Evolutionary Genomics of Prokaryotic Viruses

MART KRUPOVIČ

Institute of Biotechnology and
Department of Biosciences
Division of General Microbiology
Faculty of Biosciences
and
Viikki Graduate School in Molecular Biosciences
University of Helsinki

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biosciences, University of Helsinki, in the auditorium 3 of Info center Korona, Viikinkaari 11, Helsinki, on June 18th, 2010, at 12 noon

HELSINKI 2010

Supervisor

Professor Dennis H. Bamford
Department of Biosciences
University of Helsinki, Finland

Reviewers

Docent Alexander Plyusnin
Department of Virology
University of Helsinki, Finland

Doctor David Prangishvili
Department of Microbiology
Pasteur Institute, France

Opponent

Professor Roger W. Hendrix
Department of Biological Sciences
University of Pittsburgh, USA

ISBN 978-952-10-6282-7 (paperback)
ISBN 978-952-10-6283-4 (PDF)

ISSN 1795-7079

Yliopistopaino, Helsinki University Printing House
Helsinki 2010

ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I. Krupovič M, Bamford DH. 2007. Putative prophages related to lytic tailless marine dsDNA phage PM2 are widespread in the genomes of aquatic bacteria. *BMC Genomics*. 8:236.

II. Krupovič M, Bamford DH. 2008. Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum *Euryarchaeota*. *Virology*. 375:292-300.

III. Krupovič M, Forterre P, Bamford DH. 2010. Comparative analysis of the mosaic genomes of tailed archaeal viruses and proviruses suggests common themes for virion architecture and assembly with tailed viruses of bacteria. *J Mol Biol*. 397:144-60.

The publications are referred to in the text by their roman numerals.

SUMMARY

Evolutionary history of biological entities is recorded within their nucleic acid sequences and can (sometimes) be deciphered by thorough genomic analysis. In this study we sought to gain insights into the diversity and evolution of bacterial and archaeal viruses. Our primary interest was pointed towards those virus groups/families for which comprehensive genomic analysis was not previously possible due to the lack of sufficient amount of genomic data. During the course of this work twenty-five putative proviruses integrated into various prokaryotic genomes were identified, enabling us to undertake a comparative genomics approach. This analysis allowed us to test the previously formulated evolutionary hypotheses and also provided valuable information on the molecular mechanisms behind the genome evolution of the studied virus groups.

TABLE OF CONTENTS

A. LITERATURE REVIEW

1. Introduction

Viruses (from the Latin *virus* meaning *toxic* or *poison*) are relatively small biological entities that pursue a parasitic life-style. All viruses described thus far obligatory rely on host cells for reproduction. Outside the cell viruses are more or less inert particles (called virions). However, upon encountering a susceptible cell and delivering the genome into the cell interior the inert virion transforms into an actively replicating entity, which follows the genetically determined reproduction program. Viruses infect hosts from all three domains of cellular life (*Bacteria*, *Archaea* and *Eukarya*) and outnumber the cellular organisms by at least one order of magnitude (Bergh et al., 1989; Suttle, 2007). The total number of virus particles in the biosphere is astonishingly high ($> 10^{31}$). As a result, viruses present a major factor controlling the number, diversity and evolution of living creatures on our planet and have an immense impact on global biogeochemical cycles (Suttle, 2007; Rohwer and Thurber, 2009).

The hallmark feature of viruses, which distinguishes them from other mobile parasitic elements, such as plasmids and transposons, is that they form virions to move between the cells. Viruses come in a variety of shapes and sizes – filamentous, icosahedral, pleomorphic, etc. (Fauquet et al., 2005). Some morphotypes are common to hosts from all three domains of cellular life (Bamford et al., 2005a; Krupovič and Bamford, 2008), while others seem to be restricted to a specific domain (Prangishvili et al., 2006a). The simplest virions consist of a nucleic acid (with as little as one or two encoded genes) and a protein shell encapsidating the genome. Capsids of more complex viruses are often multilayered, composed of a number of different proteins (some of which might be post-translationally modified) and might contain lipids as a structural component (Huiskonen and Butcher, 2007). The capsid mediates recognition of the host cell, but also ensures protection of the genetic material while a virus faces harsh extracellular environments. The nucleic acids utilized by viruses to encode the genetic information are much more diverse than those of their hosts. Viral genomes might consist of either RNA (ribonucleic acid) or DNA (deoxyribonucleic acid) molecules, which can be either single- or double-stranded (ss or ds). Some viruses modify the bases in their genomes, e.g. by methylation, glucosylhydroxymethylation or else by using uracil instead of thymine (Warren, 1980).

Cells in the three domains of life differ from each other in a number of aspects, such as structure and compartmentalization, DNA replication and antiviral defense mechanisms, etc. These differences are mirrored in their viruses. In order to replicate successfully within a host cell a virus has to develop strategies to traverse sophisticated cell envelope structures, evade the defense and often hijack cellular DNA replication machinery. Inevitably, these virus-host system-related specificities translate into differences in the infection cycles of different viruses. Nevertheless, general steps of the reproduction cycle are more or less common to all viruses: (i) attachment to the host cell and internalization of the viral nucleic acid, (ii) transcription, translation and

genome replication, (iii) virion assembly, and (iv) progeny release (Abedon, 2006; Pe'ery and Mathews, 2007). Some viruses are able to delay the steps ii-iv by establishing a lysogeny of their hosts (Campbell, 2006).

2. Virus structure

Since the discovery of viruses by Dmitri Ivanowsky and Martinus Beijerinck in the 1890s (Ivanowsky, 1892; Beijerinck, 1898), a great number of viruses has been isolated and characterized, revealing that they are structurally very diverse. A capsid ensures protection of the viral genetic material when the virus is exposed to the extracellular environment in the search for a new host. However, at the same time, as soon as the cell susceptible to the virus is identified, the capsid has to bind to the cell, disassemble and to ensure the delivery of the viral genome into the replication-competent intracellular environment. This means that the capsid has to satisfy both requirements – stability, on one hand, and lability, on the other, i.e. it has to be *metastable*. Viruses came up with a multiple solutions to this dilemma. However, a common principle, which applies for all viruses, is that virions are constructed from multiple copies of a limited number of different capsid proteins (often a single protein type). The rationale behind this strategy, as suggested by Crick and Watson (1956), is the limited coding capacity of viral genomes. As a result it was predicted that “These small protein molecules then aggregate around the ribonucleic acid in a regular manner, which they can only do in a limited number of ways if they are to use the same packing arrangement repeatedly. Hence small viruses are either rods or spheres” (Crick and Watson, 1956). Indeed, regular capsids obey either helical or icosahedral symmetry rules.

2.1 Helical capsids

Helical capsids are constructed from a single type of capsid proteins arrayed around the viral nucleic acid, i.e. the capsid proteins form an extended proteinatious tube with the viral genome positioned in the central cavity. Such an arrangement can produce rigid rods as well as long flexible filaments (Fig. 1A). The ends of a helical capsid might be sealed with minor capsid proteins. Perhaps the most extensively studied viruses with helical capsids are Tobacco mosaic virus (TMV; Klug, 1999) and bacterial viruses of the *Inoviridae* family, such as fd and M13 (Marvin, 1998; Russel and Model, 2006). In both bacterial and plant viruses the length of the viral filament does not depend on the packing of the capsid proteins but rather on the size of the genome (Klug, 1999; Russel and Model, 2006). Viruses with helical capsids are not only found in bacteria and eukaryotes, but also in archaea (Prangishvili et al., 2006a). Notably, helical capsids can be formed with different types of nucleic acids: rod-shaped and filamentous viruses of plants have linear positive-sense

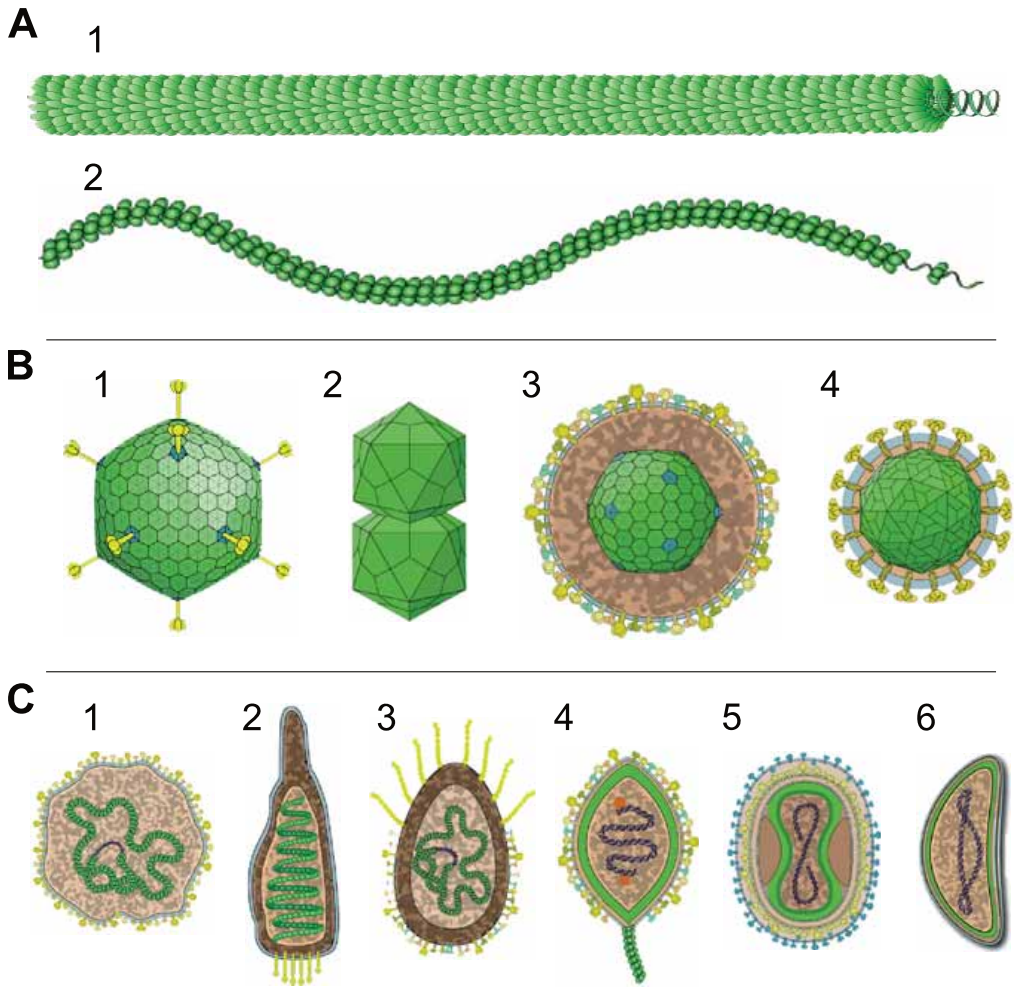

Figure 1. Examples of different virion morphotypes. (A) Helical capsids. 1, Rigid rod-shaped virion (e.g., *Virgaviridae*); 2, Flexible filamentous virion (e.g., *Potyviridae*). (B) Icosahedral capsids. 1, "Naked" icosahedral virion (e.g., *Adenoviridae*); 2, Twinned capsid (e.g., *Geminiviridae*); 3, Enveloped icosahedral virion (e.g., *Herpesviridae*; between the membrane and the capsid is the amorphous tegument layer); 4, Enveloped icosahedral virion without a tegument layer (e.g., *Togaviridae*). (C) Asymmetric and complex virions. 1, Pleomorphic virion (e.g., *Plasmaviridae*); 2, Bottle-shaped virion (e.g., *Ampullaviridae*); 3, Droplet-shaped virion (e.g., *Guttaviridae*); 4, Lemon-shaped virion (e.g., *Fuselloviridae*); 5, Oval- or brick-shaped virion (e.g., *Poxviridae*); 6, Reniform virion (e.g., *Ascoviridae*). Viruses are not drawn to scale. All pictures were adapted with permission from ViralZone, Swiss Institute of Bioinformatics (www.expasy.ch/viralzone).

ssRNA genomes (Koonin and Dolja, 1993), bacterial inoviruses have circular ssDNA genomes (Russel and Model,2006), while rod-shaped and filamentous archaeal viruses of the families *Rudiviridae* and *Lipothrixviridae*, re-

spectively, possess linear dsDNA genomes (Prangishvili et al., 2006a). Unlike in rudiviruses, filamentous capsids of lipothrixviruses are covered with a lipid membrane. However, structural analysis has revealed that the major capsid proteins of viruses from the two archaeal families display the same topology, suggesting a common ancestry (Goulet et al., 2009; Szymczyna et al., 2009).

2.2 Icosahedral capsids

Viral capsids can also be built using icosahedral symmetry (Fig. 1B). In geometry, an icosahedron is a platonic solid that is characterized by 20 identical triangular faces that come together at 12 vertices (5 faces meet at each vertex) and 30 edges. An icosahedron is built on the rotational 2-3-5 symmetry of the solid, with two-, three-, and fivefold rotational symmetry axes passing through its edges, faces, and vertices, respectively. Icosahedral arrangement is arguably the optimal way of forming a closed capsid shell from small identical protein subunits to enclose a maximal volume (Caspar and Klug, 1962). Since proteins are asymmetric in nature and perhaps none has a triangular appearance, the triangular face of the icosahedral viral capsid is formed of a minimum of three identical subunits, i.e. the minimal number of capsomers is 60.

The genome size, and consequently the volume the genome will occupy, varies tremendously among different viruses. For example, the icosahedral capsid of Satellite tobacco necrosis virus (STNV), one of the simplest capsids (Jones and Liljas, 1984), can accommodate only ~1.2 kb of linear ssRNA (Ysebaert et al., 1980), whereas the capsid of the Mimivirus, the largest viral capsid known to date (Xiao et al., 2009), has enough room for the ~1.2 Mb dsDNA genome (Raoult et al., 2004). Obviously, to pack larger genomes, viruses have to build larger capsids. Icosahedral viruses accomplish this by at least three different strategies (Krupovič and Bamford, 2008). The most obvious way to build a larger capsid is to increase the dimensions of the building block, the capsomer (Fig. 2, strategy 1). For example, bacterial virus PRD1 (*Tectiviridae*) and human adenovirus (*Adenoviridae*) use the same number and arrangement of trimeric capsomers to build their icosahedral capsids (Benson et al., 1999). In addition, the capsomers in the two viruses are structurally related. However, the width of the adenoviral capsomer is 97 Å compared with 80 Å in PRD1. This allows the adenovirus to build a virion with a diameter of 920 Å compared to the 740 Å diameter of PRD1 (Benson et al., 1999). Another notable difference between PRD1 and adenovirus is the lack of the internal membrane in the latter. As a result, the adenovirus virion accommodates a genome nearly twice longer than that of PRD1. The second possibility is to increase the number of the capsomers *per se* to create larger triangular faces of the icosahedral shell (Fig. 2, strategy 2). For example, *Paramecium bursaria* Chlorella virus type 1 (PBCV-1; *Phycodnaviridae*), which is also structurally related to PRD1 (Nandhagopal et al., 2002), uses 1,680 trimeric capsomers, compared to 240 capsomers in the PRD1 virion (Yan et al., 2000; Abrescia et al., 2004). Consequently, the diameters of PBCV-1 and

PRD1 virions differ considerably (~190 versus 74 nm). Yet another strategy, utilized by icosahedral viruses to increase the internal capsid volume, is to combine two or more of smaller icosahedral capsids into a continuous structure (Fig. 2, strategy 3). This unusual solution is employed by plant geminiviruses (Fig. 1B), with capsids constructed from two incomplete T = 1 icosahedra joined together to form twinned particles (Zhang et al., 2001; Böttcher et al., 2004). Surprisingly, aberrant virions of geminiviruses might consist not only of a single icosahedron, but also of three incomplete icosahedra joined together (Casado et al., 2004; Jovel et al., 2007). Notably, the valency of the capsid was found to correlate with the length of the packed nucleic acid, since noninfectious isometric T = 1 capsids encapsidated not more than half of the wild-type genome (Casado et al., 2004).

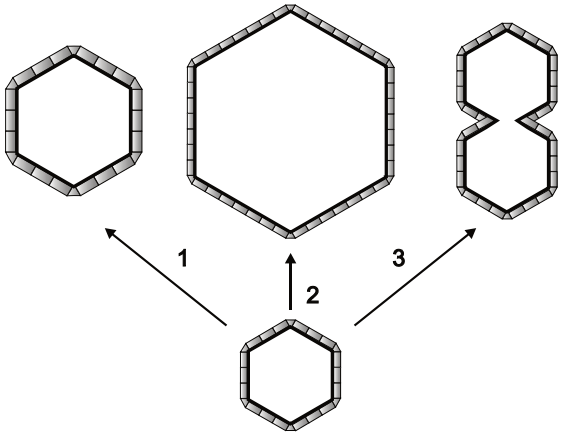

Figure 2. Three strategies utilized by icosahedral viruses to increase the internal volume of the capsid (denoted 1–3). Capsomers are depicted as grey rectangles. 1) Increasing the dimensions of the capsomers. Note that the number of capsomers in the two viruses is the same. 2) Increasing the number of copies of the capsomers used to build a capsid. 3) Association of two capsids into a twinned particle. Note that only isometric capsids are considered here.

Very often an icosahedral protein capsid is not the only layer of the virion between the exterior and the genetic material. Icosahedral virions might also possess a lipid membrane. The latter can be on the outside of the capsid, like in eukaryotic herpesviruses and bacterial cystoviruses, or beneath the protein capsid, like in bacterial virus PRD1 (Huiskonen and Butcher, 2007). Viruses with the external membrane are collectively called enveloped viruses (Fig. 1B). Some complex icosahedral viruses, such as African swine fever virus (*Asfarviridae*), possess two layers of the lipid membrane, one on the outside (envelope) and one on the inside (internal membrane) of the icosahedral protein capsid (Tulman et al., 2009). Peculiarly, the presence of the external lipid envelope is not a stable feature, which would be characteristic to a given virus family. For example, *Emiliania huxleyi* virus 86, a genuine member of the *Phycodnaviridae* family, is an enveloped virus, while other members of this family, such as PBCV-1, are not (Mackinder et al., 2009; Wilson et al., 2009).

Packing of the viral genome into the capsid in small icosahedral viruses occurs, similarly to that in helical capsids, by co-assembly of the nucleic acid with the capsid protein subunits. For example, assembly of the T = 3 capsid of bacterial ssRNA virus MS2 (*Leviviridae*) starts with the binding of a capsid protein dimer to a stem-loop structure in the genomic ssRNA. This triggers subsequent binding of additional capsid protein dimers until the capsid is com-

pleted (Stockley et al., 2007). More complex icosahedral viruses usually form an empty procapsid (or prohead), which undergoes a subsequent genome packaging and maturation. The assembly process is often assisted by various scaffolding proteins (Dokland, 1999). Genome packaging in such a systems is coupled to NTP hydrolysis by virus-encoded packaging motors (Burroughs et al., 2007; Sun et al., 2010) or, like in the case of phiX174 and related microviruses, to genome replication (Fane et al., 2006).

2.3 Asymmetric and complex virions

Not all viruses construct helical or icosahedral capsids. Morphology of enveloped viruses, which are devoid of a regular capsid, is often non-constant (i.e. they are pleomorphic) and the virion shape might depend on a host cell type, growth conditions, etc. For example, virions of the same strain of influenza virus might display diverse morphologies, from roughly spherical to long filamentous particles (Schmitt and Lamb, 2005). Other viruses form very large and complex particles that are bacilliform or reniform, e.g., ascoviruses (*Ascoviridae*; Federici et al., 2009), or else oval and brick-shaped as those of poxviruses (*Poxviridae*; Condit et al., 2006; Fig. 1C). Examination of the viral diversity associated with the domain *Archaea* revealed an unexpected morphological richness of archaeal viruses (Prangishvili et al., 2006a). In addition to helical and icosahedral virions, that are also common to viruses infecting hosts in the other two domains of life, *Bacteria* and *Eukarya*, archaeal viruses display a number of unique morphotypes (Fig. 1C). These include bottle-shaped (*Ampullaviridae*), droplet-shaped (*Guttaviridae*), lemon- or spindle-shaped (*Fuselloviridae*).

3. Evolutionary genomics of prokaryotic viruses

With the advent of genomic and, even more so, post-genomic era the way genomes are studied has changed dramatically. Modern genomics is primarily computational – although profoundly useful in the past, DNA-DNA hybridization, as a method for homology detection, is hardly ever used in modern genomics. Accumulation of immense amounts of genomic data led to the emergence of new disciplines, such as *Bioinformatics* and *Computational Biology*. Although the two terms are often used interchangeably, there is a substantial difference between them. The goal of *Bioinformatics* is engineering of computational methods that enable researchers to analyze, store and visualize the data extracted from genomes, i.e. *Bioinformatics* deals with the infrastructure needed to understand the biology of genomes. In contrast, *Computational Biology* is a research field that utilizes computational methods to generate biologically important results, i.e. the primary goal is the understanding of the genome biology itself and thus has little to do with engineering (Galperin and Koonin, 2003; Semple, 2003). *Computational Biology* is a full-fledged field of biology, which had (and still has) an undeniable impact on development

of experimental biology. Obviously, computational analyses would not be possible without experimentally acquired primary data. However, it is also true that generation of huge amounts of primary sequence data would be useless in the absence of computational utilities that we enjoy today. The latter not only enable researchers to manage and comprehend the data but also to draw biologically relevant conclusions that as a feedback guide for further experimental investigations. Besides, *Computational Biology* enables to answer questions that otherwise can hardly be resolved by experimental means. For example, computational analysis of hundreds of prokaryotic genomes revealed not only a number of features characteristic to these genomes, but also major trends and principles that drive their evolution (Koonin and Wolf, 2008).

In the beginning of this century accumulation of viral genome sequences started to gain its pace (Brüssow and Hendrix, 2002) and there are currently well over two thousand publicly available viral genome sequences in the GenBank. Such a wealth of the genomic data allows not only to get insights into biology of individual viruses, but also to grasp the bigger picture of virus evolution by comparing genome and protein sequences of more distantly related viruses. Comparative analysis of viral genomes with the intent of getting insights into virus evolution is referred here as to Evolutionary genomics of viruses.

3.1 Bacterial viruses

Based on the type of nucleic acid and virion morphology bacterial viruses (also known as bacteriophages or phages) are classified by International Committee on Taxonomy of Viruses (ICTV) into one order and ten families (Fauquet et al., 2005). Genomes of bacterial viruses can be either RNA or DNA and may vary in size tremendously, e.g. ssRNA chromosome of levivirus GA is less than 3,5 kb (Inokuchi et al., 1986), while the dsDNA genome of myovirus G is 497,5 kb (Hendrix, 2009). Table 1 summarizes characteristics of bacterial virus families.

Table 1. Overview of bacterial virus families.

Order Family	Capsid morphology	Additional features	Genome type	Examples
Caudovirales				
Myoviridae	icosahedral	tail (contractile)	dsDNA, L	T4
Siphoviridae	icosahedral	tail (long non-contractile)	dsDNA, L	λ
Podoviridae	icosahedral	tail (short non-contractile)	dsDNA, L	T7
Tectiviridae	icosahedral	internal membrane	dsDNA, L	PRD1
Corticoviridae	icosahedral	internal membrane	dsDNA, C	PM2
Plasmaviridae	pleomorphic	enveloped	dsDNA, C	L2
Microviridae	icosahedral	nonenveloped	ssDNA, C	ϕX174
Inoviridae	filamentous	long flexible or short rigid	ssDNA, C	M13
Cystoviridae	icosahedral	enveloped, multilayered	dsRNA, L, S	ϕ6
Leviviridae	icosahedral	nonenveloped	ssRNA, L	MS2

L, linear; C, circular; S, segmented.

3.1.1 Tailed dsDNA viruses

Vast majority of characterized bacterial viruses belong to the order Caudovirales (Ackermann, 2007), which unites tailed dsDNA viruses of three families – Myoviridae (contractile tails), Siphoviridae (long non-contractile tails) and Podoviridae (short non-contractile tails). Virions of viruses in Caudovirales consist of an icosahedral capsid (or head), which can be either isometric or elongated (prolate), and a tail – an appendage attached to one of the icosahedral vertexes (Hendrix and Casjens, 2005). The capsids are not enveloped and enclose linear dsDNA genomes. The tails are responsible for host cell recognition, attachment and subsequent delivery of the viral DNA across the cell envelope (Poranen and Domanska, 2008). Assembly of tailed virions is a complex process, which starts with the construction of empty procapsids that undergo maturation followed by genome packaging and tail attachment (Steven et al., 2005b). At the end of the infection cycle all tailed viruses characterized so far lyse the host cell (Young and Wang, 2006). However, temperate viruses can establish a lysogeny of their hosts by either integrating into the cellular chromosome (provirus, e.g., siphovirus λ) or replicating inside the host cytoplasm as plasmids (episome, e.g., myovirus P1).

As mentioned above, tailed viruses represent the dominant morphotype among characterized bacterial viruses – of more than 5500 prokaryotic viruses, examined under electron microscope by the end of 2006, 96% were tailed (Ackermann, 2007). This dominance is also reflected in the number of complete genome sequences for these viruses – tailed virus genomes constitute 78% (450 out of 579) of all prokaryotic virus complete genome sequences publicly available in the GenBank (as of 26.04.2010). The number gets even higher when genomic sequences of proviruses are included. Most (if not all) of the sequenced bacterial genomes contain at least one provirus or its remnants; in some cases proviruses might constitute as much as 10–20% of the bacterial chromosome (Casjens, 2003). This abundance of genomic data led to a series of informative in-depth comparative analyses that contributed tremendously to our current understanding of the genetic diversity and evolution of these viruses (Hendrix et al., 1999; Pedulla et al., 2003; Kwan et al., 2005; Kwan et al., 2006; Hatfull et al., 2010).

Diversity in genome size and content. There is little doubt that all tailed bacterial viruses share common ancestry (Baker et al., 2005; Fokine et al., 2005). However, viral genome length within this supergroup varies extensively. The extreme case is the nearly 50-fold difference between the 11,6 kb genome of podovirus P1 infecting mycoplasmas (Tu et al., 2001) and the genome of the giant myovirus G (497,5 kb; Hendrix, 2009). Since protein-coding genes are generally tightly packed in viral genomes (typically they occupy >90% of the genome in tailed viruses; Hendrix, 2002), it is natural that the number of encoded genes is proportional to the size of the genome. As a result tailed viruses with larger genomes encode higher diversity of functions compared with those that possess smaller genomes. Such additional functions might include, for example, DNA

replication and modification, nucleotide metabolism, or like in the case of the above mentioned myovirus G, aminoacyl tRNA synthesis (Hendrix, 2009).

In general, the genome size differences observed in evolutionary related viruses can arise as a result of (i) expansion; viruses with larger genomes arise from the smaller ones by acquisition of new genes or (ii) reduction; viruses with smaller genomes are descendants of an ancestor with a larger genome that underwent reduction. Although some scientists are in favor of the genome reduction hypothesis (Raoult et al., 2004), recent sequence analyses indicate that tailed bacterial viruses with large genomes evolved from a much smaller ancestor (Filée and Chandler, 2008). Three major mechanisms are accountable for such genome expansion: (i) multiple gene or genomic segment duplications, (ii) lateral acquisition of cellular and viral genes, and (iii) dissemination of diverse mobile genetic elements (Filée and Chandler, 2008). The latter hypothesis is also more plausible when the two possibilities (genome expansion versus reduction) are compared in the framework of viral capsid architecture and genome packaging mechanism (Hendrix, 2009).

Sequence divergence. Tailed viruses are an extremely divergent supergroup, and homologous nucleotide and protein sequences of these viruses often bear no recognizable similarity. For example, the major capsid protein (MCP) sequences of more distantly related tailed viruses often display only 10–20% identity in pair-wise comparisons. Nevertheless, it became apparent that all these MCPs (for which structural information is available) have the same basic topology, exemplified by gp5 of HK97 (Wikoff et al., 2000; Jiang et al., 2003; Baker et al., 2005; Fokine et al., 2005; Effantin et al., 2006; Agirrezabala et al., 2007). This divergence of homologous sequences is mainly due to genetic drift. However, not all viral genes are equally prone to accumulation of point mutations. A notable example is the large subunit of the terminase, an enzyme that cleaves a concatameric viral DNA into genome-length units and powers their encapsidation into preformed procapsids. It is perhaps the most conserved genuine viral protein specific to tailed bacterial viruses and herpesviruses (Casjens, 2003; Rao and Feiss, 2008). Although the overall sequence similarity between different terminases is not significant, the enzymatic nature of these proteins demands presence of several highly conserved motifs that do not tolerate substitutions (Rao and Feiss, 2008). However, comparison of tailed virus sequences and identification of distant homologues is not an impossible task. This is mainly due to two reasons: (i) existence of sophisticated bioinformatic sequence-comparison methods and (ii) a constantly expanding database of available (pro)viral genome and protein sequences (Hendrix, 2002). The combination of the two enables to uncover relationships between seemingly unrelated sequences through sequential comparison of multiple intermediate homologues.

Modular genome organization and mosaicism. Typically, genes of tailed viruses are organized into functional modules, i.e., genes encoding proteins that function in the same process are clustered together and are regulated by common promoters (Brüssow and Desiere, 2001). For example, genes of lambdoid viruses are usually organized into modules responsible for DNA packaging, virion morphogenesis, establishment of lysogeny and host cell lysis, genome replication and recombination as well as regulation

of transcription (Brüssow and Desiere, 2001). Therefore, the viral chromosome can be seen as an assortment of such modules. Gene order in these functional modules is often well conserved, although their position with respect to each other may vary in different virus groups (Hendrix, 2002). The module for virion structure and assembly in siphoviruses is a good example of conserved gene order (Brüssow and Desiere, 2001; Casjens, 2003; Hatfull et al., 2010). Although sequence similarity between genes present in this module in distantly related viruses is often beyond recognition, the order of the genes is generally the same.

When compared, genomes of tailed phages display mosaic pattern (Juhala et al., 2000; Proux et al., 2002; Pedulla et al., 2003). The continuity of high level pairwise similarity between two viral genomes is often disrupted by regions that display only distant, if any, similarity to each other. A unit in such genetic mosaics is typically a single gene or, less frequently, parts of genes encoding protein domains. However, not all functional modules are equally susceptible to gene substitution or insertion/deletion. In general, genes encoding interacting proteins do co-evolve. Such are, for example, modules encoding virion structural proteins, especially those responsible for capsid formation (Hendrix, 2003). In some cases, viral genomes are shaped by matching entire functional modules, as in the case of siphovirus HK97, which shares the capsid assembly module with myovirus SfV, whereas the tail assembly module is clearly related to that of siphovirus λ (Lawrence et al., 2002).

Mechanisms of evolution. Evolution of tailed virus genomes is mainly driven by illegitimate, rather than homologous, recombination occurring at essentially randomly distributed positions within the genome (Juhala et al., 2000; Pedulla et al., 2003). Most new variants generated in such process are expected to be nonviable and hence being eliminated by natural selection. However, in rare circumstances, recombinant genomes may produce viable progeny and lead to productive infection. The success of such recombinants depends on preservation of essential viral functions and maintenance of proper genome length. Therefore, the survivors usually bear neutral recombination joints. The genetic exchange might occur between two viruses co-infecting a single bacterial cell or between a virus and a resident provirus. The latter scenario is reasoned to be more likely (Hatfull, 2006), because there is no immediate pressure on the genome length of the integrated provirus nor there is urgency to form functional progeny, i.e., the provirus might reside in the bacterial chromosome for as long as viable combination is achieved (possibly involving multiple recombination events with different partners).

Although the most inventive, the illegitimate recombination, as mentioned above, is expected to be the least successful, i.e., viable viruses are produced at very low frequency. However, in the light of current information on the size and dynamics of the global population of tailed viruses (Hendrix, 2002; Suttle, 2007), even such rare events are, in fact, frequent enough to account for the observed diversity of tailed viruses. It has been estimated that 10^9–10^{10} fully functional novel genome variants are produced via nonhomologous recombination every second on a global scale (Hendrix, 2009). Once the newly created mosaic joints prove to be advantageous they can rapidly spread within the population of closely related viruses by homologous recombination

(Hendrix, 2003). To sum up, genomes of tailed bacterial viruses evolve by a combination of genetic drift, extensive illegitimate recombination, and homologous recombination (Hendrix, 2009).

3.1.2 Other bacterial viruses

The wealth of the genomic data available for tailed bacteriophages and related proviruses is counterbalanced by the scarcity of those available for other bacterial viruses. Single genome sequences represent some virus families, such as *Corticoviridae* and *Plasmaviridae* (Table 1). Consequently, understanding of the genome evolution for these viruses is far from satisfactory. Below are described the few studies on comparative genomics of bacterial viruses other than those from the order *Caudovirales*.

Tectiviridae. The characteristic tectiviral virion consists of an icosahedral protein capsid, which covers a protein-rich membrane vesicle. The latter encloses a linear dsDNA genome of ~15 kb (Grahn et al., 2006). The genome possesses inverted terminal repeats and covalently attached terminal proteins that prime the genome replication carried out by the virus-encoded type B DNA polymerase. Tectiviruses can be divided into two groups – those infecting Gram-negative bacteria (exemplified by bacteriophage PRD1) and those of Gram-positive hosts (exemplified by bacteriophage Bam35).

Genome sequences are available for several PRD1-like as well as Bam35-like viruses (Saren et al., 2005; Sozhamannan et al., 2008). Although representatives of the two groups are extremely similar to each other structurally (Abrescia et al., 2004; Laurinmäki et al., 2005), their genomes virtually share no sequence conservation; only four proteins of Bam35 could be linked to those of PRD1 using bioinformatic methods (Ravantti et al., 2003). Within the groups, however, the genomes are highly similar. For example, comparative analysis of the six available PRD1-like genomes revealed that they share 91.9% – 99.8% nucleotide sequence identity, despite the fact that the six viruses have been isolated independently at geographically remote locations worldwide (Saren et al., 2005). This is in sharp contrast to the genomics of tailed dsDNA bacteriophages, where genetic diversity is immense (as discussed above). Bam35-like tectiviruses infecting different *Bacillus* species are also genetically closely related to each other, although not to the extent of PRD1-like viruses (Sozhamannan et al., 2008). Bam35-like phages infecting the same host were found to be genetically closer to each other than to related viruses infecting different hosts (Sozhamannan et al., 2008). This suggests that Bam35-like tectiviruses co-evolve with their hosts. Notable difference between the PRD1-like and Bam35-like tectiviruses is that the latter possess the ability to lysogenize their hosts and replicate their genomes intracellularly as linear plasmids (Strömsten et al., 2003; Gaidelytė et al., 2005). Due to the same reasons, discussed above for tailed bacteriophages, the pace and mode of evolution of temperate Bam35-like viruses might differ from those of their virulent siblings. However, due to scarcity of the genomic data for tectiviruses, this possibility remains to be tested.

Despite the lack of extensive sequence conservation, the overall genome organization of PRD1-like tectiviruses is very similar to that of their Bam35-like relatives (Ravantti et al., 2003). In both groups genomes follow a modular organization, also characteristic to tailed dsDNA viruses (Brüssow and Desiere, 2001), with functionally related genes grouped into distinct clusters that are regulated by common promoters (Saren et al., 2005; Sozhamannan et al., 2008). However, unlike tailed bacteriophages, the mosaic mode of genome evolution is not characteristic to tectiviruses, with the exception of non orthologous replacement of lysin-encoding genes in Bam35-like viruses (Sozhamannan et al., 2008). It therefore appears that tectiviral genomes evolve mainly through accumulation of point mutations and seldom acquisition of new genes. The striking near-identity between PRD1-like genomes was suggested to be a result of their "optimal genome-level organization and structure, in which any change decreases fitness" (Saren et al., 2005). The overall genome organization and virion architecture suggest that PRD1-like and Bam35-like tectiviruses had a common ancestor, which existed prior to the diversification of the Gram-negative and Gram-positive bacteria more than a billion years ago (Ravantti et al., 2003; Laurinmäki et al., 2005).

Microviridae. Family *Microviridae* unites small icosahedral viruses with circular ssDNA genomes replicated via rolling-circle (RCR) mechanism (Table 1). The type member of the family, bacteriophage ϕX174, has been extensively studied structurally, biochemically and genetically (Fane et al., 2006). Microviruses have been isolated from diverse hosts, such as enterobacteria, *Bdellovibrio*, *Spiroplasma*, and *Chlamydia* (the latter three are intracellular parasitic bacteria).

Based on their genome and virion organization microviruses are divided into two distinct groups – those infecting enterobacteria and those replicating in obligate parasitic bacteria (Garner et al., 2004). The latter group has been recently assigned by the ICTV into a subfamily *Gokushovirinae* (Gokusho: Japanese for very small) within the family *Microviridae* (http://www.ictvonline.org). Like in the case of tectiviruses, the two groups share little genome and protein sequence conservation, but overall genome organization is similar. The ϕX174-like viruses possess slightly larger genomes than the viruses from the other group (5.3–6.2 kb versus 4.5 kb). The genome size difference is reflected in the absence of genes for the major spike protein (gpG) and the external scaffolding protein (gpD) in gokushoviruses (Garner et al., 2004). A large number (>40) of genomes from closely related ϕX174-like microviruses infecting *Escherichia coli* has been sequenced (Rokyta et al., 2006). A phylogenetic analysis of these sequences suggested relatively recent emergence of gene D in the ϕX174-like clade with subsequent spread of the gene in the population of related viruses. It also revealed at least two other horizontal transfer events between the clades that probably occurred by homologous recombination (Rokyta et al., 2006). However, unlike in tailed dsDNA viruses, the illegitimate recombination was not found to have any substantial contribution to the genome evolution of microviruses.

Interestingly, comparison of gokushovirus genome sequences revealed that *Bdellovibrio*-infecting virus ϕMH2K is more similar to some *Chlamydia* viruses than the *Chlamydia*-infecting mi-

croviruses are related to each other (Brentlinger et al., 2002). In addition, it was noted that microviruses encode preserved open reading frames (ORFs) nested within overlapping genes. Point mutations were suggested to accrete in such ORFs until a gene, termed "cretin" (for "accrete in"), encoding a beneficial function is produced, as might be the case for lysis gene *E* of ϕX174 (Brentlinger et al., 2002). Consequently, it was suggested that occasional species jumping and genetic drift are the two major mechanisms behind the evolution of microviruses. Such inherently different mode of genome evolution for microviruses when compared to tailed dsDNA bacteriophages was reasoned to be a result of i) strictly lytic lifestyle; ii) small genome size restrained by the capsid dimensions; iii) low abundance of the double-stranded replicative form of the viral genome inside the host cell, which should reduce the frequency of recombination (Brentlinger et al., 2002; Fane et al., 2006; Rokyta et al., 2006).

***Inoviridae*.** Similarly to microviruses, inoviruses possess small circular ssDNA genomes (4.5 – 9.0 kb) replicated via RCR mechanism, but their virions are built with helical rather than icosahedral symmetry (Fig. 1; Table 1). The family contains two genera, *Inovirus* and *Plectrovirus*. Inoviruses were found to infect diverse Gram-negative bacteria as well as one Gram-positive species (Russel and Model, 2006), while plectroviruses infect mollicutes, such as *Spiroplasma* and *Acholeplasma* (Maniloff and Dybvig, 2006). Viruses from the two genera show virtually no genome or protein sequence similarity and differ considerably in morphology: inovirus virions are generally long and flexible, while those of plectroviruses are short rods. Unlike other bacterial viruses, with the exception of plasmavirus L2 (Maniloff and Dybvig, 2006), inoviruses do not lyse host cells, but leave them by extrusion coupled to virion assembly (Russel and Model, 2006). All characterized plectroviruses and some (but not all) inoviruses are capable of lysogenizing their hosts. This ability has been acquired on multiple independent occasions, since not all members of the family are able to integrate into the host chromosome, and those that are, utilize different genome integration strategies. *Spiroplasma* plectroviruses SpV1-R8A2B and SpV1-C74, and *Neisseria* inovirus Nf encode distinct transposases belonging to IS*30*, IS*3* and IS*110*/IS*492* families, respectively (Melcher et al., 1999; Kawai et al., 2005). Interestingly, the transposase genes in the two plectroviruses, but not in the inovirus Nf, have replaced an RCR initiation protein-coding gene, which is otherwise present in the related SpV-like virus, SVTS2 (Sha et al., 2000). As a result, genome replication in the transposase-encoding plectroviruses occurs via transposition mechanism with the encapsidated genome representing a circular transposition intermediate (Melcher et al., 1999). Another temperate inovirus ϕRSM1 infecting *Ralstonia solanacearum* encodes a recombinase of the resolvase/invertase family and utilizes a host tRNA gene for integration (Kawasaki et al., 2007). An even more peculiar strategy for genome integration is utilized by CTXϕ-like inoviruses infecting *Vibrio* sp. and possibly also inoviruses of *Xanthomonas* sp. (McLeod et al., 2005). The cholera toxin-encoding CTXϕ is arguably "smarter than the average phage" (Blakely, 2004), since it hijacks cellular XerC/D enzymes for site-specific genome integration into the host chromosome and does not encode a recombinase of its own (Huber and Waldor, 2002; McLeod et al., 2005).

Genomes of inoviruses typically display modular organization with genes encoding proteins responsible for genome replication, virion morphogenesis and structure grouped into clusters (Russel and Model, 2006). However, unlike in tailed viruses and tectiviruses, gene expression is not temporal but occurs concurrently and is mainly regulated by differences in the strength and accessibility of the ribosome-binding sites, unequal codon usage, etc. (Russel and Model, 2006). Nevertheless, it appears that evolution of inovirus genomes is governed by very similar mechanisms as those operating in tailed dsDNA viruses (Lawrence et al., 2002). As mentioned above, inoviruses differ in their genome length. This is mainly a result of lateral acquisition of new genes from diverse sources. For example, a homologue of the cellular outer membrane channel EpsD, which is used for virion extrusion, is encoded by *E. coli*-infecting inoviruses (gene IV in M13), but not in *Vibrio*-infecting phages, such as CTXϕ, that utilize the cellular protein for the same purpose (Davis et al., 2000). Another notable example is the operon *ctxAB* ofbacteriophage CTXϕ encoding the cholera toxin (Waldor and Mekalanos, 1996). Inoviruses from both genera seem to enjoy frequent nonorthologous replacements within the module responsible for genome replication (Stassen et al., 1992; Waldor et al., 1997; Sha et al., 2000; Lawrence et al., 2002). In addition, homologous recombination and intergenome rearrangements have been reported to play a role in the evolution of inoviruses (Peeters et al., 1985; Lawrence et al., 2002; Russel and Model, 2006).

Cystoviridae. Members of this family possess tripartite dsRNA genomes enclosed into an icosahedral double-layered protein capsid, which is surrounded by a membrane envelope (Poranen et al., 2005). The innermost capsid, the so-called polymerase complex or procapsid, is delivered into the cell interior upon infection and is essential for genome replication and transcription (Poranen et al., 2005; Mindich, 2006). Cystoviruses infect phytopathogenic pseudomonads, such as *Pseudomonas syringae*. Several genome sequences of cystoviruses have been determined, including that of bacteriophage ϕ6, the type member of the family (Mindich, 2006).

Cystoviral genomes are organized in a modular fashion. However, unlike in DNA viruses with non-segmented genomes, functionally related proteins are generally encoded on different genomic segments. The largest of the three segments, L, encodes proteins of the procapsid, segment M contains genes for the receptor-binding complex, whereas the smallest of the three segments, S, encodes the procapsid shell protein, the major membrane protein and proteins responsible for host cell lysis (Poranen et al., 2005). Gene expression is temporal and is achieved by a combination of unequal stability and efficiency of production of the mRNAs from different genomic segments (Mindich, 2006).

The general mode of genome evolution in cystoviruses is similar to that of DNA bacteriophages, but the molecular basis is different. Illegitimate recombination is well documented for these viruses and occurs with high frequency by template switching during the synthesis of a nascent RNA strand within the polymerase complex (Mindich, 1996). Such recombination usually involves three identical bases at the crossover site, but can range from 0 to 12 nucleotides (Qiao et al., 1997). Homologous recombination has also been reported

to occur *in vitro* for bacteriophage ϕ8, which is the most distant isolated relative of ϕ6 (Onodera et al., 2001). However, this type of recombination appeared extremely rare in wild populations of cystoviruses (Silander et al., 2005) and was reasoned to play only a minor role (if any) in the evolution of cystoviruses due to low probability of incorporation of two molecules of the same segment type into a single procapsid (Mindich, 2006). Comparaive analysis of cystoviral genomes has also revealed cases of intergenome rearrangements, such as disruption of synteny in the L segment of ϕ8, where gene *7* is located at the 3' terminus instead of the 5' terminus, position observed for this gene in all the other cystoviruses (Sun et al., 2003). In addition, cystoviruses are able to acquire genetic material from different sources, which seems to occur via non-specific incorporation of heterologous transcripts into the procapsid followed by recombination with the viral RNA via template switching (Onodera et al., 2001). Notably, incorporation of native transcripts containing viral packaging signals was found to be only 10 times more efficient, suggesting that occasional incorporations of cellular transcripts or those of other viruses might be rather frequent. Analysis of the recently sequenced genome of cystovirus ϕ2954 (Qiao et al., 2010) revealed another mechanism contributing to the evolution of cystoviruses and also common to other bacterial viruses. Gene *5*, encoding the muralytic enzyme P5 in all members of the *Cystoviridae*, is replaced in ϕ2954 by a non-orthologous gene, which is homologous to the *flgJ* gene of *P. syringae* encoding a flagellar protein FlgJ with a peptidoglycan hydrolase activity. The newly acquired FlgJ homologue in ϕ2954 is thus likely to be a functional equivalent of P5, found in other cystoviruses (Qiao et al., 2010). Finally, genomic segment reassortment in wild populations of cystoviruses as well as in the laboratory settings occurs at extremely high frequencies, exceeding the rate of nucleotide substitutions (Onodera et al., 2001; Silander et al., 2005). This suggests that coinfection of the same host cell with multiple cystoviruses occurs frequently in their natural habitat and that once a new beneficial function is acquired it will be quickly spread within the population by genomic segment reassortment.

Leviviridae. Leviviruses are among the smallest viruses known. They possess linear ssRNA genomes enclosed into icosahedral protein shells (van Duin and Tsareva, 2006). Based on serological cross-reactivity, genome size and organization the family is divided into two genera: *Levivirus* and *Allolevivirus*. Leviviruses, with ~3.5 kb genomes, encode four proteins: capsid protein, maturation protein, lysis protein and an RNA-dependent RNA polymerase (RdRp). Notably, the lysis protein is encoded by an ORF, which overlaps with its 5' and 3' regions with the capsid and RdRp genes, respectively. The gene seemingly originated by a mechanism similar to that responsible for the emergence of the lysis gene of microvirus ϕX174 (see above), i.e. a preexisting genomic region has accumulated mutations until a beneficial function was achieved (Brentlinger et al., 2002; Klovins et al., 2002). Alloleviviruses possess somewhat longer genomes (~4.2 kb) and instead of the lysis protein encode two variants of the capsid protein: normal-sized and an extended version generated by an occasional (6% frequency) read-through of the stop codon (van Duin and Tsareva, 2006). The host cell lysis in allolevivirus infections is carried out by the maturation protein (Bernhardt et al., 2001).

Most of the isolated members of the *Leviviridae* infect *E. coli* and are F pilus-specific. Genome sequences for thirty such phages are currently available (15 for leviviruses and 15 for alloleviviruses). Their analysis revealed no apparent recombination events or deviation from the general genome organization characteristic to leviviruses and alloleviviruses (Friedman et al., 2009). Despite low accuracy of the RdRp, genomes within each of the two genera of the *Leviviridae* share more than 50% nucleotide identity (Friedman et al., 2009). Such sequence stability among the *E. coli*-infecting RNA viruses might be a result of their adaptation to the "needs of the day" (van Duin and Tsareva, 2006). Indeed, genome sequences of more distantly related ssRNA viruses infecting hosts other than *E. coli* display more divergence. Complete genome sequence of bacteriophage AP205 infecting *Acinetobacter* species revealed genomic organization different from the two layouts consistently found in the ssRNA coliphages (Klovins et al., 2002).

Most of the current understanding on the evolvability of leviviruses comes from results obtained during numerous laboratory manipulations with these viruses (van Duin and Tsareva, 2006). Such studies revealed that leviviruses are capable of overcoming the *in vitro* introduced detrimental defects by a number of mechanisms. Such studies have indicated that i) frequent base substitutions lead to adaptation to the changing environmental condition (Betancourt, 2009), ii) deletions as well as duplications occurring by random recombination are readily introduced into the ssRNA of leviviruses in response to various deliberate changes to the genome sequence and structure (Olsthoorn and van Duin, 1996; Licis et al., 2000), iii) genomic defects can be repaired by illegitimate as well as less frequent homologous recombination occurring as a result of transesterification and template switching, respectively (Palasingam and Shaklee, 1992; Chetverin et al., 2005).

3.2 Archaeal viruses

The domain Archaea is divided into two major phyla: Euryarchaeota and Crenarchaeota (Woese et al., 1990). Viruses infecting organisms from the two phyla share virtually no sequence similarity and are often discussed separately (Prangishvili et al., 2006a; Prangishvili et al., 2006b). All archaeal viruses characterized thus far contain DNA genomes, which can be either linear or circular. Table 2 summarizes characteristics of archaeal viruses. Most of archaeal virus families contain only few members, while some are represented by a single virus species (Prangishvili et al., 2006a). Therefore, only those virus groups that contain more than one member and for which evolutionary conclusions can be attempted from comparative genomic analysis will be discussed here.

Table 2. Overview of archaeal viruses.

Family	Capsid morphology	Additional features	Genome type	Examples
Crenarchaeal viruses				
Fuselloviridae	spindle-shaped	short appendages	dsDNA, C	SSV1
Bicaudaviridae	spindle-shaped	two tails	dsDNA, C	ATV
Guttaviridae	droplet-shaped	multiple thin fibers	dsDNA, C	SNDV
Ampullaviridae	bottle-shaped	short filaments at the broader end	dsDNA, L	ABV
Globuloviridae	spherical	helical nucleoprotein	dsDNA, L	PSV
Lipothrixviridae	filamentous	enveloped, flexible or rigid	dsDNA, L	AFV1
Rudiviridae	stiff rods	three terminal fibers	dsDNA, L	SIRV1
STIV	icosahedral	internal membrane	dsDNA, C	
STSV1	spindle-shaped		dsDNA, C	
Euryarchaeal viruses				
Myoviridae	icosahedral	tail (contractile)	dsDNA, L	ϕCh1
Siphoviridae	icosahedral	tail (long non-contractile)	dsDNA, L	ψM2
Salterprovirus	spindle-shaped		dsDNA, L	His1
PAV1	spindle-shaped	enveloped (?)	dsDNA, C	
A3 VLP	oblate		dsDNA, C	
HRPV-1	pleomorphic	enveloped	ssDNA, C	
HHPV-1	pleomorphic	enveloped	dsDNA, C	
SH1	icosahedral	internal membrane	dsDNA, L	

L, linear; C, circular; crenarchaeal and euryarchaeal viruses not assigned to families but displaying features distinguishing them from viruses belonging to approved families are separated by a broken line.

3.2.1 Crenarchaeal viruses

Comparative genomic analysis showed that crenarchaeal viruses from different families posses a small pool of common genes. However, some viruses, such as those from *Globuloviridae*, share no genes with other known viruses (Prangishvili et al., 2006b). Most of the common genes encode proteins involved in genetic processes, such as DNA replication, recombination or transcription regulation, and are likely to be exchanged horizontally between unrelated viruses or viruses and their hosts (Prangishvili et al., 2006b).

Fuselloviridae. Fuselloviruses possess lemon- or spindle-shaped virions with short filamentous appendages at one pointed end (Fig. 1) and infect hyperthermophilic hosts from crenarchaeal genera *Sulfolobus* and *Acidianus* (Prangishvili et al., 2006a; Redder et al., 2009). The circular dsDNA genomes of fuselloviruses (13.7–24.1 kb) are positively supercoiled and have been shown to integrate site-specifically into the host chromosome (Muskhelishvili et al., 1993; Held and Whitaker, 2009). Peculiarly, the integrase gene of fusellovirus-

es is partitioned upon recombination with the host chromosome into two fragments that flank the integrated provirus (She et al., 2004). Thirteen fuselloviruses genomes (9 viral and 4 proviral) are currently available (Held and Whitaker, 2009; Redder et al., 2009).

Comparative genomics has revealed that, despite of substantial variation in genetic content, fuselloviruses share a set of 12-13 "core" genes (Held and Whitaker, 2009; Redder et al., 2009). Phylogenetic analysis of these genes pointed to the co-evolution of fuselloviruses with their hosts in a biogeographic context (Held and Whitaker, 2009). It also became apparent that homologous recombination is prevalent among fuselloviruses and seems to generally occur within protein-coding genes rather than intergenic regions (Redder et al., 2009). Although heterogeneity in genome length of these viruses suggests that horizontal gene transfer plays a significant role in fusellovirus evolution, its extent has not been estimated yet.

Fuselloviruses display a peculiar relationship with pRN-like plasmids of *Sulfolobus* (Lipps, 2006). The two types of mobile elements seem to exchange genes. For example, it has been concluded that pRN-like plasmid pXZ1 developed the ability to integrate into the host chromosome by lateral acquisition of the integrase gene from a fusellovirus SSV4 (Peng, 2008). Another plasmid, pSSVx, was found to be a virus-plasmid chimera which, in addition to the typical pRN-specific gene content, contains two viral genes (Arnold et al., 1999). The two viral genes ensure encapsidation of the plasmid into virions upon superinfection of the plasmid-containing *Sulfolobus* cells with a fusellovirus SSV1 and enable dissemination of the plasmid in a virus-like fashion (Arnold et al., 1999).

***Lipothrixviridae* and *Rudiviridae*.** Helical archaeal viruses with linear dsDNA genomes are classified into two families, *Rudiviridae* and *Lipothrixviridae* (Prangishvili et al., 2006a). While rudivirus virions are stiff and rod-like, those of lipothrixviruses are flexible filaments. At least for rudiviruses, the length of the filamentous virion is proportional to the length of the genomic DNA (Vestergaard et al., 2008b). Unlike rudiviruses, lipothrixvirus filaments are enveloped (Prangishvili et al., 2006a). Despite these differences rudiviruses share at least nine genes with lipothrixviruses suggesting that viruses from the two families descended from a common ancestor (Peng et al., 2001; Prangishvili et al., 2006b). This proposal has been recently strengthened by the high-resolution structural analysis, which revealed that the DNA-binding major capsid proteins of rudiviruses and lipothrixviruses share the same fold (Goulet et al., 2009).

Based on the differences in the virion terminal structures as well as variation in gene content and organization family *Lipothrixviridae* is subdivided into four genera: Alpha-, Beta-, Gamma-, and Deltalipothrixvirus (Prangishvili et al., 2006a). Comparative analysis of lipothrixviruses revealed that linear genomes of viruses from one genus can undergo intergenomic recombination with those from other genera (Vestergaard et al., 2008a). Lipothrixviruses and rudiviruses display a considerable variation in the gene order and content. It has been noticed that rudiviruses accumulate new unique genes, with no similarity matches in public databases, at the termini of their linear genomes (Vestergaard et al., 2008b). Similar distribution of conserved versus non-conserved genes has been observed in the linear dsDNA genomes of poxviruses (Lefkowitz et al., 2006).

Comparative genomic analysis revealed that rudiviruses and lipothrixviruses not only exchange genes horizontally with each other but also are capable of gene acquisition from their hosts (Peng et al., 2001). For example, genes for dUTPase, flavin-dependent thymidylate synthase (ThyX) and Holliday junction resolvase were likely transferred into viral genomes from cellular organisms (Peng et al., 2001; Prangishvili et al., 2006b).

STIV-like viruses. The overall virion organization of *Sulfolobus* turreted icosahedral virus, STIV, resembles that of bacterial tectiviruses and corticoviruses. The STIV virion consists of an icosahedrally organized proteinaceous capsid that surrounds a protein-rich lipid membrane, enclosing the circular dsDNA genome of 17.6 kb (Rice et al., 2004; Maaty et al., 2006). Interestingly, the major capsid proteins of tectivirus PRD1, corticovirus PM2 and the archaeal STIV were found to possess the same structural fold (Benson et al., 1999; Khayat et al., 2005; Abrescia et al., 2008). In addition, the three viruses encode homologous genome packaging enzymes (Strömsten et al., 2005).

Recently, the genome sequence of a close STIV relative, STIV2, has been determined (Happonen et al., 2010). Sequence comparison revealed the overall similar genome organization for the two viruses; however several insertions/deletions were also observed (Happonen et al., 2010). Notably, STIV, but not STIV2, carries a gene which has homologues in nearly all rudiviruses. A small membrane protein encoded by this gene is implicated in virion release at the end of infection cycle (Quax et al., 2010). The presence of the homologous genes in rudiviruses and STIV is in accordance with the experimental evidence for similar virion release mechanisms utilized by the two groups of viruses (Bize et al., 2009; Brumfield et al., 2009). Consequently, it was suggested that STIV acquired its virion release gene relatively recently from *Sulfolobus*-infecting rudiviruses (Quax et al., 2010). Similar horizontal flux of lysis genes has been also observed for evolutionary unrelated viruses infecting bacteria, such as tectivirus PRD1 and tailed bacteriophages (Krupovič et al., 2008).

3.2.2 Euryarchaeal viruses

The diversity of virion morphotypes associated with euryarchaeal hosts is considered to be less pronounced than that of crenarchaeal viruses (Prangishvili et al., 2006a; Porter et al., 2007). Indeed, all currently isolated viruses of euryarchaea fall into three morphological groups: head-and-tail (resembling bacterial viruses in the order *Caudovirales*), tailless icosahedral, and spindle shaped/pleomorphic (Porter et al., 2007; Roine et al., 2010). However, recent sampling of the viral diversity in saltern ponds, a beloved habitat of halophilic archaea, unveiled that our current knowledge on the diversity of euryarchaeal viruses is far from complete; viral morphotypes previously thought to be specific to crenarchaeal viruses are also present in hypersaline environments (David Prangishvili, personal communication). Although head-and-tail viruses represent the majority of reported halovirus isolates, they seem to constitute only a minor part of euryarchaeal viruses (Prangishvili et al.,

2006a). Comparative genomics of archaeal head-and-tail viruses is one of the topics of this dissertation and will be discussed in the Results and Discussion section.

Spindle-shaped/pleomorphic viruses. Viruses with spindle-shaped appearance have been isolated from hyperthermophilic and halophilic euryarchaeal hosts of orders Thermococcales (PAV1 virus) and Halobacteriales (viruses His1 and His2), respectively (Geslin et al., 2003; Bath et al., 2006). These viruses displayed plasticity in virion morphology, which ranged from spindle-shaped to elongated and flattened (for PAV1; Geslin et al., 2003) or oblate and pleomorphic (for His2; Bath et al., 2006). Similarly, preparations of the predominantly pleomorphic haloarchaeal virus HHPV-1 also contain a fraction of spindle-shaped particles (Roine et al., 2010).

The circular dsDNA genome of PAV1 did not show similarity to other archaeal spindle-shaped viruses or indeed to any known virus (Geslin et al., 2007). Haloarchaeal virus His1 was first assigned to the family *Fuselloviridae*, which includes crenarchaeal viruses of similar morphology (Bath and Dyall-Smith, 1998). However, the isolation of the second spindle-shaped haloarchaeal virus His2 and analysis of the two genomes made it obvious that the gross morphology is the only common characteristic between crenarchaeal fuselloviruses and euryarchaeal viruses His1/2 (Bath et al., 2006). Consequently, viruses His1 and His2 were reassigned to a genus *Salterprovirus*. Both viruses contain linear dsDNA genomes with inverted terminal repeats and terminal proteins. Both encode type B DNA polymerases and their genomes are likely to be replicated in a protein-primed manner (Bath et al., 2006). The two viruses are only distantly related to each other with generally no similarity between their corresponding protein sequences (Bath et al., 2006).

Genome sequence of the recently characterized pleomorphic haloarchaeal virus HHPV-1 unexpectedly revealed that a block of genes, encoding virion structural proteins of HHPV-1, is collinear and shares significant similarity with those of the spindle-shaped virus His2 (Roine et al., 2010). The predicted genes in His2 genome are organized into two modules, which are seemingly transcribed in the opposite directions, towards the termini of the linear genome from a central region (Bath et al., 2006). Interestingly, His2 genes that share sequence similarity with those of HHPV-1 occupy almost entirety of one of these modules. The other module encodes proteins implicated in His2 genome replication (DNA polymerase and a number of proteins with potential zinc finger DNA-binding motifs; Bath et al., 2006). Such module is not present in the circular HHPV-1 genome, which instead encodes a putative rolling-circle replication initiation protein at equivalent position (Roine et al., 2010). It therefore seems that His2 or HHPV-1 donated a block of structural protein-coding genes to a linear and circular plasmid, respectively, thereby transforming a plasmid into a virus. A similar evolutionary scenario has also been proposed to be accountable for the origin of plant geminiviruses (Krupovič et al., 2009) as well as bacterial siphovirus N15 (Ravin et al., 2000; Lawrence et al., 2002).

The circular ssDNA genome of another haloarchaeal pleomorphic virus, HRPV-1, presented an even greater

amazement. Comparative analysis revealed that it is remarkably similar to the dsDNA genome of HHPV-1, both in gene content and organization, leaving no doubt about evolutionary relationship between these ss- and dsDNA viruses (Pietilä et al., 2009; Roine et al., 2010). In addition, the pattern of virion structural proteins was also very similar for HHPV-1 and HRPV-1 (Roine et al., 2010). It has also been noticed that the protein pattern is reminiscent to that of the pleomorphic ssDNA bacteriophage L175 (Dybvig et al., 1985), suggesting that the three viruses might be evolutionary related (Pietilä et al., 2009; Roine et al., 2010). Regrettably, the genome sequence of L175 is not yet available.

The HHPV-1/HRPV-1 example suggests that there is a certain liberty in the choice of the replicative form of the genome that will be encapsidated into mature virions. A reminiscent situation is observed with eukaryotic retrotranscribing viruses: members of the *Retroviridae* family encapsidate the ssRNA replicative intermediate, while hepadnaviruses (family *Hepadnaviridae*) seem to have chosen the dsDNA form (Steven et al., 2005a). Nevertheless, such a close relationship between viruses with different types of nucleic acids (ds versus ssDNA) is unprecedented among prokaryotic viruses and shakes the foundation of the current virus taxonomy (Roine et al., 2010).

Icosahedral tailless viruses. Haloarchaeal virus SH1 is the only representative of this morphotype among euryarchaeal viruses (Porter et al., 2005). Virion organization of SH1 resembles that of bacterial tectiviruses (Table 2; see also above). The linear dsDNA genome also has inverted terminal repeats and covalently attached terminal proteins (Bamford et al., 2005b; Porter and Dyall-Smith, 2008). Unexpectedly, in contrast to all other known viruses with this genome type, SH1 does not encode an identifiable DNA polymerase (Bamford et al., 2005b). Although analysis of the SH1 protein sequences did not reveal considerable relationship to other archaeal viruses (Bamford et al., 2005b), it did point to a connection between SH1 and bacteriophage P23-77 (circular dsDNA) infecting *Thermus thermophilus* (Jalasvuori et al., 2009). The two viruses share a block of genes encoding the putative genome packaging ATPase and the two major capsid proteins. In line with this finding, the evolutionary link between SH1 and P23-77 has also been deduced from the structural analysis, which showed very similar organization (T = 28) of their virions (Jaatinen et al., 2008).

4. Virus lineage hypothesis: deep evolutionary connections

Viruses are fast evolving organisms, with the nucleotide substitution rates, especially for ssRNA and ssDNA viruses, generally being higher than those observed for cellular organisms (Duffy et al., 2008; Holmes, 2010). On top of that, virus–virus as well as virus-host genome recombination has a considerable effect on viral genome evolution leading to emergence of new chimeric variants. As a result, the kinship between viruses that have a common ancestor but have diverged in a more distant past often cannot be deduced from the direct comparison of their genomic or protein sequences. However,

deep evolutionary relationships between seemingly unrelated viruses are of great interest.

The curiosity of virologists on this matter was at least in part satisfied by conclusions drawn from the accumulating body of structural data on both individual viral proteins and entire virions. It became apparent that certain viruses infecting phylogenetically distant hosts, which sometimes reside in different domains of life, are remarkably similar structurally despite the lack of recognizable sequence conservation (Bamford et al., 2002). Furthermore, available structural information revealed that there are several such groups of viruses that are structurally similar to each other within the group, but are considerably different between the groups. The similarity within each group was found to stretch from the overall virion organization and principles of virion assembly to the structure of the major virion building blocks and often genome packaging machineries (Bamford et al., 2005a). In order to explain such structural relationships between viruses, infecting hosts as distant as human beings and *E. coli*, a virus lineage hypothesis has been put forward (Bamford et al., 2002). According to this hypothesis, the observed structural similarity between the viruses is a result of a common descent. Consequently, viruses of a certain lineage presently infecting hosts in different domains of life once shared an ancestor which preyed on the last universal cellular ancestor (LUCA); as cellular organisms diversified so did their viruses by co-evolving with their hosts. Existence of several distinct viral lineages suggests independent origins for their corresponding progenitor viruses (Bamford et al., 2002). Structural information available so far allowed identification of four viral lineages that unite viruses infecting hosts in different domains of life; additional viral lineages are likely to be uncovered in the near future (Bamford et al., 2005a; Krupovič and Bamford, 2008). Table 3 presents an overview of the four viral lineages.

Table 3. Viral lineages.

Viral lineage	Bacterial	Archaeal		Eukaryotic
		Cren-	Eury-	
Double β-barrel	*Tecti-, Cortico-*	STIV	–	*Adeno-, Phycodna-, Irido-, Asfar-, Pox-, Asco-,* Mimivirus, Sputnik, Mavericks
Single β-barrel	P23-77	–	SH1	–
HK97-Herpes	*Myo-, Sipho-, Podo-*	–	*Myo-, Sipho-*	*Herpes-*
ϕ6-Reo	*Cysto-*	–	–	*Reo-, Toti-*

–, not determined (yet).

B. AIMS OF THIS STUDY

The prokaryotic viruses were for a long time considered only as useful tools for molecular biology or else as model systems for understanding different biological phenomena relevant to cellular organisms, such as, for instance, membrane biogenesis. Only recently it was appreciated that viruses constitute an integral and, perhaps it will not be an exaggeration to say, the major part of the biosphere (Hendrix, 2002; Suttle, 2007). The studies on the diversity, distribution and evolution of the virosphere are therefore *per se* of great significance and fundamental interest.

The aim of this study was to gain insights into the diversity and evolution of underrepresented (only few isolated representatives of a particular virus group) prokaryotic viruses, both bacterial and archaeal. A special attention was paid to the molecular mechanisms of their genome evolution as well as to the relationships between viruses infecting phylogenetically distant hosts. In order to increase the number of genomic sequences available for comparative analysis, the wealth of genomic data available in public databases was exploited. As a result, a number of proviruses (viral genomes integrated into cellular chromosomes) related to the group of viruses under study were identified and a comparative genomics approach was undertaken. The specific aims of this study were:

1. To gain insights into the diversity and evolution of corticoviruses (I). Until this study the knowledge on the family *Corticoviridae* was derived exclusively from the analyses carried out on the sole isolated member of this viral family.

2. To identify new possible members of the double β-barrel viral lineage in the domain Archaea and to improve understanding on the diversity of these viruses (II).

3. To enrich the genomic pool of archaeal viruses related to tailed dsDNA bacteriophages of the order Caudovirales; to gain insights into their phylogenetic distribution; and finally, to verify the genomic-level relationship between these bacterial and archaeal viruses (III).

C. METHODS

All genomic sequences described in this study were downloaded from the GenBank at the NCBI (National Center for Biotechnology Information) and analyzed using various bioinformatic tools listed in Table 4.

Table 4. Bioinformatic tools used in this study.

Program	Website for application/download	Reference
Genomic context analysis		
Vector NTI Suite v8.0	http://www.invitrogen.com/	Invitrogen, Inc.
CLC Main Workbench v5.0	http://www.clcbio.com/index.php?id=92	CLC Bio, Inc.
Homology detection		
PSI-BLAST	http://blast.ncbi.nlm.nih.gov/	Altschul et al., 1997
HHpred	http://toolkit.lmb.uni-muenchen.de/hhpred	Söding, 2005
FFAS03	http://ffas.ljcrf.edu/	Friedberg et al., 2006
3D-Jury	http://meta.bioinfo.pl/	Ginalski et al., 2003
Sequence alignments		
PROMALS	http://prodata.swmed.edu/promals/	Pei et al., 2007
CLUSTALW	http://www.ebi.ac.uk/clustalw/	Thompson et al., 1994
Protein domain analysis		
TMHMM	http://www.cbs.dtu.dk/services/TMHMM-2.0/	Krogh et al., 2001
InterProScan	http://www.ebi.ac.uk/InterProScan/	Zdobnov and Apweiler, 2001
IS sequence analysis		
ISFinder database	www-is.biotoul.fr/	Siguier et al., 2006
Structural modeling		
MODELLER v9.6	http://www.salilab.org/modeller/	Marti-Renom et al., 2000
Model quality verification		
MolProbity	http://molprobity.biochem.duke.edu/	Lovell et al., 2003
ProSA-web	http://prosa.services.came.sbg.ac.at/prosa.php	Wiederstein and Sippl, 2007
Visualization of structures		
VMD v1.8.7	http://www.ks.uiuc.edu/research/vmd/	Humphrey et al., 1996

D. RESULTS AND DISCUSSION

1. Identification of proviruses

Cellular organisms usually harbor at least one provirus per genome (Casjens, 2003). Therefore, to identify novel viral genomes and consequently to enrich the genomic dataset for the comparative analysis, we took advantage of the ample (and constantly increasing) genomic information available for diverse bacterial and archaeal organisms by scanning them for the presence of certain proviruses.

It became apparent that horizontal exchange of genes between diverse viruses as well as between viruses and their hosts is a prevalent phenomenon shaping the genomes of both parties and therefore has a profound effect on their evolution (Hendrix, 2002). However, according to the viral lineage hypothesis (see above), genes encoding virion structural components as well as the genome packaging machinery are inherited vertically within a given lineage of viruses (Bamford et al., 2005a; Krupovič and Bamford, 2008). Accordingly, horizontal exchange of genetic information between different viral lineages is expected to be not equally unrestricted for functionally different genes. Therefore, to identify proviruses related to a certain viral group we used as seeds sequences of proteins encoded by genes that were most likely to be inherited vertically within that particular group/lineage of viruses. In the first part of this study (I) we attempted to identify relatives of bacterial corticovirus PM2. Therefore, the sequence of PM2 major capsid protein (MCP) was used as a seed in iterative database searches. In the second study (II), the goal was to identify members of the double β-barrel viral lineage associated with the domain Archaea. To achieve that, archaeal genomes were searched for the presence of genes encoding homologues of the MCP of crenarchaeal virus STIV, at that time the only known virus containing a double β-barrel capsid protein and infecting archaea (Khayat et al., 2005). In the third study (III), the proviruses related to the tailed bacterial and archaeal dsDNA viruses of the order Caudovirales were identified in archaeal genomes by searching for homologues of the large subunit of the terminase (TerL), an ATPase powering the viral genome packaging into preformed empty procapsids in members of the Caudovirales and *Herpesviridae* (Rao and Feiss, 2008). Notably, double β-barrel capsid proteins as well as TerL ATPases are strictly virus-specific and do not have close homologues in the cellular proteome (Koonin et al., 2006).

Once all the homologues of the selected viral proteins encoded in bacterial (I) and archaeal (II, III) genomes were identified, the genomic context for each positively identified homologous gene was thoroughly analyzed. As a result, we identified thirteen putative proviruses related to corticovirus PM2 (I), two proviruses encoding putative double β-barrel MCPs and integrated into euryarchaeal genomes (II), and nine proviruses related to members of the Caudovirales in various archaeal genomes (III).

It should be noted that in these studies we did not consider the functionality of the identified proviruses. The

presence of the proviruses in bacterial and archaeal genomes *per se* was taken as a sufficient evidence of these cellular organisms having been in contact with the corresponding viruses. Consequently, the evolutionary insights gained from comparative genomic analysis of the identified proviruses were assumed to reflect the evolution of their free-living relatives.

2. Corticoviral elements

2.1 Comparative analysis of corticoviral genomes

Bacterial virus PM2 is currently the sole member assigned to the family *Corticoviridae* (Bamford and Bamford, 2006). Its overall virion morphology resembles that of tectiviruses (Table 1), except that the genome of PM2 is highly supercoiled circular dsDNA molecule of ~10 kb (Gray et al., 1971; Männistö et al., 1999). Recently, structural analysis revealed that the MCP and the penton protein of PM2 have the same general topology as the corresponding proteins of tectivirus PRD1 (Abrescia et al., 2008). In addition, the two viruses encode homologous packaging ATPases (Strömsten et al., 2005) and hence belong to the same structure-based viral lineage (Krupovič and Bamford, 2008). The PM2 genome is replicated in proximity of the cytoplasmic membrane via a rolling-circle (RCR) mechanism initiated by the phage-encoded replication initiation protein P12 (Brewer, 1978; Männistö et al., 1999). The PM2 genome is organized into three operons (two early and one late) regulated by phage-encoded transcription factors (Männistö et al., 2003). Interestingly, the organization of the two early PM2 operons was found to be similar to those required for the maintenance of the *Pseudoalteromonas* plasmid pAS28 (Fig. 3; Männistö et al., 2003). One of these two operons shares significant sequence similarity with the corresponding region in pAS28 and was therefore suggested to represent a recent horizontal acquisition in PM2 genome from a pAS28-like plasmid (Männistö et al., 2003). Notably, the replication proteins in PM2 and pAS28 are nonhomologous, although they occupy equivalent positions (Fig. 3).

PM2 infects marine *Pseudoalteromonas* species (Kivelä et al., 1999) and is strictly virulent under laboratory conditions; at the end of infection cycle the host cells are disrupted via mechanism which differs considerably from those described for other dsDNA bacteriophages, including tectivirus PRD1 (Krupovič et al., 2007; Krupovič et al., 2008). Therefore, identification of proviruses related to PM2 was somewhat unexpected (I). All PM2 genes can be categorized into four functionally distinct groups or modules encoding (i) proteins responsible for structural components of the virion (proteins P1–P10), (ii) regulation of transcription (P13–P16), (iii) genome replication (RCR initiation protein P12), (iv) proteins involved in cell lysis (P17, P18). Figure 3 shows the genomic alignment of a selection of PM2-like proviruses identified in Article I. Comparative genomics of the thirteen identified proviruses revealed that genes encoding the MCP (gene *II*), structural protein P7 (gene *VII*), and the putative genome packaging ATPase (gene *IX*) were invariably present in all proviral genomes. In contrast, genes of the other three modules (replication, regulation

of transcription, host lysis) were found to be frequently substituted with functionally equivalent but nonhomologous genes. For example, some PM2-like proviruses were found to encode typical endolysin genes, homologous to those encoded by numerous tailed dsDNA bacteriophages. PM2 genome, on the other hand, does not carry a gene for an apparent endolysin homologue (Krupovič et al., 2007). Similarly, some corticoviral elements code for transcriptional regulators homologous to those of PM2 (Männistö et al., 2003), while others encode clearly distinct transcriptional factors. For example, a corticoelement of *Vibrio splendidus* instead of PM2-like transcriptional regulators encodes a homologue of a typical CI repressor (Fig. 3). Most surprising, however, was the observed diversity of replication proteins encoded by different PM2-like proviruses. Corticoviral elements apparently can rely on at least four distinct strategies for genome replication. While majority of identified provi-

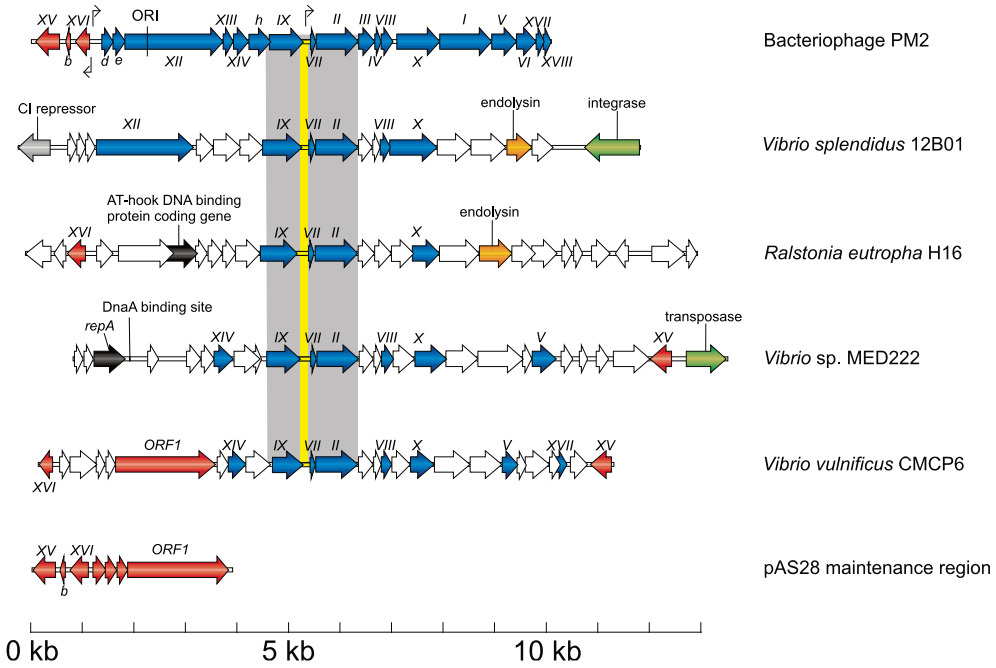

Figure 3. Genetic organization of the selected corticoviral elements, bacteriophage PM2, and the maintenance region of the pAS28 plasmid. Genes encoding structural and nonstructural proteins (Roman numerals) and ORFs (lower case letters) are visualized as arrows indicating the direction of transcription. The origin of replication (ORI) and the three promoters (kinked arrows) are indicated. Genes in the PM2 genome, as well as in corticoviral elements related to those of the pAS28 plasmid, are coloured red. The viral "self" core, containing genes *IX* (putative genome packaging ATPase-coding gene), *VII*, and *II* (major capsid protein-coding gene) is shaded grey. The late promoter region of phage PM2 and the noncoding region matching its position in all putative proviruses is coloured yellow. Recombinase- and endolysis-coding genes are coloured green and orange, respectively. Genes encoding replication-associated proteins (other than XII- and ORF1-type) are shown in black.

ruses encode PM2-like RCR initiation proteins, some were found to contain genes for replication proteins homologous to those encoded by different plasmids (ORF1 of *Pseudoalteromonas* plasmid pAS28, RepA of *Pseudomonas* plasmid pRA2, and C1 of *Ralstonia* plasmid pJTPS1; Fig. 3). This observation indicates that the history of replication genes can be (and often is) independent from that of genes involved in other viral functions (Krupovič and Bamford, 2009).

Conservation of genes encoding structural virion components, and especially the MCP and the genome packaging ATPase, in PM2-like proviruses suggests that once (and if ever) induced they are likely to produce virions resembling those of PM2. Consequently, of the four functional modules constituting corticoviral genomes only the one for virion structure seems to be maintained vertically, while the other three are prone to horizontal flux. This observation supports the structure-based viral lineage hypothesis (Bamford et al., 2005a) and suggests that those who seek to classify viruses should perhaps primarily concentrate on the virus-encoded features defining the virion structure and only then on those involved in virus-host interactions and genome replication.

2.2 Abundance and distribution of corticoviruses

Thirteen proviruses identified in this Study (Article I) were integrated into the genomes of eleven bacteria belonging to the phylum *Proteobacteria* (classes β and γ). Notably, all the organisms that harbor PM2-like proviruses were found to be residents of aquatic habitats, either marine or fresh-water. Previous transmission electron microscopy (TEM) studies indicate that tailed dsDNA bacteriophages represent the predominant morphotype in marine environments (Wommack and Colwell, 2000). It was therefore interesting to compare the relative abundance of PM2-like versus tailed phage-like proviruses in the genomes of aquatic bacteria (at the time of the study there were 269 aquatic bacterial genomes in the GenBank). Unexpectedly, proviruses related to marine tailed bacteriophages were not identified more frequently than PM2-like proviruses, when the same identification strategy was employed (i.e., searches against aquatic bacterial genomes using the MCP and TerL sequences of tailed viruses as seeds). Although this observation is not a definitive prove of PM2-like viruses being as plentiful as tailed bacteriophages in the aquatic environment, it suggests that their abundance as well as the impact on the ecology of aquatic microorganisms might be underestimated.

During the last three years, which passed since the publication of Article I, a number of new bacterial genomes has been sequenced. A quick analysis of this new dataset revealed that the number of PM2-like proviruses has almost tripled; at the moment, nearly forty corticoelements integrated into diverse bacterial genomes can be identified (our unpublished observation). Most (but not all) of these new elements are found in different *Vibrio* strains and seem to be relatively closely related to the *Vibrio* elements described in Article I. Interestingly, four putative PM2-like proviruses are present in the genome of *Opitutaceae* bacterium TAV2 (phylum *Verru-*

comicrobia), which was isolated from the gut of a wood-feeding termite (NCBI Genome Project ID: 36791). Notably, *Verrucomicrobia* are phylogenetically distant from *Proteobacteria*. Another peculiar bacterium harboring in its genome a corticoelement (our unpublished observation) is *Teredinibacter turnerae* T7901, an intracellular endosymbiont of marine wood-boring shipworms (Yang et al., 2009). In addition, metagenomic study of methylamine microcosms revealed presence of free-living PM2-like viruses that are most closely related to the proviruses of *Methylobacillus flagellatus* KT (Article I) and were suggested to infect *Methylotenera mobilis* (Kalyuzhnaya et al., 2008). These observations suggest that the host range of corticoviruses might not be restricted to aquatic microorganisms and is possibly reaching way beyond proteobacterial hosts.

3. Double β-barrel viruses of Archaea

The most populated viral lineage described to date is the double β-barrel lineage (Table 3; Krupovič and Bamford, 2008). The majority of viruses in this lineage possess icosahedral protein capsids with an internal membrane enclosing dsDNA genomes. The capsids are built of pseudohexameric capsomers formed of the trimeric major capsid protein displaying the characteristic double β-barrel topology (Fig. 4; Krupovič and Bamford, 2008). In addition, all members of the lineage, except adenoviruses, encode homologous genome packaging ATPases of the FtsK/HerA superfamily (Strömsten et al., 2005). The lineage contains a number of members infecting *Bacteria* and *Eukarya*, but only a single representative, *Sulfolobus* turreted icosahedral virus (STIV), in the domain *Archaea* (Table 3). STIV infects a crenarchaeal host (Rice et al., 2004). In order to see whether the double β-barrel lineage could be extended to the second major archaeal phylum, *Euryarchaeota*, we searched the available archaeal genomes for the presence of putative proviruses related to STIV and other members of this viral lineage (Article II). We identified two homologues of the STIV MCP encoded in the genomes of *Thermococcus kodakarensis* KOD1 (order Thermococcales) and *Methanococcus voltae* A3 (order Methanococcales), both belonging to the phylum *Euryarchaeota*. Thorough inspection of the genomic context surrounding these two genes revealed that they were embedded within putative proviruses, which were called TKV4 and MVV, respectively. In both cases, the second hallmark gene of the double β-barrel lineage viruses, coding for the putative genome packaging ATPase, was found close to the MCP-coding gene. The two proviruses have apparently used tRNA genes for integration; whereas MVV has utilized the 3'-distal region of a tRNA gene as the attachment site, TKV4 has recombined with the 5'-distal part of the corresponding gene of *T. kodakarensis* KOD1. The tRNA genes often serve as integration targets for prokaryotic viruses (Reiter et al., 1989). However, recombination with the 5'-distal regions of these genes has not been previously reported.

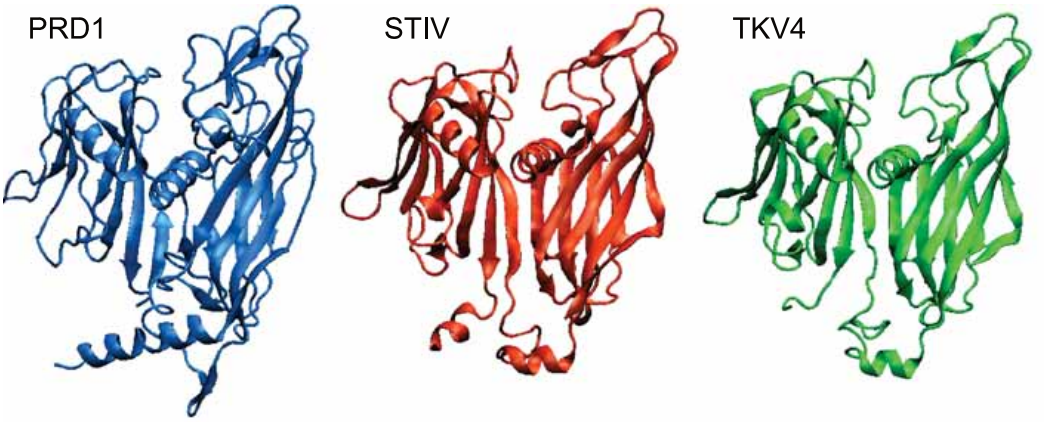

Figure 4. Double β-barrel major capsid proteins of tectivirus PRD1 (blue; Benson et al., 1999), *Sulfolobus* turreted icosahedral virus (STIV; red; Khayat et al., 2005), and the putative euryarchaeal provirus TKV4 (green; Article II).

Analysis of the genetic content revealed that STIV, TKV4, and MVV are only distantly related to each other and do not share other genes than those for the MCPs and putative packaging ATPases. Like in the case of corticoelements, archaeal double β-barrel (pro)viruses were found to encode different types of transcriptional regulators and replication proteins. For example, MVV encodes a typical RCR initiation protein, while TKV4 genome replication apparently relies on the MCM helicase. The replication genes are, however, at equivalent positions within the two genomes – upstream of the ATPase/MCP-coding genes. Such gene order (replication–ATPase–MCP) is common to all prokaryote-infecting members of the double β-barrel lineage (Krupovič and Bamford, 2008). Interestingly, when a profile created from the alignment of STIV and TKV4 MCP sequences was used during the PSI-BLAST analysis, the MCP of GIL16, a tectivirus infecting *Bacillus* sp. (Verheust et al., 2005), was found among the low-scoring hits. This result could not be obtained with the STIV sequence alone, indicating that the MCP sequence of TKV4 is a bridge between archaeal and bacterial double β-barrel virus MCPs. This suggests that one day we might have a sufficient number of MCP sequences to relate distant members of this viral lineage through sequence-sequence comparisons without the necessity to rely on the structural information.

As a bonus in Article II was the identification of another provirus, which was adjacent to MVV and apparently utilized the same attachment site. Notably, it has been shown previously that *M. voltae* A3 produce oblate virus-like particles (A3 VLP) containing ~23 kb circular dsDNA (Table 2; Wood et al., 1989). The same study has also revealed that the genome of that VLP is present in an integrated form on the cellular chromosome. Although the nucleotide sequence of A3 VLP DNA has not been determined, the restriction map was created (Wood et al., 1989). Surprisingly, the *in silico* generated restriction pattern of the MVV neighbor is remarkably similar to the pattern experimentally obtained for A3 VLP. Therefore, based on the very similar genome

length and nearly identical restriction pattern we concluded that the provirus located next to MVV in the *M. voltae* A3 genome represents the integrated form of the previously isolated A3 VLP.

Identification of proviruses TKV4 and MVV extends the double β-barrel viral lineage to the phylum *Euryarchaeota*. Genomic distinctiveness of the archaeal double β-barrel (pro)viruses implies that their association with archaeal hosts is ancient, likely predating the divergence of the archaeal domain into *Crenachaeota* and *Euryarchaeota*.

4. Comparative genomics of tailed dsDNA viruses of Archaea

Tailed dsDNA bacteriophages of the order Caudovirales and eukaryotic herpesviruses form another structure-based lineage (Baker et al., 2005; Bamford et al., 2005a). The similarity between tailed bacteriophages and herpesviruses was brought to light by extensive functional and structural studies on the two viral systems. It extends from basic principles of virion assembly and maturation to the topology of the major capsid proteins as well as to similar genome packaging machineries composed of homologous proteins (Dokland, 1999; Baker et al., 2005; Steven et al., 2005b; Duda et al., 2006; Rao and Feiss, 2008). Notably, all these features are very different from those characteristic to viruses in the double β-barrel lineage as should be expected if the progenitor viruses of the two lineages have originated independently (Bamford et al., 2005a). As shown in Tables 2 and 3, tailed dsDNA viruses resembling bacterial viruses of the order Caudovirales also infect euryarchaeal hosts. However, detailed structural, functional and sequence information for these viruses is not available. Thus their origin, mechanisms of genome evolution as well as relationship to tailed bacteriophages and herpesviruses remains obscure. To shed light on these issues, we undertook the comparative genomic approach (Article III).

Contrasting the wealth of genomic information available for tailed viruses infecting bacteria (448 viral and many more proviral genomes), until Article III the number of available genomes for tailed archaeal viruses was limited to one proviral and five viral genomes (Prangishvili et al., 2006b). Notably, head-and-tail viruses seem to be restricted to the phylum Euryarchaeota (Prangishvili et al., 2006a), where they were isolated from hosts falling into just two out of eight classes (*class* is a taxonomic rank fitting between *phylum* and *order*), namely, *Halobacteria* and *Methanobacteria*. To enrich the genomic data pool and to gain insights into the distribution of head-and-tail viruses in Archaea, we scanned the available archaeal genomes for the presence of proviruses related to tailed viruses of bacteria and archaea. As a result, nine "tailed" proviruses were identified. This allowed for an exhaustive comparative genomic analysis of this viral group.

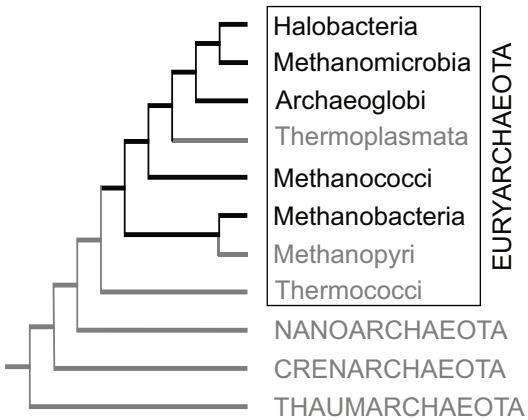

Figure 5. Distribution of tailed dsDNA viruses in the domain *Archaea*. Schematic representation of a consensus archaeal phylogeny based on recent phylogenomic analyses (Brochier-Armanet et al., 2008). Archaeal phyla are written in capital letters. The eight recognized taxonomic classes of Euryarchaeota are boxed. Tree branches corresponding to the euryarchaeal classes that are known to be infected by tailed viruses or contain related proviruses are shown in black. The phylogeny was kindly provided by Dr. Simonetta Gribaldo.

Host-range of tailed archaeal viruses. Our analysis revealed that the host-range of head-and-tail viruses is wider than previously suspected (Fig. 5). We identified "tailed" proviruses in the genomes of archaea belonging to classes *Halobacteria*, *Methanobacteria*, *Methanococci*, *Methanomicrobia*, and *Archaeoglobi* (the genome of *Archaeoglobus profundus* DSM 5631 was deposited into the GenBank only after the publication of Article III; therefore, the associated provirus Apro-Pro1 has not been described in Article III). Consequently, organisms falling into five out of eight classes of *Euryarchaeota* (Fig. 5) are (or were) under attack of head-and-tail viruses. Furthermore, based on the sequence similarity and gene content, these archaeal (pro)viruses could be divided into distinct groups that follow the taxonomic grouping of their archaeal host species. This observation strongly suggests a long-lasting co-evolution of "tailed" (pro)viruses with their hosts.

Genome organization and mechanisms of evolution. Comparative analysis revealed a modular organization of the head-and-tail archaeal (pro)virus genomes, very similar to that found in their bacterial relatives (Brüssow and Desiere, 2001). Most evident was the clustering of genes encoding proteins for virion morphogenesis and structure and also for genome packaging. As mentioned above, based on the gene content tailed archaeal (pro)viruses could be clustered into distinct groups. It became apparent that even within a group of related viruses the position of these functional modules with respect to each other may vary considerably, likely due to intergenome rearrangement events.

Comparative analysis exposed mosaic character of viral genomes within different groups, with a number of gene substitution events. Some genes are replaced with homologues from distantly related viruses. For example, genes for genome packaging proteins and capsid formation are clearly homologous and collinear in the two methanobacterial proviruses (Article I). However, the downstream lying genes for tail tape measure proteins have their closest counterparts in viruses outside the methanobacterial virus group. Other genes are substituted with nonhomologous, but functionally equivalent genes. This is the case for replication protein-coding genes of some halobacterial and methanococcal proviruses in which MCM helicase-coding genes

seem to be replaced with genes for distinct replication proteins. Illegitimate recombination is the likely cause of such replacements. In addition, obvious cases of horizontal gene acquisition from the host were detected; in one case, fusion of the incoming cellular gene to the preexisting viral gene led to the emergence of a novel chimeric unit. Previous studies have also presented evidence for the homologous recombination between closely related tailed viruses of archaeal (Tang et al., 2004).

Results presented in Article III, in combination with those obtained by others, uncovered the basic mechanisms of genome evolution of tailed archaeal viruses. It therefore became apparent that diversification, homologous and illegitimate recombination, and intergenome rearrangements are the major mechanisms of evolution of not only tailed bacteriophages (Hendrix et al., 1999; Casjens, 2005), but their archaeal counterparts as well.

Relationship to tailed dsDNA viruses of bacteria and eukaryotic herpesviruses. As mentioned above, tailed bacteriophages and herpesviruses encode common determinants for virion assembly, maturation and structure as well as genome packaging. Comparative genome analysis revealed a set of conserved genes in tailed archaeal (pro)viruses. This set among other genes includes those for the major capsid protein, prohead protease, portal and the terminase complex. The very same set of genes underlies the structural similarity between tailed bacteriophages and herpesviruses. Notably, products of these genes in the archaeal viruses were found to share significant sequence similarity with the corresponding proteins of tailed bacteriophages. In addition, fold prediction and structural modeling experiments suggest that the major capsid proteins of tailed archaeal viruses from all four groups adopt the same topology as the equivalent proteins of tailed bacterial viruses and herpesviruses. The molecular principles of virion assembly and maturation as well as virion structure of tailed archaeal viruses are thus expected to be very similar to those of tailed bacterial viruses and eukaryotic herpesviruses, suggesting a common ancestry of these viruses.

E. CONCLUDING REMARKS

Comparative genomics of tailed dsDNA bacteriophages has yielded valuable information on their evolution and genetic diversity (Hendrix et al., 1999; Hatfull et al., 2010). Unfortunately, genomic data is not equally abundant for prokaryotic viruses belonging to other families. Neither there was much effort devoted to isolation of new members of these scarcely populated viral families. For example, family *Corticoviridae* still includes a single member, which was isolated more than forty years ago (Espejo and Canelo, 1968).

In this study we sought to gain insights into the diversity and evolution of such underrepresented prokaryotic viruses. To live in the post-genomic era and not to exploit the wealth of constantly increasing database of complete bacterial and archaeal genomes would be inexcusable. Indeed, targeted provirus searches (i.e., for genomes of a certain viral group) proved to be very useful; not only they allowed for obtaining genomic information on specific viral groups, but also provided more insights into the host range and phylogenetic distribution of these viruses. As a result, a number of PM2-like proviruses was identified, putting to the end the forty years of solitude for PM2 (Article I). Similarly, proviruses encoding the double β-barrel capsid proteins were found not to be restricted to crenarchaeal hosts but to be associated with euryarchaeal hosts as well (Article II). In addition, the host range of tailed archaeal viruses was found to be considerably wider than previously suspected, covering the majority of taxonomic classes of *Euryarchaeota* (Article III).

Comparative genomic analysis of the identified proviruses and free-living related viruses was very useful for understanding their evolution. Viral genomes consist of several functional modules (structure, replication, gene expression, host lysis, etc.) and their evolution proceeds by shuffling of these modules via illegitimate recombination with other (often evolutionary unrelated) viruses, plasmids, or even host chromosome. Such a recombination within a given viral group (or lineage) seems to be fairly unrestricted, i.e. virtually any gene, or in the case of genes for interacting proteins blocks of genes, can undergo lateral transfer from one virus to another (Hendrix, 2009). However, our analysis shows that such horizontal exchange between distinct viral lineages is not equally unobstructed for different types of genes/modules. Comparative genomics of corticoelements provides a good illustration of this point. Three out of four functional modules comprising the genomes of corticoviruses are frequently exchanged for functionally equivalent but nonhomologous units from other viruses and plasmids. Most notable is the freelance nature of genes encoding replication proteins. Despite the fact that these genes are usually found among the most conserved ones at the sequence level, their evolution seems to occur independently from that of other viral functions. The only determinants that are consistently retained within the group of PM2-like viruses are those responsible for the virion construction. A similar mode of genome evolution is characteristic also to viruses from other families.

These observations might have an effect on the way we see and define viruses. Recently, the “network of life” view became very popular in evolutionary biology for both cellular organisms

(Doolittle and Bapteste, 2007; Koonin and Wolf, 2008) and viruses (Koonin et al., 2009). The proponents of this view argue that the evolution of biological entities occurs in a network-like fashion and that classical tree-like evolutionary pattern is irrelevant due to the rampant horizontal gene transfer (HGT). However, as rightly pointed out by Gribaldo and Brochier (2009), “HGT events represent only one aspect of genome evolution and do not deny the history of organisms [viruses], but rather, are understandable only in the framework of such a history”. Our results indicate that despite the extensive shuffling of functional modules responsible for various genetic processes and interaction with the host, a set of genes defining the structural character of a virion is restricted to and preserved in a given virus lineage and therefore can be used to dissect deep vertical relationships between evolutionary distant viruses.

F. ACKNOWLEDGEMENTS

This work was carried out in the Finnish Academy Centre of Excellence in Virus Research (2006-2011), Institute of Biotechnology and Department of Biosciences, at the University of Helsinki, under the supervision of Professor Dennis Bamford.

I am immensely grateful to Dennis for many things, especially for the opportunity to work in his group and for all the support in social and scientific matters. Dennis, thank you for being a 'fair guy' and for your (generally) liberal attitude, which allowed me to pursue my personal research interests under your skillful mentorship.

Docent Alexander Plyusnin and Dr. David Prangishvili are thanked for carefully reviewing the thesis manuscript and for their valuable comments. Docents Benita Westerlund-Wikström and Tero Ahola are thanked for their support and helpful advice on my work as members of my follow-up group.

The Viikki Graduate School in Molecular Biosciences (VGSB) is thanked for the financial support provided in the course of my PhD training period as well as for the opportunity to attend a number of useful courses.

I am indebted to my former supervisor Professor Rimantas Daugelavičius for introducing me to the virus world and generously sharing his expertise in electrochemistry.

I am extremely grateful to Professor Jaana Bamford for showing me where and how to run a "BLAST" search and to Dr. Janne Ravantti for his scintillating optimism and encouraging discussions about the value of bioinformatics and computational biology.

I want to thank my collaborators from all the projects I had a pleasure to participate in.

I express my gratitude to all the former and present members of the Bamford laboratory, especially those from the "small" office. Elina, I am extremely grateful for all the advices and help I got from you throughout these years as well as for all the discussions, which were very enjoyable.

My coffee-breaks in the lab would not be half as pleasant without the lively conversations I had with Virginija Cvirkaitė-Krupovič, Greta Daujotaitė Violeta Manole, Simonas Laurinavičius, and Andrius Buivydas. I also enjoyed tremendously the company of Andrius and Miao Jiang on the squash court. Matti Jalasvuori is thanked for all the wonderful discussions we had as well as for his courageous behaviour when we got lost on a golf course somewhere in Scotland. I admire your vivid imagination and cool attitude. I want to thank all the afore mentioned people as well as my friends from the Vilnius University, especially Petras Gagilas, Gintarė Kubilickaitė and Eglė Aleksaitė, for their support and friendship, which are of immense value to me.

My warmest appreciation goes to my mother, Ala Krupovič, and my wonderful parents in law. Ma, I would have never become what I am without you.

Most of all, I am grateful to my colaborator, my best friend, my wife, Virginija, and our wonderful son Matas. Thank you for your angelic patience, versatile support and limitless love. I love you both.

Mart Krupovič
Helsinki, May 2010

G. REFERENCES

Abedon ST. 2006. Phage Ecology. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 37-46.

Abrescia NG, Cockburn JJ, Grimes JM, et al. 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature*. 432:68-74.

Abrescia NG, Grimes JM, Kivelä HM, Assenberg R, Sutton GC, Butcher SJ, Bamford JK, Bamford DH, Stuart DI. 2008. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Mol Cell*. 31:749-761.

Ackermann HW. 2007. 5500 Phages examined in the electron microscope. *Arch Virol*. 152:227-243.

Agirrezabala X, Velazquez-Muriel JA, Gomez-Puertas P, Scheres SH, Carazo JM, Carrascosa JL. 2007. Quasi-atomic model of bacteriophage T7 procapsid shell: insights into the structure and evolution of a basic fold. *Structure*. 15:461-472.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389-3402.

Arnold HP, She Q, Phan H, Stedman K, Prangishvili D, Holz I, Kristjansson JK, Garrett R, Zillig W. 1999. The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol Microbiol*. 34:217-226.

Baker ML, Jiang W, Rixon FJ, Chiu W. 2005. Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol*. 79:14967-14970.

Bamford DH, Bamford JK. 2006. Lipid-containing bacteriophage PM2, the type-organism of *Corticoviridae*. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 171-175.

Bamford DH, Burnett RM, Stuart DI. 2002. Evolution of viral structure. *Theor Popul Biol*. 61:461-470.

Bamford DH, Grimes JM, Stuart DI. 2005a. What does structure tell us about virus evolution? *Curr Opin Struct Biol*. 15:655-663.

Bamford DH, Ravantti JJ, Ronnholm G, Laurinavičius S, Kukkaro P, Dyall-Smith M, Somerharju P, Kalkkinen N, Bamford JK. 2005b. Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon *Haloarcula hispanica*. *J Virol*. 79:9097-9107.

Bath C, Cukalac T, Porter K, Dyall-Smith ML. 2006. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, *Salterprovirus*. *Virology*. 350:228-239.

Bath C, Dyall-Smith ML. 1998. His1, an archaeal virus of the *Fuselloviridae* family that infects *Haloarcula hispanica*. *J Virol*. 72:9392-9395.

Beijerinck MW. 1898. Concerning a contagium vivum fluidum as a cause of the spot-disease of tobacco leaves. *Verh Akad Wetensch*, *Amsterdam, II*. 6:3-21.

Benson SD, Bamford JK, Bamford DH, Burnett RM. 1999. Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell*. 98:825-833.

Bergh O, Borsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature*. 340:467-468.

Bernhardt TG, Wang IN, Struck DK, Young R. 2001. A protein antibiotic in the phage Qbeta virion: diversity in lysis targets. *Science*. 292:2326-2329.

Betancourt AJ. 2009. Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. *Genetics*. 181:1535-1544.

Bize A, Karlsson EA, Ekefjard K, Quax TE, Pina M, Prevost MC, Forterre P, Tenaillon O, Bernander R, Prangishvili D. 2009. A unique virus release mechanism in the Archaea. *Proc Natl Acad Sci U S A*. 106:11306-11311.

Blakely GW. 2004. Smarter than the average phage. *Mol Microbiol*. 54:851-854.

Böttcher B, Unseld S, Ceulemans H, Russell RB, Jeske H. 2004. Geminate structures of African cassava mosaic virus. *J Virol*. 78:6758-6765.

Brentlinger KL, Hafenstein S, Novak CR, Fane BA, Borgon R, McKenna R, Agbandje-McKenna M. 2002. *Microviridae*, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium *Bdellovibrio bacteriovorus*. *J Bacteriol*. 184:1089-1094.

Brewer GJ. 1978. Membrane-localized replication of bacteriophage PM2. *Virology*. 84:242-245.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the

Thaumarchaeota. *Nat Rev Microbiol*. 6:245-252.

Brumfield SK, Ortmann AC, Ruigrok V, Suci P, Douglas T, Young MJ. 2009. Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus *Sulfolobus* turreted icosahedral virus. *J Virol*. 83:5964-5970.

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

Brüssow H, Hendrix RW. 2002. Phage genomics: small is beautiful. *Cell*. 108:13-16.

Burroughs AM, Iyer LM, Aravind L. 2007. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn*. 3:48-65.

Campbell A. 2006. General aspects of lysogeny. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 66-73.

Casado CG, Javier Ortiz G, Padron E, Bean SJ, McKenna R, Agbandje-McKenna M, Boulton MI. 2004. Isolation and characterization of subgenomic DNAs encapsidated in "single" T = 1 isometric particles of Maize streak virus. *Virology*. 323:164-171.

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

Casjens SR. 2005. Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol*. 8:451-458.

Caspar DL, Klug A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*. 27:1-24.

Chetverin AB, Kopein DS, Chetverina HV, Demidenko AA, Ugarov VI. 2005. Viral RNA-directed RNA polymerases use diverse mechanisms to promote recombination between RNA molecules. *J Biol Chem*. 280:8748-8755.

Condit RC, Moussatche N, Traktman P. 2006. In a nutshell: structure and assembly of the Vaccinia virion. *Adv Virus Res*. 66:31-124.

Crick FH, Watson JD. 1956. Structure of small viruses. *Nature*. 177:473-475.

Davis BM, Lawson EH, Sandkvist M, Ali A, Sozhamannan S, Waldor MK. 2000. Convergence of the secretory pathways for cholera toxin and the filamentous phage, CTXphi. *Science*. 288:333-335.

Dokland T. 1999. Scaffolding proteins and their role in viral assembly. *Cell Mol Life Sci*. 56:580-603.

Doolittle WF, Bapteste E. 2007. Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci U S A*. 104:2043-2049.

Duda RL, Hendrix RW, Huang WM, Conway JF. 2006. Shared architecture of bacteriophage SPO1 and herpesvirus capsids. *Curr Biol*. 16:R11-13.

Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*. 9:267-276.

Dybvig K, Nowak JA, Sladek TL, Maniloff J. 1985. Identification of an enveloped phage, mycoplasma virus L172, that contains a 14-kilobase single-stranded DNA genome. *J Virol*. 53:384-390.

Effantin G, Boulanger P, Neumann E, Letellier L, Conway JF. 2006. Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. *J Mol Biol*. 361:993-1002.

Espejo RT, Canelo ES. 1968. Properties of bacteriophage PM2: a lipid-containing bacterial virus. *Virology*. 34:738-747.

Fane BA, Brentlinger KL, Burch AD, Hafenstein SL, Moore E, Novak CR, Uchiyama A. 2006. ϕX174 et al., the *Microviridae*. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 129-145.

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA. 2005. VIIIth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier–Academic Press.

Federici BA, Bideshi DK, Tan Y, Spears T, Bigot Y. 2009. Ascoviruses: superb manipulators of apoptosis for viral replication and transmission. *Curr Top Microbiol Immunol*. 328:171-196.

Filée J, Chandler M. 2008. Convergent mechanisms of genome evolution of large and giant DNA viruses. *Res Microbiol*. 159:325-331.

Fokine A, Leiman PG, Shneider MM, Ahvazi B, Boeshans KM, Steven AC, Black LW, Mesyanzhinov VV, Rossmann MG. 2005. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proc Natl Acad Sci U S A*. 102:7163-7168.

Friedberg I, Jambon M, Godzik A. 2006. New avenues in protein function prediction. *Protein Sci*. 15:1527-1529.

Friedman SD, Genthner FJ, Gentry J, Sobsey MD, Vinje J. 2009. Gene mapping and phylogenetic analysis of the complete genome from 30 single-stranded RNA male-specific coliphages (family Leviviridae). *J Virol*. 83:11233-11243.

Gaidelytė A, Jaatinen ST, Daugelavičius R, Bamford JK, Bamford DH. 2005. The linear double-stranded DNA of phage Bam35

enters lysogenic host cells, but the late phage functions are suppressed. *J Bacteriol*. 187:3521-3527.

Galperin MY, Koonin EV. 2003. Preface. In: MY Galperin, EV Koonin, editors. Frontiers in Computational Genomics. Wymondham: Caister Academic Press p. ix-xi.

Garner SA, Everson JS, Lambden PR, Fane BA, Clarke IN. 2004. Isolation, molecular characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydophila pecorum*. *Virus Genes*. 28:207-214.

Geslin C, Gaillard M, Flament D, Rouault K, Le Romancer M, Prieur D, Erauso G. 2007. Analysis of the first genome of a hyperthermophilic marine virus-like particle, PAV1, isolated from *Pyrococcus abyssi*. *J Bacteriol*. 189:4510-4519.

Geslin C, Le Romancer M, Erauso G, Gaillard M, Perrot G, Prieur D. 2003. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "*Pyrococcus abyssi*". *J Bacteriol*. 185:3888-3894.

Ginalski K, Elofsson A, Fischer D, Rychlewski L. 2003. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*. 19:1015-1018.

Goulet A, Blangy S, Redder P, Prangishvili D, Felisberto-Rodrigues C, Forterre P, Campanacci V, Cambillau C. 2009. *Acidianus* filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage. *Proc Natl Acad Sci U S A*. 106:21155-21160.

Grahn AM, Butcher SJ, Bamford JK, Bamford DH. 2006. PRD1 - dissecting the genome, structure and entry. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 161-170.

Gray HB, Jr., Upholt WB, Vinograd J. 1971. A buoyant method for the determination of the superhelix density of closed circular DNA. *J Mol Biol*. 62:1-19.

Gribaldo S, Brochier C. 2009. Phylogeny of prokaryotes: does it exist and why should we care? *Res Microbiol*. 160:513-521.

Happonen LJ, Redder P, Peng X, Reigstad LJ, Prangishvili D, Butcher SJ. 2010. Hot familial relationships in acid-loving archaeal viruses. *J Virol*.

Hatfull GF. 2006. Mycobacteriophages. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 602-620.

Hatfull GF, Jacobs-Sera D, Lawrence JG, et al. 2010. Comparative genomic analysis of 60 mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *J Mol Biol*. 397:119-143.

Held NL, Whitaker RJ. 2009. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ Microbiol*. 11:457-466.

Hendrix RW. 2002. Bacteriophages: evolution of the majority. *Theor Popul Biol*. 61:471-480.

Hendrix RW. 2003. Bacteriophage genomics. *Curr Opin Microbiol*. 6:506-511.

Hendrix RW. 2009. Jumbo bacteriophages. *Curr Top Microbiol Immunol*. 328:229-240.

Hendrix RW, Casjens SR. 2005. Caudovirales. In: CM Fauquet, MA Mayo, J Maniloff, U Desselberger, LA Ball, editors. VIIIth Report of the International Committee on Taxonomy of Viruses. San Diego: Elsevier-Academic Press. p. 35-42.

Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A*. 96:2192-2197.

Holmes EC. 2010. Evolution in health and medicine Sackler colloquium: The comparative genomics of viral emergence. *Proc Natl Acad Sci U S A*. 107 Suppl 1:1742-1746.

Huber KE, Waldor MK. 2002. Filamentous phage integration requires the host recombinases XerC and XerD. *Nature*. 417:656-659.

Huiskonen JT, Butcher SJ. 2007. Membrane-containing viruses with icosahedrally symmetric capsids. *Curr Opin Struct Biol*. 17:229-236.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph*. 14:33-38.

Inokuchi Y, Takahashi R, Hirose T, Inayama S, Jacobson AB, Hirashima A. 1986. The complete nucleotide sequence of the group II RNA coliphage GA. *J Biochem*. 99:1169-1180.

Ivanowsky D. 1892. Concerning the mosaic disease of the tobacco plant. *St. Petersburg Acad Imp Sci Bull*. 35:67-70.

Jaatinen ST, Happonen LJ, Laurinmaki P, Butcher SJ, Bamford DH. 2008. Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic *Thermus thermophilus*. *Virology*. 379:10-19.

Jalasvuori M, Jaatinen ST, Laurinavičius S, Ahola-Iivarinen E, Kalkkinen N, Bamford DH, Bamford JK. 2009. The closest relatives of icosahedral viruses of thermophilic bacteria are among viruses and plasmids of the halophilic archaea. *J Virol*. 83:9388-9397.

Jiang W, Li Z, Zhang Z, Baker ML, Prevelige PE, Jr., Chiu W. 2003. Coat protein fold and

maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nat Struct Biol*. 10:131-135.

Jones TA, Liljas L. 1984. Structure of satellite tobacco necrosis virus after crystallographic refinement at 2.5 Å resolution. *J Mol Biol*. 177:735-767.

Jovel J, Preiss W, Jeske H. 2007. Characterization of DNA intermediates of an arising geminivirus. *Virus Res*. 130:63-70.

Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol*. 299:27-51.

Kalyuzhnaya MG, Lapidus A, Ivanova N, et al. 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol*. 26:1029-1034.

Kawai M, Uchiyama I, Kobayashi I. 2005. Genome comparison *in silico* in *Neisseria* suggests integration of filamentous bacteriophages by their own transposase. *DNA Res*. 12:389-401.

Kawasaki T, Nagata S, Fujiwara A, Satsuma H, Fujie M, Usami S, Yamada T. 2007. Genomic characterization of the filamentous integrative bacteriophages ϕRSS1 and ϕRSM1, which infect *Ralstonia solanacearum*. *J Bacteriol*. 189:5792-5802.

Khayat R, Tang L, Larson ET, Lawrence CM, Young M, Johnson JE. 2005. Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc Natl Acad Sci U S A*. 102:18944-18949.

Kivelä HM, Männistö RH, Kalkkinen N, Bamford DH. 1999. Purification and protein composition of PM2, the first lipid-containing bacterial virus to be isolated. *Virology*. 262:364-374.

Klovins J, Overbeek GP, van den Worm SH, Ackermann HW, van Duin J. 2002. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J Gen Virol*. 83:1523-1533.

Klug A. 1999. The tobacco mosaic virus particle: structure and assembly. *Philos Trans R Soc Lond B Biol Sci*. 354:531-535.

Koonin EV, Dolja VV. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol*. 28:375-430.

Koonin EV, Senkevich TG, Dolja VV. 2006. The ancient Virus World and evolution of cells. *Biol Direct*. 1:29.

Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res*. 36:6688-6719.

Koonin EV, Wolf YI, Nagasaki K, Dolja VV. 2009. The complexity of the virus world. *Nat Rev Microbiol*.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 305:567-580.

Krupovič M, Bamford DH. 2008. Virus evolution: how far does the double beta-barrel viral lineage extend? *Nat Rev Microbiol*. 6:941-948.

Krupovič M, Bamford DH. 2009. Does the evolution of viral polymerases reflect the origin and evolution of viruses? *Nat Rev Microbiol*. 7:250.

Krupovič M, Cvirkaitė-Krupovič V, Bamford DH. 2008. Identification and functional analysis of the *Rz*/*Rz1*-like accessory lysis genes in the membrane-containing bacteriophage PRD1. *Mol Microbiol*. 68:492-503.

Krupovič M, Daugelavičius R, Bamford DH. 2007. A novel lysis system in PM2, a lipid-containing marine double-stranded DNA bacteriophage. *Mol Microbiol*. 64:1635-1648.

Krupovič M, Ravantti JJ, Bamford DH. 2009. Geminiviruses: a tale of a plasmid becoming a virus. *BMC Evol Biol*. 9:112.

Kwan T, Liu J, DuBow M, Gros P, Pelletier J. 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc Natl Acad Sci U S A*. 102:5174-5179.

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

Laurinmäki PA, Huiskonen JT, Bamford DH, Butcher SJ. 2005. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure*. 13:1819-1828.

Lawrence JG, Hatfull GF, Hendrix RW. 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. *J Bacteriol*. 184:4891-4905.

Lefkowitz EJ, Wang C, Upton C. 2006. Poxviruses: past, present and future. *Virus Res*. 117:105-118.

Licis N, Balklava Z, Van Duin J. 2000. Forced retroevolution of an RNA bacteriophage. *Virology*. 271:298-306.

Lipps G. 2006. Plasmids and viruses of the thermoacidophilic crenarchaeote *Sulfolobus*. *Extremophiles*. 10:17-28.

Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. 2003. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins*. 50:437-450.

Maaty WS, Ortmann AC, Dlakic M, et al. 2006. Characterization of the archaeal thermophile *Sulfolobus* turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. *J Virol*. 80:7625-7635.

Mackinder LC, Worthy CA, Biggi G, Hall M, Ryan KP, Varsani A, Harper GM, Wilson WH, Brownlee C, Schroeder DC. 2009. A unicellular algal virus, *Emiliania huxleyi* virus 86, exploits an animal-like infection strategy. *J Gen Virol*. 90:2306-2316.

Maniloff J, Dybvig K. 2006. Mycoplasma phages. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 636-652.

Männistö RH, Grahn AM, Bamford DH, Bamford JK. 2003. Transcription of bacteriophage PM2 involves phage-encoded regulators of heterologous origin. *J Bacteriol*. 185:3278-3287.

Männistö RH, Kivelä HM, Paulin L, Bamford DH, Bamford JK. 1999. The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. *Virology*. 262:355-363.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 29:291-325.

Marvin DA. 1998. Filamentous phage structure, infection and assembly. *Curr Opin Struct Biol*. 8:150-158.

McLeod SM, Kimsey HH, Davis BM, Waldor MK. 2005. CTXphi and *Vibrio cholerae*: exploring a newly recognized type of phage-host cell relationship. *Mol Microbiol*. 57:347-356.

Melcher U, Sha Y, Ye F, Fletcher J. 1999. Mechanisms of spiroplasma genome variation associated with SpV1-like viral DNA inferred from sequence comparisons. *Microb Comp Genomics*. 4:29-46.

Mindich L. 1996. Heterologous recombination in the segmented dsRNA genome of bacteriophage ϕ6. *Semin Virol*. 7:389–397.

Mindich L. 2006. Phages with segmented double-stranded RNA genomes. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 197-207.

Muskhelishvili G, Palm P, Zillig W. 1993. SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. *Mol Gen Genet*. 237:334-342.

Nandhagopal N, Simpson AA, Gurnon JR, Yan X, Baker TS, Graves MV, Van Etten JL, Rossmann MG. 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci U S A*. 99:14758-14763.

Olsthoorn RC, van Duin J. 1996. Evolutionary reconstruction of a hairpin deleted from the genome of an RNA virus. *Proc Natl Acad Sci U S A*. 93:12256-12261.

Onodera S, Sun Y, Mindich L. 2001. Reverse genetics and recombination in Phi8, a dsRNA bacteriophage. *Virology*. 286:113-118.

Palasingam K, Shaklee PN. 1992. Reversion of Q beta RNA phage mutants by homologous RNA recombination. *J Virol*. 66:2435-2442.

Pe'ery T, Mathews MB. 2007. Viral conquest of the host cell. In: DM Knipe, editor. Field's Virology. Philadelphia: Lippincott Williams and Wilkins. p. 169–208.

Pedulla ML, Ford ME, Houtz JM, et al. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell*. 113:171-182.

Peeters BP, Peters RM, Schoenmakers JG, Konings RN. 1985. Nucleotide sequence and genetic organization of the genome of the N-specific filamentous bacteriophage IKe. Comparison with the genome of the F-specific filamentous phages M13, fd and f1. *J Mol Biol*. 181:27-39.

Pei J, Kim BH, Tang M, Grishin NV. 2007. PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res*. 35:W649-652.

Peng X. 2008. Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of *Sulfolobus* and the implications for plasmid survival. *Microbiology*. 154:383-391.

Peng X, Blum H, She Q, Mallok S, Brugger K, Garrett RA, Zillig W, Prangishvili D. 2001. Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology*. 291:226-234.

Pietilä MK, Roine E, Paulin L, Kalkkinen N, Bamford DH. 2009. An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol Microbiol*. 72:307-319.

Poranen MM, Domanska A. 2008. Virus entry to bacterial cells. In: BW Mahy, MH Van Regenmortel, editors. Encyclopedia of Virology: Academic Press. p. 365-370.

Poranen MM, Tuma R, Bamford DH. 2005. Assembly of double-stranded RNA bacteriophages. *Adv Virus Res*. 64:15-43.

Porter K, Dyall-Smith ML. 2008. Transfection of haloarchaea by the DNAs of spindle and round haloviruses and the use of transposon mutagenesis to identify non-essential regions. *Mol Microbiol*. 70:1236-1245.

Porter K, Kukkaro P, Bamford JK, Bath C, Kivela HM, Dyall-Smith ML, Bamford DH. 2005. SH1: A novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology*. 335:22-33.

Porter K, Russ BE, Dyall-Smith ML. 2007. Virus-host interactions in salt lakes. *Curr Opin Microbiol*. 10:418-424.

Prangishvili D, Forterre P, Garrett RA. 2006a. Viruses of the Archaea: a unifying view. *Nat Rev Microbiol*. 4:837-848.

Prangishvili D, Garrett RA, Koonin EV. 2006b. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res*. 117:52-67.

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

Qiao X, Qiao J, Mindich L. 1997. An in vitro system for the investigation of heterologous RNA recombination. *Virology*. 227:103-110.

Qiao X, Sun Y, Qiao J, Di Sanzo F, Mindich L. 2010. Characterization of Phi2954, a newly isolated bacteriophage containing three dsRNA genomic segments. *BMC Microbiol*. 10:55.

Quax TE, Krupovič M, Lucas S, Forterre P, Prangishvili D. 2010. The *Sulfolobus* rod-shaped virus
2 encodes a prominent structural component of the unique virion release system in Archaea. *Virology*. In press.

Rao VB, Feiss M. 2008. The bacteriophage DNA packaging motor. *Annu Rev Genet*. 42:647-681.

Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM. 2004. The 1.2-megabase genome sequence of Mimivirus. *Science*. 306:1344-1350.

Ravantti JJ, Gaidelytė A, Bamford DH, Bamford JK. 2003. Comparative analysis of bacterial viruses Bam35, infecting a gram-positive host, and PRD1, infecting gram-negative hosts, demonstrates a viral lineage. *Virology*. 313:401-414.

Ravin V, Ravin N, Casjens S, Ford ME, Hatfull GF, Hendrix RW. 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. *J Mol Biol*. 299:53-73.

Redder P, Peng X, Brugger K, et al. 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environ Microbiol*. 11:2849-2862.

Reiter WD, Palm P, Yeats S. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res*. 17:1907-1914.

Rice G, Tang L, Stedman K, Roberto F, Spuhler J, Gillitzer E, Johnson JE, Douglas T, Young M. 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci U S A*. 101:7716-7720.

Rohwer F, Thurber RV. 2009. Viruses manipulate the marine environment. *Nature*. 459:207-212.

Roine E, Kukkaro P, Paulin L, Laurinavičius S, Domanska A, Somerharju P, Bamford DH. 2010. New, closely related haloarchaeal viral elements with different nucleic acid types. *J Virol*. 84:3682-3689.

Rokyta DR, Burch CL, Caudle SB, Wichman HA. 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol*. 188:1134-1142.

Russel M, Model P. 2006. Filamentous phage. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 146-160.

Saren AM, Ravantti JJ, Benson SD, Burnett RM, Paulin L, Bamford DH, Bamford JK. 2005. A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. *J Mol Biol*. 350:427-440.

Schmitt AP, Lamb RA. 2005. Influenza virus assembly and budding at the viral budozone. *Adv Virus Res*. 64:383-416.

Semple CA. 2003. Meet the parents: bioinformatics and computational biology. *Heredity*. 91:542–543.

Sha Y, Melcher U, Davis RE, Fletcher J. 2000. Common elements of spiroplasma plectroviruses revealed by nucleotide sequence of SVTS2. *Virus Genes*. 20:47-56.

She Q, Shen B, Chen L. 2004. Archaeal integrases and mechanisms of gene capture. *Biochem Soc Trans*. 32:222-226.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*. 34:D32-36.

Silander OK, Weinreich DM, Wright KM, O'Keefe KJ, Rang CU, Turner PE, Chao L.

2005. Widespread genetic exchange among terrestrial bacteriophages. *Proc Natl Acad Sci U S A*. 102:19009-19014.

Söding J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 21:951-960.

Sozhamannan S, McKinstry M, Lentz SM, et al. 2008. Molecular characterization of a variant of *Bacillus anthracis*-specific phage AP50 with improved bacteriolytic activity. *Appl Environ Microbiol*. 74:6792-6796.

Stassen AP, Schoenmakers EF, Yu M, Schoenmakers JG, Konings RN. 1992. Nucleotide sequence of the genome of the filamentous bacteriophage I2-2: module evolution of the filamentous phage genome. *J Mol Evol*. 34:141-152.

Steven AC, Conway JF, Cheng N, Watts NR, Belnap DM, Harris A, Stahl SJ, Wingfield PT. 2005a. Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res*. 64:125-164.

Steven AC, Heymann JB, Cheng N, Trus BL, Conway JF. 2005b. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr Opin Struct Biol*. 15:227-236.

Stockley PG, Rolfsson O, Thompson GS, Basnak G, Francese S, Stonehouse NJ, Homans SW, Ashcroft AE. 2007. A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol*. 369:541-552.

Strömsten NJ, Bamford DH, Bamford JK. 2005. In vitro DNA packaging of PRD1: a common mechanism for internal-membrane viruses. *J Mol Biol*. 348:617-629.

Strömsten NJ, Benson SD, Burnett RM, Bamford DH, Bamford JK. 2003. The *Bacillus thuringiensis* linear double-stranded DNA phage Bam35, which is highly similar to the *Bacillus cereus* linear plasmid pBClin15, has a prophage state. *J Bacteriol*. 185:6985-6989.

Sun S, Rao VB, Rossmann MG. 2010. Genome packaging in viruses. *Curr Opin Struct Biol*. 20:114-120.

Sun Y, Qiao X, Qiao J, Onodera S, Mindich L. 2003. Unique properties of the inner core of bacteriophage phi8, a virus with a segmented dsRNA genome. *Virology*. 308:354-361.

Suttle CA. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol*. 5:801-812.

Szymczyna BR, Taurog RE, Young MJ, Snyder JC, Johnson JE, Williamson JR. 2009. Synergy of NMR, computation, and X-ray crystallography for structural biology. *Structure*. 17:499-507.

Tang SL, Nuttall S, Dyall-Smith M. 2004. Haloviruses HF1 and HF2: evidence for a recent and large recombination event. *J Bacteriol*. 186:2810-2817.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673-4680.

Tu AH, Voelker LL, Shen X, Dybvig K. 2001. Complete nucleotide sequence of the mycoplasma virus P1 genome. *Plasmid*. 45:122-126.

Tulman ER, Delhon GA, Ku BK, Rock DL. 2009. African swine fever virus. *Curr Top Microbiol Immunol*. 328:43-87.

van Duin J, Tsareva N. 2006. Single-stranded RNA phages. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 175-196.

Verheust C, Fornelos N, Mahillon J. 2005. GIL16, a new gram-positive tectiviral phage related to the *Bacillus thuringiensis* GIL01 and the *Bacillus cereus* pBClin15 elements. *J Bacteriol*. 187:1966-1973.

Vestergaard G, Aramayo R, Basta T, et al. 2008a. Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. *J Virol*. 82:371-381.

Vestergaard G, Shah SA, Bize A, Reitberger W, Reuter M, Phan H, Briegel A, Rachel R, Garrett RA, Prangishvili D. 2008b. *Stygiolobus* rod-shaped virus and the interplay of crenarchaeal rudiviruses with the CRISPR antiviral system. *J Bacteriol*. 190:6837-6845.

Waldor MK, Mekalanos JJ. 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*. 272:1910-1914.

Waldor MK, Rubin EJ, Pearson GD, Kimsey H, Mekalanos JJ. 1997. Regulation, replication, and integration functions of the *Vibrio cholerae* CTXphi are encoded by region RS2. *Mol Microbiol*. 24:917-926.

Warren RA. 1980. Modified bases in bacteriophage DNAs. *Annu Rev Microbiol*. 34:137-158.

Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 35:W407-410.

Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. 2000. Topologically linked protein rings in the

bacteriophage HK97 capsid. *Science*. 289:2129-2133.

Wilson WH, Van Etten JL, Allen MJ. 2009. The *Phycodnaviridae*: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol*. 328:1-42.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*. 87:4576-4579.

Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev*. 64:69-114.

Wood AG, Whitman WB, Konisky J. 1989. Isolation and characterization of an archaebacterial viruslike particle from *Methanococcus voltae* A3. *J Bacteriol*. 171:93-98.

Xiao C, Kuznetsov YG, Sun S, Hafenstein SL, Kostyuchenko VA, Chipman PR, Suzan-Monti M, Raoult D, McPherson A, Rossmann MG. 2009. Structural studies of the giant mimivirus. *PLoS Biol*. 7:e92.

Yan X, Olson NH, Van Etten JL, Bergoin M, Rossmann MG, Baker TS. 2000. Structure and assembly of large lipid-containing dsDNA viruses. *Nat Struct Biol*. 7:101-103.

Yang JC, Madupu R, Durkin AS, et al. 2009. The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS One*. 4:e6085.

Young R, Wang I. 2006. Phage lysis. In: R Calendar, editor. The Bacteriophages. New York: Oxford University Press. p. 104–125.

Ysebaert M, van Emmelo J, Fiers W. 1980. Total nucleotide sequence of a nearly full-size DNA copy of satellite tobacco necrosis virus RNA. *J Mol Biol*. 143:273-287.

Zdobnov EM, Apweiler R. 2001. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 17:847-848.

Zhang W, Olson NH, Baker TS, Faulkner L, Agbandje-McKenna M, Boulton MI, Davies JW, McKenna R. 2001. Structure of the Maize streak virus geminate particle. *Virology*. 279:471-477.