AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Molecular Analysis of *Arthrobacter* Myovirus vB_ArtM-ArV1: We Blame It on the Tail

Laura Kaliniene,[a] Eugenijus Šimoliūnas,[a] Lidija Truncaitė,[a] Aurelija Zajančkauskaitė,[a] Juozas Nainys,[b] Algirdas Kaupinis,[c] Mindaugas Valius,[c] Rolandas Meškys[a]

Department of Molecular Microbiology and Biotechnology[a] and Proteomics Centre,[c] Institute of Biochemistry, Life Sciences Centre, Vilnius University, Vilnius, Lithuania; Department of Eukaryote Gene Engineering, Institute of Biotechnology, Life Sciences Centre, Vilnius University, Vilnius, Lithuania[b]

ABSTRACT This is the first report on a myophage that infects *Arthrobacter*. A novel virus, vB_ArtM-ArV1 (ArV1), was isolated from soil using *Arthrobacter* sp. strain 68b for phage propagation. Transmission electron microscopy showed its resemblance to members of the family *Myoviridae*: ArV1 has an isometric head (~74 nm in diameter) and a contractile, nonflexible tail (~192 nm). Phylogenetic and comparative sequence analyses, however, revealed that ArV1 has more genes in common with phages from the family *Siphoviridae* than it does with any myovirus characterized to date. The genome of ArV1 is a linear, circularly permuted, double-stranded DNA molecule (71,200 bp) with a GC content of 61.6%. The genome includes 101 open reading frames (ORFs) yet contains no tRNA genes. More than 50% of ArV1 genes encode unique proteins that either have no reliable identity to database entries or have homologues only in *Arthrobacter* phages, both sipho- and myoviruses. Using bioinformatics approaches, 13 ArV1 structural genes were identified, including those coding for head, tail, tail fiber, and baseplate proteins. A further 6 ArV1 ORFs were annotated as encoding putative structural proteins based on the results of proteomic analysis. Phylogenetic analysis based on the alignment of four conserved virion proteins revealed that *Arthrobacter* myophages form a discrete clade that seems to occupy a position somewhat intermediate between myo- and siphoviruses. Thus, the data presented here will help to advance our understanding of genetic diversity and evolution of phages that constitute the order *Caudovirales*.

IMPORTANCE Bacteriophages, which likely originated in the early Precambrian Era, represent the most numerous population on the planet. Approximately 95% of known phages are tailed viruses that comprise three families: *Podoviridae* (with short tails), *Siphoviridae* (with long noncontractile tails), and *Myoviridae* (with contractile tails). Based on the current hypothesis, myophages, which may have evolved from siphophages, are thought to have first emerged among Gram-negative bacteria, whereas they emerged only later among Gram-positive bacteria. The results of the molecular characterization of myophage vB_ArtM-ArV1 presented here conform to the aforementioned hypothesis, since, at a glance, bacteriophage vB_ArtM-ArV1 appears to be a siphovirus that possesses a seemingly functional contractile tail. Our work demonstrates that such "chimeric" myophages are of cosmopolitan nature and are likely characteristic of the ecologically important soil bacterial genus *Arthrobacter*.

KEYWORDS *Arthrobacter*, *Myoviridae*, *Siphoviridae*, bacterial viruses, bacteriophage evolution, tailed viruses, vB_ArtM-ArV1

Received 5 January 2017 **Accepted** 23 January 2017

Accepted manuscript posted online 25 January 2017

Citation Kaliniene L, Šimoliūnas E, Truncaitė L, Zajančkauskaitė A, Nainys J, Kaupinis A, Valius M, Meškys R. 2017. Molecular analysis of *Arthrobacter* myovirus vB_ArtM-ArV1: we blame it on the tail. J Virol 91:e00023-17. https://doi.org/10.1128/JVI.00023-17.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

Address correspondence to Laura Kaliniene, laura.kaliniene@bchi.vu.lt.

L.K. and E.S. contributed equally to this article.

Strains of the Gram-positive bacterial genus *Arthrobacter* are widely distributed in the environment, especially in soil, and have been found to be among the predominant members of culturable aerobic soil bacteria (1, 2). The prevalence of *Arthrobacter* strains is likely due to their nutritional versatility as well as their pronounced resistance to desiccation, long-term starvation, and environmental stress (2, 3). While quite a number of strains of the genus *Arthrobacter* have been the subject of extensive studies (4–8), relatively little is known about their predators in nature, i.e., bacteriophages (phages).

Bacteriophages not only are extremely diverse but also are ubiquitous in the biosphere (9). They play a key role in microbial ecology (9, 10) and even have been suggested to be a driving force in maintaining genetic diversity of the bacterial community (9–12). Based on the classification system approved by the International Committee on Taxonomy of Viruses (ICTV), phages are classified according to their core genetic material and virion morphology. More than 95% of known phages are double-stranded DNA (dsDNA)-containing tailed viruses that comprise the order *Caudovirales* (13). On the basis of tail morphology, members of the *Caudovirales* are further subdivided into three families: *Podoviridae* (with short noncontractile tails), *Siphoviridae* (with long noncontractile tails), and *Myoviridae* (with long contractile tails) (14). With the advent of next-generation sequencing and molecular biology, it has become increasingly evident that the current phage classification system, which does not reflect genomic and proteomic data, is in need of revision, since some phages (e.g., podovirus P22 and siphovirus lambda), while having closely related genomes, are separated into different families based on the tail morphology alone (15). Since it is thought that tailed phages originated in the early Precambrian Era and very likely evolved from a common ancestor (16), more and more such "unrelated and yet related" bacterial viruses will likely be sequenced and/or isolated in the future.

The phage population represents a vast reservoir of unexplored genetic diversity. Therefore, unsurprisingly, the number of publications on phage genomics is growing rapidly, and so is the number of new phage genome sequences in GenBank, which already contains more than 3,000 tailed phage genomes. In the case of *Arthrobacter* phages, however, the situation is rather different. As summarized in reference 17, the majority of reports about this particular group of bacterial viruses (podoviruses and siphoviruses exclusively) were published more than 30 years ago and focused mostly on the isolation and/or partial characterization of bacteriophages active on laboratory strains or soil isolates of *Arthrobacter*. With the exception of two podoviruses, AN25S-1 and AN29-R-2 (18), all *Arthrobacter* phages described to date belong to the family *Siphoviridae* (17). As mentioned above, bacteria of the genus *Arthrobacter* are ubiquitous in soil environments (19). However, the first *Arthrobacter* phage genome, that of siphovirus vB_ArtM-ArV2 (ArV2), became available in the GenBank nucleotide sequence database and was published only in 2014 (17). Bacteriophages ArV2 and vB_ArtM-ArV1, a myovirus described in this study, were isolated from soil in Lithuania. The annotated genome sequence of the latter phage was deposited in the NCBI database in 2014. Since then, 42 *Arthrobacter* siphoviruses joined ArV2 and vB_ArtM-ArV1 in GenBank, followed by 8 *Myoviridae* genomes (those of phages Galaxy [accession no. KU160644], PrincessTrina [KU160660], TaeYoung [KU160668], Martha [KU160656], Jawnski [KU160651], Brent [KT365401], Sonny [KU160665], and [BarretLemon KU647629]). Nevertheless, at the time of this writing, siphovirus ArV2 remains the only *Arthrobacter* phage whose morphology and molecular characteristics have been published thus far.

In this study, we report the isolation and characterization of the *Arthrobacter* species-infecting myovirus vB_ArtM-ArV1, subsequently referred to by its shorter common laboratory name, ArV1. Morphological characterization of ArV1 places this phage in the family *Myoviridae*. Comparative genome sequence analysis, however, indicates that ArV1 has more in common with siphoviruses than it does with any well-known myovirus. Given that this is the first report on the molecular characterization of *Arthrobacter* myoviruses, the results presented here not only offer a glimpse into the

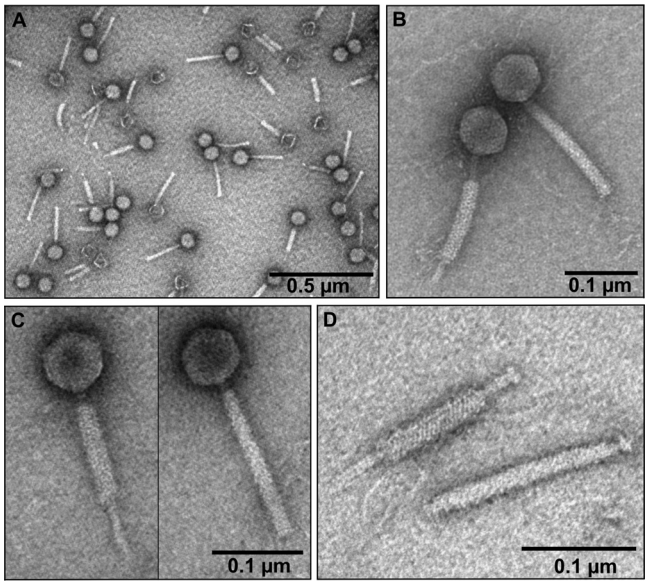

FIG 1 Electron micrographs of *Arthrobacter* phage ArV1. (A) CsCl-purified ArV1 virions. (B and C) ArV2 particle with contracted (left panels) and extended (right panels) tails. (D) ArV1 tail with contracted (top) and extended (D) tail sheath.

biology of this particular group of bacterial viruses but also provide new insights into the evolutionary relationships between different phage families.

RESULTS

Virion morphology. Transmission electron microscopy (TEM) observations of ArV1 (Fig. 1) revealed a particle that fits the A1 morphotype in Bradley's classification (20). Based on morphological characteristics, phage ArV1 belongs to the family *Myoviridae* and is characterized by an isometric head (diameter, 74.13 ± 5.7 nm [$n = 30$]) and a contractile tail (191.82 ± 11.4 nm in length [$n = 40$] and 18.94 ± 7.6 nm in width [$n = 16$]). Notably, the TEM analysis revealed a rather unusual mode of ArV1 tail sheath contraction. As seen in Fig. 1, the contracted tail sheath could take different positions with respect to the nucleocapsid: closer to the phage head (Fig. 1C) or at the distal part of the tail (Fig. 1B). Also, although neither the base plate nor the tail fibers were clearly visible by TEM, the genes coding for the base plate and tail fiber components had been detected by bioinformatics approaches and/or by proteomics analysis (see below).

Host range and physiological characteristics. To determine the optimal conditions for phage propagation and further phage experiments, the effect of temperature on the efficiency of plating (EOP) was investigated. The EOP of ArV1 was examined in the temperature range of 18 to 37°C, and the test revealed that the phage has an optimum temperature for plating of around 28°C. Notably, despite varying the host strain, the temperature, or the concentration of the soft agar, plaques produced by ArV1 (Fig. 2) were not uniform in size and varied from 0.1 to 0.5 mm in diameter. As observed with other soilborne phages (13, 17, 21), bacteriophage ArV1 failed to reproduce after inoculation into liquid bacterial culture; therefore, the one-step growth experiment was not performed.

In total, 51 bacterial strains (Table 1) were used to determine the host range of ArV1. Phage ArV1 was capable of infecting 6 phylogenetically related *Arthrobacter* species strains out of 32 used during this experiment. Also, 19 bacteria other than *Arthrobacter* were subjected to plaque assay yet found to be resistant to this phage, suggesting that the host range of ArV1 is limited to *Arthrobacter* only.

Overview of the phage ArV1 genome. Phage ArV1 has a linear dsDNA genome consisting of 71,200 bp with a G+C content of 61.6%, which is similar to that observed for *Arthrobacter* (22). The results of PCR and restriction-digestion analyses (data not

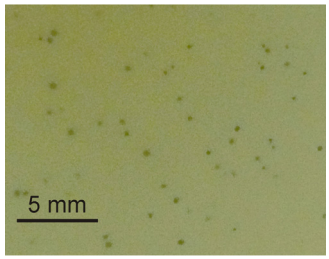

FIG 2 Plaques formed by ArV1 on a lawn of *Arthrobacter* sp. strain 68b after 2 days (48 h) of incubation at 28°C.

shown) revealed that the ArV1 genome is a circularly permuted molecule. As is the case with other tailed dsDNA phages (23), the genome of ArV1 is extremely compact, with ~95% of the DNA sequence representing coding sequences. The genome sequence analysis revealed that ArV1 has a total of 101 probable protein-encoding genes (with an average size of 679 bp) and no genes for tRNA (Fig. 3). Also, there is a marked asymmetry in the distribution of the genes on the two DNA strands. In total, 97 ArV1 open reading frames (ORFs) have been predicted to be transcribed from the same DNA strand, and only 4 ORFs have been found on the opposite strand.

The results of BLASTP analysis revealed that 44% of ArV1 genes encode unique proteins that either have no reliable identity (E values of ≥0.0001) to database entries (9 ORFs) or have homologues in bacteriophage PrincessTrina only (35 ORFs). In addition, 15 ArV1 ORFs encode proteins whose homologues are found exclusively in *Arthrobacter* phages (both sipho- and myoviruses) and bacteria. Using bioinformatics approaches, a putative function was assigned to 30 ArV1 ORFs, including 13 genes coding for virion morphogenesis-related proteins, as well as 15 genes associated with DNA metabolism and packaging. None of the predicted ArV1 proteins showed sequence homology with antibiotic resistance determinants, virulence factors, or integration-related proteins.

Structural proteins. Using bioinformatics approaches, 13 ArV1 structural genes were identified, including those coding for head (ORF02, ORF08, and ORF09), tail (ORF15, ORF16, and ORF18), tail fiber (ORF22, ORF23, and ORF25), and baseplate (ORF19, ORF20, ORF32, and ORF33) proteins, as well as 3 ORFs coding for virion morphogenesis-related proteins, namely, ORF12 (Phage_tail_S superfamily tail completion protein, pfam05069), ORF03 (putative capsid maturation protease), and ORF07 (Mu-like_Pro superfamily protein, cl19864). A further 6 ArV1 ORFs were annotated as putative structural proteins based on the results of proteomic analysis (Table 2). Therefore, in total, 22 ArV1 gene products were assigned as proteins involved in packaging, structure, and morphogenesis.

As was observed with other myoviruses (24), three ArV1 gene products, the major head (gp09), tail sheath (gp15), and tail tube (gp16) proteins, are the main building blocks of the ArV1 virion (Fig. 4). A BLASTP search against the nonredundant NCBI protein database revealed that the gene product of ArV1 ORF09 exhibits similarity to the major capsid proteins of *Gordonia*, *Mycobacterium*, and *Bacillus* siphoviruses and has recognizable homologues in only five, as yet unpublished, myoviruses, namely, *Arthrobacter* phages PrincessTrina, Brent, Jawnski, BarretLemon, and Martha. ArV1 gp15, which has been predicted to belong to the phage tail sheath protein family (pfam04984), shows similarity to homologous proteins in seven *Arthrobacter*-infecting myoviruses yet also has a homologue in *Gordonia* siphovirus GMA6 (25). Also, HHPred showed that ArV1 gp15 is homologous to the sheath proteins of *Escherichia coli* phage T4 (HHPred probability, 100%; E value, 1.2e−50) and R-type pyocin from *Pseudomonas aeruginosa* (HHPred probability, 100%; E value, 1.2e−44). The putative tail tube protein ArV1 gp16 has no homology to known tail tube proteins; however, it shares 33% amino acid (aa) sequence identity (E value, 1e−26) with the putative major tail protein GMA6_30 (gb|AKL88311.1|), which belongs to a T4-like virus tail tube protein family

TABLE 1 Bacterial strains used in this study

Strain or species[a]	Relevant characteristics	Source or reference
Acinetobacter baumannii 46		E. Suziedeliene
Acinetobacter gen. sp. 13 23		E. Suziedeliene
***Arthrobacter alkaliphilus* DSM 23368**	Type strain	DSMZ
Arthrobacter aurescens DSM 20116	Type strain	DSMZ
Arthrobacter chlorophenolicus DSM 12829	Type strain	DSMZ
Arthrobacter citreus DSM 20133	Type strain	DSMZ
Arthrobacter crystallopoetes DSM 20117	Type strain	DSMZ
Arthrobacter defluvii DSM 18782	Type strain	DSMZ
Arthrobacter gandavensis DSM 15046	Type strain	DSMZ
Arthrobacter globiformis DSM 20124	Type strain	DSMZ
***Arthrobacter histidinolovorans* DSM 20115**	Type strain	DSMZ
***Arthrobacter ilicis* DSM 20138**	Type strain	DSMZ
Arthrobacter koreensis DSM 16760	Type strain	DSMZ
Arthrobacter luteolus DSM 13067	Type strain	DSMZ
Arthrobacter methylotrophus DSM 14008	Type strain	DSMZ
Arthrobacter nicotinovorans DSM 420	Type strain	DSMZ
***Arthrobacter nitroguajacolicus* DSM 15232**	Type strain	DSMZ
Arthrobacter oxydans DSM 20119	Type strain	DSMZ
***Arthrobacter* sp. strain 68b DSM 103156**	Environmental isolate	81
***Arthrobacter* sp. strain 68 m**	Environmental isolate	Laboratory collection
Arthrobacter sp. strain 83b	Environmental isolate	Laboratory collection
Arthrobacter sp. strain 85	Environmental isolate	Laboratory collection
Arthrobacter sp. strain 94	Environmental isolate	Laboratory collection
Arthrobacter sp. strain 96	Environmental isolate	Laboratory collection
Arthrobacter sp. strain 25DMP1	Environmental isolate	82
***Arthrobacter* sp. strain 25DOT1**	Environmental isolate	82
Arthrobacter sp. strain IN13	Environmental isolate	83
Arthrobacter sp. strain PY11	Environmental isolate	84
Arthrobacter sp. strain PY21	Environmental isolate	85
Arthrobacter sp. strain PY22	Environmental isolate	85
Arthrobacter sp. strain PRH1	Environmental isolate	85
Arthrobacter sp. strain VM22	Environmental isolate	84
Arthrobacter sp. strain VP3	Environmental isolate	84
***Arthrobacter ureafaciens* DSM 20126**	Type strain	DSMZ
Citrobacter freundii		E. Suziedeliene
Enterobacter cloacae		E. Suziedeliene
Erwinia carotovora 8982		E. Suziedeliene
Escherichia coli B^E	*sup*^{0}	L. W. Black
Escherichia coli BL21(DE3)	F^- *dcm ompT hsdS*(r_B^- m_B^-) *gal* λ(DE3)	Avidis
Escherichia coli DH5α	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *hsdR17*(r_K^- m_K^+) λ^-	Pharmacia
Escherichia coli DH10β	F^- *endA1 recA1 galE15 galK16 nupG rpsL* Δ*lacX74* ϕ80*lacZ*ΔM15 *araD139* Δ(*ara leu*)*7697 mcrA* Δ(*mrr-hsdRMS-mcrBC*) λ^-	Invitrogen
Klebsiella pneumoniae 279		E. Suziedeliene
Klebsiella sp. strain KV-3	Veterinary isolate, Amp^r Str^r Tet^r Kan^s Gm^s Nc^s $Cl^{r/s}$	86
Kocuria palustris DSM 11925	Type strain	DSMZ
Kribbella catacumbae DSM 19601	Type strain	DSMZ
Nesterenkonia aethiopica DSM 17733	Type strain	DSMZ
Pseudomonas aeruginosa PAO1		E. Suziedeliene
Rothia aeria DSM 14556	Type strain	DSMZ
Salmonella enterica serovar Typhimurium 292		E. Suziedeliene
Solitalea canadensis DSM 3403	Type strain	DSMZ
Yaniella soli DSM 22211	Type strain	DSMZ

[a]ArV1-sensitive strains are in bold.

(pfam06841). Also, HHpred analysis revealed that gp16 of ArV1 may be related to the HCP1 family protein from the type VI secretion system (T6SS) of *Acinetobacter baumannii* (HHPred probability, 97%; E value, 0.009) and the major tail protein gpV of *E. coli* phage lambda (HHPred probability, 82%; E value, 12). In type VI secretion systems, HCP1 family proteins are responsible for the formation of the tube (26). Notably, in the case of all three ArV1 gene products discussed above, there was a close match between the position of the protein on the gel and the predicted molecular mass of the protein (Fig. 4), suggesting that these peptides are not subjected to extensive posttranslational modifications during ArV1 virion maturation.

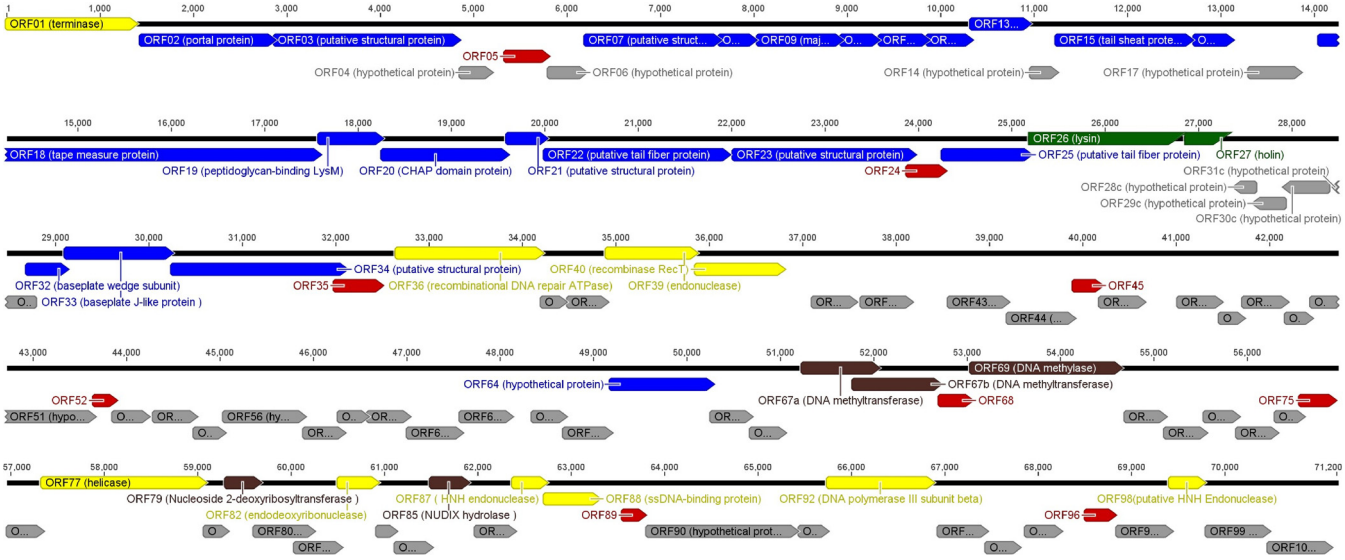

FIG 3 Functional genome map of bacteriophage ArV1. The coding capacity of the ArV1 genome is shown. Functions are assigned according to the characterized ORFs in the NCBI database and/or MS/MS analysis. The color code is as follows: yellow, DNA replication, recombination, repair, and packaging; brown, transcription, translation, and nucleotide metabolism; blue, structural proteins; green, lysis; gray, ORFs of unknown function; red, ArV1-specific ORFs that encode unique proteins with no reliable identity to database entries.

As mentioned above, while neither the baseplate nor the tail fibers are clearly visible by TEM (Fig. 1), four ArV1 genes coding for baseplate proteins (gp19, gp20, gp32, and gp33) as well as three ORFs coding for tail fiber/spike components (gp22, gp23, and gp25) have been identified by bioinformatics approaches. Proteomics analysis revealed that all of these proteins are present in the virion of ArV1, although gp25 and gp32 have been identified only by filter-aided sample preparation (FASP). Notably, a predicted collagen-like (pfam01391) protein, gp25, has been annotated as a putative tail fiber protein based on a weak similarity to the putative tail fiber proteins of *Enterobacterium* myovirus 4MG (32% aa sequence identity; E value, 0.018) and *Cronobacter*

TABLE 2 Structural ArV1 proteins identified by MS

Gene	Protein characteristics			
	Function	Molecular mass (kDa)	No. of peptides	Sequence coverage (%)
ORF02[a]	Portal protein	52.044	57	83.72
ORF03[a]	Putative structural protein	71.187	63	87.15
ORF07[a]	Putative structural protein	50.905	17	43.96
ORF08[a]	Putative structural protein	12.544	14	82.54
ORF09[a]	Major capsid protein	37.602	46	74.59
ORF10	Putative structural protein	16.293	10	78.21
ORF11	Putative structural protein	18.157	9	69.82
ORF12[a]	Phage virion morphogenesis protein	17.738	9	64.21
ORF13	Putative structural protein	22.851	7	42.93
ORF15[a]	Tail sheath protein	49.106	126	91.95
ORF16[a]	Putative tail tube protein	14.357	11	88.89
ORF18[a]	Tape measure protein	122.323	73	51.51
ORF19[a]	Peptidoglycan-binding LysM	24.767	34	96.32
ORF20[a]	CHAP domain-containing protein	52.172	33	74.80
ORF21	Putative structural protein	14.356	12	90.30
ORF22[a]	Putative tail fiber protein	68.336	61	88.99
ORF23[a]	Putative structural protein	66.881	96	88.98
ORF25[a]	Putative tail fiber protein	31.741	11	67.42
ORF32[a]	Baseplate wedge subunit	14.742	6	77.86
ORF33[a]	Baseplate J-like protein	39.086	25	63.75
ORF34	Putative structural protein	65.772	67	84.98
ORF64	Hypothetical protein	40.121	19	51.89

[a]Gene encoding putative structural ArV1 protein identified by bioinformatics approaches.

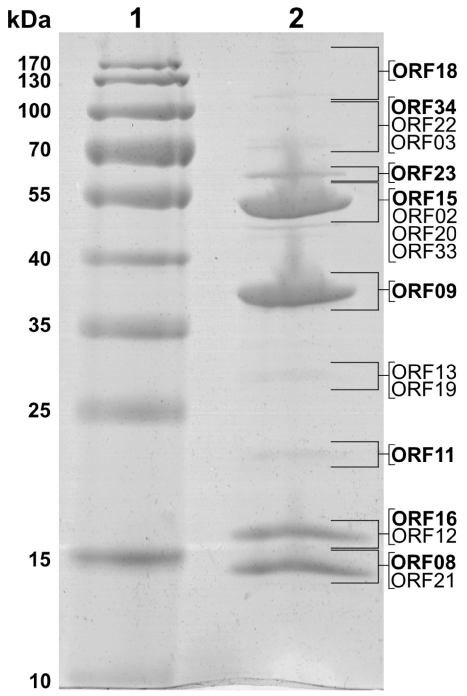

FIG 4 SDS-PAGE of ArV1 virion proteins. Lanes: 1, molecular mass marker Page Ruler prestained protein ladder (Thermo Fisher); 2, phage ArV1 structural proteins. Relative migrations of molecular mass marker proteins are indicated on the left. Proteins identified by MS/MS are indicated on the right.

myovirus vB_CsaM_GAP31 (32% aa sequence identity; E value, 0.022). Also, while no amino acid sequence similarity has been identified between the baseplate proteins of characterized myoviruses and ArV1 gp32, the latter protein has been predicted to belong to the GPW_gp25 superfamily (cl01403/pfam04965), which includes a structural component of the outer wedge of the baseplate of T4 (T4 gp25) that has acidic lysozyme activity.

The baseplate protein ArV1 gp19 contains a C-terminal LysM superfamily domain (cl23764) and is similar to LysM and cell wall-binding domain-containing proteins from a wide range of diverse bacteria (128 hits) as well as phages active against *Arthrobacter* (9 hits) or *Clostridium* (2 hits). Notably, it has been reported recently that the LysM domain is present in many phage baseplates, and it is often fused to the tail tube initiator (27). Based on its genomic position and the presence of LysM, ArV1 gp19 is a probable candidate for a tail tube initiator. Baseplate_J superfamily (cl01294) protein ArV1 gp33, which is an orthologue of T4 gp6 (HHpred probability, 100%; E value, 1.9e−35), shows similarity exclusively to homologous proteins in *Arthrobacter*-infecting myoviruses and *Gordonia* siphovirus GMA6. Homologues to ArV1 gp20 are only found in five *Arthrobacter* myoviruses (PrincesTrina, Brent, BarretLemon, Jawnski, and Sonny), yet gp20 itself is an orthologue of the baseplate hub protein gp44 of *Enterobacterium* myovirus Mu (HHpred probability, 99.64%; E value, 2.1e−15) and has a C-terminal peptidase domain (HHpred probability, 99.8%; E value, 1.8e−20).

The gene product of ArV1 ORF22 (654 aa) has homologues in 7 *Arthrobacter* myoviruses and shares one region of sequence similarity (aa 300 to 500) with various proteins, mostly hypothetical proteins, from siphoviruses that infect *Arthrobacter* or *Gordonia*. The tail fiber protein of ArV1, gp23, has a C-terminal Peptidase_S74 domain (pfam13884), a conserved chaperone domain that is commonly found in endosialidases. Phage endosialidases are usually present on the virus particle in the form of tail spikes or tail fibers (28, 29) and have been found in *E. coli* K1- or K92-specific podoviruses, siphoviruses, and the myophage phi92 (29, 30).

As seen in Fig. 3, all ArV1 structural genes except ORF64 are found within a large, ~30-kb cluster located just downstream from the gene for the terminase large subunit

(gp01). Bioinformatics approaches have yielded no putative function for ORF64, whose predicted amino acid sequence (362 aa) shows similarity to that of the hypothetical protein PRINCESSTRINA_79 (31% aa identity) yet has no detectable homology to any other entries in the public databases. Nevertheless, the gene product of ORF64 has been identified by both liquid chromatography-tandem mass spectrometry (LC-MS/MS) of the structural ArV1 proteins separated by SDS-PAGE and FASP (Table 2), suggesting that gp64 may indeed be the component of the phage particle. Also, as seen in Fig. 4, ArV1 gp64 migrates with a molecular mass slightly higher than that predicted by the sequence alone (40.1 kDa), which may be due to the posttranslational modifications or to remaining secondary and higher-order structures. Notably, ORF64 lies in a genomic region that contains a cluster of ORFs of unknown function (Fig. 4). Although none of these ORFs have been identified by proteomics approaches, there is a possibility that at least some of these putative genes of bacteriophage ArV1 encode proteins that participate in virion formation.

Packaging. The packaging machine of tailed phages usually consists of two essential components: a portal ring and a terminase complex (31). Most characterized terminases consist of a small subunit (TerS) involved in DNA recognition and a large terminase subunit (TerL) containing the ATPase and the endonuclease activities (32). ArV1 ORF02 was annotated as a portal protein based on its similarity to the corresponding proteins from *Arthrobacter*, *Gordonia*, and *Mycobacterium* phages and was detected by proteomics approaches as well (Fig. 4; Table 2). Based on the results of bioinformatics analysis, the gene product of ArV1 ORF01 is a TerL, since it has been predicted to belong to the terminase-like family (pfam03237) and shows detectable similarity to other phage-encoded TerL proteins. In contrast, no gene for ArV1 TerS was identified by homology searches. The small terminase subunits of tailed viruses generally are small proteins that contain an N-terminal DNA-binding domain and are expressed from a gene upstream of the large terminase gene (32, 33). On a circular genome map of ArV1, the gene upstream of ORF01 is the hypothetical protein-encoding ArV1 ORF100, which has no amino acid sequence homology with any of the phage proteins except for a hypothetical protein, PRINCESSTRINA_2. Nevertheless, the presence of a putative N-terminal helix-turn-helix domain (HTH_18, PF12833), identified by MOTIF-Search analysis, suggests that ArV1 gp100 may be a novel small terminase subunit.

DNA RRR. The genome of bacteriophage ArV1 contains no homologues to characterized DNA polymerase genes, suggesting that this phage most likely takes advantage of the replication machinery of the host cell. However, bioinformatics analysis identified a set of ArV1 replication, recombination, and repair (RRR) genes that, unlike structural genes of ArV1, are apparently scattered throughout the genome.

The product of ArV1 ORF92 has been predicted to belong to the Beta_clamp superfamily (cl21574) and, among phage genomes, has a reliable homologue only in *Arthrobacter* phage PrincessTrina. In contrast, the single-stranded DNA binding (SSB) protein ArV1 gp88, belonging to the RPA_2b-aaRSs_OBF_like superfamily (cl09930), shows similarity to SSB proteins from a variety of organisms. Another RRR protein, ArV1 gp77, contains a conserved C-terminal helicase domain (pfam00271) and has been predicted to belong to the SSL2 superfamily (superfamily II DNA or RNA helicase, COG1061).

According to the literature, a clear correlation between the size of the genome and self-sufficiency of viral DNA replication exists. Phages with large genome size (>140 kb) tend to encode their own replication machinery, whereas relatively small viruses (<100 kb) more often than not employ a replicative polymerase of the host cell (34). Such "less self-sufficient" dsDNA bacteriophages, e.g., *E. coli* phage λ and *Bacillus subtilis* phage SPP1, encode a subset of proteins, including an origin-binding protein (OBP), that all play pivotal roles in the recruitment of the host DNA replication machinery (35, 36). Bioinformatics analysis failed to identify the gene for an OBP of ArV1. However, as seen in Fig. 3, a large number of the genes for putative proteins of unknown function/origin

surround those ArV1 RRR genes that have been identified by bioinformatics approaches. Therefore, it is possible that the genome of this particular myovirus encodes a yet-unrecognized OBP. On the other hand, since replicative helicases have been suggested to be involved in the recognition of replication origins (37), there is a slight possibility that ArV1 recruits replicative DNA polymerase of the host cell by using its own helicase and SSB protein complex.

Bioinformatics analysis also allowed the identification of the phage-encoded recombination system in the genome of ArV1. The gene ArV1 ORF39 codes for a protein that has been predicted to belong to the YqaJ family of exonucleases (pfam09588) and is followed by a gene for a RecT superfamily protein (cl04285). The genome of ArV1 also contains a gene for a recombinational DNA repair ATPase, which has an AAA_23 domain (pfam13476) located in its N terminus. The recombinational cassette of ArV1 is very similar to that found in *Mycobacterium* siphophage Giles (38). According to van Kessel and coauthors (38), expression of the mycobacteriophage-encoded exonuclease and recombinase substantially enhances recombination frequencies in both fast- and slow-growing mycobacteria. Thus, there is a possibility that myovirus ArV1 employs a phage-encoded recombination system to enhance the recombination frequency in reportedly slow-growing *Arthrobacter* spp. (3).

Other ArV1 ORF products possibly involved in DNA recombination include endodeoxyribonuclease (gp82), that is similar to a crossover junction endodeoxyribonuclease RuvA of *Catenulispora acidiphila* (35% aa identity; E value, 2e−05), and two putative HNH endonucleases, gp87 and gp98, which show similarity to an HNH endonuclease of *Gordonia* phage Yvonnetastic and a putative HNH endonuclease of *Bacillus* phage Shanette, respectively. It has been reported that HNH endonucleases promote the lateral transfer of their own coding regions and flanking DNA between genomes by a recombination-dependent process termed homing (39, 40). Hence, as such, these site-specific DNA endonucleases may be, at least partially, responsible for apparent mosaic architecture of the genome of ArV1.

Nucleotide metabolism and DNA modification. Based on the amino acid sequence similarity, two gene products of ArV1 were assigned as nucleotide metabolism enzymes. Bacteriophage ArV1 ORF79 codes for a nucleoside 2-deoxyribosyltransferase that is similar to a nucleoside 2-deoxyribosyltransferase from *Aeromonas jandaei* (44% identity; E value, 2e−19), while the product of ArV1 ORF85 is a Nudix hydrolase that shows similarity to a corresponding protein in *Planomonospora sphaerica* (43% identity; E value, 2e−17). Nudix hydrolases hydrolyze a wide range of organic pyrophosphates, including nucleoside di- and triphosphates, with various degrees of substrate specificity (41).

To protect the genomic DNA from restriction endonucleases of the host cell, phages employ various strategies, including adenine and cytosine methylation (42). Based on the results of bioinformatics analysis, three putative methylases are present in the genome of ArV1. The protein encoded by ArV1 ORF67a is a putative Dcm methylase that is similar to that from *Brachybacterium* sp. strain SW0106-09 (46% identity; E value, 1e−34). The nucleotide sequence of ORF67a is partially overlapped by ArV1 ORF67b, which also codes for a Dcm superfamily methylase, a homologue of the cytosine methyltransferase from *Gulosibacter molinativorax* (44% identity, E value, 2e−55). Also, the genome of ArV1 contains a gene (ORF69) for a putative AdoMet_MTase superfamily (cl17173) protein.

The restriction digestion analysis revealed that the DNA of ArV1 seems not to be Dam methylated (data not shown), suggesting that the gene product of ORF69 is likely not functional. However, since ArV1 DNA is resistant to cleavage by EcoRII and NotI (enzymes that are sensitive to Dcm and CpG methylation, respectively), at least one cytosine methylase is probably active in ArV1. Notably, since gp67a shares similarity to the N terminus of DNA methylase from *Mycobacterium* siphovirus Florinda (44.2% identity; E value, 3.6e−48), whereas gp67b is similar to the C terminus of the afore-

mentioned protein (46% identity; E value, 1.1e−21), there is a strong possibility that the cytosine methylase of ArV1 functions as a heterodimeric enzyme.

Lysis cassette. According to Young and colleagues (43), all dsDNA phages accomplish host lysis using a muralytic enzyme (known as an endolysin) and a holin, a small membrane protein that permeabilizes the membrane at a programmed time. The lysis cassette of ArV1 comprises two genes, ORF26 and ORF27. The gene product of ORF26 (550 aa) shows homology to different endolysins and contains five conserved domains, namely, an N-terminal CHAP domain (pfam05257), a central region conserved between glycosyl hydrolase family 25 (GH25, pfam01183) proteins, and three C-terminal LysM domains (pfam01476), whereas based on its length (152 aa) and three predicted transmembrane regions, gp27 is a plausible candidate for a type I holin (43).

Phylogenetic analysis. As discussed above, TEM analysis revealed that the virion of ArV1 has a long contractile tail characteristic of *Myoviridae*. However, based on the results of the genome sequence analysis, bacteriophage ArV1 has no myoviral relatives except for a small group of yet-unpublished *Arthrobacter*-infecting myoviruses, whose genomes have been released by GenBank only recently. Therefore, to unravel evolutionary relationships between ArV1 and other tailed phages, a number of phylogenetic reconstruction methods have been used.

A novel "head-neck-tail"-based classification method proposed by Lopes and coauthors (44) classified ArV1 as a myovirus of "type 1 (cluster 5)" (Fig. 5). Based on the data presented by VIRFAM, "neck type 1" phages adopt the structural organization of the *Siphoviridae* phage SPP1 neck, while cluster 5 is composed strictly of siphophages, of which all but two, *Streptomyces* phages phi-C31 and phi-BT1, infect *Proteobacteria*. Notably, VIRFAM analysis classified *Arthrobacter* bacteriophages PrincessTrina, Martha, BarretLemon, Sonny, Brent, Jawnsky, and TaeYoung as the neck type 1 myoviruses yet failed to assign them to a particular cluster. Also, *Arthrobacter* phage Galaxy, annotated as *Myoviridae* in the NCBI database, was recognized as a type 1 siphovirus belonging to cluster 10. While the results of VIRFAM analysis correlated well with those obtained by other homology-based approaches, suggesting that bacteriophage ArV1 has much more in common with phages from the family *Siphoviridae* than it does with any myovirus characterized to date, the phylogenetic position of ArV1 with respect to other tailed bacteriophages, especially *Myoviridae* phages, remained unclear.

Unlike the 16S rRNA gene in *Bacteria* and *Archaea*, there is no universally accepted phylogenetic marker gene for tailed bacteriophages (45). However, comparative phage genomics suggest that two proteins, the terminase large subunit and the portal protein, can be recognized by sequence/structure similarity to be encoded by most sequenced tailed phages (46–48), and they have often been used as markers in phylogenetic analysis (49–51). Another protein whose amino acid sequence alignment has often been used in phylogenetics is the major capsid/head protein (52, 53). Also, a tail sheath protein presumably exclusive to myoviruses has been suggested as a potential marker for this specific viral family (45). Therefore, four phylogenetic trees based on the alignment of the ArV1 TerL (Fig. 6), portal (Fig. 7), tail sheath (Fig. 8), and major head (Fig. 9) protein sequences with those returned by BLASTP homology searches were constructed. Notably, in the case of the portal, major capsid, and tail sheath proteins, due to the near absence of myoviral representatives, the amino acid sequences of the corresponding proteins from *Campylobacter* phage CP220, *Bacillus* phage SPO1, and *Enterobacterium* phages P2, T4, and Mu (representing myoviral subfamilies *Eucampyvirinae*, *Spounavirinae*, *Peduovirinae*, *Tevenvirinae*, and *Vequintavirinae*, respectively) were included in the alignment as an outgroup. As seen in Fig. 6 to 8, *Arthrobacter* myophages form a discrete clade in all four phylogenetic trees and, in most cases, seem to occupy a somewhat intermediate position between myo- and siphoviruses.

To obtain a more detailed picture of the phylogenetic relationships within the group of *Arthrobacter*-infecting myoviruses, the annotated PrincessTrina, Martha, BarretLemon, Sonny, Brent, Jawnsky, TaeYoung, and Galaxy genomes were downloaded from the NCBI

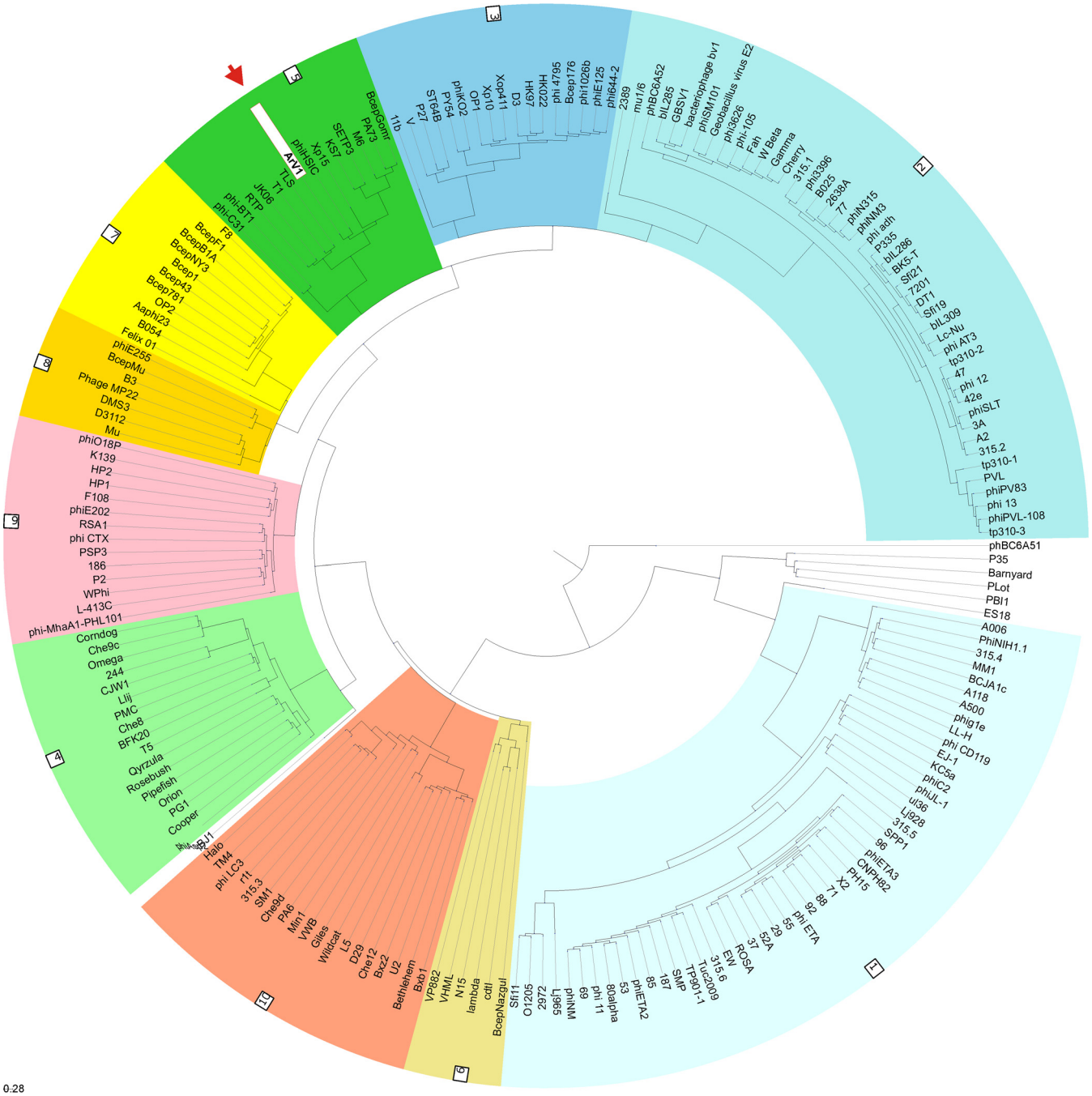

FIG 5 VIRFAM-generated clustering of ArV1 with the phages sharing the most similar head-neck-tail modules. Different type 1 phage clusters are highlighted by different background colors. ArV1 is indicated by the red arrow and white background.

database and compared to that of ArV1. As seen in Fig. 10A, Mauve alignment of ArV1 and PrincessTrina revealed that these two phages share a high level of sequence homology and have almost identical genome organizations. When the genomes from all seven *Arthrobacter* myoviruses were aligned (note that Mauve failed to perform the alignment when the genome sequence of Galaxy was included), several regions of nucleotide sequence similarity were identified that, in ArV1, covered the head and tail morphogenesis genes (Fig. 10B).

Since the results of homology searches indicate that the similarity between ArV1 and other *Arthrobacter* phages listed above is not limited to the head and tail morphogenesis genes, the CoreGenes3.5 comparison program (54), which proved to be

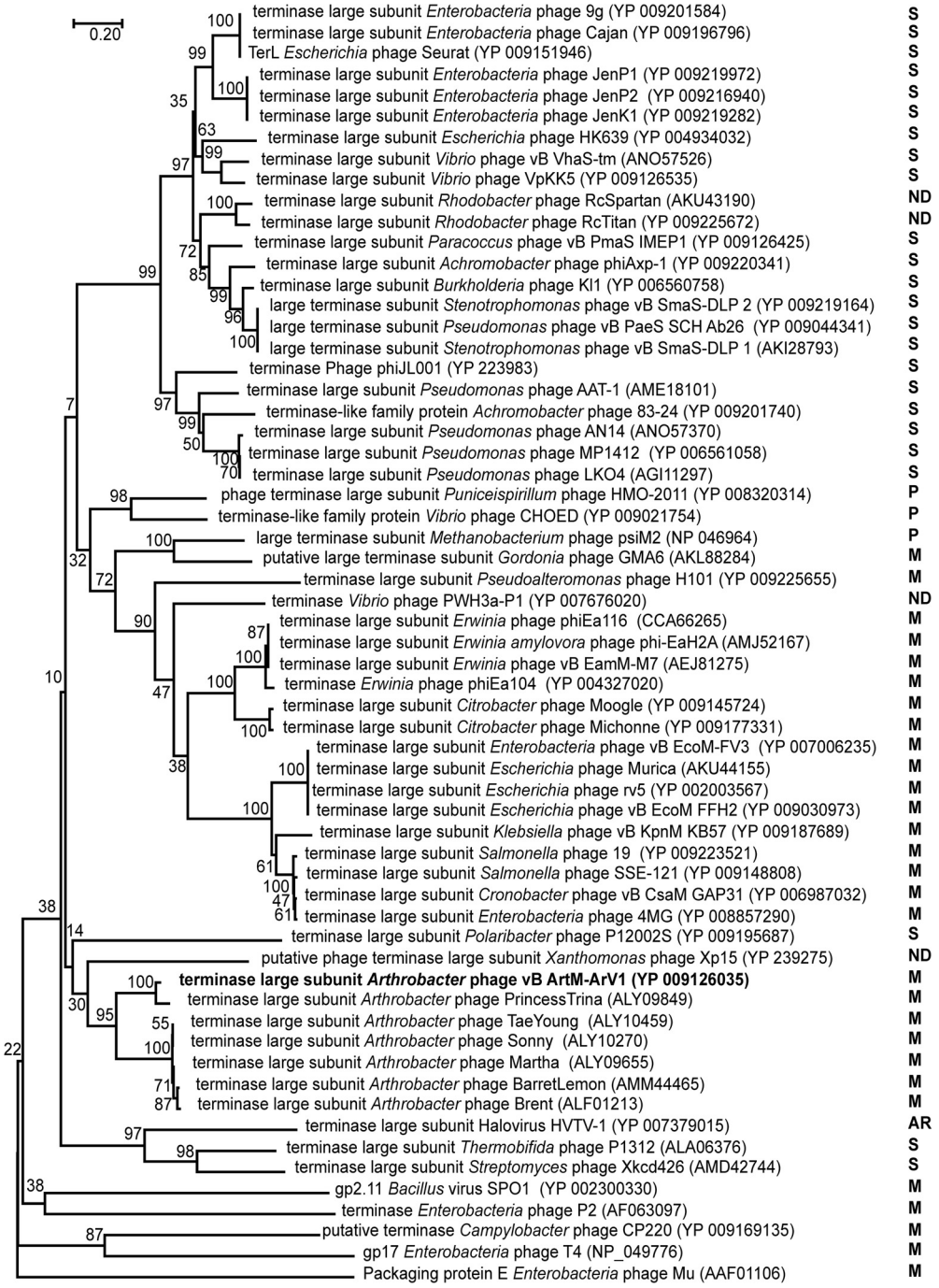

FIG 6 Relationships of terminase large subunits across diverse phage types. The numbers at the nodes indicate the bootstrap probabilities. S, *Siphoviridae*; M, *Myoviridae*; P, *Podoviridae*; ND, not determined; AR, *Archaea*.

valuable in phage taxonomic analysis (55, 56), was used to compare the phage proteomes using ArV1 as a reference phage. In total, 32 core genes (32% of the ArV1 coding capacity) belonging to different functional groups (see Table S1 in the supplemental material) were identified by CoreGenes3.5.

DISCUSSION

This is the first report on the molecular characterization of a myophage that infects bacteria of the genus *Arthrobacter*. Myovirus ArV1 was isolated and sequenced in 2013. At that time, the functional annotation of ArV1 genome was complicated due to the

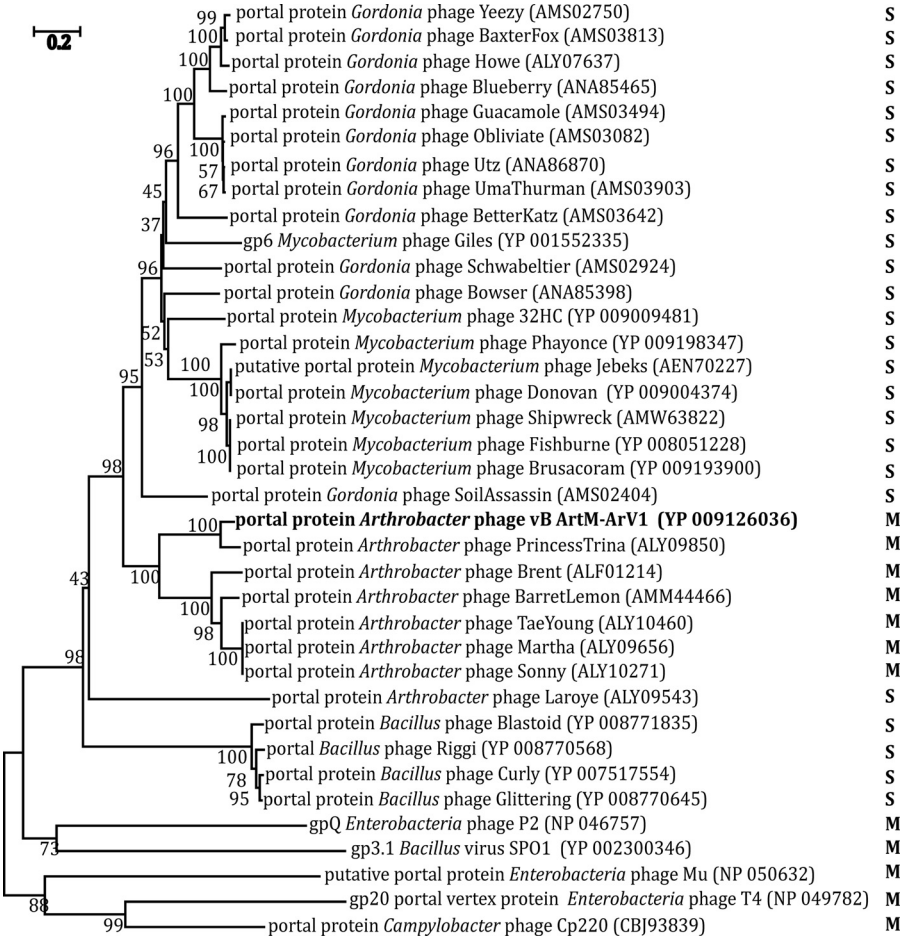

FIG 7 Neighbor-joining tree analysis based on the alignment of the amino acid sequences of the portal protein. The numbers at the nodes indicate the bootstrap probabilities. S, *Siphoviridae*; M, *Myoviridae*; ND, not determined.

fact that more than 60% of its genes were ORFans. Moreover, at the time, the majority of ArV1 ORFs whose functions had been inferred by homology searches (including most of the genes coding for virion structural proteins) showed no reliable homology to any other myoviral sequences in the databases. In fact, the results of the genome sequence

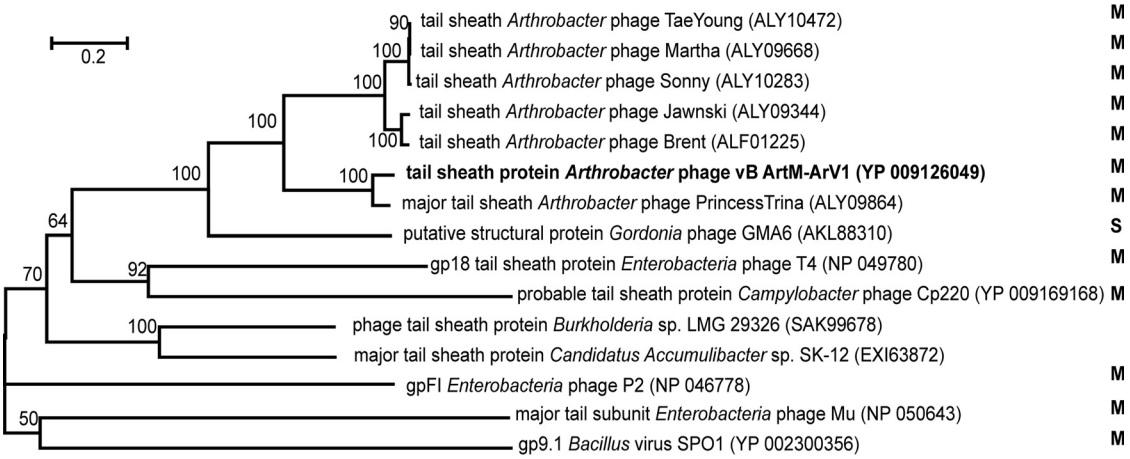

FIG 8 Neighbor-joining tree analysis based on the alignment of the amino acid sequences of the tail sheath proteins. The numbers at the nodes indicate the bootstrap probabilities. S, *Siphoviridae*; M, *Myoviridae*.

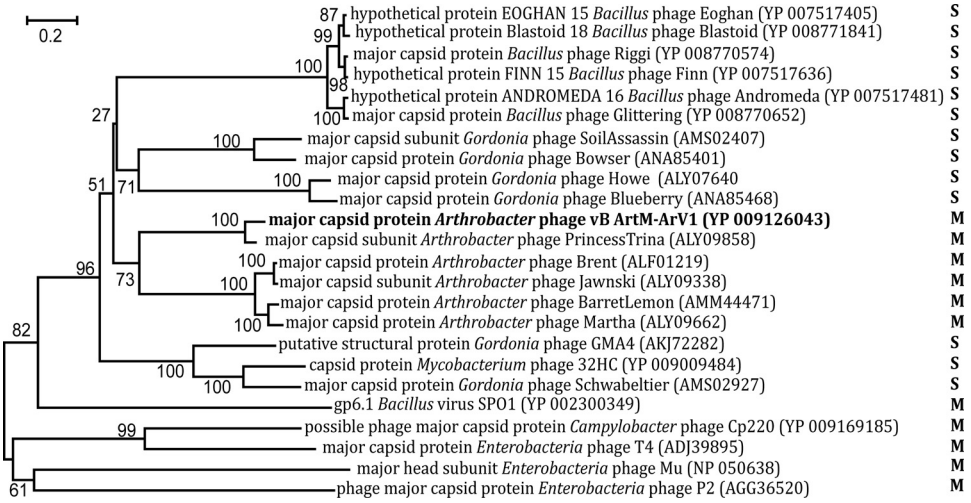

FIG 9 Neighbor-joining tree analysis based on ClustalW alignment of the major capsid protein sequences. The numbers at the nodes indicate the bootstrap probabilities. S, *Siphoviridae*; M, *Myoviridae*.

analysis revealed that ArV1 had more in common with siphoviruses than it did with any known myovirus. Since 2013, the genomes of 8 new *Arthrobacter* myophages have become available in the databases, together with quite a number of genome sequences of various myoviruses infecting *Actinobacteria* other than *Arthrobacter*. Yet, while the annotation of the ArV1 genome has become, perhaps, more accurate, not much has changed from the point of view of phylogeny, except that now ArV1, together with other myophages infecting *Arthrobacter*, comprises a group of odd myoviruses that share more genes/proteins with phages from the family *Siphoviridae* than they do with those belonging to *Myoviridae*. Nevertheless, the micrographs obtained by TEM indicate that ArV1 is just an ordinary myovirus. Or is it?

As seen in Fig. 1, for a myovirus with a medium-sized genome (71 kb), the tail of Arv1 seems to be excessively long (~192 nm) and exhibits a rather unusual mode of sheath contraction. According to the literature, while siphoviruses usually possess longer tails than myoviruses, the length of the tail in both types of phages is determined by the length of the tape measure protein (TMP) (57). While the average lengths of the TMPs of siphophages and myophages are ~1,200 and ~800 residues, respectively, bacteriophages with larger genomes tend to code for longer TMPs than those with smaller genomes (58). It has been well documented that in siphoviruses, TMP plays an important role in the DNA injection process by forming a channel through the host cell membrane for phage genome entry and protecting the DNA from degradation by periplasmic endonucleases (58–60). Also, the TMPs of several siphoviruses (e.g., *Escherichia coli* phage T5) have been implicated in the local degradation of the peptidoglycan (61). In the case of myophages, the function of TMP in the DNA injection has, until recently, been unclear. However, last year significant advances were made in the characterization of TMP functionality for myophage T4 (62). According to Hu and coauthors (62), during the translocation of phage DNA from the virion capsid into the cytoplasm of the host cell, TMP of the myovirus T4 is ejected from the tail tube, after which it likely participates in the formation of the transmembrane channel for the passage of viral DNA. These results suggest that the TMPs of contractile and noncontractile tails function in a similar manner. Gene 18 of bacteriophage ArV1 codes for a TMP of 1,177 residues, and its homologues (Fig. 11) are present in as many as 137 siphoviruses (mostly *Mycobacterium* and *Propionibacterium* phages) but in only 1 myovirus (*Arthrobacter* phage PrincessTrina). An intriguing theory regarding phage evolution has been raised by Davidson and coauthors, stating that compared to the myoviruses of Gram-negative bacteria, viruses with contractile tails are a relatively new addition to the population of phages infecting Gram-positive cells and that myophages

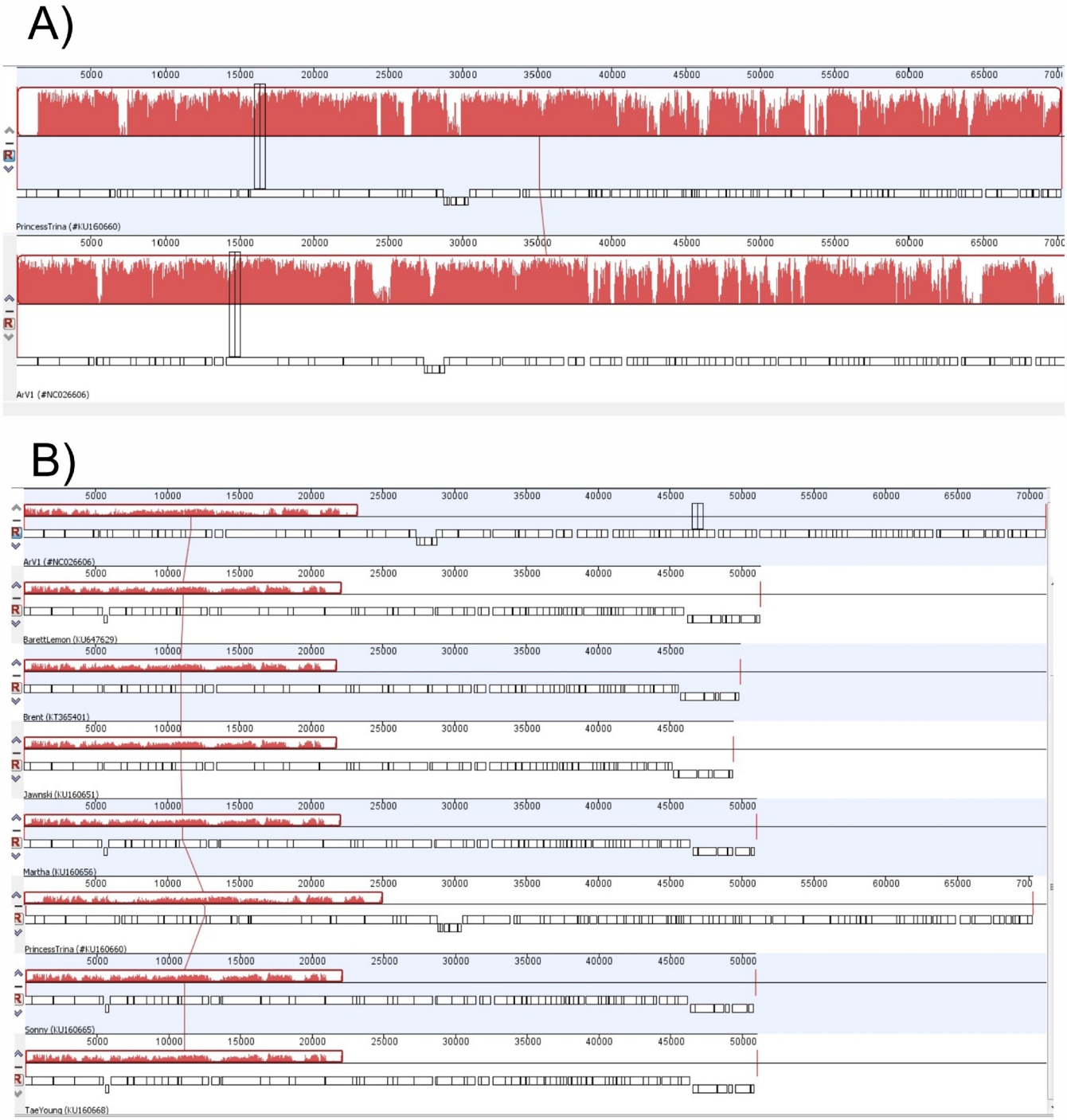

FIG 10 Progressive Mauve whole-genome alignment generated using Geneious Pro v5.5.6. (A) Alignment of ArV1 and PrincessTrina genome sequences. (B) ArV1 genome sequence alignment with the genome sequences from 7 *Arthrobacter* phages. Red blocks represent aligned regions, and similarity is indicated by the height of the bars.

likely evolved from siphophages (58). If so, then perhaps ArV1, after the addition of the tail sheath and contractile mechanism to the original noncontractile tail design, preserved the ancestral TMP/tail length, which likely not only increased the chance of contacting a suitable host cell in the terrestrial environments but also could be beneficial in penetrating the thick wall of the Gram-positive cell.

As mentioned above, the electron micrographs obtained by TEM (Fig. 1) showed that the contracted tail sheath of ArV1 can be located at both proximal (the usual

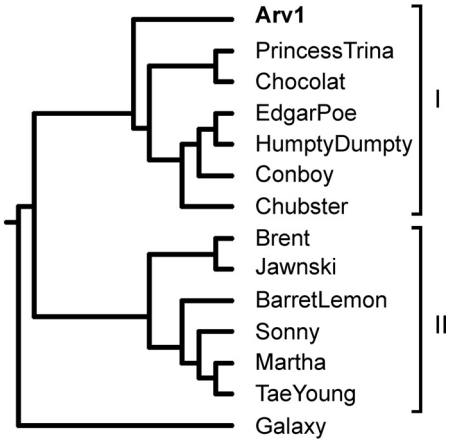

FIG 11 mVISTA-generated phylogenetic tree based on MLAGAN alignment of 14 arthrobacterial phage genome sequences available in GenBank.

location for the contacted sheath of myoviruses) and distal regions of the tail. To our knowledge, a similar phenomenon was first reported in 2005 for *Listeria monocytogenes* phage 01761 (63). Over the years, reports on three more myoviruses, namely, *Burkholderia pseudomallei* phages KS5 and KS14 (64), and ST2 (65), with unusually positioned contracted sheaths have been published. Notably, while both *Listeria monocytogenes* and *Burkholderia pseudomallei* are pathogenic to humans, these two bacteria, just like *Arthrobacter*, are found in soil, suggesting that perhaps the aforementioned phenomenon is not unusual for terrestrial phages. However, the question of why this is inherent to only a few myoviruses remains.

According to the literature, all known contractile phage tail-like structures, such as tailed phages, the type VI secretion system (T6SS), R-type pyocins, and phage tail-like protein translocation structures, are multimeric protein complexes that share an architecture and a set of conserved building blocks (26, 66, 67). All these contractile machineries consist of the main tube, the outer sheath, and the baseplate (26). For many years, bacteriophage T4, the most thoroughly investigated myovirus, served as a model system for all contractile phage tail-like structures. However, quite a number of myoviruses, especially those with smaller genomes, have tails significantly less complex than that of T4 (27, 68). Therefore, by analyzing the available genomic and protein structural data for three well-studied myoviruses (T4, P2, and Mu), Leiman and Shneider identified a minimal set of 12 proteins that comprise a functional contractile phage tail (66) Based on the data provided in Results, we tentatively predict that at least 9 ArV1 structural proteins represent the aforementioned conserved set (Table 3), and this suggests, in accordance with the TEM results, that bacteriophage ArV1 has a "simple" baseplate likely somewhat similar to that of Mu. As was the case with the *Mycobacterium* and *Thermus* phages analyzed by Büttner and colleagues (27), two conserved contractile tail proteins corresponding to T4 gp5 and T4 gp53/gp7 appear to have no recognizable homologues in ArV1. Nevertheless, HHPred revealed a low-probability link between the N terminus of ArV1 gp22 (aa 23 to 93/654) and the N-terminal OB-fold domain of P2 gpV (aa 18 to 93/211). In the case of the C terminus of ArV1 gp22 (aa 568 to 653/654), HHPred returned high-probability hits to the receptor-binding domain of *Lactococcus* phage p2 RBP (aa 171 to 263/264). Based on this, we designated ArV1 gp22 as the tail fiber protein. Nevertheless, a possibility remains that this protein is actually a functional analogue of T4 gp5. As seen in Table 3, two ArV1 proteins, gp12 and gp13, are candidates for components of the tail terminator complex. Notably, both proteins show no sequence similarity with tail terminators of known myophages. However, BLASTP predicted that ArV1 gp12 has a conserved Phage_tail_S family domain (aa 31 to 130/156; pfam05069), whereas ArV1 gp13 has a weak HHPred match (probability of 20%; aa 1 to 122/205) to the N-terminal

TABLE 3 Putative orthologues of conserved T4, P2, and Mu tail proteins[a] predicted in ArV1

	Protein in[b]:			
Category	**T4**	**P2**	**Mu**	**ArV1**
Baseplate hub	gp27	gpD	**gp44**	gp20
Spike/needle	gp5	gpV	gp45	—
Baseplate	**gp25**	gpW	gp46	gp32
	gp6	gpJ	gp47	gp33
	gp53/gp7	gpI	gp48	—
Receptor-binding protein	gp12	gpH	gp49	gp22/gp23
Tail tube initiator	gp48/gp54	gpU	gp43	gp19
Tape measure	gp29	gpT	gp42	gp18
Tail tube	gp19	gpFII	gp40	gp16
Tail sheath	**gp18**	gpFI	gp39	gp15
Tube terminator	gp3	gpS	gp37	gp12/gp13
Sheath terminator	gp15	gpR	gp38	gp12/gp13

[a]Based on the data from reference 66.
[b]HHPred hits are in bold. —, not determined.

part of the head-to-tail connector protein gpU from *E. coli* siphophage. The Phage_tail_S family domain is represented by P2 protein S (69), which is thought to act in tail completion and stable head joining. The gene product gpU, which permanently halts tail polymerization and forms the interface for head attachment, has been shown to be related to phage Mu gp37, which is thought to be a tail terminator protein based on its genomic position and the phenotype of mutants with mutations in the gene encoding it (66, 70, 71). In the case of phage T4, to prevent depolymerization before the tail attaches to the head, the tail tube and sheath are capped by the terminator proteins gp3 and gp15, respectively. gp15 has been shown to interact with the neck protein gp14, the tail tube terminator (gp3), and the last row of gp18 molecules. In the contracted tail, the last interaction helps to maintain the integrity of the tail in its contracted form (72). Taken together, the data on the analysis of ArV1 structural proteins presented in the previous section and summarized in Table 3 indicate that while the myophage ArV1 does, in fact, share certain components with the tails of the 3 most-researched myoviruses, T4, P2, and Mu, the inner core of the ArV1 tail is much more related to the *Siphoviridae* family. Following this observation, we hypothesize that the reason why in nearly 20% of ArV1 virions with contracted tails the sheaths are located at the distal region of the tail (near the baseplate) is because protein-protein interactions between ArV1 sheath and the tail core proteins are likely weak enough to be occasionally lost during phage purification procedures.

As mentioned above, the NCBI list of all fully sequenced phage genomes had, for quite some time, contained only one sequence belonging to *Arthrobacter*-infecting myophages, the genome sequence of bacteriophage ArV1. Later, the pace of *Arthrobacter* phage whole-genome sequence determination rapidly accelerated due to an integrated research and education program, SEA-PHAGES, jointly administered by Graham Hatfull's group at the University of Pittsburgh and the Howard Hughes Medical Institute's Science Education division (http://seaphages.org). Thus, since 2014, the genomes of eight *Arthrobacter* myoviruses have joined ArV1 in the GenBank database. At the time of this writing, GenBank has released five more genome sequences related to that of ArV1: three *Arthrobacter* myoviruses, Chubster (KX670786.1), Conboy (KX522650.1), and Chocolat (KX670787.1), as well as two *Arthrobacter* siphoviruses, HumptyDumpty (KX855962.1) and EdgarPoe (KX855961.1). Comparative genome sequence analysis indicated that all the aforementioned *Arthrobacter* phages, with the exception of myophage Galaxy (which we strongly suspect is not a myovirus), are related to each other, with various degrees of DNA sequence similarity (data not shown), and that they can be grouped into two clusters (Fig. 12). Cluster I comprises seven *Arthrobacter* phages with medium-sized genomes (70 ± 1 kb), while six *Arthrobacter* myoviruses with smaller genomes (50 ± 1 kb) fall into cluster II. Within each

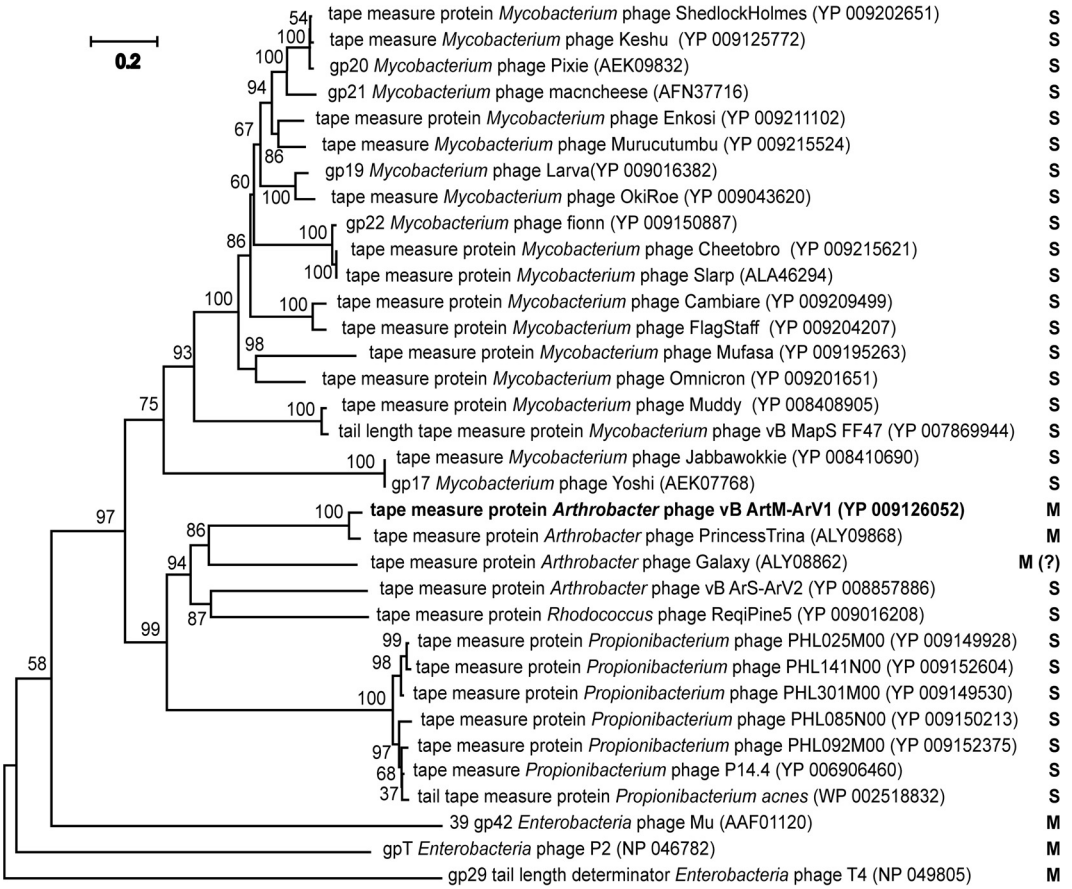

FIG 12 Neighbor-joining tree analysis based on ClustalW alignment of the tape measure protein sequences. The numbers at the nodes indicate the bootstrap probabilities. S, *Siphoviridae*; M, *Myoviridae*.

cluster, the genomes share a high degree of sequence similarity that spans more than 60% of the genome length. Between the clusters, similarity is much less extensive, with regions with evident nucleotide sequence similarity mostly spanning the structural module of each given genome. Nevertheless, as discussed in Results, a set of 32 proteins that belong to different functional groups (e.g., morphogenesis, DNA replication, lysis, etc.) is conserved in all *Arthrobacter* myoviruses listed above. Notably, while phage ArV1 was isolated in Lithuania, all other *Arthrobacter* myophages discussed in this paper were isolated in the United States, hinting at the diverse and cosmopolitan nature of ArV1-like viruses.

Our experience with ArV1 revealed that *Arthrobacter* myoviruses and, apparently, myophages infecting *Actinobacteria* other than *Arthrobacter* as well, are not easy to classify. If one omits purification in CsCl gradients and examines only a limited number of phage particles, virions with contracted tails can be easily overlooked. Consequently, especially if the phage genome sequence analysis is performed by using conventional homology searches only (e.g., BLAST and FASTA), such a virus may be misclassified as a siphovirus. The NCBI database currently contains the genome sequences of three siphoviruses (*Arthrobacter* phages HumptyDumpty and EdgarPoe as well as *Gordonia* siphovirus GMA6) that we strongly suspect are myoviruses. Since in the case of *Actinobacteria* phages, tail protein sequence relationships are nearly impossible to detect using conventional homology searches, we urge our fellow phage biologists to use HHPred or, at the very least, VIRFAM, which can be used for a fast and reliable initial phage assignment. Also, we would like to advocate the usage of the phage nomenclature proposed by Kropinski and colleagues (73). After all, phage names such as "MollyDolly" or "HappySunday," while undeniably beautiful, have no informational

value and are of no use when dealing with extensive lists of hits returned by homology searches. In contrast, by following the phage-naming method proposed in reference 73, two most important characteristics of the virus are being provided: the name of the host and the morphology of the phage particle.

In conclusion, based on an elegant review written by D. Veesler and C. Cambillau (74), the apparent common structure of contractile and noncontractile tails suggests that both siphoviruses and myoviruses evolved from the common progenitor phage. If so, what was this protophage? Was it a siphophage-like virus, one of whose proteins evolved to interact with the tail tube, thus giving rise to *Myoviridae*? Or, perhaps, was it a myophage-like progenitor that, for reasons unknown, lost its sheath to give rise to the family *Siphoviridae*? If one wishes to come closer to unraveling the mysteries of phage evolution, the results of ArV1 genome sequence analysis suggest that *Arthrobacter*-infecting myoviruses may be a good model object to begin with.

MATERIALS AND METHODS

Phages and bacterial strains. Phage ArV1 was isolated from soil samples collected in Vilnius (Lithuania) using *Arthrobacter* sp. strain 68b as the host for phage propagation and phage growth experiments. The bacterial strains used in this study for host range determination are described in Table 1. For phage experiments, bacteria were cultivated in Luria-Bertani broth (LB) or LB agar at 28°C. Bacterial growth was monitored turbidimetrically by reading the optical density at 600 nm (OD_{600}). An OD_{600} of 1.0 corresponded to 9×10^8 *Arthrobacter* sp. strain 68b cells/ml.

Phage techniques. Phage isolation and propagation were carried out as described previously (17). The determination of the efficiency of plating (EOP) was performed as described by Kaliniene and colleagues (75). High-titer phage stocks were diluted and plated in duplicate. Plates incubated at 18, 20, 22, 24, 26, 28, 30, 32, 34, and 37°C were read after 18 to 96 h of incubation. The temperature at which the largest number of plaques formed was taken as the standard for the EOP calculation.

TEM. CsCl density gradient-purified phage particles were diluted to approximately 10^{11} PFU/ml with distilled water, 5 μl of the sample was directly applied on the carbon-coated nitrocellulose grid, excess liquid was drained with filter paper before staining with two successive drops of 2% uranyl acetate (pH 4.5), and the sample was dried and examined in Morgagni 268(D) transmission electron microscope (FEI, Hillsboro, OR, USA).

DNA isolation and restriction analysis. Aliquots of phage suspension (10^{11} to 10^{12} PFU/ml) were subjected to phenol-chloroform extraction and ethanol precipitation as described by Carlson and Miller (76). Isolated phage DNA was subsequently subjected to genome sequencing and restriction digestion analysis.

Restriction digestion was performed with BamHI, Bpu1102I, BspTI, EcoRI, EcoRII, EcoRV, HindIII, KpnI, MboI, NdeI, NheI, NotI, PstI, PvuII, SnaBI, SspI, VspI, XbaI, and XhoI restriction endonucleases (Thermo Fisher Scientific, Lithuania) according to the supplier's recommendations. DNA fragments were separated by electrophoresis in a 0.8% agarose gel containing ethidium bromide. Restriction analysis was performed in triplicate to confirm the results.

Genome sequencing and analysis. The complete genome sequence of ArV1 was determined using Illumina DNA sequencing technology (BaseClear, the Netherlands). Open reading frames (ORFs) were predicted with Glimmer v3.02 (https://www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi) and Geneious Pro v5.5.6. (Biomatters, Auckland, New Zealand). Analysis of the genome sequence was performed using BLAST, PSI-BLAST, and Megablast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) as well as Transeq and Clustal Omega (http://www.ebi.ac.uk) and HHPred, HHblits, and HHsenser (77, 78), Also, tRNAscan-SE 1.21 (http://lowelab.ucsc.edu/tRNAscan-SE/) was used to search for tRNAs. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5 (79), mVISTA (80), and VIRFAM (44).

Analysis of structural proteins. Analysis of ArV1 virion structural proteins was performed as described previously (17).

Accession number(s). The complete genome sequence of *Arthrobacter* bacteriophage ArV1 was deposited in the EMBL nucleotide sequence database (http://www.ebi.ac.uk/ena/data/view/KM879463) and in GenBank under accession number KM879463.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JVI.00023-17.

SUPPLEMENTAL FILE 1, PDF file, 0.2 MB.

ACKNOWLEDGMENT

This work was supported by the Research Council of Lithuania (project no. MIP-042/2012).

REFERENCES

1. Cacciari I, Lippi D. 1987. Arthrobacters: successful arid soil bacteria. A review. Arid Soil Res Rehabil 1:1–30.
2. Jones D, Keddie RM. 2006. The genus *Arthrobacter*, p 945–960 In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (ed), The prokaryotes. Springer, New York, NY.
3. Goodfellow M. 2012. Phylum XXVI. Actinobacteria phyl. nov, p 33–2088. *In* Goodfellow M, Ǩampfer P, Busse H-J, Trujillo ME, Suzuki K-I, Ludwig W, Whitman WB (ed), Bergey's manual of systematic bacteriology. Springer, New York, NY.
4. Hagedorn C, Holt JG. 1975. A nutritional and taxonomic survey of *Arthrobacter* soil isolates. Can J Microbiol 21:353–361. https://doi.org/10.1139/m75-050.
5. Roh SW, Sung Y, Nam YD, Chang HW, Kim KH, Yoon JH, Jeon CO, Oh HM, Bae JW. 2008. *Arthrobacter soli* sp. nov., a novel bacterium isolated from wastewater reservoir sediment. J Microbiol 46:40–44. https://doi.org/10.1007/s12275-007-0239-8.
6. Scheublin TR, Leveau JH. 2013. Isolation of *Arthrobacter* species from the phyllosphere and demonstration of their epiphytic fitness. Microbiologyopen 2:205–213. https://doi.org/10.1002/mbo3.59.
7. Pindi PK, Manorama R, Begum Z, Shivaji S. 2010. *Arthrobacter antarcticus* sp. nov., isolated from an Antarctic marine sediment. Int J Syst Evol Microbiol 60:2263–2266. https://doi.org/10.1099/ijs.0.012989-0.
8. Hanbo Z, Changgun D, Qiyong S, Weimin R, Tao S, Lizhong C, Zhiwei Z, Bin H. 2004. Genetic and physiological diversity of phylogenetically and geographically distinct groups of *Arthrobacter* isolated from lead-zinc mine tailings. FEMS Microbiol Ecol 49:333–341. https://doi.org/10.1016/j.femsec.2004.04.009.
9. Clokie MR, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. Bacteriophage 1:31–45. https://doi.org/10.4161/bact.1.1.14942.
10. Abedon ST. 2008. Phages, ecology, evolution, p 1–28. *In* Abedon ST (ed), Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses. Cambridge University Press, Cambridge, UK.
11. Mann NH. 2005. The third age of phage. PLoS Biol 3:e182. https://doi.org/10.1371/journal.pbio.0030182.
12. Salmond GP, Fineran PC. 2015. A century of the phage: past, present and future. Nat Rev Microbiol 13:777–786. https://doi.org/10.1038/nrmicro3564.
13. Ackermann HW. 2001. Frequency of morphological phage descriptions in the year 2000. Arch Virol 146:843–857. https://doi.org/10.1007/s007050170120.
14. Kutter E, Sulakvelidze A. 2004. Bacteriophages: biology and applications. CRC Press, Boca Raton, FL.
15. Nelson D. 2004. Phage taxonomy: we agree to disagree. J Bacteriol 186:7029–7031. https://doi.org/10.1128/JB.186.21.7029-7031.2004.
16. Ackermann HW. 1998. Tailed bacteriophages: the order caudovirales. Adv Virus Res 51:135–201. https://doi.org/10.1016/S0065-3527(08)60785-X.
17. Šimoliūnas E, Kaliniene L, Stasilo M, Truncaitė L, Zajančkauskaitė A, Staniulis J, Nainys J, Kaupinis A, Valius M, Meškys R. 2014. Isolation and characterization of vB_ArS-ArV2—first *Arthrobacter* sp. infecting bacteriophage with completely sequenced genome. PLoS One 9:e111230. https://doi.org/10.1371/journal.pone.0111230.
18. Brown DR, Holt JG, Pattee PA. 1978. Isolation and characterization of *Arthrobacter* bacteriophages and their application to typing of soil *Arthrobacter*. Appl Environ Microbiol 35:185–191.
19. Niewerth H, Schuldes J, Parschat K, Kiefer P, Vorholt JA, Daniel R, Fetzner S. 2012. Complete genome sequence and metabolic potential of the quinaldine-degrading bacterium *Arthrobacter* sp. Rue61a. BMC Genomics 13:534. https://doi.org/10.1186/1471-2164-13-534.
20. Bradley DE. 1967. Ultrastructure of bacteriophages and bacteriocins. Bacteriol Rev 31:230–314.
21. Serwer P, Hayes SJ, Zaman S, Lieman K, Rolando M, Hardies SC. 2004. Improved isolation of undersampled bacteriophages: finding of distant terminase genes. Virology 329:412–424. https://doi.org/10.1016/j.virol.2004.08.021.
22. Monnet C, Loux V, Gibrat JF, Spinnler E, Barbe V, Vacherie B, Gavory F, Gourbeyre E, Siguier P, Chandler M, Elleuch R, Irlinger F, Vallaeys T. 2010. The *Arthrobacter arilaitensis* Re117 genome sequence reveals its genetic adaptation to the surface of cheese. PLoS One 5:e15489. https://doi.org/10.1371/journal.pone.0015489.
23. Hendrix RW. 2002. Bacteriophages: evolution of the majority. Theor Popul Biol 61:471–480. https://doi.org/10.1006/tpbi.2002.1590.
24. Fokine A, Rossmann MG. 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4:e28281. https://doi.org/10.4161/bact.28281.
25. Dyson ZA, Tucci J, Seviour RJ, Petrovski S. 2015. Lysis to kill: evaluation of the lytic abilities, and genomics of nine bacteriophages infective for *Gordonia* spp. and their potential use in activated sludge foam biocontrol. PLoS One 10:e0134512. https://doi.org/10.1371/journal.pone.0134512.
26. Sarris PF, Ladoukakis ED, Panopoulos NJ, Scoulica EV. 2014. A phage tail-derived element with wide distribution among both prokaryotic domains: a comparative genomic and phylogenetic study. Genome Biol Evol 6:1739–1747. https://doi.org/10.1093/gbe/evu136.
27. Büttner CR, Wu Y, Maxwell KL, Davidson AR. 2016. Baseplate assembly of phage Mu: defining the conserved core components of contractile-tailed phages and related bacterial systems. Proc Natl Acad Sci U S A 113:10174–10179. https://doi.org/10.1073/pnas.1607966113.
28. Stirm S, Bessler W, Fehmel F, Freund-Mölbert E. 1971. Bacteriophage particles with endo-glycosidase activity. J Virol 8:343–346.
29. Schwarzer D, Browning C, Stummeyer K, Oberbeck A, Mühlenhoff M, Gerardy-Schahn R, Leiman PG. 2015. Structure and biochemical characterization of bacteriophage phi92 endosialidase. Virology 477:133–143. https://doi.org/10.1016/j.virol.2014.11.002.
30. Mühlenhoff M, Stummeyer K, Grove M, Sauerborn M, Gerardy-Schahn R. 2003. Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. J Biol Chem 278:12634–12644. https://doi.org/10.1074/jbc.M212048200.
31. Rao VB, Feiss M. 2008. The bacteriophage DNA packaging motor. Annu Rev Genet 42:647–681. https://doi.org/10.1146/annurev.genet.42.110807.091545.
32. Rao VB, Feiss M. 2015. Mechanisms of DNA packaging by large double-stranded DNA viruses. Annu Rev Virol 2:351–378. https://doi.org/10.1146/annurev-virology-100114-055212.
33. Petrovski S, Seviour RJ, Tillett D. 2011. Characterization of the genome of the polyvalent lytic bacteriophage GTE2, which has potential for biocontrol of *Gordonia*-, *Rhodococcus*-, and *Nocardia*-stabilized foams in activated sludge plants. Appl Environ Microbiol 77:3923–3929. https://doi.org/10.1128/AEM.00025-11.
34. Kazlauskas D, Krupovic M, Venclovas Č. 2016. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. Nucleic Acids Res 44:4551–4564. https://doi.org/10.1093/nar/gkw322.
35. Seco EM, Zinder JC, Manhart CM, Lo Piano A, McHenry CS, Ayora S. 2013. Bacteriophage SPP1 DNA replication strategies promote viral and disable host replication in vitro. Nucleic Acids Res 41:1711–1721. https://doi.org/10.1093/nar/gks1290.
36. Mensa-Wilmot K, Seaby R, Alfano C, Wold MC, Gomes B, McMacken R. 1989. Reconstitution of a nine-protein system that initiates bacteriophage lambda DNA replication. J Biol Chem 264:2853–2861.
37. Costa A, Onesti S. 2008. The MCM complex: (just) a replicative helicase? Biochem Soc Trans 36:136–140. https://doi.org/10.1042/BST0360136.
38. van Kessel JC, Marinelli LJ, Hatfull GF. 2008. Recombineering mycobacteria and their phages. Nat Rev Microbiol 6:851–857. https://doi.org/10.1038/nrmicro2014.
39. Belfort M, Derbyshire V, Cousineau B, Lambowitz A. 2002. Mobile introns: pathways and proteins, p 761–783. *In* Craig N, Craigie R, Gellert M, Lambowitz A (ed), Mobile DNA II. ASM Press, Washington, DC.
40. Friedrich NC, Torrents E, Gibb EA, Sahlin M, Sjöberg BM, Edgell DR. 2007. Insertion of a homing endonuclease creates a genes-in-pieces ribonucleotide reductase that retains function. Proc Natl Acad Sci U S A 104:6176–6181. https://doi.org/10.1073/pnas.0609915104.
41. McLennan AG. 2006. The Nudix hydrolase superfamily. Cell Mol Life Sci 63:123–143. https://doi.org/10.1007/s00018-005-5386-7.
42. Warren RA. 1980. Modified bases in bacteriophage DNAs. Annu Rev Microbiol 34:137–158. https://doi.org/10.1146/annurev.mi.34.100180.001033.
43. Young I, Wang I, Roof WD. 2000. Phages will out: strategies of host cell lysis. Trends Microbiol 8:120–128. https://doi.org/10.1016/S0966-842X(00)01705-4.
44. Lopes A, Tavares P, Petit MA, Guérois R, Zinn-Justin S. 2014. Automated identification of tailed bacteriophages and classification according to

their neck organization. BMC Genomics 15:1027. https://doi.org/10.1186/1471-2164-15-1027.
45. Adriaenssens EM, Cowan DA. 2014. Using signature genes as tools to assess environmental viral ecology and diversity. Appl Environ Microbiol 80:4470–4480. https://doi.org/10.1128/AEM.00878-14.
46. Casjens SR. 2008. Diversity among the tailed-bacteriophages that infect the *Enterobacteriaceae*. Res Microbiol 159:340–348. https://doi.org/10.1016/j.resmic.2008.04.005.
47. Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW. 2005. The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. J Bacteriol 187:1091–1104. https://doi.org/10.1128/JB.187.3.1091-1104.2005.
48. Mitchell MS, Matsuzaki S, Imai S, Rao VB. 2002. Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. Nucleic Acids Res 30:4009–4021. https://doi.org/10.1093/nar/gkf524.
49. Johnson MC, Tatum KB, Lynn JS, Brewer TE, Lu S, Washburn BK, Stroupe ME, Jones KM. 2015. *Sinorhizobium meliloti* Phage ΦM9 defines a new group of T4 superfamily phages with unusual genomic features but a common T=16 capsid. J Virol 89:10945–10958. https://doi.org/10.1128/JVI.01353-15.
50. Brewer TE, Stroupe ME, Jones KM. 2014. The genome, proteome and phylogenetic analysis of *Sinorhizobium meliloti* phage ΦM12, the founder of a new group of T4-superfamily phages. Virology 450-451:84–97. https://doi.org/10.1016/j.virol.2013.11.027.
51. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, Sullivan MB, Woyke T, Wommack KE, Stepanauskas R. 2015. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. ISME J 9:2386–2399. https://doi.org/10.1038/ismej.2015.48.
52. Comeau AM, Krisch HM. 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol Biol Evol 25:1321–1332. https://doi.org/10.1093/molbev/msn080.
53. Shen M, Le S, Jin X, Li G, Tan Y, Li M, Zhao X, Shen W, Yang Y, Wang J, Zhu H, Li S, Rao X, Hu F, Lu S. 2016. Characterization and comparative genomic analyses of *Pseudomonas aeruginosa* phage PaoP5: new members assigned to PAK_P1-like viruses. Sci Rep 6:34067. https://doi.org/10.1038/srep34067.
54. Turner D, Reynolds D, Seto D, Mahadevan P. 2013. CoreGenes3.5: a webserver for the determination of core genes from sets of viral and small bacterial genomes. BMC Res Notes 6:140. https://doi.org/10.1186/1756-0500-6-140.
55. Lavigne R, Darius P, Summer EJ, Seto D, Mahadevan P, Nilsson AS, Ackermann HW, Kropinski AM. 2009. Classification of Myoviridae bacteriophages using protein sequence similarity. BMC Microbiol 9:224. https://doi.org/10.1186/1471-2180-9-224.
56. Adriaenssens EM, Edwards R, Nash JH, Mahadevan P, Seto D, Ackermann HW, Lavigne R, Kropinski AM. 2015. Integration of genomic and proteomic analyses in the classification of the *Siphoviridae* family. Virology 477:144–154. https://doi.org/10.1016/j.virol.2014.10.016.
57. Belcaid M, Bergeron A, Poisson G. 2011. The evolution of the tape measure protein: units, duplications and losses. BMC Bioinformatics 12(Suppl 9):S10. https://doi.org/10.1186/1471-2105-12-S9-S10.
58. Davidson AR, Cardarelli L, Pell LG, Radford DR, Maxwell KL. 2012. Long noncontractile tail machines of bacteriophages, p 115–142. *In* Rossman MG, Rao VB (ed), Viral molecular machines. Springer-Verlag, New York, NY.
59. Bohm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, Rigaud JL. 2001. FhuA-mediated phage genome transfer into liposomes: a cryo-electron tomography study. Curr Biol 11:1168–1175. https://doi.org/10.1016/S0960-9822(01)00349-9.
60. Cumby N, Reimer K, Mengin-Lecreulx D, Davidson AR, Maxwell KL. 2015. The phage tail tape measure protein, an inner membrane protein and a periplasmic chaperone play connected roles in the genome injection process of *E. coli* phage HK97. Mol Microbiol 96:437–447. https://doi.org/10.1111/mmi.12918.
61. Boulanger P, Jacquot P, Plançon L, Chami M, Engel A, Parquet C, Herbeuval C, Letellier L. 2008. Phage T5 straight tail fiber is a multifunctional protein acting as a tape measure and carrying fusogenic and muralytic activities. J Biol Chem 283:13556–13564. https://doi.org/10.1074/jbc.M800052200.
62. Hu B, Margolin W, Molineux IJ, Liu J. 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. Proc Natl Acad Sci U S A 112:E4919-4928. https://doi.org/10.1073/pnas.1501064112.
63. Loessner MJ, Calendar RL. 2006. The *Listeria* bacteriophages, pp593–602. *In* Calendar RL (ed), The bacteriophages, 2nd ed. Oxford University Press, New York, NY.
64. Lynch KH, Stothard P, Dennis JJ. 2010. Genomic analysis and relatedness of P2-like phages of the *Burkholderia cepacia* complex. BMC Genomics 11:599. https://doi.org/10.1186/1471-2164-11-599.
65. Yordpratum U, Tattawasart U, Wongratanacheewin S, Sermswan RW. 2011. Novel lytic bacteriophages from soil that lyse *Burkholderia pseudomallei*. FEMS Microbiol Lett 314:81–88. https://doi.org/10.1111/j.1574-6968.2010.02150.x.
66. Leiman PG, Shneider MM. 2012. Contractile tail machines of bacteriophages, p 93–114. *In* Rossman MG, Rao VB (ed), Viral molecular machines. Springer-Verlag, New York, NY.
67. Kube S, Wendler P. 2015. Structural comparison of contractile nanomachines. AIMS Biophys 2:88–115. https://doi.org/10.3934/biophy.2015.2.88.
68. Maxwell KL, Fatehi Hassanabad M, Chang T, Paul VD, Pirani N, Bona D, Edwards AM, Davidson AR. 2013. Structural and functional studies of gpX of *Escherichia coli* phage P2 reveal a widespread role for LysM domains in the baseplates of contractile-tailed phages. J Bacteriol 195:5461–5468. https://doi.org/10.1128/JB.00805-13.
69. Linderoth NA, Julien B, Flick KE, Calendar R, Christie GE. 1994. Molecular cloning and characterization of bacteriophage P2 genes R and S involved in tail completion. Virology 200:347–359. https://doi.org/10.1006/viro.1994.1199.
70. Smith ML, Avanigadda LN, Liddell PW, Kenwright KM, Howe MM. 2010. Identification of the J and K genes in the bacteriophage Mu genome sequence. FEMS Microbiol Lett 313:29–32. https://doi.org/10.1111/j.1574-6968.2010.02128.x.
71. Pell LG, Liu A, Edmonds L, Donaldson LW, Howell PL, Davidson AR. 2009. The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. J Mol Biol 389:938–951. https://doi.org/10.1016/j.jmb.2009.04.072.
72. Yap ML, Rossmann MG. 2014. Structure and function of bacteriophage T4. Future Microbiol 9:1319–1327. https://doi.org/10.2217/fmb.14.91.
73. Kropinski AM, Prangishvili D, Lavigne R. 2009. Position paper: the creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. Environ Microbiol 11:2775–2777. https://doi.org/10.1111/j.1462-2920.2009.01970.x.
74. Veesler D, Cambillau C. 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev 75:423–433. https://doi.org/10.1128/MMBR.00014-11.
75. Kaliniene L, Klausa V, Truncaite L. 2010. Low-temperature T4-like coliphages vB_EcoM-VR5, vB_EcoM-VR7 and vB_EcoM-VR20. Arch Virol 155:871–880. https://doi.org/10.1007/s00705-010-0656-6.
76. Carlson K, Miller E. 1994. Experiments in T4 genetics, p 419–483. *In* Karam JD (ed), Bacteriophage T4. ASM Press, Washington DC.
77. Alva V, Nam SZ, Söding J, Lupas AN. 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. https://doi.org/10.1093/nar/gkw348.
78. Söding J, Biegert A, Lupas AN. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Suppl 2):W244–W248. https://doi.org/10.1093/nar/gki408.
79. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. https://doi.org/10.1093/molbev/msr121.
80. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res 32(Suppl 2):W273–W279. https://doi.org/10.1093/nar/gkh458.
81. Stanislauskienė R, Rudenkov M, Karvelis L, Gasparavičiūtė R, Meškienė R, & Ccaron;asaite V, Meškys R. 2011. Analysis of phthalate degradation operon from *Arthrobacter* sp. 68b. Biologija 57:45–54.
82. Kutanovas S, Rutkienė R, Urbelis G, Tauraitė D, Stankevičiūtė J, Meškys R. 2013. Bioconversion of methylpyrazines and pyridines using novel pyrazines-degrading microorganisms. Chemija 24:67–73.
83. Gasparavičiūtė R, Kropa A, Meškys R. 2006. A new *Arthrobacter* strain utilizing 4-hydroxypyridine. Biologija 4:41–45.
84. Semėnaitė R, Gasparavičiūtė R, Duran R, Precigou S, Marcinkevičienė L,

Bachmatova I, Meškys R. 2003. Genetic diversity of 2-hydroxypyridine-degrading soil bacteria. Biologija 2:27–29.
85. Stanislauskiene R, Gasparaviciute R, Vaitekunas J, Meskiene R, Rutkiene R, Casaite V, Meskys R. 2012. Construction of *Escherichia coli-Arthrobacter-Rhodococcus* shuttle vectors based on a cryptic plasmid from *Arthrobacter rhombi* and investigation of their application for functional screening. FEMS Microbiol Lett 327:78–86. https://doi.org/10.1111/j.1574-6968.2011.02462.x.
86. Šimoliūnas E, Kaliniene LTruncaitė L, Zajančkauskaitė A, Staniulis Kaupinis J, Ger A, Valius M, Meškys M, ·· R. 2013. *Klebsiella* phage vB_KleM-RaK2—a giant singleton virus of the family *Myoviridae*. PLoS One 8:e60717. https://doi.org/10.1371/journal.pone.0060717.