Virology 477 (2015) 155–163

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

A structured annotation frame for the transposable phages: A new proposed family "Saltoviridae" within the *Caudovirales*

Chantal Hulo [a], Patrick Masson [a], Philippe Le Mercier [a], Ariane Toussaint [b,*]

[a] Swiss-Prot Group, Swiss Institute of Bioinformatics, Centre Médical Universitaire, CH-1211 Geneva 4, Switzerland
[b] Université Libre de Bruxelles, Génétique et Physiologie Bactérienne (LGPB), 12 rue des Professeurs Jeener et Brachet, 6041 Charleroi (Gosselies), Belgium

ARTICLE INFO

Article history:
Received 29 July 2014
Returned to author for revisions
2 October 2014
Accepted 6 October 2014
Available online 12 December 2014

Keywords:
Phage protein annotation
Phage taxonomy
Caudovirales
"Saltoviridae"
"Myosaltovirinae"
"Siphosaltovirinae"

ABSTRACT

Enterobacteriophage Mu is the best studied and paradigm member of the transposable phages. Mu-encoded proteins have been annotated in detail in UniProtKB and linked to a controlled vocabulary describing the various steps involved in the phage lytic and lysogenic cycles.

Transposable phages are ubiquitous temperate bacterial viruses with a dsDNA linear genome. Twenty-six of them, that infect α, β and γ-proteobacteria, have been sequenced. Their conserved properties are described. Based on these characteristics, we propose a reorganization of the *Caudovirales*, to allow for the inclusion of a "Saltoviridae" family and two newly proposed subfamilies, the "Myosaltovirinae" and "Siphosaltovirinae". The latter could temporarily be included in the existing *Myoviridae* and *Siphoviridae* families.

Introduction

Escherichia coli bacteriophage Mu was discovered and named in the early 1960 by A.L (Larry) Taylor, when he readily recognized the phage mutagenic properties and demonstrated that this resulted from the phage capacity to integrate and block expression of many if not any gene in the chromosome of its bacterial host (Taylor (1963), reviewed in Symonds et al. (1987) and Harshey (2012)). Over 20 years, genetic and molecular characterization of Mu revealed its unique properties, which among other developments lead to the engineering of a number of very widely used "genetic tools" for *E. coli* and other Enterobacteria (Akhverdyan et al., 2011; Ferrieres et al., 2010 for a recent development; Van Gijsegem et al., 1987). In the middle 80's, the phage became the paradigm for the biochemical in vitro deciphering of the molecular mechanism of transposition catalyzed by the "DDE recombinases" (reviewed in Harshey, 2012). This paved the way for the understanding of retroviral integration, V(D)J recombination and many other DNA transposition reactions and culminated with the resolution of the 3D structure of a transposase-Mu ends DNA complex (Montano et al., 2012). Along a similar line, the Mu host range variation system, which consists of gene coding for a serine-based site-specific recombinase (Gin) that catalyzes the inversion of a phage DNA segment covering the phage tail fibers (the G region), became the focus of in vitro studies with its close orthologues, the *Salmonella* antigenic variation system H-Hin (Johnson, 1991) and the γ−δ transposon resolvase Res (Grindley et al., 2006).

Transposable phages started to become a distinct clade when Victor Krylov isolated *Pseudomonas* mutator phages B3 and D3112 (Krylov et al., 1980; Akhverdian et al., 1985). More recently genomic sequencing led to the recognition of a growing group not only of phages but also of many prophages in many different bacterial taxons (Lima-Mendez et al., 2011 and references therein). Analysis of the genomic content of all these (pro)phages shows that although the general genome organization, the proteins required for transposition and its regulation and a few head proteins are conserved only within the group, the rest of the head, the base plate, the tail and the fiber proteins have orthologues in other bacteriophages (Lima-Mendez et al., 2011; Toussaint, 2013).

In view of the large potential of transposable phages for engineering genetic tools and the probable very large impact they have on the evolution of their host, it is important that these genomes are readily recognized and correctly annotated. Therefore, within the general frame of the development of an ontology for phage protein annotation, Mu-encoded proteins have been annotated in UniProtKB and in ViralZone (http://viralzone.expasy.org); (Hulo et al., 2011). The viral particle and the various steps involved in the phage lytic and lysogenic cycles have been illustrated in the general frame of viral biological processes (http://viralzone.expasy.org/all_by_species/507.html), and linked to a controlled vocabulary describing the various

* Corresponding author. Tel.: +32 2 3546290; fax: +32 2 650 97 70.
E-mail address: Ariane.Toussaint@ulb.ac.be (A. Toussaint).

http://dx.doi.org/10.1016/j.virol.2014.10.009
0042-6822/

developmental steps. In this paper we shall review this information with the aim to discuss how it will help to recognize and robustly annotate transposable phages and prophages.

Despite the fact that the group includes phages with flexible and contractile tails (siphoviruses and myoviruses respectively), we propose that the group is recognized and organized into a single taxonomic family among the order *Caudovirales.*

The Mu genome and lytic cycle

In the viral particle, the Mu genome is a ~40 kbp linear double stranded DNA, consisting of 38 kbp of actual viral DNA, flanked by variable host DNA sequences 100–150 bp on one side defined as left, and 1–1.5 kbp on the right end (Fig. 1). This particular configuration results from packaging of viral DNA by a regular full head mechanism, but which proceeds on randomly integrated copies spread through the host genome (Fig. 2). These accumulate during replication of the viral DNA (Fig. 1 and below for more details). The 38 kbp phage moiety ends display a 2 bp 5'-TG-CA-3' inverted repeat (Kahmann and Kamp, 1979). It flanks 56 ORF's, 36 of which correspond to genes coding for proteins identified by classical genetics, most of which have been assigned with a function (and hence a biological process (Fig. 1 and http://www.uniprot.org/uniprot/?query=AF083977+AND+phage&sort=score for the functional annotation of Mu proteins and references). The 3D structure of only a few of these proteins has been resolved (the DNA binding moiety of Repc, Ner, MuA, and MuB, the transcriptional activator Mor/GemB, gene product Mu44 the central hub component of the base plate, part of the baseplate puncturing device Mu45 and MuGin (see Suppl. Table 1 for PDB IDs and references)).

Eight functional blocks can be recognized along the genome (Fig. 1): from left to right the early regulatory block (genes *repc-ner*), the integration–replication block (*A-B* genes), a stretch of genes, most of which have unknown functions originally called "semi-essential region" (because mutants in these genes were viable though somewhat debilitated/attenuated), the lysis, head, tail and fibers blocks and finally the host restriction evasion and modification block (*com-mom*). Besides Repc and Ner, which regulate the early genes (first three blocks), two regulators, Mor and MuC respectively, positively regulate the middle (4th block) and late (4 latter blocks) genes at the transcriptional level. Com positively regulates Mom synthesis at the translational level.

Important protein binding sequences are the phage ends *attL* and *attR* recognized by MuA, the *pac* site where packaging initiates, operators *O1-O3*/*IAS* recognized by Repc, Ner and MuA, the repressor/latency promoter p_{CM} and early promoter P_e in *O1-O3*, the middle promoter P_m downstream from *mor*, the late promoters P_{lys}, P_I and P_P, the central gyrase binding site *SGS*, the IR inverted repeat flanking the fiber genes cassette recognized by Gin and the P_{mom} promoter (see their location in Fig. 1). How these genes/gene products and sites articulate the lytic cycle, lysogeny switch and the known interactions with the host is described in Figs. 2, 3 and 4.

Mu transposition

The most notable characteristic of Mu is its mode of integration and replication by conservative and replicative transposition respectively. Briefly, both processes involve synapses of the phage DNA ends by a MuA DDE-recombinase hexamer, which binds to 3 sites each in *attL* and *attR* and to the *IAS*. After infection, the phage structural MuN protein binds the viral DNA ends, most probably during ejection, helping to bring the ends together to generate a non-covalent circular DNA substrate for MuA binding (Gloor and Chaconas, 1986; Harshey and Bukhari, 1983). After induction and during replication, viral DNA ends are brought in close proximity by host gyrase binding to the *SGS* located in the middle of the viral genome (Oram et al., 2006 and references therein). Host nucleoid associated proteins HU and IHF are also essential for synapses. They respectively bind to the viral left end and the IHF biding site in the operator/*IAS* region. A MuA tetramer is then retained in the assembled protein–DNA complex, forming the stable synaptic complex (SSC) transpososome, and a nick is introduced at each of the phage ends (TG–CA repeat) to generate 3'-OH residues. MuB binds host DNA, allowing for interaction of the transpososome with the target DNA. MuA catalyzes ligation of the 3'-OH ends to target DNA, at a 5 bp interval on the two attacked strands (staggered cut). This releases free 3'-OH ends at the target, which are used for DNA repair synthesis (conservative transposition leading to integration) or PriA-primed DNA

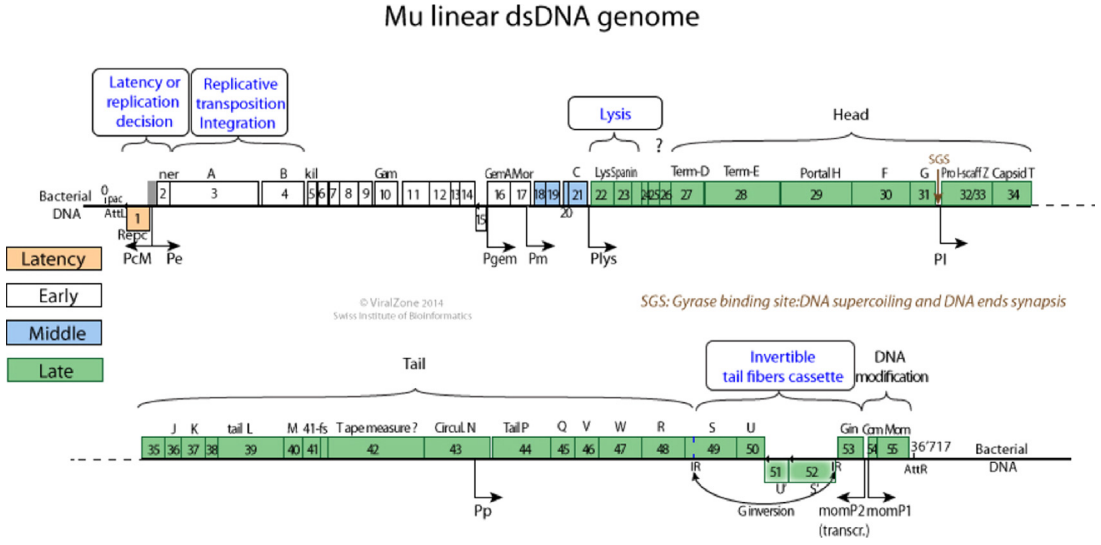

Fig. 1. Genomic map of bacteriophage Mu. http://viralzone.expasy.org/all_by_species/4356.html. See text for details. Segments of the viral genome that constitute transcriptional units (latency from P_{cM}, early from P_e, middle from p_m and late from P_{lys}, P_I, P_P and P_{mom}) are shown in different colors. The random host sequences flanking the ends are lost upon integration. The viral and prophage DNA and genetic maps are collinear (Abelson et al., 1973). *SGS* is the central high affinity host gyrase binding site, which is important for end synapses and transpososome assembly (Oram et al., 2006).

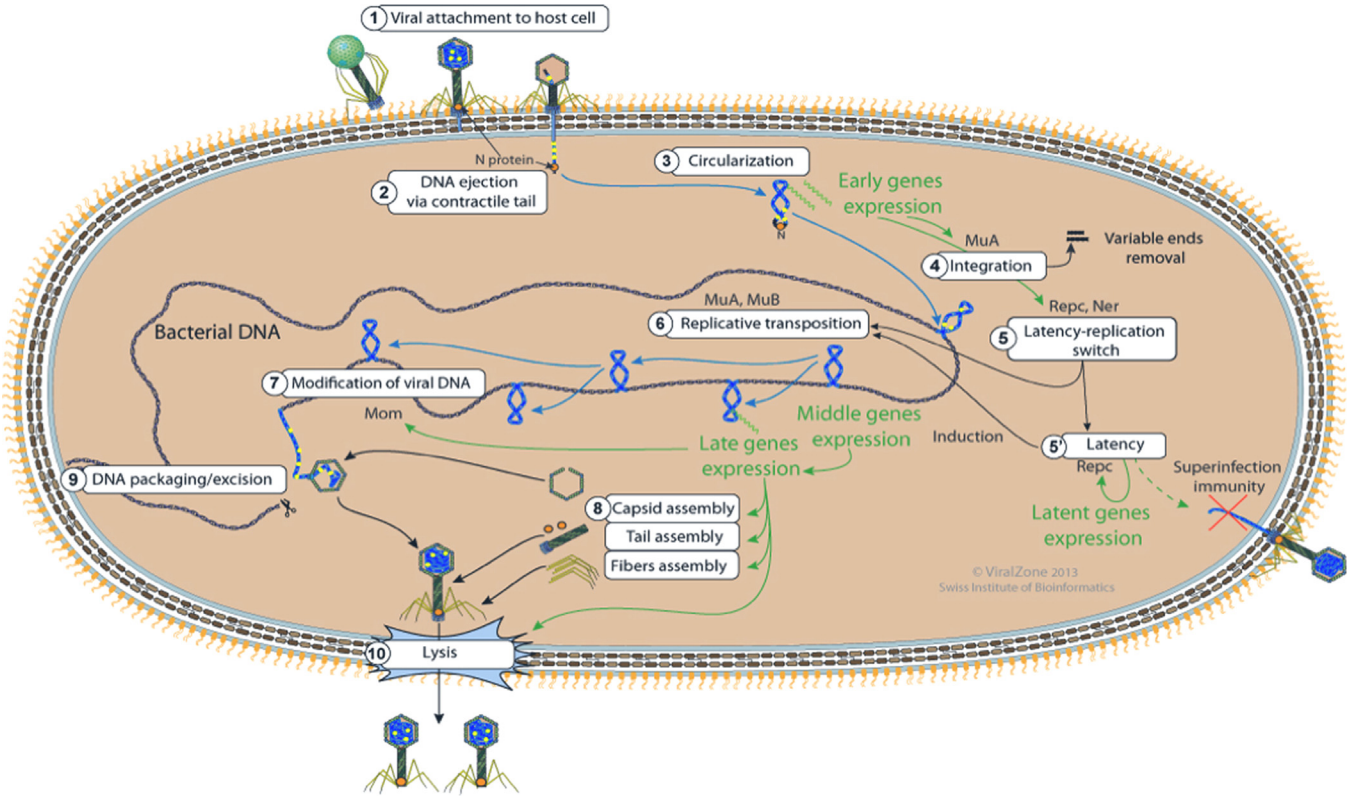

Fig. 2. The Mu lytic cycle (http://viralzone.expasy.org/all_by_species/4277.html). 1. The virion fibers attach to the host cell surface lipopolysaccharides (LPS) thereby initiating infection. 2. Upon binding to the host cell surface, the baseplate changes its conformation and triggers sheath contraction, driving the rigid internal tail tube through the cell envelope leading to viral genome entry. See text for details about Mu viral DNA properties. 3. The MuN protein, which is present in the virion, is ejected with and binds to the viral DNA to circularize it in a non covalent form.. 4. Early transcription ensures synthesis of at least the Repc and Ner repressors and DDE recombinase MuA, which catalyzes integration. Flanking bacterial sequences are removed from the viral genome during integration. Integration of the Mu genome is required before any decision between latency and lytic replication is occuring. 5. The Repc/Ner repressors ratio determines whether the phage enters latency or lytic cycle (replication). Repc represses the early promoter P_e thereby establishing latency. Ner represses Repc expression thereby promoting early genes expression leading to the onset of viral replication, and P_e to keep early protein expression independent of copy-number. 6. Upon Repc inactivation, Ner, the DDE recombinase MuA and the target binding/activator ATPase MuB are expressed. MuA catalyzes transpososome assembly, cutting and ligation of the viral genome ends to target host DNA. This viral-host DNA structure is resolved by target-primed replication leading to two integrated copies of the viral genome in a process called replicative transposition. Target selection depends upon MuB. Successive rounds of replicative transposition lead to the accumulation of up to 50–100 copies of the viral into the host genome, which as a result, is profoundly reorganized. 7. Late transcription allows for the expression of an adenine modification enzyme, which makes the phage and its host DNA resistant to some restriction enzymes. New viral particles are thus partially protected from host restriction during subsequent infections. 8. Structural genes are expressed in the late phase leading to the assembly of: (a) empty capsids: the portal protein and the viral protease MuI probably form the early 25S initiator complex. The scaffolding protein MuZ and the capsid protein MuT are added to the initiator complex to form the immature procapsid. The viral protease cleaves the portal protein making the procapsids competent for DNA packaging. (b) fibers: Mu can change its fibers specificity through a site-specific inversion of a genome segment catalyzed by the virally encoded invertase Gin. This genome segment codes for two alternate sets of tail fibers/fibers assembly proteins thereby expanding the host range of the phage. (c) tails: tail tube proteins polymerize on the tape-measure protein and the tail terminator Mu37 stops the assembly once it has reached the correct length. A similar process allows polymerization of the sheath protein around the tail tube. 9. Maturation of viral DNA proceeds from the viral *pac* site located near the Mu left end. The bacterial DNA is cut 50–150 bp upstream of the left of the integrated Mu genome. The second cut occurs once the phage head has been filled. Since there is a bit more space in the procapsid than required for the viral genome, 1–1.5 kbp of bacterial DNA downstream of the right end of the Mu genome is also packaged with the viral genome. Since each Mu genome is packaged from a different site in the bacterial genome, the host DNA on Mu ends is unique in each phage head such that, in a Mu lysate, the whole host genome is represented. 10. The newly synthesized virions are released by lysis.

replication (replicative transposition or transpositional replication, see Figs. 2 and 3). In both cases, the 5 base pairs single stranded DNA resulting from the staggered cut target is filled in, generating the duplication of 5 base pairs, which as a result flanks the replicated phage genome copies. After conservative integration, the transposition intermediate is repaired by transposase-mediated resection of the 5' flaps attached to the ends of the incoming phage genome, followed by filling the remaining 5 bp gaps at each end by recombinational repair (Jang et al., 2012). During replicative transposition, depending on the way the phage DNA nicked 3'-OH ends attack the target, the segment between the integrated viral genome and the target will be either deleted (binding to the same strand) or inverted (binding to the opposite strand, shown in Fig. 3). In case transposition targets a distinct DNA replicon, a plasmid for instance, the donor and target replicons will be fused by means of two copies of Mu in the same orientation (replicon fusion, Shapiro, 1969).

This particular mode of replication directs a later step of phage assembly, packaging, which has to proceed on integrated replicas of phage DNA. This and features that are a consequence of those modes of replication and packaging, as the presence of variable host DNA sequences at viral DNA ends, are expected to be conserved feature in transposable phages and are outlined below.

Transposable phages in 2014

E. coli phage D108 and *Pseudomonas* phages B3 and D3112 were the first Mu-like phages described, although their genome sequences and phenotypes were only recently published (Braid et al., 2004; Kim et al., 2012; Kropinski A.M et al. deposited in 2009 see sequence accession number in Table 1). B3 revealed a slightly different genome organization, with an inversion in the early region such that *mor* is the left end gene, the absence of

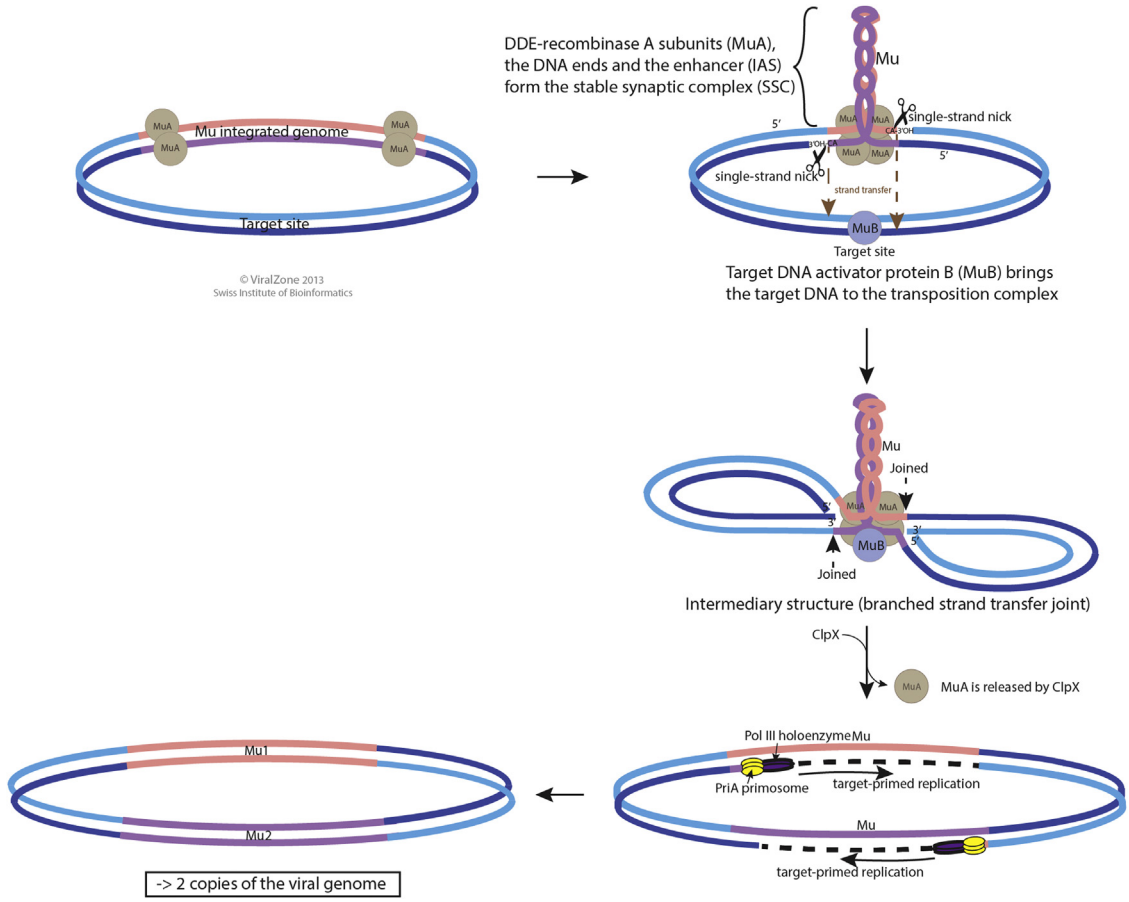

Fig. 3. Mu conservative and replicative transposition. http://viralzone.expasy.org/all_by_species/4017.html. The figure summarizes data accumulated through the years by in vitro studies using mini-Mus, i.e. Mu derivatives that only retain Mu transposase binding sites. The overlapping of the *IAS* with the operator region allows for Repc to regulate transpososome assembly by competitive binding (see text for more details). The whole viral genome is represented here with the *SGS* at the apex of the supercoiled Mu DNA with its ends synapsed. See text for some additional details.

conservation of the immunity repressor, the tail structure and the tail fibers and no invertible region (D3112 and B3 are *Siphoviridae*, which infect *Pseudomonas*). More transposable phage candidates, with various levels of similarity with Mu, were sequenced more recently, forming so far a group of 26 viruses (Tables 1 and 2). Accessibility to complete bacterial genome sequences also allowed for the search of Mu-related prophages, some of which have been propagated as viral particles (see for instance Fouts et al. (2005), Parker et al. (2006)), some of which have been shown to generate chromosomal rearrangements (Scott et al., 2007). All together these form now a group of viruses infecting or residing as prophages in representatives of all proteobacterial phyla and some Firmicutes (Lima-Mendez et al., 2011; Toussaint, 2013).

Sequence comparisons between all these genomes allow for listing a set of properties that characterize the group, members of which would all benefit from an annotation consistent with that described above for Mu.

Conserved features of the genome – how to recognize a transposable phage

The vast majority (possibly all) of the sequenced transposable phage genomes include a “phage” moiety 37–39 kbp long (Vibrio phage Martha 12B12 standing as an exception), delimited by a 5'TG-CA3' repeat, and variable ends (VE) of random host DNA segments. Transposable prophages have this terminal repeat usually flanked by a direct repeat of 5 bp of the host target sequence, a landmark of integration by conservative transposition. These features can be verified by viral genome sequencing, by analysis of restriction enzyme digests of viral DNA (end fragments are heterogeneous and hence, depending on the enzyme used, either spread over the gel and do not appear or appear as a fuzzy band) and by electron microscopy (Hsu and Davidson, 1974). The existence of VE's should of course be kept in mind when finishing and assembling the viral genome sequence. As seen in Table 1, not all genomes display the expected TG–CA end repeat. This may be real but could as well be due to missing terminal fragments in the clones sequenced, due to their heterogeneity.

Integration is a pre-requisite for replication. Therefore, although this has yet to be demonstrated even for Mu, the lytic-lysogeny regulatory switch (detailed in Fig. 4) ought to operate after the first integration event for all phages in the group.

After induction of a transposable prophage, viral-host DNA junctions remain present and detectable (by Southern blot hybridization or by PCR), while new junctions progressively accumulate as a result of rounds of replicative transposition (Ljungquist and Bukhari, 1979; Rehmat and Shapiro, 1983).

As a result of random site selection for integration, lysogens include a significant number of mutants (e.g. 1% auxotrophs in the case of *E. coli* lysogenic for Mu, only catabolic mutants for D3112). This property so far experimentally verified for only a fraction of the concerned phages (Table 1), should be used as a criterion for transposable phage identification.

The conserved modes of transpositional replication and packaging from integrated viral DNA copies are correlated with conserved phage proteins e.g. the DDE recombinase (a transposase), the associated regulatory ATPase, the portal protein, probable portal associated proteins and an early protein of unknown function (Lima-

Table 1
Completely sequenced transposable phages and some of their recognizable features.

Phage name	Length (bp)	ca_tg nucleotides	Aux	VE	Genome organization	NCBI sequence ID	Tail
Enterobacteria phage Mu	36,717	Y	Y	Y	Mu	NC_000929.1	M
Escherichia phage D108	37,235	Y	Y	Y	Mu	NC_013594.1	M
Burkholderia phage BcepMu	36,748	Y			B3	NC_005882.1	M
Burkholderia phage KS10	37,635	No?			PhiE255-like	NC_011216.1	M
Burkholderia phage phiE255	37,446	Y			PhiE255	NC_009237.1	M
Haemophilus phage SuMu	37,151	No?			Mu	NC_019455.1	M
Rhizobium phage RR1-B	37,378	Probably			Mu	NC_021557.1	M
Vibrio phage Martha 12B12	33,277	No?			Mu	NC_021070	M?
Rhodobacter phage RC1	39,573	Probably			Mu	NC_020839.1	S
Rhodobacter phage RcapMu	39,283	Y			Mu	NC_016165.1	S
Pseudomonas phage MP22	36,409	Y			Mu	NC_009818.1	S
Pseudomonas phage MP29	36,632	Y			Mu	NC_011613.1	S
Pseudomonas phage MP38	36,885	Y			Mu	NC_011611	S
Pseudomonas phage MP42	36,847	Y			Mu	NC_018274.1	S
Pseudomonas phage B3	38,439	Y	Y	Y	B3	NC_006548.1	S
Pseudomonas phage D3112	37,611	Y	Y[a]	Y	Mu	NC_005178.1	S
Pseudomonas phage DMS3	36,415	Y			Mu	NC_008717.1	S
Pseudomonas phage JBD5	37,740	Y			Mu	NC_020202.1	S
Pseudomonas phage JBD18	39,014	Probably			B3	JX495041.1	S
Pseudomonas phage JBD24	37,095	Y			Mu	NC_020203.1	S
Pseudomonas phage JBD25	39,552	No?			B3	JX495042.1	S
Pseudomonas phage JBD26	37,840	Y			Mu	JN811560.1	S
Pseudomonas phage JBD30	36,947	Y			Mu	NC_020198.1	S
Pseudomonas phage JBD67	38,232	Probably			B3	JX495043.1	S
Pseudomonas phage JBD88a	36,429	Probably			Mu	NC_020200.1	S
Pseudomonas phage LPB1	36,814	Y			Mu	HE584812.1	S

bp: base-pairs; Aux: induces auxotrophic mutations; VE: has variable ends of host DNA; M: contractile tail; S: non-contractile tail; No?: not visible on sequence but could be a sequencing problem.

[a] D3112: see Krylov et al. (1980), Pleteneva et al. (1999), Rehmat and Shapiro (1983) for a detailed analysis of D3112 induced mutagenesis and fate of prophage ends and internal restriction fragments after prophage induction.

Table 2
Myoviruses and siphoviruses in the newly proposed taxonomic family of "Slatoviridae".

Myosaltovirinae	Siphosaltovirinae
Enterobacteria phage Mu (2)[a]	Pseudomonas phage MP22 (1)
Escherichia phage D108 (2)	Pseudomonas phage D3112 (1) (AC)
Haemophilus phage SuMu (1, 2)	Pseudomonas phage DMS3 (1) (AC)
Burkholderia phage KS10 (2)	Pseudomonas phage JBD24 (1) (AC)
Pseudomonas phage PhiE255 (3, 4)	Pseudomonas phage JBD26 (1) (AC)
Burkholderia phage BcepMu (3, 4)	Pseudomonas phage JBD30 (1) (AC)
Rhizobium phage RR1-B (1, 4)	Pseudomonas phage JBD5 (1) (AC)
Vibrio phage Martha 12B12 (2)	Pseudomonas phage LPB1 (1)
	Pseudomonas phage MP29 (1) (AC)
	Pseudomonas phage MP38 (1)
	Pseudomonas phage MP42 (1)
	Pseudomonas phage JBD88a (1) (AC)
	Pseudomonas phage JBD25 (1, 3)
	Pseudomonas phage JBD18 (1, 3)
	Pseudomonas phage B3 (3)
	Pseudomonas phage JBD67 (1, 3)
	Rhodobacter phage RcapMu (1, 2, 4)
	Rhodobacter phage RC1 (1, 2)

[a] Parenthesis indicate the cluster(s) to which the phage belongs (see text and Fig. 6 for more details), (AC) codes for anti-CRISPR proteins.

Mendez et al., 2011; Toussaint, 2013), orthologues of MuA, B, H, F, G Mu36 and GemA respectively. These and the genome organization are the main conserved features when considering the whole set of transposable (pro)phages identified so far.

Non-conserved parts of transposable phage genomes

As further described below, conservation varies widely within the group and there are features that, although at first view they appear key to viral development, are not conserved. This is for instance the case of the MuN, which appears essential for transpososome assembly before random integration. Only half of the phages in Table 1 code for a MuN orthologue (ACLAME family:vir_proph:136 and Suppl. Table 2 genomes/prot-families). Viral genome ends synapses is thus likely to follow alternative routes for the other half. Around this process of transpososome assembly, another feature established for Mu appears unique. The similarity between the N-terminal operator/*IAS* binding domains of immunity repressor Repc (an orfan protein so far, except for its similarity with MuA) and MuA does not seem to exist in other phages in the group (data not shown). Transpososome assembly and its regulation may thus also follow different routes, which remain to be elucidated.

Another particularity of Mu is the presence of a middle operon, which is absent in most if not all other genomes in the set (data not shown).

Genome features shared with other phage clades

As listed above, conserved features are quite limited. Many proteins encoded by transposable phages have orthologues in other phages as well as in bacterial chromosomes, as first illustrated by the discovery of the similarities between Mu Gin and the Salmonella Hin phase variation serine-based site-specific recombinases. These proteins catalyze inversion of unrelated DNA segments. If one looks at families of conserved phage proteins as those in the ACLAME database (http://aclame.ulb.ac.be/), it is obvious that besides the specific proteins described above, whether regulatory or structural, a large fraction of them belong to very large families, many of which can be assigned with a function including interaction with the bacterial host, evasion of host defenses, positive and negative transcriptional regulation, lysis, head and tail assembly and structures. Mu tail is for instance related to that of phage P27, and in general, contractile tails of transposable (pro)phages are similar with those of other phages (Lima-Mendez et al., 2011). Flexible tails

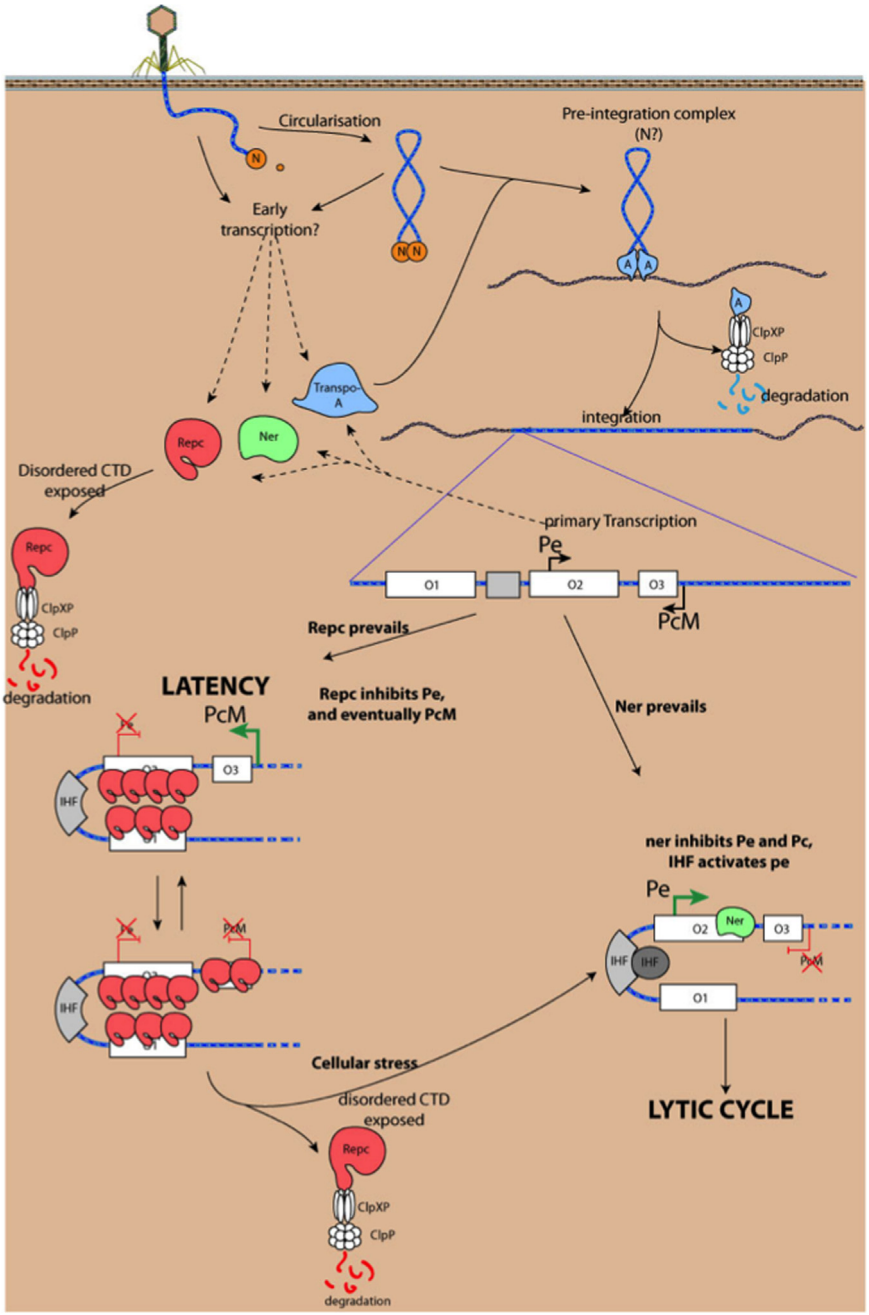

Fig. 4. The lytic-lysogeny switch after Mu infection. http://viralzone.expasy.org/all_by_species/4516.html. The figure summarizes data obtained over many years in vivo and in vitro on immunity repressor Repc, its sequence, mutants, conformational changes, DNA binding properties and interaction with the host ClpXP protein. As mentioned in the text, the switch has to occur after the first integration event required for both lysogenization and lytic replication. Since integration requires MuA synthesis that is driven by P_e, and Repc is transcribed from P_{cM}, a switch ought to occur from P_e to P_{cM} directed transcription. There is yet no detailed information available on how this is occurring. Binding of Integration Host Factor (IHF) to its cognate site, bends the *O1-O3* region (not represented here), enhancing Repc binding. The host HNS and Fis proteins also influence Repc binding to the operators (Ranquet et al., 2005 and references therein).

are more homogenous (most of these phages infect the same host *Pseudomonas aeruginosa*) but nevertheless similar to tails of non-transposable phages as well (illustrated by ACLAME family:vir_proph:534 and 629 for instance). Many head proteins (e.g. terminase large subunit, major capsid potein) also belong to protein families that include non-transposable phage proteins (ACLAME family:vir_proph:60 for instance).

Maybe less surprising because they ensure recognition of the host, host range/tail fiber proteins do the same. The G region of Mu has orthologues e.g. in phage P1 and the cryptic *E. coli* K12 prophage e14 (Hiestand-Nauer and Iida, 1983; Plasterk and van de Putte, 1985). More unexpected, the wider host variation system DGRE, first identified on *Bordetella* phage BBP1 (Doulatov et al., 2004) is present on some transposable prophage genomes at the same location occupied by the G region in Mu (Toussaint, 2013). Nevertheless, a comprehensive analysis of correlations between tail fibers of phages from various taxonomic groups infecting the same host has yet to be achieved.

A new taxonomic grouping for transposable phages?

Despite the relatively low number of characteristic features unique to transposable phages, because of the large impact they must have on the stability and evolution of their hosts as a result of their mode of integration and replication that introduce insertional mutations and

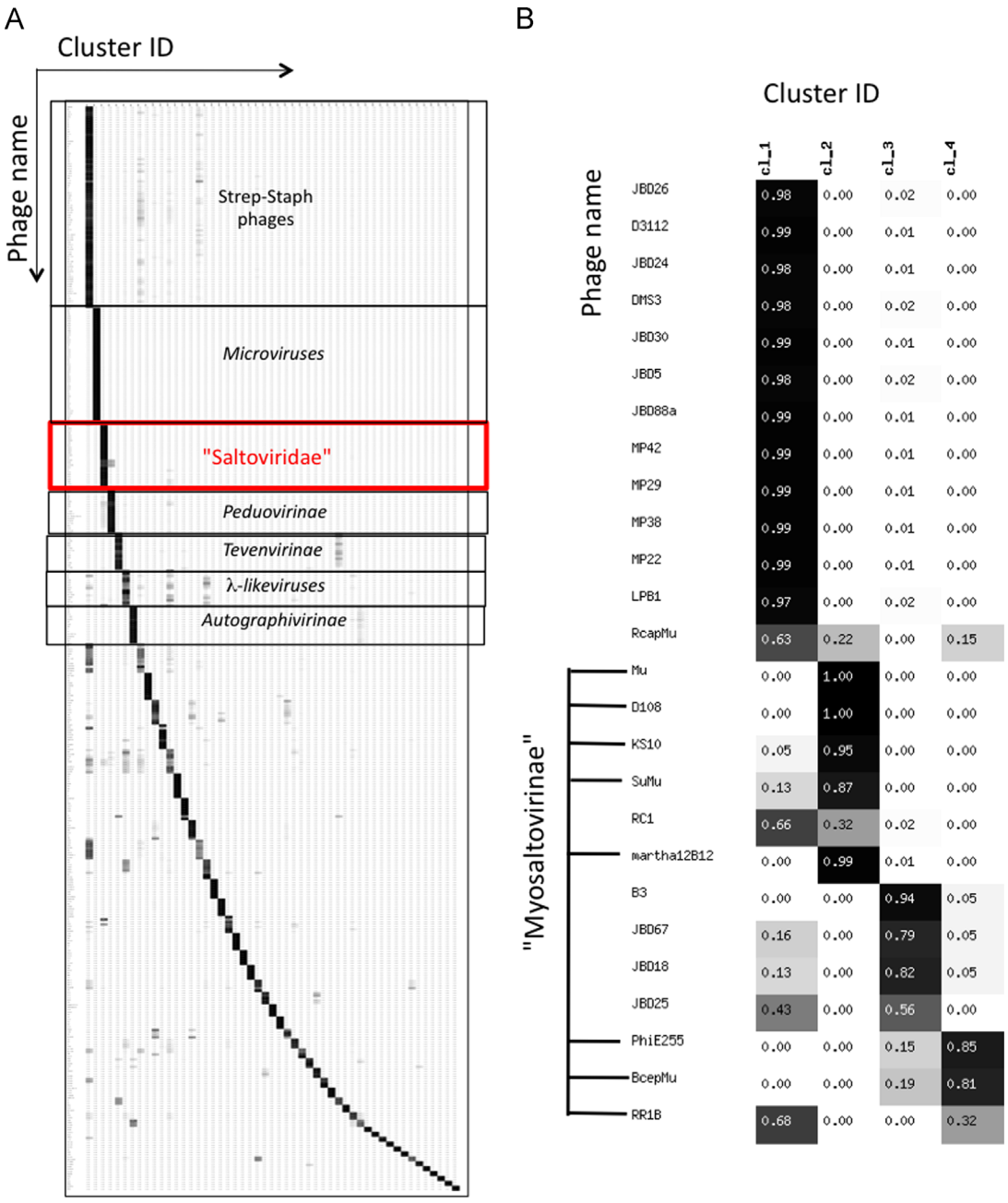

Fig. 5. Heatmap representation of transposable phages relationships with other phages and among themselves. A: Heatmap representation of the result of fuzzy clustering of 476 phage genomes, 26 of which are the transposable phages listed in Table 1. Each genome was expressed as the set of ACLAME families of proteins matching the genome encoded proteins (Supp. Table 1, see text for more details on the protocol). Over 50 clusters were obtained. The gray boxes clearly emphasize the belonging of some genomes to several clusters. Transposable phages ("Saltoviridae") form a single group (Cluster 3), but some display similarity with phages in Cluster 4 (*Peduovirinae*). B: Heatmap representation of the result of fuzzy clustering of the 26 transposable phages listed in Table 1. The method used is the same as in A. *Pseudomonas* infecting phages with Mu-like genome organization and non-contractile tails group in Cluster 1, those with B3 organization in Cluster 3. Phages with a contractile tail are in Clusters 2 and 4. Several genomes clearly split between several clusters. "Myosaltovirinae" are indicated, the other phages belong to the "Siphovirinae" subfamily.

chromosomal rearrangements, we think that these phages deserve recognition as a single taxonomic family among the order *Caudovirales*. Despite the fact that transposable phages include representatives within siphoviruses and myoviruses, we propose a new family "Saltoviridae" (from the latin *Saltus*, jump and salto jumping figures in French), to include the subfamilies "Siphosaltovirinae" and "Myosaltovirinae". The latter two could transiently be part of the *Siphoviridae* and *Myoviridae* families respectively.

Why would this be beneficial? One reason we think is that a significant number of bacterial genomes contain one and often more than one copy of a transposable prophage. Even in the absence of transposase expression, these become portable regions of homology for Rec mediated recombination. A clonal bacterial culture will thus evolve to contain chromosomes with different organizations in various proportions due to recombination between these prophages. This appears to generate problems for assembling genomic sequences. A quick survey of partially assembled genomes of bacteria containing transposable prophage(s) reveals that the prophage ends are usually missing, while all internal fragments pile up into a single contig! If the prophage could be readily identified as a "Saltoviridae", thanks to a robust annotation based on that of Mu and other paradigm phages and their characteristic features, possible genome heterogeneity and assembly problems could be pinpointed more easily.

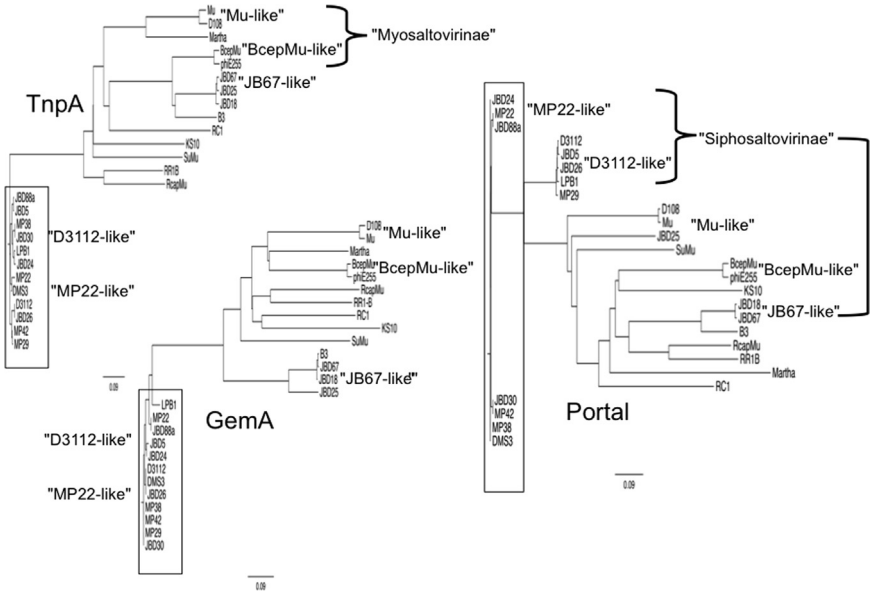

Fig. 6. "Saltoviridae" subfamilies as they appear defined by protein–protein relationships using transposases TnpA, MuGemA-like and portal proteins. The sets of 26 GemA orthologues, DDE recombinases and portal proteins were aligned using ClustalW2 (EMBL-EBI webserver, (Larkin et al., 2007)) with default parameters. The trees were generated from those alignments, using Phylogenetic Tree (same webserver), which provides neighbour-joining trees without distance corrections, again default parameters. The limits between groups are set arbitrarily.

The "Saltoviridae" family

As mentioned earlier, 26 sequenced viruses form the "Saltoviridae" family, (Table 1), including 8 "Myosaltovirinae" and 18 "Siphovirinae" all of the latter infecting *Pseudomonas*, eight of which code for the only anti-CRISPR proteins described so far (Bondy-Denomy et al., 2013).

The level of similarity within the group is extremely variable, going from over 90% nucleotide identity between Mu and D108 or MP22 and JBD88a to around 55% overall amino-acid identity between for instance the Mu and B3 portal proteins (data not shown) but with a well conserved overall genome organization (Sup Fig. 1).

As a preliminary analysis to support the existence of the "Saltoviridae" family and subfamilies, we used the 2-step fuzzy clustering method described by Lima-Mendez and co-workers (Lima-Mendez et al., 2011) and available online at http://rsat.ulb.ac.be/rsat/index_neat.html, (Lima-Mendez, 2012). The method generates a reticulate classification where each phage is associated with a membership vector, which quantitatively characterizes its membership to a set of clusters. We used the protein families defined in ACLAME version 0.4, which includes proteins from 457 phages infecting various bacteria. Each viral genome was represented by its set of protein families. Proteins coded by 19 transposable phages not included in ACLAME version 0.4 were compared to all phage and predicted prophage proteins in the database. The family corresponding to each best hit was retained and used to represent the 19 genomes by their protein families (Suppl. Table 2). These were added to the 457 set. The result shows that the "Saltoviridae" indeed form a well separated group, as do other well known families (Fig. 5A). When the same method is applied to the sole 26 "Saltoviridae" genomes (Fig. 5B), 4 clusters emerge. Clusters 1 and 3 include the *Pseudomonas* siphoviruses and Cluster 2 and 4 the myoviruses. This fits our proposed "Myosaltovirinae" and "Siphosaltovirinae" subfamilies. Several genomes are more or less mosaic and partition between different clusters.

To have a preliminary idea of possible genera in the subfamilies, multiple sequence alignments were generated for the most conserved proteins, transposases, portal proteins and GemA and used to build trees. As shown in Fig. 6 some consistent groups appear, that could constitute genera. The largest one includes *Pseudomonas* infecting "Siphosaltvirinae", among which, D3112. Elsewhere in this issue, Adriaenssens et al. propose a whole series of new siphovirus subfamilies, including a "D3112-like" subfamily, which, in the present International Committee for Virus Taxonomy (ICTV) schema, obviously corresponds to this large possible genus.

Conclusions

A robust system for annotating phage and prophage-encoded proteins was so far not available. ViralZone now offers a set of definitions included in a structured ontology to be soon incorporated in GO (Giglio et al., 2009). It will hopefully contribute to a more homogeneous and informative annotation of phage and prophage proteins in a near future.

The presence of 26 sequenced phage genomes with characteristic features for being related to Mu by their mode of transpositional replication and packaging from integrated genome copies, calls for the creation of a proper taxonomic family. Because two morphological phage types belong to the group, which is not unique to this group among the *Caudovirales*, we propose the new "Saltoviridae" family, with subfamilies of "Myosaltovirinae" and "Siphosaltovirinae". In order to accommodate this classification, although this goes against the traditional existence of the *Myoviridae* and *Siphoviridae* families, we propose to move the morphotypes one step down in the phage taxonomic hierarchy. Such a move would, we think, help to reconcile a significant number of discrepancies between phylogeny and morphology that has been rampant in the field for years (e.g. Lawrence et al., 2002). Resolving this problem will also resolve inconsistencies in annotations and organization in databases such as GeneBank, Conserved Domains or Pfam, the primary structure of which rests on taxonomy.

Acknowledgments

Work in Geneva benefited from the Swiss Federal Government through the State Secretariat for Education.

Appendix A. Supplementary materials

Supplementary materials associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2014.10.009.

References

Abelson, J., Boram, W., Bukhari, A.I., Faelen, M., Howe, M., Metlay, M., Taylor, A.L., Toussaint, A., Van de Putte, P., Westmaas, G.C., Wijffelman, C.A., 1973. Summary of the genetic mapping of prophage Mu. Virology 54, 90–92.

Akhverdian, V.Z., Khrenova, E.A., Reulets, M.A., Gerasimova, T.V., Krylov, V.N., 1985. Characteristics of phages-transposons of *Pseudomonas aeruginosa* belonging to 2 groups distinguished by DNA–DNA homology. Genetika 21, 735–747.

Akhverdyan, V.Z., Gak, E.R., Tokmakova, I.L., Stoynova, N.V., Yomantas, Y.A., Mashko, S.V., 2011. Application of the bacteriophage Mu-driven system for the integration/amplification of target genes in the chromosomes of engineered Gram-negative bacteria–mini review. Appl. Microbiol. Biotechnol. 91, 857–871.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., Davidson, A.R., 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429–432.

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

Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., Simons, R.W., Zimmerly, S., Miller, J.F., 2004. Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. Nature 431, 476–481.

Ferrieres, L., Hemery, G., Nham, T., Guerout, A.M., Mazel, D., Beloin, C., Ghigo, J.M., 2010. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. J. Bacteriol. 192, 6418–6427.

Fouts, D.E., Mongodin, E.F., Mandrell, R.E., Miller, W.G., Rasko, D.A., Ravel, J., Brinkac, L.M., DeBoy, R.T., Parker, C.T., Daugherty, S.C., Dodson, R.J., Durkin, A.S., Madupu, R., Sullivan, S.A., Shetty, J.U., Ayodeji, M.A., Shvartsbeyn, A., Schatz, M.C., Badger, J.H., Fraser, C.M., Nelson, K.E., 2005. Major structural differences and novel potential virulence mechanisms from the genomes of multiple campylobacter species. PLoS Biol. 3, e15.

Giglio, M.G., Collmer, C.W., Lomax, J., Ireland, A., 2009. Applying the gene ontology in microbial annotation. Trends Microbiol. 17, 262–268.

Gloor, G., Chaconas, G., 1986. The bacteriophage Mu N gene encodes the 64-kDa virion protein which is injected with, and circularizes, infecting Mu DNA. J. Biol. Chem. 261, 16682–16688.

Grindley, N.D., Whiteson, K.L., Rice, P.A., 2006. Mechanisms of site-specific recombination. Annu. Rev. Biochem. 75, 567–605.

Harshey, R.M., 2012. The Mu story: how a maverick phage moved the field forward. Mob. DNA 3, 21.

Harshey, R.M., Bukhari, A.I., 1983. Infecting bacteriophage mu DNA forms a circular DNA–protein complex. J. Mol. Biol. 167, 427–441.

Hiestand-Nauer, R., Iida, S., 1983. Sequence of the site-specific recombinase gene cin and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBO J. 2, 1733–1740.

Hsu, M.T., Davidson, N., 1974. Electron microscope heteroduplex study of the heterogeneity of Mu phage and prophage DNA. Virology 58, 229–239.

Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., Le Mercier, P., 2011. ViralZone: a knowledge resource to understand virus diversity. Nucleic Acids Res. 39, D576–D582.

Jang, S., Sandler, S.J., Harshey, R.M., 2012. Mu insertions are repaired by the double-strand break repair pathway of *Escherichia coli*. PLoS Genet. 8, e1002642.

Johnson, R.C., 1991. Mechanism of site-specific DNA inversion in bacteria. Curr. Opin. Genet. Dev. 1, 404–411.

Kahmann, R., Kamp, D., 1979. Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. Nature 280, 247–250.

Kim, S., Rahman, M., Kim, J., 2012. Complete genome sequence of *Pseudomonas aeruginosa* lytic bacteriophage PA1O which resembles temperate bacteriophage D3112. J. Virol. 86, 3400–3401.

Krylov, V.N., Bogush, V.G., Ianenko, A.S., Kirsanov, N.B., 1980. *Pseudomonas aeruginosa* bacteriophages with DNA structure similar to the DNA structure of Mu1 phage. II. Evidence for similarity between D3112, B3, and B39 bacteriophages: analysis of DNA splits by restriction endonucleases, isolation of D3112 and B3 recombinant phages. Genetika 16, 975–984.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

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.

Lima-Mendez, G., 2012. Reticulate classification of mosaic microbial genomes using NeAT website. Methods Mol. Biol. 804, 81–91.

Lima-Mendez, G., Toussaint, A., Leplae, R., 2011. A modular view of the bacteriophage genomic space: identification of host and lifestyle marker modules. Res. Microbiol. 162, 737–746.

Ljungquist, E., Bukhari, A.I., 1979. Behavior of bacteriophage Mu DNA upon infecton of *Escherichia coli* cells. J. Mol. Biol. 133, 339–357.

Montano, S.P., Pigli, Y.Z., Rice, P.A., 2012. The mu transpososome structure sheds light on DDE recombinase evolution. Nature 491, 413–417.

Oram, M., Travers, A.A., Howells, A.J., Maxwell, A., Pato, M.L., 2006. Dissection of the bacteriophage Mu strong gyrase site (SGS): significance of the SGS right arm in Mu biology and DNA gyrase mechanism. J. Bacteriol. 188, 619–632.

Parker, C.T., Quinones, B., Miller, W.G., Horn, S.T., Mandrell, R.E., 2006. Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic elements similar to those present in *C. jejuni* strain RM1221. J. Clin. Microbiol. 44, 4125–4135.

Plasterk, R.H., van de Putte, P., 1985. The invertible P-DNA segment in the chromosome of *Escherichia coli*. EMBO J. 4, 237–242.

Pleteneva, E.N., Mit'kina, L.N., Burkal'tseva, M.V., Krylov, V.N., 1999. Comparison of the frequency of imi-r resistant *Pseudomonas aeruginosa* mutants generated by infection of the bacteria with various bacteriophage-transposons. Genetika 35, 886–890.

Ranquet, C., Toussaint, A., de Jong, H., Maenhaut-Michel, G., Geiselmann, J., 2005. Control of bacteriophage mu lysogenic repression. J. Mol. Biol. 353, 186–195.

Rehmat, S., Shapiro, J.A., 1983. Insertion and replication of the *Pseudomonas aeruginosa* mutator phage D3112. Mol. Gen. Genet. 192, 416–423.

Scott, A.E., Timms, A.R., Connerton, P.L., Loc Carrillo, C., Adzfa Radzum, K., Connerton, I.F., 2007. Genome dynamics of *Campylobacter jejuni* in response to bacteriophage predation. PLoS Pathog. 3, e119.

Shapiro, J.A., 1969. Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. J. Mol. Biol. 40, 93–105.

Symonds, N., Toussaint, A., van de Putte, P., Howe, M.M., 1987. In: Phage, Mu, N. Symonds, A.T., van de Putte, P., Howe, M.M. (Eds.), Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Taylor, A.L., 1963. Bacteriophage-induced mutation in *E. coli*. Proc. Natl. Acad. Sci. USA 50, 1043–1051.

Toussaint, A., 2013. Transposable Mu-like phages in Firmicutes: new instances of divergence generating retroelements. Res. Microbiol. 164, 281–287.

Van Gijsegem, F., Toussaint, A., Casadaban, M., 1987. Mu as a genetic tool. In: N. Symonds, A.T., van de Putte, P., Howe, M.M. (Eds.), Phage Mu. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, pp. 215–250.