Merabishvili, M., Zizi, M., and Pirnay, J.P. (2007) European regulatory conundrum of phage therapy. *Future Microbiol.* **2**: 485-491.

Vitiello, C.L., Merril, C.R., and Adhya, S. (2005) An amino acid substitution in a capsid protein enhances phage survival in mouse circulatory system more than a 1000-fold. *Virus Res.* **114**: 101-103.

Vos, M., Birkett, P.J., Birch, E., Griffiths, R.I., and Buckling, A. (2009) Local adaptation of bacteriophages to their bacterial hosts in soil. *Science* **325**: 833.

Waldor, M.K., Mekalanos, J.J. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**: 1910-1914.

Wang, H., Grimes, S., Anderson, D.L., and Serwer, P. (2004) Terminal protein-induced stretching of bacteriophage phi29 DNA. *J Microsc.* **213**: 172-179.

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

Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Wang, J., Bao, Q., Kong, H., and Hu, S. (2005) Complete genome sequence of bacteriophage T5. *Virology* **332**: 45-65.

Wang, P.W., Chu, L., and Guttman, D.S. (2004) Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. *J Bacteriol.* **186**: 400-410.

Weigel, C., and Seitz, H. (2006) Bacteriophage replication modules. *FEMS Microbiol Rev.* **30**: 321-381.

Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., King, J., Hatfull, G.F., Lawrence, J.G., and Hendrix, R.W. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting marine Prochlorococcus and Synechococcus. *Environ Microbiol*. **9**: 1675-1695.

Wilhelm, K., and Rüger, W. (1992) Deoxyuridylate-hydroxymethylase of bacteriophage SPO1. *Virology* **189**: 640-646.

Wilhelm, S. W. , Weinbauer, M. G. , Suttle, C. A. , Pledger, R. J., and Mitchell, D. L. (1998)Measurements of DNA damage and photoreactivation imply that most viruses in marine surface waters are infective. *Aquat Microb Ecol.* **14**: 215–222.

Wilhelm, S.W., Carberry, M.J., Eldridge, M.L., Poorvin, L., Saxton, M.A., and Doblin, M.A. (2006) Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of g20 genes. *Appl Environ Microbiol.* **72**: 4957-4963.

Williamson, K.E., Radosevich, M., Smith, D.W., and Wommack, K.E. (2007) Incidence of lysogeny within temperate and extreme soil environments. *Environ Microbiol.* 9: 2563–2574.

Williamson, K.E., Wommack, K.E., Radosevich, M. (2003) Sampling natural viral communities from soil for culture-independent analyses. *Appl Environ Microbiol* **69**: 6628-6633.

Williamson, S.J., McLaughlin, M.R., and Paul, J.H. (2001) Interaction of the PhiHSIC virus with its host: lysogeny or pseudolysogeny? *Appl Environ Microbiol.* **67**: 1682-1688.

Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., Andrews-Pfannkoch, C., Fadrosh, D., Miller, C.S., Sutton, G., Frazier, M., and Venter, J.C. (2008) The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* **3**: e1456.

Woese, C.R. (2002) On the evolution of cells. *PNAS USA* **99**: 8742-8747.

Wolfgang, M.C., Kulasekara, B.R., Liang, X., Boyd, D., Wu, K., Yang, Q., et al. (2003) Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *PNAS USA* **100**: 8484-8489.

Wommack, K.E., and Colwell, R. (2000) Virioplankton: Viruses in aquatic ecosystems. *Microbiol Mol Biol Rev.* **64**: 69-114.

Wommack, K.E., Ravel, J., Hill, R.T., Chun, J., and Colwell, R.R. (1999) Population dynamics of chesapeake bay virioplankton: total-community analysis by pulsed-field gel electrophoresis. *Appl Environ Microbiol* **65**: 231-240.

Wösten, M.M. (1998) Eubacterial sigma-factors. *FEMS Microbiol Rev.* **22**: 127-150.

Wright, A., Hawkins, C.H., Anggård, E.E., and Harper, D.R. (2009) A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clin Otolaryngol.* **34**: 349-357.

Wu, H., Sampson, L., Parr, R., and Casjens, S. (2002). The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. *Mol Microbiol.* **45**: 1631-1646.

Xu, M., Arulandu, A., Struck, D.K., Swanson, S., Sacchettini, J.C., and Young, R. (2005) Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. *Science* **307**: 113-117.

Xu, M., Struck, D.K., Deaton, J., Wang, I.N., and Young, R. (2004) A signal-release sequence export and control of the phage P1 endolysin. *PNAS* **101**: 6415-6420.

Yin, Y., and Fischer, D. (2008) Identification and investigation of ORFans in the viral world. *BMC Genomics* **9**: 24.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., et al. (2007) The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**: e16.

Young, R., Wang, I.N., and Roof, W.D. (2000) Phages will out: strategies of host cell lysis. *Trends Microbiol.* **8**: 120-126.

Zegans, M.E., Wagner, J.C., Cady, K.C., Murphy, D.M., Hammond, J.H., and O'Toole, G.A. (2009) Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa. J Bacteriol.* **191**: 210-219.

Zhao, Y., Wang, K., Jiao, N., and Chen, F. (2009) Genome sequences of two novel phages infecting marine roseobacters. *Environ Microbiol.* **11**: 2055-2064.

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.

Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science* **244**: 48-52.