Available online at www.sciencedirect.com

ScienceDirect

Virology 362 (2007) 384–396

VIROLOGY

www.elsevier.com/locate/yviro

Modular architecture of the T4 phage superfamily: A conserved core genome and a plastic periphery

André M. Comeau, Claire Bertrand, Andrei Letarov [1], Françoise Tétart, H.M. Krisch *

Laboratoire de Microbiologie et Génétique Moléculaire, CNRS-UMR5100, 31062 Toulouse Cedex 9, France

Received 11 October 2006; returned to author for revision 4 December 2006; accepted 13 December 2006
Available online 7 February 2007

Abstract

Among the most numerous objects in the biosphere, phages show enormous diversity in morphology and genetic content. We have sequenced 7 T4-like phages and compared their genome architecture. All seven phages share a core genome with T4 that is interrupted by several *h*yper *p*lastic *r*egions (HPRs) where most of their divergence occurs. The core primarily includes homologues of essential T4 genes, such as the virion structure and DNA replication genes. In contrast, the HPRs contain mostly novel genes of unknown function and origin. A few of the HPR genes that can be assigned putative functions, such as a series of novel Internal Proteins, are implicated in phage adaptation to the host. Thus, the T4-like genome appears to be partitioned into discrete segments that fulfil different functions and behave differently in evolution. Such partitioning may be critical for these large and complex phages to maintain their flexibility, while simultaneously allowing them to conserve their highly successful virion design and mode of replication.

Keywords: Genome evolution; Genome plasticity; Phage; Conserved core genome; T4 superfamily

Introduction

The most astonishing feature of phages is their abundance and diversity. Recent enumerations indicate that they are probably the most prevalent entities in the biosphere (Wommack and Colwell, 2000). Although only a minute fraction of the phage population has been examined by electron microscopy (Demuth et al., 1993; Frank and Moebus, 1987; Hara et al., 1991), these display a seemingly limitless variety of morphologies. Phages manifest a corresponding heterogeneity in genome size, gene content and gene organization (Ackermann and Dubow, 1987). Here, we describe the genomes of a series of phages that share a common origin with the phage T4. The comparison between these genomes provides some interesting insights into the evolutionary mechanisms that generate genomic diversity and the constraints that limit it.

Multiple generations of molecular biologists, biochemists and geneticists have used the large and complex phage T4 as a model system. The ~169 kb genome of T4 has been sequenced and the functions of many of its ~300 genes are known (Miller et al., 2003a). Although T4 is by far the best characterized, >200 similar phages have been described (Ackermann and Krisch, 1997). These T4-like phages all share a common virion morphology—an elongated head, a contractile tail that ends with a complex base-plate with six long fibers radiating from it. About 90% of the known T4-like phages grow on *Escherichia coli* or other enterobacteria, but the remaining 10% grow on phylogenetically more distant (Kersters et al., 2003; Waterbury, 1999) bacteria (*Aeromonas*, *Vibrio*, cyanobacteria, etc.). These distant T4-like phages can vary significantly in virion morphology (Hambly et al., 2001; Tétart et al., 2001). The T4 DNA replication apparatus produces long concatenates of the phage genome that are packaged by a head-filling mechanism (Leiman et al., 2003) and, hence, any T4-type phage mutant that changes the head size also alters the genome size. Thus, a

Abbreviations: CTS, capsid targeting sequence; HPR, hyperplastic region; IP, Internal Protein; LDF, large distal fiber; LGT, lateral gene transfer; SDF, short distal fiber.

* Corresponding author. Université Paul Sabatier, CNRS-LMGM, Bât. IBCG, 118 Route de Narbonne, 31062 Toulouse Cedex 9, France. Fax: +33 5 61 33 58 86.

E-mail address: krisch@ibcg.biotoul.fr (H.M. Krisch).

[1] Current address: S.N. Winogradsky Institute of Microbiology RAS, 117312, Moscow, pr.60-letiya Octyabrya, build. 7, corp. 2, Russia.

0042-6822/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.12.031

Table 1
Characteristics of the sequenced phages versus T4

Phage	Primary host	Size (bp)	%AT	ORFs	tRNAs	Accession
T4	*Escherichia coli*	168,903	65	289	8	NC_000866
RB43	*Escherichia coli*	180,500	67	292	1	NC_007023
RB49	*Escherichia coli*	164,018	60	273	0	NC_005066
RB69	*Escherichia coli*	167,560	63	273	2	NC_004928
Aeh1	*Aeromonas hydrophila*	233,234	57	352	24	NC_005260
25	*Aeromonas salmonicida*	161,475	59	242	13	NC_008208
31	*Aeromonas salmonicida*	172,963	57	247	15	NC_007022
44RR2.8t	*Aeromonas salmonicida*	173,591	57	252	17	NC_005135

mutant T4-type phage that has a larger head also has a partially duplicated genome that can easily evolve by replacement of its duplicated genes with new sequences acquired by lateral gene transfer (LGT). Such a change in genome content can result in additional morphological variation of the phage.

Despite the diversity in the T4 family, when this study was initiated little sequence data were available for any of the divergent T4-like phages. In this, and the associated reports (Nolan et al., 2006; Petrov et al., 2006), we describe results from a high throughput sequencing project that has obtained the genomic sequences of 7 T4-like phages (Table 1). Here, we present an analysis of the genomic architecture of these phages, using two dissimilar T4-like phages, RB49 and Aeh1, to illustrate the various features of the T4-type genome architecture. RB49 is a divergent T4-type coliphage and serves as a representative of the two other sequenced coliphages (RB43 and RB69), while *Aeromonas* phage Aeh1 serves as an example of a phage infecting a non-enterobacterial host (such as phages 25, 31 and 44RR2.8t). Our analysis of these seven genomes reveals that they share a conserved core genome with T4 which is interspersed with hyperplastic regions (HPRs) where most of the sequence variation occurs. The core genomes of T4-like phage appear to be collages assembled mostly from genes originating among their closest phylogenetic neighbors within the T4 superfamily. The HPRs, however, are primarily composed of genes of unknown origin although they do contain some identifiable sequences from bacteria and unrelated phages. Thus, the evolution of the T4-like phages seems to involve predominantly vertical drift for the "core" genes involving limited LGT, but in the adjacent HPRs the situation is similar to that in other phages groups, with few limitations on LGT.

Results and discussion

The RB49 phage genome

The coliphage RB49 was isolated from a Long Island sewage treatment plant in 1964 (Russell and Huskey, 1974). Although it had previously been observed that its virion morphology was indistinguishable from T4 (Ackermann and Krisch, 1997), analysis of a few segments of the RB49 genome revealed that its sequence had diverged substantially (Desplats et al., 2002; Desplats and Krisch, 2003; Monod et al., 1997; Tétart et al., 2001). This divergence led us to propose that RB49 belonged to a new subgroup of the T4-like phage family, the PseudoT-evens (Monod et al., 1997), which are morphologically similar to the T-even phages (T2, T4, T6, etc.). The cohesion of this subgroup, as well as the phylogenetic relationships between the various other T4-like subgroups, has recently been confirmed by a rigorous phylogenetic analysis (Filée et al., 2006). The complete sequence of RB49 discussed here (Table 1) provides an overview of the divergence that has occurred among the genomes of the T4-like coliphages, including RB43 and RB69.

RB49 has a similar size and encodes a similar number of ORFs as T4 (Table 1). The RB49 genome, however, has a somewhat lower %AT content (60%) than does T4 (65%) and, strikingly, lacks any tRNA sequences (Table 1). Fig. 1 presents a schematic diagram of the organization of the RB49 genome. As was previously suggested (Monod et al., 1997), this analysis reveals that a substantial fraction of the T4 and RB49 genomes (64% of the coding sequence; 42% of the ORFs) are derived from a common ancestral sequence. The remaining third of the RB49 sequence is completely unrelated to T4. In the conserved portions, both the gene order and gene content of T4 are broadly maintained. The absence of any major rearrangements (Fig. 1) in the RB49 genome is surprising given its phylogenetic distance from T4. For example, within the 55 kb region of RB49 encoding the major virion structural module (RB49 coordinates 60–115 kb), the only examples of plasticity are the absence of three T4 mobile elements and the presence of a few small ORFs of mostly unknown function. A significant feature of the architecture of the RB49 genome becomes evident when one analyzes its 131 ORFs that have homology to phage-encoded genes (Fig. 2). The best blastp hits of 89% ($n=116$) of these ORFs are to genes belonging to the T-even and the PseudoT-evens phages, while 7% ($n=9$) belong to the more phylogenetically distant SchizoT-even phages and the remaining 5% ($n=6$) derive from phages not related to T4. All but one of these six "foreign" phage ORFs appear to be unique to RB49 since they have no homology ($E<0.05$) to any sequences in the other T4-type phages (Supplementary Table 2). The sole exception is RB49 ORF229 which is most closely related to a phosphatase from a siphovirus, but which has homology to the SchizoT-even vibriophage KVP40 ($E=10^{-5}$). The RB49 genome is thus a collage of phage sequences, but the vast majority of these genes (95%) come from within the T4 superfamily. Interestingly, the phage ϕ1, isolated in the Republic of Georgia, exhibits ~95% nucleotide similarity to RB49 yet still shows a few gene exchanges (http://www.phage.bioc.tulane.edu/). Similar close phylogenetic relations exist between Aeh1 and phage 65, and among phages 25, 31 and 44RR2.8t (http://www.phage.bioc.tulane.edu/). As would be expected, most of the limited gene shuffling among these close phylogenetic neighbors has occurred outside of the core genome. Such closely related phages may eventually provide some interesting insights into the first steps of genome divergence.

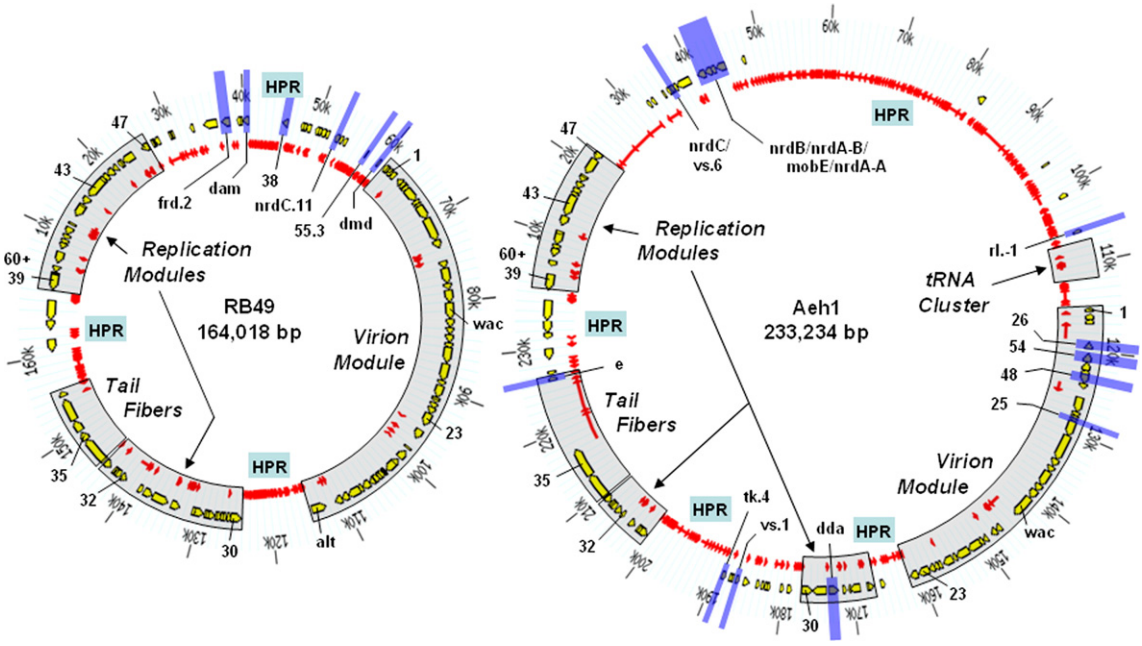

Fig. 1. Genome organization of RB49 and Aeh1. Genome maps of the phages are drawn to scale and, in each, the outer rings show the scales (in kb), the middle rings (yellow arrows) the T4 homologues, and the inner rings (red arrows) the ORFs with no homology to T4. T4 genes that have been rearranged (radiating blue bars) are indicated in their new locations. Various T4 genes are given as landmarks on the outside of the rings and the hyperplastic regions (HPRs) are indicated. The virion modules (see Fig. 5), tail fiber genes (see Fig. 6), replication modules and tRNA cluster (for Aeh1) are indicated by gray boxes.

Genome features that are conserved between T4 and RB49

Although less conserved than the virion structural module, other modules of T4-like phages also share common ancestries, for example, the DNA replication module (Petrov et al., 2006). The least conserved modules are those responsible for DNA modification and the control of transcription in T4, nevertheless most of the differences in genome content between T4 and RB49 involve homologues of non-essential T4 genes. Only four of the 62 essential T4 genes (Miller et al., 2003a) are missing in RB49—*42*, *e*, *asiA* and *motA*. Gene *42* is a dCMP hydroxymethylase required for the synthesis of modified cytosine residues. The absence of *g42* from the RB49 genome is consistent with the lack of such modified cytosines in RB49 DNA (Monod et al., 1997). The lack of the *asiA* and *motA* gene homologues was predictable since these functions are involved in the middle-mode of T4 transcription which appears to be absent in RB49. No mot-like middle-mode promoters sequences have been identified in RB49 (Desplats et al., 2002). Finally, there is no homologue in the RB49 genome of the T4 gene *e*, the soluble lysozyme/endolysin required for cell lysis. However, RB49 ORF102 encodes an unrelated phage endolysin (phage P27-like, $E=10^{-35}$) that presumably replaces this function.

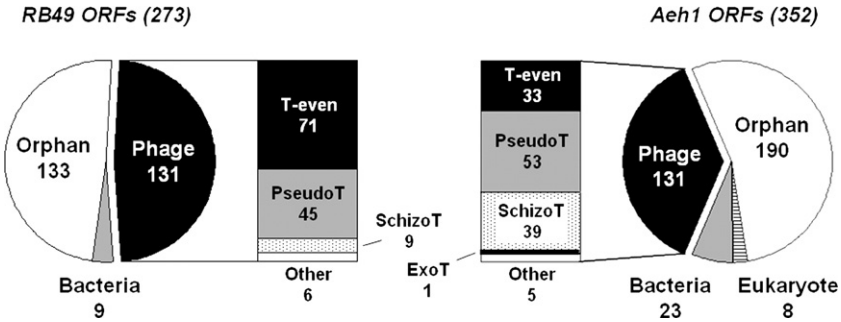

Fig. 2. Distribution of ORF identities in RB49 and Aeh1. For each phage, the number of ORFs with homology (blastp values $E<0.05$) to the various biological groups (Phage, Bacteria and Eukaryotes) is indicated, as well as the number ORFs without known homologues (database Orphans). The phage ORFs are further subdivided on the basis of their best blastp hit with respect to the different phylogenetic subgroups of the T4 superfamily (Desplats and Krisch, 2003; Filée et al., 2006) or to non-T4-related phages ("Other"). The T-even phages are all coliphages very closely related to T4, the PseudoT-evens (e.g., RB49) comprise the subgroup of T4-types that have diverged significantly from the T-evens (some of which infect non-enterobacterial hosts), the SchizoT-evens (e.g., Aeh1) are yet more divergent phages that infect host distant from *E. coli* (e.g., *Aeromonas*, *Vibrio*) and the ExoT-evens include the T4-type phages (e.g., S-PM2) that have diverged the most from T4 and infect hosts that are very distant from *E. coli* (such as cyanobacteria).

Genome features that differ between T4 and RB49

One notable difference between the T4 and RB49 genomes is the latter's paucity of mobile elements. There are six *seg* sequences in T4 and none in RB49. Similarly, there are 3 introns and 5 *mob* sequences in T4, but no similar sequences are found in RB49. However, RB49 is predicted to have an HNH(AP2)-type (Magnani et al., 2004) endonuclease gene (ORF040; Supplementary Table 2). The T4 *seg* and *mob* sequences encode different structural classes of homing endonucleases (Miller et al., 2003a). Except for phage 25 that has three, all the other phages in Table 1 either have none or only one *seg* or *mob* sequences, indicating that, in contrast to the example of T4 itself, such types of mobile elements are not generally numerous in the T4 superfamily.

Although T4 encodes seven tRNA sequences, there are none in RB49. The exact function of phage-encoded tRNAs in the phage-infected cell is still not clear. They could be involved either in the adaptation of the host translation apparatus to the demands of the phage codon usage pattern (Mosig, 1994) or in the rapid extinction of host specific translation immediately after phage infection (Miller et al., 2003a). The RB49 genome, however, clearly demonstrates that such phage-induced modification of the translation apparatus is not necessary for this phage to successfully grow in *E. coli*.

There are no sequences in the RB49 genome with homology to the T4-like IP (internal protein) genes. Such sequences encode small, basic proteins that are encapsulated in large numbers into the phage head and injected with the viral genome in the earliest stage of infection. It is believed that most, if not all IPs, are involved in phage subversion of host macromolecular biosynthesis (Comeau and Krisch, 2005). In some cases, it has been shown that IPs counteract host defensive mechanisms against phage infection (Tock and Dryden, 2005). All of the characterized T4-like IP sequences have a conserved 10 aa capsid targeting sequence (CTS) at the N-terminus of the protein which is cleaved when it is packaged in the head (Mullaney and Black, 1996; Repoila et al., 1994). There are no RB49 proteins that have the consensus N-terminal signal sequence that would make them IP candidates. Furthermore, Monod et al. (1997) have shown that the RB49 virion has no small abundant proteins in the size range of the T4 IPs. RB49 could lack such proteins or, alternatively, in RB49 such functions may be unrelated to the T4 IP sequences. Curiously, the lack of IP candidates in RB49 appears to be unique since all of the other phages analyzed in this project either have T4-like IPs or good candidates for a novel class of IPs (see below).

There are three segments of the RB49 genome containing substantial blocks of genes of unknown function (Fig. 1). Since such regions in the T4-like genomes appear to be highly variable in their gene content, we call them hyperplastic regions (HPR). Unlike bacterial pathogenicity islands (Hacker and Carniel, 2001), these regions have a uniform GC content that is indistinguishable from the remainder of the genome (data not shown). The largest (~35 kb) of the HPRs in RB49 is located between RB49 coordinate 26 and 61 kb and there are two additional smaller-sized ones of ~10 kb centered around coordinates 120 and 160 kb. In T4, the analogous HPR segments contain blocks of uncharacterized genes, most of which are probably dispensable. Even among the closely related T-even phages, the HPRs are highly variable, suggesting that they are the preferred targets for frequent gene shuffling.

The Aeh1 phage genome

The large *Aeromonas hydrophila* phage Aeh1 was isolated from a sewage treatment plant in Wisconsin (Chow and Rouf, 1983). On the basis of its morphology, Aeh1 was classified as one of the few then known T4-like phages that infected non-enterobacterial hosts (Ackermann and Krisch, 1997). Subsequent analyses of its sequences (Desplats and Krisch, 2003; Tétart et al., 2001) revealed that Aeh1 belonged to another subgroup of T4-like phages that have diverged more from T4 than the PseudoT-evens. The diagnostic property of these phages is a larger, more elongated head and a bigger genome than T4. This subgroup was named the SchizoT-evens (Tétart et al., 2001) and, so far, includes phages that infect species of *Aeromonas* and *Vibrio*.

The larger Aeh1 head (137×84 nm) allows it to package ~65 kb more DNA than can be accommodated in the T4 head (110×80 nm). The 233 kb genome of Aeh1 is approximately 40% larger than that of T4, with 63 additional ORFs and three times the number of tRNAs (Table 1). The %AT content of Aeh1 is lower (57%) than either RB49 (60%) or T4 (65%), but nowhere near as low as that of its *Aeromonas* hosts (%AT of 37–43; Farmer, 1999). Aeh1 contains slightly fewer T4 homologues than RB49 (98 versus 115), but it contains more than twice as many novel ORFs with significant homology to bacterial or other phage genes (Fig. 2). Aeh1's greater divergence from the T4 genome is also manifested by its genome organization (Fig. 1). Compared to T4, there are a number of gene rearrangements and insertions/deletions in the Aeh1 genome, including several regions where the T4 genes have been replaced by large blocks of novel ORFs. For example, the ~70 kb of sequence found within the HPR segment between 45 and 115 kb is largely responsible for the increase in size of the Aeh1 genome. The adjacent major virion structural module (115–161 kb) has diverged in its sequence from T4 somewhat more than the analogous segment in RB49, but the T4 gene order and composition is basically preserved. The vast majority (96%) of the 131 ORFs in the Aeh1 genome that come from phage sources appear to originate within the T4 superfamily (Fig. 2). Unlike the situation in RB49, however, the SchizoT-even group contributes a substantial number of the best Aeh1 similarities (30%), along with the PseudoT-evens (40%; some of which are phages that infect *Aeromonas*) and finally the T-evens (25%). Only one Aeh1 ORF is closely related to a sequence from the ExoT-even phages. Thus, the PseudoT- and SchizoT-even phages contribute 70% of Aeh1's ORFs while fewer come from the T-evens coliphages and only a handful of ORFs (5/131) derive from non-T4 phages (Fig. 2). Only two of these "foreign" phage ORFs (058 and 238; Supplementary Table 2) appear to be unique to Aeh1 and not shared by any of the other T4-type phages. The other three ORFs are closely

related to *Salmonella* phages and vibriophages, but have weaker hits to PseudoT-even phages, indicating a possible transfer of a T4-like gene through other intermediates.

Genome features that are conserved between T4 and Aeh1

As is the case with RB49, the replication module of Aeh1 also has a common ancestry with T4. Aeh1 has fewer T4-like regulation genes than RB49 and, similarly, appears to lack a cytosine modification pathway. There are 11 essential T4 genes that lack homologues in Aeh1—*42* (dCMP hydroxymethylase), *67* (prohead core), *33* and *motA* (both involved in the regulation of transcription), *56* (dN[T/D]Pase), *57A* (tail fiber chaperone), *t* (holin) and 4 tail assembly/structural genes (*51*, *27*, *28*, *29*). Although there is no BLAST-significant homologue of *g33* in Aeh1, an ORF at a similar location contains several small groups of amino acids which suggest a *g33* function (E.P. Geiduschek, personal communication). Of the 4 tail genes, only gene *29*, the tail tape measure protein, is likely to be critical and, based on its size and position in the genome, Aeh1 ORF198 is a possible replacement.

Unlike RB49, Aeh1 encodes 24 tRNAs (Table 1), 7 of which have the same anti-codons as those in T4. As mentioned previously, there is debate as to the function(s) of phage-encoded tRNAs (Kunisawa, 2000). A possible role for them is to permit more efficient translation of certain phage genes, often those that are non-essential, which have very different codon usages (Kunisawa, 1992). Indeed, a comparative analysis of phage and host codon usage patterns shows that the T4 codons for which there are phage tRNAs are all enriched (average of ~3-fold). While RB69 shows a 3.5-fold enrichment of phage tRNA-corresponding codons, the four *Aeromonas* phages all show a lower ~1.5-fold enrichment. The unusually large number of tRNAs in Aeh1 could be somehow necessitated by its much larger genome compared to the other T4-likes. The 245 kb phage KVP40 genome also encodes a large number of tRNAs (Miller et al., 2003b). Since such large phage genomes contain a multitude of genes of presumably disparate foreign origins, this may require a more complex suite of tRNAs. Alternatively, adaptation for growth in non-*Escherichia* hosts may be the determining factor since the coliphages in Table 1 contain fewer tRNAs (<9) than the phages infecting non-enteric hosts (12 to 24).

Aeh1 appears to have only one T4-like mobile element, a homologue of the *mobE* mobile endonuclease which is present in some of the other T4-like phages. Interestingly, this *mobE*-like gene splits Aeh1's nucleotide reductase α subunit (*nrdA*) in two. These cistrons, along with the β subunit (*nrdB*), are all located together, but at a different locus (Fig. 1, coordinates 40–45 kb) than in the T4 genome.

There is a single candidate for a T4-related IP gene in Aeh1. ORF081 encodes a T4-like capsid targeting sequence (CTS) at its N-terminus (Fig. 3A). The T4 consensus CTS is N′-MKT (Y/F)(Q/K)E(Y/F)(I/L)XE-C′ (Mullaney and Black, 1996; Repoila et al., 1994) and the ORF081 sequence is 80% similar. Importantly, the T4 *g21* protease recognition/cleavage site (L/I)XE (Black et al., 1994) that is required for the maturation of the IPs and other capsid-associated proteins is located immediately downstream of the CTS. The putative Aeh1 IP would be similar in size to T4 IP3 and is slightly alanine/lysine-rich. Various other potential IP genes exist in the seven phages sequenced in this study (Fig. 3A), as well as a few candidates in the more distant T4-like vibriophage KVP40 (Miller et al., 2003b) and the cyanophage P-SSM2 (Sullivan et al., 2005). All of these potential IPs show the conserved protease cleavage site and most have the other characteristics of T4-like IP sequences (small, basic, A/K-rich), however none is located at the T4 IP loci. The sole exception is RB69 which has 8 IP candidates and the majority of these are found in clusters which correspond to the two IP loci in T4 (Fig. 3B). Curiously, RB69's IP1 has four additional residues and six differences in its N-terminal sequence compared to the T4 IP1 even though the proteins are nearly identical after the cleavage point (72/76 conserved residues).

The three *Aeromonas* phages 25, 31 and 44RR2.8t have no detectable T4-like IP capsid targeting sequences, however, several of their ORFs have small, conserved N-terminal motifs that may indicate a new class of divergent IP analogues (Fig. 4A). These putative IPs are moderately A/K-rich, but unlike the T4-like IPs, their p*I*s are rather acidic. About half have the standard (L/I)XE recognition/cleavage site, with the other half showing less efficient, but potentially functional (Black et al., 1994), *g21* protease recognition sites ([Q/N/F/Y]XE). Some potential cleavage sites are less certain for these IPs, including the possibility of two cleavage sites within the ORF006 IPs. Two phage 25 candidate IPs (ORF002/004) have no easily discernable cleavage sites, indicating that they may have novel cleavage sites (such as the conserved FSE motif). Two of the ORF002 candidate IPs have 40 residues separating the CTS from the potential cleavage site, creating much smaller matured IPs or, as above, indicating novel cleavage motifs. This latter scenario, implying altered specificity of the *g21* protease, seems less likely given the conservation of *g21* among the seven T4-like phages (>65% protein similarity to T4) and given the ubiquitous presence of the standard (L/I)XE cleavage site in other proteins processed by *g21*, such as the major capsid protein (*g23*) and Alt protein (host RNA polymerase modifier).

This new class of potential IP sequences is grouped together near rIIA at the beginning of the genome map (Fig. 4B). In T4, the corresponding region contains two genes of unknown function and a split version of the topoisomerase gene (39+60), which is intact in the other phages (Fig. 4B). Further analyses will be required to verify the IP-like encapsidation of these gene products, but this collection of putative IPs suggests a general capability of the T4-like phages to acquire and evolve multiple genes families with a similar structure to the IPs that could be involved in adapting these phages to varying environments/hosts. In this regard, it is important to note the considerable variability in the size and sequence of this new class of IPs, a situation reminiscent of the hypervariable sequences of the IP genes in other T-evens (Repoila et al., 1994).

It is unclear why the phage head of Aeh1 is larger than that of T4. One small virion head gene, *g67* (prohead core protein), that is essential in T4 is absent and another, *g68* (prohead core

A

Phage	ORF	N-terminal Sequence	aa	%A+K	pI
Consensus sequence		YK YL MKTFQEFIXE▼			
T4	IP1	MKTFKEFTSTTTPVSTITE▼AT	95	23	9.7
	IP2	MKTYQEFIAE▼AR	100	31	10.0
	IP3	MKTYQEFIAE▼AS	193	27	10.2
RB43	200	MKSFKELISE▼SV	202	22	8.6
	241	MKSFKDFVSLDE▼AL	277	22	10.0
RB69	IP1	MEPKMKTYKEFISPVANVSTLTE▼AT	99	23	9.5
	arn.3	MKSFKERLELLDLALSRETPESLAE▼KF	153	11	4.4
	123	MKTYQEFIAAVTLSAEQKAAIEE▼GK	109	28	10.3
	124	MKTYKEFIAE▼SA	90	13	6.8
	125	MKMYQEFIIE▼AD	86	19	4.6
	127	MKTYMNFYQNLSLNE▼AT	93	22	8.5
	149	MKTYAEFLTE▼AA	146	29	10.4
	156	MKTYQEFIAE▼AT	108	19	10.0
Aeh1	081	MKSLADFLAE▼SH	186	18	7.0
KVP40	150	MNRMELKTFDEFKNSDELKE▼AT	183	23	9.8
	311	MKLKTFDEYHKLVE▼KV	59	29	10.3
P-SSM2	114	MRTFQEFLQLCE▼GG	169	20	10.6
	115	MKSFKQFLEE▼AY	86	24	10.6
	194	MKSFSQFLTE▼TT	424	15	6.0
	211	MKSLSQFIIE▼TQ	573	25	11.1

B

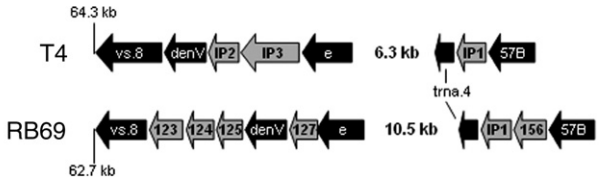

Fig. 3. T-even-like Internal Proteins (IPs) sequences in the various genomes. (A) Characteristics of the putative IPs. The residues of the N-terminal capsid targeting sequence (CTS) identical to the T-even consensus sequence (Mullaney and Black, 1996; Repoila et al., 1994) are shaded in black, while similar residues are in gray. The critical *g21* protease recognition sites (Black et al., 1994), (L/I)XE▼, are also shaded in black at each known and presumed location, with the cleavage points indicated by the downwards triangles. T-even IPs are generally alanine/lysine (A/K)-rich and usually have basic isoelectric points (p*I*s). The average percent A/K of the entire proteomes of these phages varies between 12.5 and 14.6% and the potential IP A/K values are shaded in gray if they are below this average. Non-basic p*I*s are similarly shaded. (B) The genomic loci of the RB69 candidate IP genes. The two IP clusters, both located within an HPR, contain the majority of RB69's IP candidates and their organization is comparable to that found in T4. T4 gene homologues are in black and known/candidate IP genes are in gray. The genome coordinates (in kb) of the clusters are indicated.

protein), is ~20% larger in size. The major capsid gene, *g23*, and also *g24*, which encodes the minor head vertex protein, are both roughly the same size as their T4 homologues. Although all the head protein sequences have diverged significantly (~50–60% identity) from their T4 homologues, the gene *24* protein has diverged the most (only 30% identity). Interestingly, mutational studies of T4 indicate that single amino acid changes in gene *23* can cause the production of variably sized "giant" phage heads (Cummings et al., 1977) and bypass the need for the *g24* vertex protein (Fokine et al., 2006). The question of capsid morphology is a complex issue given that simple sequence comparisons among these divergent phages can only give us limited information. The recent structural determination of *g24* (Fokine et al., 2005), showing similarities among different phage groups, highlights the importance of moving beyond sequence homology to analyze the "viral self" (Bamford et al., 2002) to establish structural/functional relationships within lineages. The availability of the T4-like genomes presented in this project now makes possible a more in-depth analysis of capsid structure relationships, a task we have recently undertaken using protein homology modeling (Comeau and Krisch, unpublished data).

Genome features that differ between T4 and Aeh1

The most significant difference between Aeh1 and T4 is the additional ~70 kb of DNA that is located in the large HPR segment at 45–115 kb. More than half (54%) of the proteins encoded by the Aeh1 genome have no homologues in the databases (Fig. 2) and presumably many of these encode functions that help this phage to adapt to hosts that inhabit aquatic environments. A similar situation exists in the cyanophage S-PM2 where photosynthesis genes have been found in an analog of an HPR (Mann et al., 2005). LGT most certainly has played a role in such adaptation and this may have been facilitated, in part, by certain particularities of the T4 phage recombination and gene expression systems. T4 recombines most efficiently early in infection (Krisch et al.,

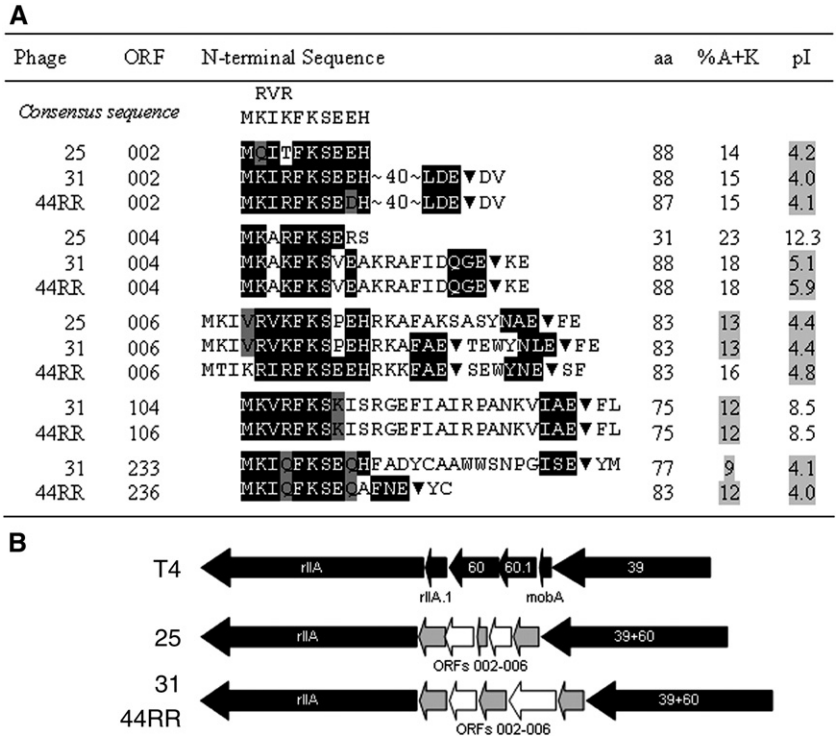

Fig. 4. New class of Internal Proteins (IP) candidate ORFs. (A) Characteristics of candidates for a new class of IPs found in the *Aeromonas* phages. Details are as in Fig. 3, except that less efficient *g21* protease recognition sites ([Q/N/F/Y]XE▼) are also included. Some potential cleavage sites are less certain for these IPs, including the possibility of two cleavage sites within the ORF006 IPs. The ORF002 IPs either have no potential cleavage sites (phage 25) or have 40 residues separating the CTS from the potential site (phages 31 and 44RR2.8t). The average percent A/K of the entire proteomes of the three listed phages is 13.9%. (B) The major genomic locus of the candidate *Aeromonas* phage IP genes. One cluster, located within an HPR, contains the majority of the IPs (ORFs 002/004/006). Phages 31 and 44RR2.8t have identical topologies in this region. The comparable region in T4 contains genes of unknown function and the topoisomerase gene is split (39+60). T4 gene homologues are in black, candidate IP genes are in gray, and proteins of unknown origin are in white.

1972) before the host genome is degraded, and this could facilitate the acquisition of host genes by the phage. Additionally, the T4 recombination system requires as little as 50 bp of homology, and perhaps less, to recombine at reasonable frequencies (Singer et al., 1982). Thus, "semi-legitimate" recombination perhaps relying on the small, but generally conserved regulatory signals in intergenic regions (promoters, translation initiation regions, transcription terminators, etc.) could have mediated the acquisition of such foreign genes (Repoila et al., 1994). Once these host genes are acquired, the T4 expression system could exploit their endogenous bacterial promoters since these sequences are very similar to the early promoter sequences of the T4-type phages (Desplats et al., 2002; Nolan et al., 2006). In the opposite direction, LGT from phage to bacteria would also be facilitated by the fact that both T4 early promoters and T4 "cryptic" promoters are extremely efficient in diverse bacterial species (Belin et al., 1987; Frey et al., 1988). Perhaps the T4 recombination system also employs the small patches of quasi-homology in the HPRs to randomly shuffle the gene content of the HPRs among the T4-type phages (see Repoila et al., 1994).

Aside from the ORFs with homology to T4 genes, there are 33 sequences in Aeh1 with homology to 13 other phages (Supplementary Table 2). Thirteen of these 33 homologues are closest to KVP40 followed by 7 from the coliphage RB43 and 2 each from the *Aeromonas* phages 25 and 65. The remainder of the homologues originate from a panoply of phages: *Aeromonas* phages 31 and 44RR2.8t, coliphages RB49 and T7, vibriophage VP16C, *Salmonella* phages 5 and Felix01, *Pseudomonas* phage F116 and *Prochlorococcus* phage P-SSM2. This list illustrates the apparent ease of gene transfer between phages infecting relatively distant hosts (e.g., enterobacteria and cyanobacteria). As usual, the majority of the novel phage sequences encode unknown functions.

Besides these phage-like ORFs, 23 of the remaining identified Aeh1 ORFs (Fig. 2) appear to be homologues of bacterial genes (Supplementary Table 2). The range of bacterial species represented is large, suggesting that the subversion of bacterial genes by the T4 phage is frequent and promiscuous (Comeau and Krisch, 2005). Twenty-two bacterial species are represented that range from *Aeromonas*, *Streptococcus*, *Clostridium*, *Mycoplasma*, *Halorhodospira* (photosynthetic γ-proteobacteria) and *Gloeobacter* (cyanobacterium). Many of these genes encode enzymes that are implicated in DNA/RNA modifications, energy cycling, lipid biosynthesis and signaling. For example, the vibriophage KVP40 contains a novel pyridine nucleotide (NAD^+) salvage pathway (Miller et al., 2003b) and Aeh1 also has homologues of many of these enzymes. This pathway could have some particular utility in phages with large genomes.

The conserved virion structural module of the T4 superfamily

One of the most conserved segments of the T4-like genome appears to be the large module that encodes most of the structural components of the virion (Fig. 5). It appears clear that this complex structure is a major component of the "viral self" (Bamford et al., 2002) defining the T4-type phage lineage. In RB49, the only substantial differences from the T4 genome within this region are two small replacements/deletions of non-essential genes (*segC* and *D*, *5.3*) and two small insertions (totaling ~2 kb) of orphan ORFs. The two regions of the virion module containing *seg* mobile endonucleases are frequently plastic in the various T4-like phage genomes (Fig. 5), indicating perhaps that the acquisition of such mobile elements at these sites by T4 is not surprising. The Aeh1 genome has further rearrangements within this module that involve an additional ~10.5 kb of DNA. For example, there is the relocation of four homologues of T4 base-plate genes (Fig. 1) that are found elsewhere in the T4 genome. Interestingly, similar movement of base-plate genes was also found in both the vibriophage KVP40 (Miller et al., 2003b) and the T4-type cyanophage S-PM2 (Mann et al., 2005).

Some selective pressure on the genes of the virion region appears to restrain its sequence divergence among the T4-like phages. Phages as diverged as Aeh1 and the cyanophage S-PM2 still show aa identities to T4 as high as 64% and 39%, respectively, in this region. The highest levels of plasticity in the virion module occur in the C-terminal portions of the fibrous proteins encoded by *gwac* and *g12* homologues. Gene *12* is the short tail fiber that is responsible for the phage's irreversible attachment to the host cell that initiates the cascade of conformational changes in the phage tail structure which leads to DNA injection (Rossmann et al., 2004). Although the N-terminal portion of *g12* that is fixed to the base-plate structure is conserved, the C-terminal portion of the protein involved in the receptor–ligand interaction is highly variable. Similar variability is seen in the *wac* fiber which acts as an environmental sensor and infectivity control by interacting with the highly variable tail fibers (Letarov et al., 2005). The characteristic heptad periodicity of the *wac* protein's coiled-coiled protein structure is retained in all the phages, but the C-terminal domains, responsible for interacting with the highly variable tail genes, diverge completely (Letarov et al., 2005).

The extreme plasticity of the tail fiber module of the T4 superfamily

The tail fiber module is the most plastic structural region in the T4-like phage genomes (Fig. 6). Even in otherwise closely related phages such as T4, T2 and T6, the sequence and structure of this locus have diverged substantially (Riede et al., 1987; Tétart et al., 1998). All of the RB49 and Aeh1 genes in this locus have less than 50% aa identity to their T4 homologues. The first gene in this locus, *g34* (the proximal tail fiber subunit), is relatively well conserved, but the next gene, *g35*, that encodes the tail fiber hinge joint has apparently been the subject of domain shuffling. In Aeh1, for example, a T4-related N-terminal domain of this protein has been fused to a large segment of novel protein sequence, while in RB43 and KVP40 somewhat smaller fusions occurred. Similarly for the small distal fiber (SDF; *g36*), which in RB49 is a collage of various phage tail fiber domains with homology to the T4 *g36* sequence only being retained in the N-terminal domain. In Aeh1, the SDF (*g36*) and long distal fiber (LDF; *g37*) homologues are fused into a single gene that has duplicated and diverged. Gene *36* fusions, and single or duplicated S/LDF chimeras, occur frequently in the T4-like phages (Fig. 6). These chimeras generally retain only small patches of homology to the

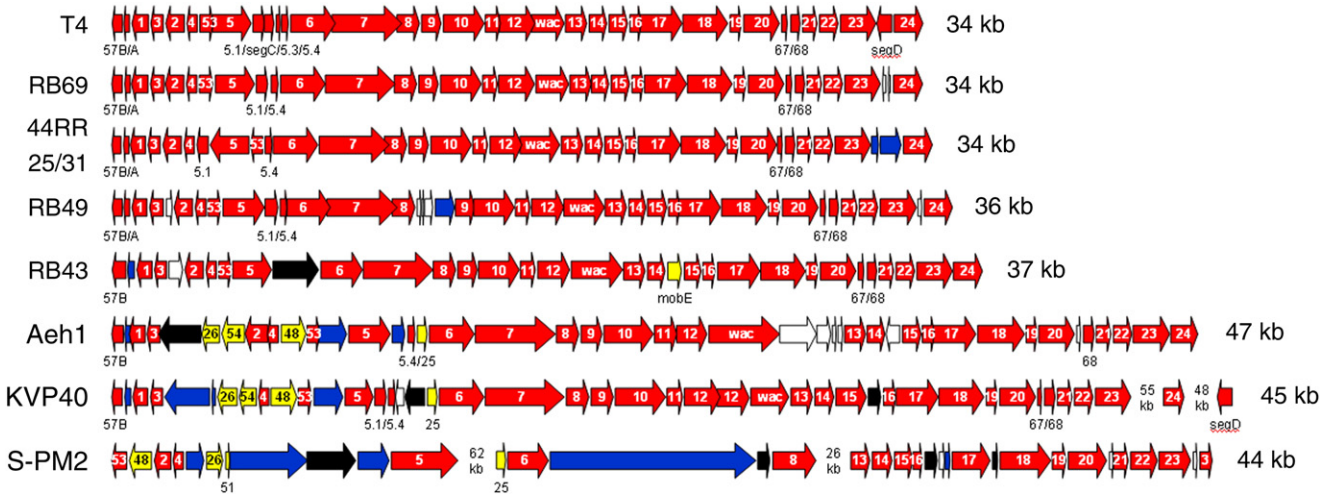

Fig. 5. Conserved virion structural modules of seven completed T4-like phages compared to T4. Vibriophage KVP40 (accession NC_005083) and cyanophage S-PM2 (NC_006820) are included as examples of the most divergent T4-type phages. Genes homologous to those in the T4 virion module are in red, while other T4 genes that have been relocated in this module are in yellow, ORFs in blue are from phages other than T4, ORFs in black are from bacteria, and ORFs in white are of unknown origin. *Aeromonas* phages 44RR2.8t, 25 and 31 have identical topologies in this module. Six overlapping T4 genes (*repEB*, *repEA*, *17′A*, *17′B*, *17″* and *21′*) are generally not present in the other phages and have been omitted for clarity. The lengths of the complete modules do not include the gaps between the virion gene clusters (for KVP40 and S-PM2).

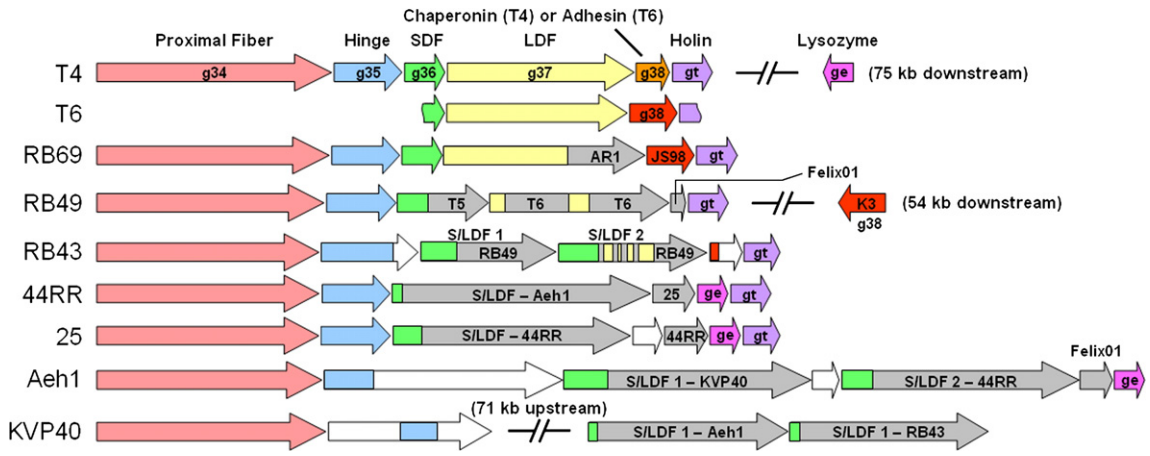

Fig. 6. Hyperplastic tail fiber modules of six completed T4-like phages compared to T4 and T6. KVP40 (accession NC_005083), a more distant ancestor, is included as an example of a more divergent phage. For each gene, regions of T4 homology (T6 homology [AF052605] for the adhesins) are indicated by the corresponding colors, regions of similarity to other phages are in gray, whereas regions of unknown origin are in white. The sources of phage homology (best blastp match) are indicated. Since 44RR2.8t is ~97% similar to *Aeromonas* phage 31, the second match is listed. RB49's *g38* homologue has been displaced to a distant genomic locus, whereas some phages do not have identifiable adhesin genes (nor *gt*/*ge* candidates for KVP40). The length of the complete T4 module (minus *ge*) is 10 kb.

T4 LDF. The LDF of RB49 is most closely related to T6, but with a sequence rearrangement. In T6, the *g38* encoded adhesin is located adjacent to *g37* and binds to the sequence at the tip of the tail fiber, thus determining the phage's specificity for host bacterial receptors (Riede et al., 1987; Tétart et al., 1998). In phage RB49, the *g38* homologue, which has a surprising >90% aa identity to the *g38* of the T-even phage K3 (P07876), has been displaced halfway around the genome into the HPR at position 45 kb. This is a striking example of the acquisition by an HPR of a gene whose function is to adapt the phage to its host. Perhaps some of the genes of unknown functions adjacent to the RB49 *g38* are also involved in this phage's host adaptation. Of the phages represented in Fig. 6, only the coliphages have T-even-like adhesin sequences, suggesting that phages with other hosts use different classes of adhesin molecules. For example, following the two S/LDF chimeras in Aeh1 there are two ORFs that might encode adhesins. The first is a database orphan and the second is an ORF related to an unknown protein of *Salmonella* phage Felix01 ($E=10^{-13}$). Although neither of these sequences is related to any known adhesin sequences, their positions in the tail fiber module make them potential adhesin candidates. The final gene in the tail fiber module of Aeh1 is not the holin gene *t*, as found in T4 and RB49, but an ortholog of the T4 lysozyme gene *e*. This appears to be a common organization among the most divergent T4-like phages.

Conclusions

The full portrait of the T4 phage superfamily that has emerged from the genomic comparisons reported here and elsewhere (Mann et al., 2005; Miller et al., 2003b; Nolan et al., 2006; Petrov et al., 2006; Sullivan et al., 2005) is that the T4 superfamily can be distilled down to a core of ~75 signature genes. This number depends on the precise subset of phages considered. For example, there are ~90 core genes shared among RB49, Aeh1 and T4, whereas the very divergent cyanophages have <45 T4-like ORFs each. Such genes constitute the backbone of the genome of the T4 superfamily and include most, but not all, of the essential genes in T4, such as the virion constituents and the DNA replication apparatus. Surprisingly, some genes of unknown function are conserved among these core functions. The genes implicated in interactions with the cell (tail fibers, tRNAs, modified bases, IPs, etc.) are much more variable and are often located within the genome's hyperplastic regions.

A patchwork genomic structure involving both HPRs and a conserved core genome appears to be a particular feature of the T4-like myoviruses. In Rohwer and Edwards' (2002) whole-genome phage phylogeny, the T4 superfamily is quite distinct from the other phages and even other myoviruses. In general, the siphoviruses appear to have a less constrained type of evolution than T4-like myoviruses since they can apparently swap large portions of their genomes with unrelated phages. Whole-genome comparisons by dot plots (Fig. 7) show that the T4-like phages all share a common ancestral sequence that is distributed about the genome. Multiple alignments of the genomes of these phages (data not shown) show that the nucleotide sequence identities remain high (>50%) in these more conserved regions. In contrast, the lambdoid phages (siphoviruses) show a much smaller ancestral core (Fig. 7) that is usually restricted to the right half of the genome, except for the phage N15 where it is in the left half. The lambdoid core genome also has a pattern of homology that is quite different from that of the T4-like phages. There are a few regions of near identity with λ and a few regions of intermediate homology (25–50% nucleotide identity), but since the core is so small, the vast majority of the genome has no homology to λ (data not shown). This could be a result of the more modular evolution among the λ-like phages. These phages, which include coliphages and mycobacteriophages (Juhala et al., 2000; Pedulla et al., 2003), can apparently easily replace essential genes with completely unrelated sequences that are functionally equivalent (Hendrix et al., 2003; Lawrence et al., 2002). This is

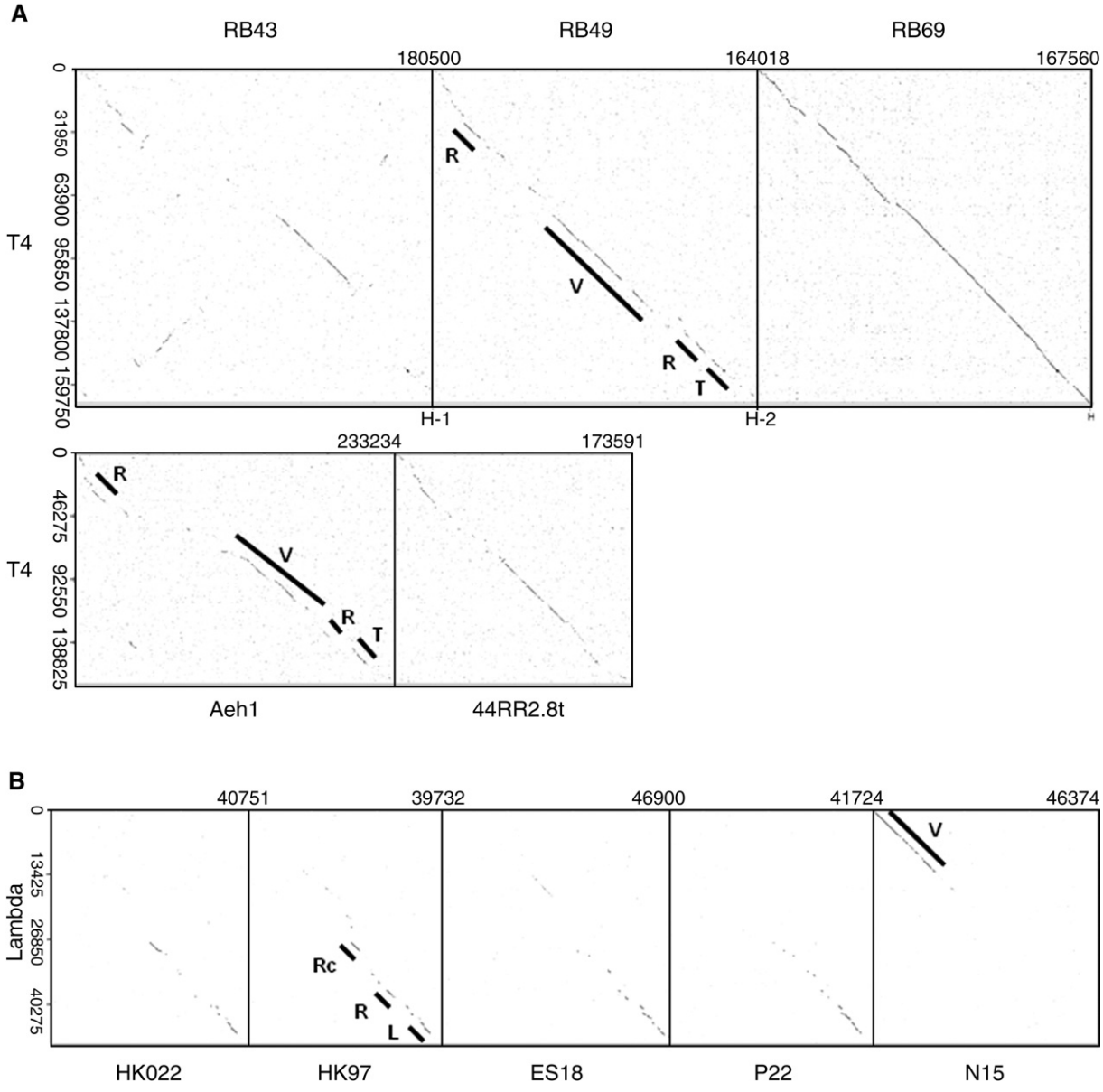

Fig. 7. Whole-genome dot plots of T4-like phages compared to lambdoid phages. Dot plots were constructed in JDotter (Brodie et al., 2004), available at http://athena.bioc.uvic.ca, with a gray-scale index of 40–80 and a sliding window of 50 bp. (A) Comparison of five of the T4-like phages sequenced in this project versus T4. Phages 25 and 31 were excluded because they are so similar (and hence give very similar patterns) to 44RR2.8t. The positions of the replication (R), virion (V) and tail (T) modules are indicated on RB49 and Aeh1 for reference. (B) Comparison of five lambdoid phages (coliphages HK022, HK97 and N15; *Salmonella* phages ES18 and P22) versus lambda. The replication (R), virion (V), recombination (Rc) and lysis (L) regions are indicated on lambda and N15 for reference. Accession numbers, lambda (NC_001416), HK022 (NC_002166), HK97 (NC_002167), N15 (NC_001901), ES18 (NC_006949) and P22 (NC_002371).

in direct contrast to the T4-like phages where there seem to be strong barriers that limit exchanges to close homologues and preserve the general synteny of the core group of genes (Filée et al., 2006). Perhaps the greater complexity of T4-like phages, with their multitude of protein–protein interactions between the constituents of the virion (Leiman et al., 2003) and replication complex (Karam and Konigsberg, 2000), severely restricts their ability to acquire functional analogues. Regardless of the explanation, the more rigid conservation of the core genes leads to a "more vertical" evolutionary pathway with restricted LGT in this part of the genomes, while LGT is rampant in the HPRs.

We suggest that the importance of LGT depends on phage complexity, with the simpler phages benefiting much more from it because of their more forgiving enzymatic and morphological architectures. For example, the two subgroups of the lipid-containing Tectiviruses, which have small genomes (~15 kb) and simple morphology, have similar genome organizations but lack significant sequence homology (Ravantti et al., 2003; Saren et al., 2005). In contrast, by constraining promiscuous exchange to their HPRs, the large and complex T4-like genomes can conserve their successful and robust core-genome-encoded "nano-machinery" (Filée et al., 2006) intact and still benefit

from considerable genetic diversity elsewhere in their genomes. We suggest that the conserved core replication and virion supermodules were originally assembled from modular constituents, but as these various elements became better adapted to each other, their modular interchangeability was sacrificed.

Future analyses of additional T4-like phage genomes should improve our understanding of how the T4 superfamily has managed to be so successful in occupying new niches in the biosphere. Among the most interesting questions that remain unanswered about the T4-type phage genomes concerns the numerous and diverse genes inhabiting the HPRs. What these genes do, how they got into the HPRs and where they originally come from are largely unsolved mysteries. We are currently investigating these questions by looking at closely related pairs of T4-type phage genomes that have just begun to diverge from each other. For example, the genomes of phages RB49 and ϕ1 are virtually identical over >93% of their sequence, but in ~10 loci they have nevertheless diverged substantially (Arbiol et al., unpublished data). This pair of phages has slightly different host ranges. The correlation of these and other phenotypic differences between them with variations in the phage genomic sequences could provide useful clues on the function and origin of these genes.

Materials and methods

Genomic DNA and sequencing

Genomic DNA was isolated from PEG-purified, high-titer phage stocks using either CsCl centrifugation or commercial kit extractions as detailed by Petrov et al. (2006). High-throughput DNA sequencing was then carried out by two commercial providers (Integrated Genomics, Chicago IL; Fidelity Systems, Gaithersburg, MD) using either a shotgun cloning approach or a clone-free "fimer" direct sequencing approach (Petrov et al., 2006; Slesarev et al., 2002). Final gap closures and corrections using standard PCR sequencing were carried out by the commercial providers, the Tulane DNA Sequencing Core Facility, and by our laboratory using the Toulouse CNRS IFR109 DNA Sequencing Platform. Completed genomes were posted as soon as working drafts became available on the project website at http://www.phage.bioc.tulane.edu. The overall accuracy of the polished genomes is estimated to be <2.5 errors per 100 kb.

Genome annotation

Automated and manual genome annotations were carried out as described by Nolan et al. (2006) and Petrov et al. (2006). Briefly, ORFs were detected using GeneMarkS (Besemer et al., 2001) and adjusted using GLIMMER (Delcher et al., 1999). Protein statistics were generated using a local implementation of the EMBOSS suite (Rice et al., 2000). All ORF homologies, and specifically the T4 classification, were determined by a permissive $E<0.05$ match in blastp against the complete database (nr) and in the conserved domain database (CDD) on the NCBI website (http://www.ncbi.nlm.nih.gov/). The permissive E value was utilized within the context of this report in order to reveal more distant homologies. However, a more stringent E value of $<10^{-4}$ was utilized for annotation of the final GenBank submissions. Finer scale analysis (especially for putative IP genes; see below) within the T4-like phages, or among the draft phage sequences, was performed using a local blast available on our T4-like genome sequencing project website (http://www.phage.bioc.tulane.edu/). Multiple sequence alignments were carried out in BioEdit v.7 (Hall, 1999) implementing clustalW (Thompson et al., 1994). Finally, tRNA genes were identified using tRNAscan-SE (Lowe and Eddy, 1997) and codon usage was analyzed using the Codon Usage Database (http://www.kazusa.or.jp/codon/; Nakamura et al., 2000). All similarity values, unless indicated otherwise, are percent protein identity.

Acknowledgments

We thank Yvette de Preval, Christine Arbiol and CNRS IFR109 for DNA sequencing in Toulouse; as well as Vasiliy Petrov, Jim Nolan and Jim Karam at the Tulane Medical School for their collaboration in the genomic sequencing and in the implementation of the online automated genome annotations. We also thank Justin Manges of the Tulane DNA Sequencing Core facility and the staff of Integrated Genomics, Inc. and Fidelity Systems, Inc. for sequencing services and technical advice. Pat Dennis of NSF and E. Peter Geiduschek provided sage advice. This work was principally supported by grant MCB-0138236 (J.D. Karam and H.M. Krisch, co-P.I.s) and in a minor way by grant EF-0333130 (J.D. Karam [P.I.] and H.M. Krisch [foreign consultant]) from the National Science Foundation, USA; as well as intramurally by the CNRS. A.L. was supported by grant RFFI 04-04-48789-a.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2006.12.031.

References

Ackermann, H.-W., Dubow, M.S., 1987. Viruses of Prokaryotes. CRC Press, Boca Raton.

Ackermann, H.-W., Krisch, H.M., 1997. A catalogue of T4-type bacteriophages. Arch. Virol. 142, 2329–2345.

Bamford, D.H., Burnett, R.M., Stuart, D.I., 2002. Evolution of virus structure. Theor. Popul. Biol. 61, 461–470.

Belin, D., Mudd, E.A., Yu, Y.-Y., Prentki, P., Krisch, H.M., 1987. Sense and antisense transcription of bacteriophage T4 gene 32: processing and stability of the mRNAs. J. Mol. Biol. 194, 231–243.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618.

Black, L.W., Showe, M.K., Steven, A.C., 1994. Morphogenesis of the T4 head. In: Karam, J.D. (Ed.), Molecular Biology of Bacteriophage T4. American Society for Microbiology Press, Washington, pp. 182–185.

Brodie, R., Roper, R.L., Upton, C., 2004. JDotter: a Java interface to multiple dotplots generated by dotter. Bioinformatics 20, 279–281.

Chow, M.S., Rouf, M.A., 1983. Isolation and partial characterization of 2 *Aeromonas hydrophila* bacteriophages. Appl. Environ. Microbiol. 45, 1670–1676.
Comeau, A.M., Krisch, H.M., 2005. War is peace—Dispatches from the bacterial and phage killing fields. Curr. Opin. Microbiol. 8, 488–494.
Cummings, D.J., Chapman, V.A., DeLong, S.S., 1977. Structural aberrations in T-even bacteriophage. IX: Effect of mixed infection on production of giant bacteriophage. J. Virol. 22, 489–499.
Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636–4641.
Demuth, J., Neve, H., Witzel, K.-P., 1993. Direct electron microscopy study on the morphological diversity of bacteriophage populations in Lake Plußee. Appl. Environ. Microbiol. 59, 3378–3384.
Desplats, C., Krisch, H.M., 2003. The diversity and evolution of the T4-type bacteriophages. Res. Microbiol. 154, 259–267.
Desplats, C., Dez, C., Tétart, F., Eleaume, H., Krisch, H.M., 2002. Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J. Bacteriol. 184, 2789–2804.
Farmer III, J.J., 1999. The family *Vibrionaceae*. In: Dworkin, M. (Ed.), The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. Springer-Verlag, New York (http://141.150.157.117:8080/prokPUB/index.htm).
Filée, J., Bapteste, E., Susko, E., Krisch, H.M., 2006. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol. Biol. Evol. 23, 1688–1696.
Fokine, A., Leiman, P.G., Shneider, M.M., Ahvazi, B., Boeshans, K.M., Steven, A.C., Black, L.W., Mesyanzhinov, V.V., Rossmann, M.G., 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.
Fokine, A., Battisti, A.J., Kostyuchenko, V.A., Black, L.W., Rossmann, M.G., 2006. Cryo-EM structure of a bacteriophage T4 gp24 bypass mutant: the evolution of pentameric vertex proteins in icosahedral viruses. J. Struct. Biol. 154, 255–259.
Frank, H., Moebus, K., 1987. An electron microscopic study of bacteriophages from marine waters. Helgol. Meeresunters. 41, 385–414.
Frey, J., Mudd, E.A., Krisch, H.M., 1988. A bacteriophage T4 expression cassette that functions efficiently in a wide range of Gram-negative bacteria. Gene 62, 237–247.
Hacker, J., Carniel, E., 2001. Ecological fitness, genomic islands and bacterial pathogenicity: a Darwinian view of the evolution of microbes. EMBO Rep. 2, 376–381.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hambly, E., Tétart, F., Desplats, C., Wilson, W.H., Krisch, H.M., Mann, N.H., 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc. Natl. Acad. Sci. U.S.A. 98, 11411–11416.
Hara, S., Terauchi, K., Koike, I., 1991. Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy. Appl. Environ. Microbiol. 57, 2731–2743.
Hendrix, R.W., Hatfull, G.F., Smith, M.C., 2003. Bacteriophages with tails: chasing their origins and evolution. Res. Microbiol. 154, 253–257.
Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299, 27–51.
Karam, J.D., Konigsberg, W.H., 2000. DNA polymerase of the T4-related bacteriophages. Prog. Nucleic Acid Res. Mol. Biol. 64, 65–96.
Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P., Stackebrandt, E., 2003. Introduction to the *Proteobacteria*. In: Dworkin, M. (Ed.), The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. Springer-Verlag, New York (http://141.150.157.117:8080/prokPUB/index.htm).
Krisch, H.M., Hamlett, N.V., Berger, H., 1972. Polynucleotide ligase in bacteriophage T4D recombination. Genetics 72, 187–203.
Kunisawa, T., 1992. Synonymous codon preferences in bacteriophage T4: a distinctive use of transfer RNAs from T4 and its host *Escherichia coli*. J. Theor. Biol. 159, 287–298.
Kunisawa, T., 2000. Functional role of mycobacteriophage transfer RNAs. J. Theor. Biol. 205, 167–170.
Lawrence, J.G., Hatfull, G.F., Hendrix, R.W., 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184, 4891–4905.
Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F., Rossmann, M.G., 2003. Structure and morphogenesis of bacteriophage T4. Cell. Mol. Life Sci. 60, 2356–2370.
Letarov, A., Manival, X., Desplats, C., Krisch, H.M., 2005. gpwac of the T4-type bacteriophages: structure, function, and evolution of a segmented coiled-coil protein that controls viral infectivity. J. Bacteriol. 187, 1055–1066.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Magnani, E., Sjölander, K., Hake, S., 2004. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in pants. Plant Cell 16, 2265–2277.
Mann, N.H., Clokie, M.R.J., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., Krisch, H.M., 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. J. Bacteriol. 187, 3188–3200.
Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003a. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156.
Miller, E.S., Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Durkin, A.S., Ciecko, A., Feldblyum, T.V., White, O., Paulsen, I.T., Nierman, W.C., Lee, J., Szczypinski, B., Fraser, C.M., 2003b. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J. Bacteriol. 185, 5220–5233.
Monod, C., Repoila, F., Kutateladze, M., Tétart, F., Krisch, H.M., 1997. The genome of the pseudo T-even bacteriophages, a diverge group that resembles T4. J. Mol. Biol. 267, 237–249.
Mosig, G., 1994. Synthesis and maturation of T4-encoded tRNAs. In: Karam, J.D. (Ed.), Molecular Biology of Bacteriophage T4. American Society for Microbiology Press, Washington, pp. 182–185.
Mullaney, J.M., Black, L.W., 1996. Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. J. Mol. Biol. 261, 372–385.
Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28, 292.
Nolan, J., Petrov, V., Bertrand, C., Krisch, H.M., Karam, J.D., 2006. Genetic diversity among five T4-like bacteriophages. Virol. J. 3, 30.
Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs, W.R., Hendrix, R.W., Hatfull, G.F., 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113, 171–182.
Petrov, V., Nolan, J., Bertrand, C., Levy, D., Desplats, C., Krisch, H.M., Karam, J., 2006. Plasticity of the gene functions for DNA replication in the T4-like phages. J. Mol. Biol. 361, 46–68.
Ravantti, J.J., Gaidelyte, A., Bamford, D.H., Bamford, J.K.H., 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.
Repoila, F., Tétart, F., Bouet, J.-Y., Krisch, H.M., 1994. Genomic polymorphism in the T-even bacteriophages. EMBO J. 13, 4181–4192.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.
Riede, I., Drexler, K., Schwarz, H., Henning, U., 1987. T-even-type bacteriophages use an adhesin for recognition of cellular receptors. J. Mol. Biol. 194, 23–30.
Rohwer, F., Edwards, R., 2002. The phage proteomic tree: a genome-based taxonomy for phage. J. Bacteriol. 184, 4529–4535.
Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., Leiman, P.G., 2004. The bacteriophage T4 DNA injection machine. Curr. Opin. Struct. Biol. 14, 171–180.

Russell, R.L., Huskey, R.J., 1974. Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. Genetics 78, 989–1014.

Saren, A.M., Ravantti, J.J., Benson, S.D., Burnett, R.M., Paulin, L., Bamford, D.H., Bamford, J.K.H., 2005. A snapshot of viral evolution from genome analysis of the *Tectiviridae* family. J. Mol. Biol. 350, 427–440.

Singer, B.S., Gold, L., Gauss, P., Doherty, D.H., 1982. Determination of the amount of homology required for recombination in bacteriophages T4. Cell 31, 25–33.

Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V., Kozyavkin, S.A., 2002. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. Proc. Natl. Acad. Sci. U.S.A. 99, 4644–4649.

Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., Chisholm, S.W., 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. PLoS Biol. 3, e144.

Tétart, F., Desplats, C., Krisch, H.M., 1998. Genome plasticity in the distal tail fiber locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J. Mol. Biol. 282, 543–556.

Tétart, F., Desplats, C., Kutateladze, M., Monod, C., Ackermann, H.-W., Krisch, H.M., 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J. Bacteriol. 183, 358–366.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tock, M.R., Dryden, D.T.F., 2005. The biology of restriction and anti-restriction. Curr. Opin. Microbiol. 8, 466–472.

Waterbury, J.B., 1999. The cyanobacteria—Isolation, purification and identification. In: Dworkin, M. (Ed.), The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. Springer-Verlag, New York (http://141.150.157.117:8080/prokPUB/index.htm).

Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.