Journal of Structural Biology 200 (2017) 343–359

Contents lists available at ScienceDirect

Journal of Structural Biology

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

This article is part of a Special Issue honoring the scientific contributions of Donald L. D. Caspar

Structure, proteome and genome of *Sinorhizobium meliloti* phage ΦM5: A virus with LUZ24-like morphology and a highly mosaic genome☆

Matthew C. Johnson[a,b,1,2], Marta Sena-Velez[a,1], Brian K. Washburn[a,c], Georgia N. Platt[a], Stephen Lu[a], Tess E. Brewer[a,3], Jason S. Lynn[a], M. Elizabeth Stroupe[a,b,*], Kathryn M. Jones[a,*]

[a] Department of Biological Science, Florida State University, Biology Unit I, 230A, 89 Chieftain Way, Tallahassee, FL 32306-4370, United States
[b] Institute of Molecular Biology, Florida State University, United States
[c] Department of Biological Science Core Facilities, Florida State University, United States

ARTICLE INFO

Keywords:
Bacteriophage
ΦM5
phiM5
$T = 7$ capsid geometry
Cryo-EM
Rhizophage tail fiber
Sinorhizobium meliloti
Alphaproteobacteria
Icosahedral
Short direct terminal repeat
Terminase
LUZ24 phage
Genome
Proteome
ΦM12
ΦM9
Podovirus

ABSTRACT

Bacteriophages of nitrogen-fixing rhizobial bacteria are revealing a wealth of novel structures, diverse enzyme combinations and genomic features. Here we report the cryo-EM structure of the phage capsid at 4.9–5.7 Å-resolution, the phage particle proteome, and the genome of the *Sinorhizobium meliloti*-infecting Podovirus ΦM5. This is the first structure of a phage with a capsid and capsid-associated structural proteins related to those of the LUZ24-like viruses that infect *Pseudomonas aeruginosa*. Like many other Podoviruses, ΦM5 is a $T = 7$ icosahedron with a smooth capsid and short, relatively featureless tail. Nonetheless, this group is phylogenetically quite distinct from Podoviruses of the well-characterized T7, P22, and epsilon 15 supergroups. Structurally, a distinct bridge of density that appears unique to ΦM5 reaches down the body of the coat protein to the extended loop that interacts with the next monomer in a hexamer, perhaps stabilizing the mature capsid. Further, the predicted tail fibers of ΦM5 are quite different from those of enteric bacteria phages, but have domains in common with other rhizophages. Genomically, ΦM5 is highly mosaic. The ΦM5 genome is 44,005 bp with 357 bp direct terminal repeats (DTRs) and 58 unique ORFs. Surprisingly, the capsid structural module, the tail module, the DNA-packaging terminase, the DNA replication module and the integrase each appear to be from a different lineage. One of the most unusual features of ΦM5 is its terminase whose large subunit is quite different from previously-described short-DTR-generating packaging machines and does not fit into any of the established phylogenetic groups.

1. Introduction

Nitrogen-fixing rhizobial bacteria that form species-specific mutualisms with host legume plants are among the most important bacteria in soils. The interaction between rhizobia and plant hosts has been actively studied for over a century, however, the interaction between these bacteria and the bacteriophages that prey upon them has received less attention until recently. New genome sequences and structural analyses have shown that many rhizophages are quite novel (Brewer et al., 2014; Crockett et al., 2015; Deak et al., 2010; Dziewit et al., 2014; Ganyu et al., 2005; Halmillawewa et al., 2015, 2014a,b; Henn et al., 2013a,b; Hodson et al., 2015; Johnson et al., 2015; Restrepo-Cordoba et al., 2014; Santamaria et al., 2014; Schouten et al., 2015; Schulmeister et al., 2009; Stroupe et al., 2014). Many sequenced rhizophages do not fit well into established phage taxonomy, which is dominated by phages that infect a limited diversity of hosts, mostly gammaproteobacteria, cyanobacteria, *Staphylococcus*, *Bacillus* and *Mycobacteria* (Adams et al., 2016). The majority of characterized rhizobial phages are Myoviruses, while only a few rhizobial Podoviruses have been studied in detail (Halmillawewa et al., 2014a; Santamaria et al., 2014; Schouten et al., 2015). Until now, no rhizobial Podoviruses have been analyzed at the structural or proteomic level. Here we report the

☆ This Special Issue was edited by Piotr Fajer, Alexei S. Soares and Kenneth Taylor and represents a Festschrift honoring Donald L. D. Caspar on the occasion of his 90th birthday, based on a meeting held at Florida State University in January 2017.

* Corresponding authors at: Institute of Molecular Biophysics, Florida State University, 91 Chieftain Way Tallahassee, FL 32306-4380, United States (M.E. Stroupe). Department of Biological Science, Florida State University, Biology Unit I, 230A, 89 Chieftain Way, Tallahassee, FL 32306-4370, United States (K.M. Jones).
E-mail addresses: mestroupe@bio.fsu.edu (M.E. Stroupe), kmjones@bio.fsu.edu (K.M. Jones).

[1] Equal contribution.
[2] Present address: Department of Biochemistry, University of Washington, Seattle, WA, United States.
[3] Present address: Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, United States.

http://dx.doi.org/10.1016/j.jsb.2017.08.005
Received 3 May 2017; Received in revised form 24 July 2017; Accepted 21 August 2017
Available online 24 August 2017
1047-8477/ © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

$T = 7$ capsid structure at 4.9 Å resolution of *Sinorhizobium meliloti* phage ΦM5, the first capsid structure of a Podovirus infecting a rhizobial bacterium. We have also determined the virus particle proteome and the 44,005 bp genome sequence.

When originally characterized, ΦM5 was found to infect *Sinorhizobium* SU47-derived strains and to be incapable of efficient generalized transduction (Finan et al., 1984). Subsequently, ΦM5 infection of *S. meliloti* 1021 was found to be dependent upon an intact lipopolysaccharide (LPS) core (Campbell et al., 2002, 2003), and on the presence of amino acids 204–205 of the outer membrane protein RopA1 (Crook et al., 2013). Our initial examination of ΦM5 by transmission electron microscopy (TEM) showed that it has a short non-contractile tail and is thus a member of the Podovirus family (Adriaenssens et al., 2017). However, the high level of genomic mosaicism and diversity within this group makes more precise classification of new Podoviruses a perpetual challenge (Grose and Casjens, 2014; Lavigne et al., 2008; Lawrence et al., 2002). Podoviruses vary based on capsid morphology (Suhanovsky and Teschke, 2015), DNA packaging strategy (Grose and Casjens, 2014), DNA replication enzymes (Weigel and Seitz, 2006), and lysis/lysogeny genes (Howard-Varona et al., 2017). The modular nature of phage genomes, due to interphage recombination, means that gene cassettes encoding proteins that accomplish these separable functions are often found in unexpected combinations (Botstein, 1980; Iranzo et al., 2016; Krupovic et al., 2011; Veesler and Cambillau, 2011; Weigel and Seitz, 2006). There are a growing number of examples of phages with extreme genomic mosaicism combining modules from surprisingly different genetic lineages (Glazko et al., 2007; Zhan et al., 2016). *Sinorhizobium meliloti* phage ΦM5 is a prime example of this type of extreme bacteriophage genomic mosaicism.

2. Materials and methods

2.1. Bacterial strains, phage isolates, and growth conditions

S. meliloti 1021 (Meade et al., 1982) was grown at 30 °C in LBMC medium (Glazebrook and Walker, 1991) or tryptone yeast medium (0.5% tryptone, 0.3% yeast extract, 10 mM $CaCl_2$) supplemented with 500 μg/mL streptomycin. Optimal production of ΦM5 virions was obtained by inoculating 10 μL of crude phage preparation into 25 mL of *S. meliloti* 1021 at an optical density at 600 nm (OD600) of 0.1–0.2. The infected culture was incubated at 30 °C overnight or until lysis was apparent, at which point it was centrifuged at 3800 × *g* for 30 min to remove cellular debris. The supernatant was extracted twice with chloroform (Finan et al., 1984). The phage lysate was stored over 1/5 vol of chloroform at 4 °C until further purification. Phage titers were monitored by plaque assay (Finan et al., 1984).

2.2. Phage purification for genomic DNA sequencing

Chloroform-extracted phage ΦM5 was concentrated and washed in an Amicon concentrator (Millipore, Billerica, MA) with a 50-kDa molecular mass cutoff. Concentrated phages were suspended in 10 mL of buffer EX from the Large Construct kit (Qiagen, Valencia, CA) and treated twice with DNase by using 1 U (80 μg) of ATP-dependent exonuclease (Qiagen) to remove *S. meliloti* genomic DNA (Johnson et al., 2015). Prior to capsid lysis, the ATP was removed to inactivate the exonuclease by washing in an Amicon concentrator with a 50-kDa molecular mass cutoff. Phages were lysed at 65 °C for 1 h in phage buffer with 0.5 M EDTA, 0.5% SDS, and 25 mg/mL proteinase K. Phage DNA was isolated by standard methods (Sambrook and Russell, 2001). After resuspension, phage DNA was treated with 1 mg of RNase A (Qiagen), and repurified by standard methods (Sambrook and Russell, 2001).

2.3. Illumina sequencing of the ΦM5 genome and genome assembly

Sequencing was performed as previously described (Johnson et al., 2015). Briefly, two separate ΦM5 DNA samples, each from a plaque-purified phage sample, were sheared to an ~860-bp average size with a Diagenode Bioruptor. Indexed libraries were constructed with an NEBNext Ultra DNA Library Prep kit for Illumina (NEB, Ipswich, MA) in accordance with the manufacturer's instructions. For each library, 1 μg of DNA was end repaired and ligated to NEBNext adapters. Fragments of ~860 bp were isolated by electrophoresis on Bio-Rad Low Range Ultra Agarose and amplified for 8–10 cycles with NEB High-Fidelity 2× master mix and NEBNext multiplex oligonucleotides. Library size distribution was measured on an Agilent Bioanalyzer high-sensitivity chip, and quantity was determined with the KAPA Biosystems Library Quantification kit. Paired-end 300-base sequence reads were generated on an Illumina MiSeq with a 600-cycle MiSeq v3 Reagent kit. Genome assembly from MiSeq reads was performed with Lasergene SeqMan Pro v. 11.2.1.25 (DNAStar, Madison, WI). Plaque isolate 1 produced a major contig of 47,408 bp with a 3760 bp circular permutation, whereas isolate 2 produced a major contig of 45,218 bp with a circular permutation of 1570 bp. Position 1 of the major contig of isolate 2 matched position 19,538 of the major contig from isolate 1. In analyzing the sequence assembly, we found a region of higher than average Illumina read coverage at the ends of the 5.1 assembly, which can be indicative of direct terminal repeats (DTRs). The sequences of genome ends were determined by restriction mapping with the enzymes *Hind*III, *Xba*I, *Blp*I, and *Alw*NI (New England Biolabs, Ipswich, MA), and direct Sanger sequencing of purified phage DNA.

2.4. ORF and sequence motif prediction and analysis

Open reading frames (ORFs) were predicted with GeneMark.hmm for prokaryotes (version 2) (Lukashin and Borodovsky, 1998), and the NCBI ORF Finder (Sayers et al., 2011). The genome was searched for tRNA sequences with tRNAScan-SE (Lowe and Eddy, 1997). Searches for promoters predicted to be recognized by *S. meliloti* 1021 sigma factors were performed using the promoter sequence motif data from Schlüter et al. (2013) and the PhiSite promoter hunter (http://www.phisite.org/main/index.php?nav=tools&nav_sel=hunter) (Klucar et al., 2010; Stano and Klucar, 2011).

2.5. Construction of genomic alignments, amino acid sequence alignments, and phylogenetic trees

The best homolog of the ΦM5 genome in public databases is a prophage found integrated into the chromosome of *Rhizobium favelukesii* LPU83 chromosome from positions 1,667,000–1,710,000. To determine the degree of genomic synteny between these sequences, the ΦM5 genome and the *Rhizobium favelukesii* LPU83 prophage were aligned with the Mauve (Darling et al., 2004) plugin in Geneious version 10 (https://www.geneious.com) (Kearse et al., 2012). For phylogenetic trees, MUSCLE multiple amino acid sequence alignments were performed in Geneious (Edgar, 2004; Kearse et al., 2012). The maximum number of iterations selected was eight, with the anchor optimization option. The trees from iterations 1 and 2 were not retained. The distance measure for iteration 1 was kmer6_6, and that for subsequent iterations was pctid_kimura. The clustering method used for all iterations was UPGMB (which is based on a combination of both the unweighted-pair group method using average linkages and neighbor joining). Un-rooted PhyML trees were constructed from the MUSCLE alignments with the PhyML plugin within Geneious (Guindon and Gascuel, 2003; Lefort et al., 2012). PhyML was performed with the LG amino acid substitution matrix (Le and Gascuel, 2008) with the proportion of invariable sites fixed and four substitution rate categories. The fast nearest-neighbor interchange tree topology search (Desper and Gascuel, 2002) was used, and 100 boot-straps were performed.

2.6. Phage purification for cryo-EM and proteomic analysis

ΦM5-infected cell lysate was prepared by inoculating 0.25 mL of fresh lysate into 375 mL of S. meliloti 1021 culture at an approximate OD_{600} of 0.2 and allowing lysis to proceed overnight. The lysate was

centrifuged at 3800 × *g* for 20 min to remove cellular debris. The supernatant was extracted once with chloroform, precipitated twice in 10% polyethylene glycol (PEG) 8000/0.5 M NaCl, and further chloroform extracted to remove PEG (Yamamoto et al., 1970). The PEG-purified phage was further concentrated on a 30 kDa MWCO concentrator (Pall Corporation, Port Washington, NY, USA). Concentrated phage was layered onto a continuous density gradient of 10–50% OptiPrep density gradient medium (Sigma-Aldrich) in gradient buffer (20 mM Tris-HCl, pH 7, 100 mM KCl, 5 mM $MgSO_4$) and centrifuged at 41,000 × *g* for 8 h in an SW-41 rotor in a Beckman Coulter Optima L-100 XP ultracentrifuge. Fractions were collected on a Brandel BR-188 Density Gradient Fractionation System with a 1.5 mL/min flow rate and 10 s per fraction. At each step, the phage was titered to monitor recovery. Fractions containing high concentrations of phage were assayed for capsid integrity by EM of samples negatively stained with 1% uranyl formate. Selected fractions were further concentrated on a 30 kDa MWCO concentrator at low speed (≤ 2000 × *g*) to prevent phage rupture.

2.7. Proteomic analysis of phage particles

ΦM5 phage was prepared as described above. The titer of infective phage was quantified and the total protein concentration was determined using Bio-Rad Protein Assay Dye Reagent (Bio-Rad, Hercules, CA). Approximately 2×10^{09} phage particles were prepared for shotgun proteomics as previously described (Brewer et al., 2014) using a Filter-Aided Sample Prep (FASP) kit (Expedeon USA, San Diego) and trypsin from porcine pancreas (Sigma-Aldrich, St. Louis), according to the manufacturer's instructions (Wisniewski et al., 2009). The shotgun proteome analysis was performed at the Florida State University Translational Science Laboratory, as previously described (Brewer et al., 2014).

Tandem mass spectra were extracted, charge state deconvoluted and deisotoped by Protein Discoverer (version 1.4) (Thermo-Scientific). All MS/MS samples were analyzed using SequestHT (version 1.4.0.288, Thermo-Scientific), X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1), and the Percolator peptide validator. Sequest and X! Tandem were used to search a list of all 199 ORFs originally predicted for ΦM5 (database file PhageM5formatted.fasta). Scaffold (version Scaffold_4.0.5, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. The peptide identification threshold was a false discovery rate of 0.1% established by the Scaffold Local FDR algorithm. Protein identification threshold was 1% FDR with a minimum of 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003). The full Scaffold protein report is in Supplemental Table 1.

The protein abundance index (PAI) for each identified phage protein was calculated as PAI = number of observed exclusive unique peptides/theoretical peptides per identified phage protein (Rappsilber et al., 2002). Unique theoretical peptides were calculated for all detected ΦM5 proteins using the MS-Digest function in the Protein Prospector program at the website http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest (Chalkley et al., 2005), with the settings trypsin digest, 2 maximum missed cleavages; peptide mass: 350–5000; minimum peptide length 6; constant modification: Carbamidomethyl (C); variable modifications: Oxidation (M) and Phospho (STY); and report multiple charges. The calculation of protein content weight percent for each identified phage protein was emPAI × MW/Σ(emPAI × MW) of all identified phage proteins × 100 (Ishihama et al., 2005). emPAI is the exponentially modified protein abundance index = $10^{PAI} - 1$ (Ishihama et al., 2005).

2.8. Cryo-EM sample preparation and data collection

Purified phage particles were applied to glow-discharged EM grids (Quantifoil 2/2), and rapidly blotted and plunged into liquid ethane using a Vitrobot (FEI, Hillsboro, OR) plunging apparatus, set to 4 °C and 100% relative humidity. Data collection was performed on a Titan Krios TEM (FEI) equipped with a DE20 direct electron detector (Direct Electron, San Diego, CA), via the Leginon automation package (Carragher et al., 2000; Shrum et al., 2012; Suloway et al., 2005). Images were recorded at 22,500x nominal magnification, with a defocus values varying from −3.5 to −1.0 μm, and a total electron dose of 60 $e^-/Å^2$.

2.9. Image processing and three-dimensional (3D) reconstruction

Early processing tasks were performed within the Appion package (Lander et al., 2009). Particles were picked with the reference-free application Dogpicker, and defocus estimation for contrast transfer function (CTF) correction was performed using Ace2 and CTFFind functions (Mallick et al., 2005; Mindell and Grigorieff, 2003). This single particle data set (15,895 particles total) was aligned and reconstructed within the Relion package (Scheres, 2012), followed by Frealign (Grigorieff, 2007). After 17 rounds of icosahedrally-symmetrized frequency refinement using Frealign, the reconstruction (thresholded to include only the highest-scoring 12,129 particles) reached a resolution of 4.9 Å at a Fourier Shell Correlation (FSC) of 0.143, and 5.7 Å at a Fourier Shell Correlation (FSC) of 0.5, as measured between the Frealign-generated half maps (Supplemental Fig. 1). Three-dimensional classification did not improve the resolution of the reconstruction. The final map was sharpened by applying a B-factor of −339.19, as calculated by EM-BFACTOR (Fernandez et al., 2008).

3. Results and discussion

3.1. The structural genes of ΦM5 are similar to those of LUZ24 phages of Pseudomonas and Φeco32 phages of Escherichia coli

Initial genome sequence analysis of ΦM5 showed only short islands of DNA sequence homology and limited overall open reading frame (ORF) synteny with characterized phages (data not shown). Although ΦM5 does not have conserved overall synteny with other characterized phages, genes predicted to be involved in related functions are arranged in modules from separate identifiable lineages. One of these modules on the sense strand of the chromosome (positions 2671–9973) contains structural ORFs that have very good homology to the LUZ24 phages that infect species of *Pseudomonas* (ORFs: M5_04 [portal/head-tail connector protein]; M5_09 [putative scaffolding protein]; M5_11 [major capsid protein]; M5_15; and M5_16 [tail tubular protein A]) (Altschul et al., 1997; Ceyssens et al., 2008; Drummond et al., 2012; Edgar, 2004). The three ORFs encoding the predicted portal, capsid and tail tubular A proteins are also similar to those of the ΦEco32 viruses that infect *E. coli* (Mirzaei et al., 2014; Savalia et al., 2008). The major capsid protein of ΦM5 (M5_11) is 37% identical to that of LUZ24 and 30% identical to ΦEco32. Fig. 1A shows an unrooted phylogenetic tree constructed by aligning a conserved internal segment of the ΦM5 major capsid protein with its orthologues from other Podoviruses (sequence information in Supplemental Table 2A). In the phylogenetic tree shown in Fig. 1A, the low bootstrap percentages at the nodes separating ΦM5 from the LUZ24 phages and the ΦEco32 phages reflect low confidence in these branch points obtained from randomized replicate phylogenies (Felsenstein, 1985). Thus, it is difficult to discern the proper relationship between ΦM5 and the LUZ24 phages and ΦEco32 phages. Morphologically, ΦM5 (Fig. 2A) resembles the C1-morphology, icosahedral LUZ24 phages (Ceyssens et al., 2008) much more strongly than it resembles the C3-morphology, elongated ΦEco32 phages (Mirzaei et al., 2014; Savalia et al., 2008). Since ΦM5 morphology and structural gene synteny suggest a closer relationship to the LUZ24 phages than to the ΦEco32 phages, a second tree was constructed to elucidate these relationships. Internal segments of the 5 conserved ORFs in genome

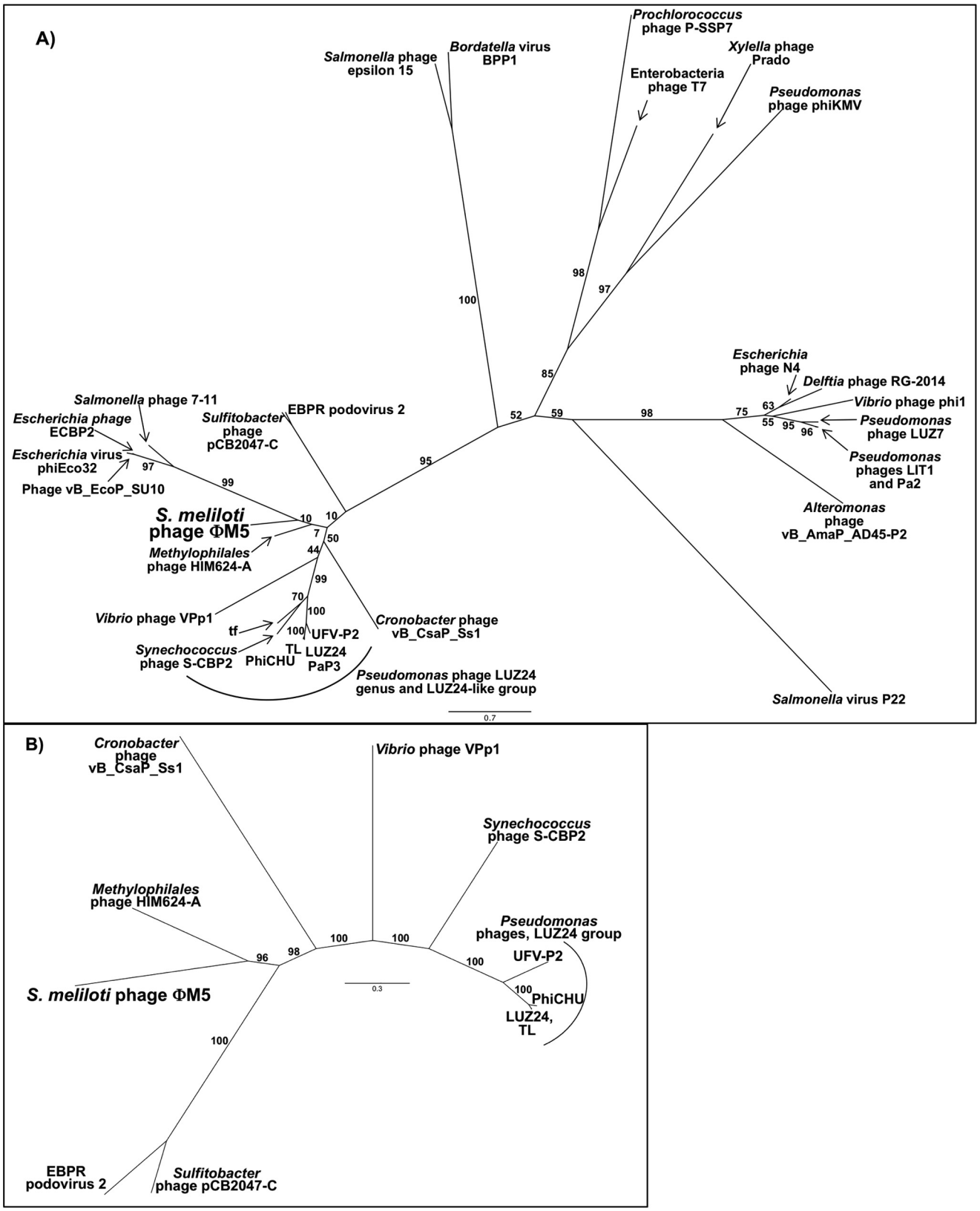


Fig. 1. An unrooted phyML tree based on major structural proteins. A) A conserved internal segment of the ΦM5 major capsid protein (M5_11, amino acids 16–318) was aligned in MUSCLE with orthologous sequences from other phages. The major capsid protein of ΦM5 (M5_11) is 37% identical to that of LUZ24 and 30% identical to ΦEco32. (See Supplemental Table 2A for protein sequence information). B) To further define the relationships between ΦM5 ORFs in the structural region of the genome and those of the LUZ24-like phages, the following sequences were concatenated and aligned with orthologous sequences from other LUZ24-like phages: M5_11, capsid; M5_04, portal protein; M5_16, tail tubular protein A; M5_09, scaffolding protein; M5_15, conserved LUZ24 phage protein). The proteins of the structural region of the ΦM5 genome are most closely related to those of *Methylophilales* phage HIM624-A (Brown et al., 2013). Other closely-related phages are *Sulfitobacter* phage ΦCB2047-C (Ankrah et al., 2014), EPBR Podovirus 2 (Skennerton et al., 2011), *Vibrio* phage VPp1 (Peng et al., 2013), *Synechococcus* phage S-CBP2 (Dekel-Bird et al., 2013; Huang et al., 2015) and *Cronobacter* phage vB_CsaP_Ss1 (Endersen et al., 2015). The bootstrap percentage shown for each branch reflects the degree of confidence in the placement of that node in the phylogenetic reconstruction. The bar indicates branch distance. (See Supplemental Table 2B for protein sequence information).

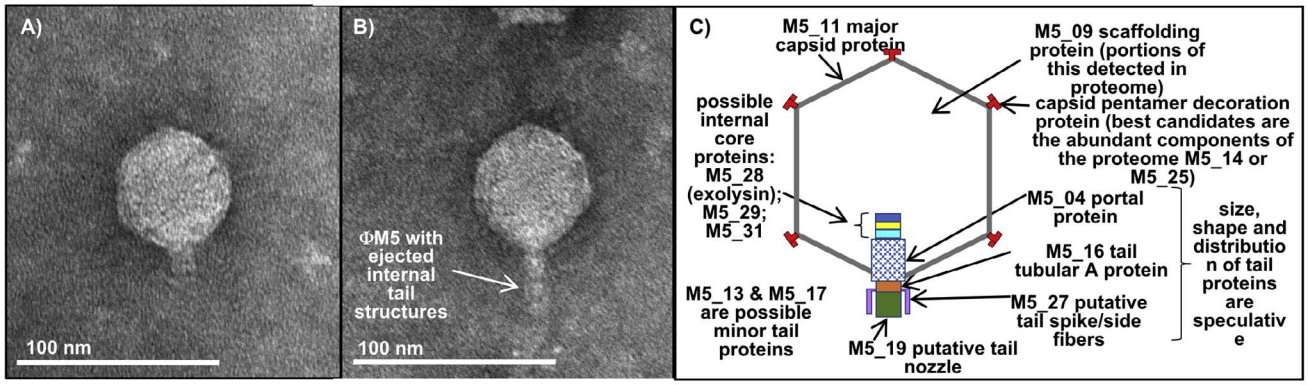


Fig. 2. TEM of the ΦM5 phage particle and predicted location of structural proteins. A) The ΦM5 capsid has C1 icosahedral morphology and is approximately 75 nm in diameter with a nearly featureless tail of 12–18 nm in length and 10–12 nm in diameter. Bar corresponds to 100 nm. B) A ΦM5 virion from which the internal phage tail proteins have been ejected in the absence of a host cell. C) Predicted positions of the structural proteins of ΦM5 shown on a schematic of a generic Podovirus.

segment 2671–9973 (Supplemental Table 2B) were concatenated, aligned, and used to produce the unrooted tree shown in Fig. 1B. This tree shows that the ΦM5 structural region genes are most similar to those of *Methylophilales* phage HIM624-A. This phage is known from a partial genome sequence obtained from a metagenomic community associated with *Trichodesmium* marine cyanobacteria (Brown et al., 2013). No information about phage morphology is available for this phage or any close relatives of ΦM5 except the LUZ24 phages and *Cronobacter* phage vB_CsaP_Ss1, which has also been observed to be a Podovirus (Endersen et al., 2015). This phylogeny based on multiple conserved predicted structural proteins and assembly proteins suggests that the ΦM5 structural module has a common ancestor with the LUZ24 phages.

The absence of structural information for many groups of phages leaves serious gaps in our knowledge of the links between capsid morphology and structural protein phylogeny. The majority of the podoviral structures that have been deposited in the EMDataBank (Lawson et al., 2016) are for phages in a relatively small number of subfamilies and genera, with a third of them in the T7 supergroup. The cryo-EM structures of T7, epsilon 15, and P22 reveal that all these Podoviruses have $T = 7$ icosahedral symmetry despite the highly diverged protein sequences of their major capsid proteins (Suhanovsky and Teschke, 2015). An important goal of modern comparative morphology is to use structural and phylogenetic data to understand how highly diverged protein sequences assume common secondary and higher order structure. The tree shown in Fig. 1A suggests that the major capsid proteins of ΦM5, the LUZ24 phages, and the ΦEco32 phages are highly diverged from the well-characterized Podoviruses such as T7, epsilon 15, and P22. No capsid structure of a LUZ24 phage or a ΦEco32 phage has yet been solved. Therefore, this structure of the ΦM5 capsid is the first from this highly-diverged branch of the Podovirus capsid phylogenetic tree.

3.2. Cryo-EM structure of the ΦM5 capsid

In ΦM5, seven HK97-like coat proteins form the characteristic $T = 7$ hexamer-plus-one asymmetric unit that lock together to complete the icosahedron (Fig. 3). The coat is relatively smooth with a small turret that sits about 5 nm above the plus-one monomer that forms the fivefold axis of symmetry (Fig. 3, upper inset). Each of the five subunits that form the turret presents four columns of density that have a right-hand twist. Each column of density is about 1 nm in diameter and fits a generic alpha helix of 20 amino acids, suggesting that the pentamer cap is made up of a protein that contains a four-helix bundle. This distinctive morphology that runs vertically up from the pentamer contrasts with the immunoglobulin-like (Ig-like) fold-containing protein observed at the pentamer of *Vibrio* phage SIO-2 where the central beta barrel of one pentamer decoration protein points to the next symmetric position (Lander et al., 2012). The identity of this turret decoration protein is not obvious from the ΦM5 proteome (Table 1), but possible candidates are M5_25 and M5_14 (see below). Despite extensive efforts at asymmetric reconstruction, the tail was not visible in the raw images and did not provide sufficient contrast to break the icosahedral symmetry. The lack of high contrast features in the tail is also evident in the negative-stained TEM shown in Fig. 2A.

ΦM5-specific coat protein features of the capsid protein account for all of the structural features of the capsid, which suggests there is no external structural accessory protein to stabilize the 75 nm-diameter capsid. In contrast, another phage of *S. meliloti*, ΦM9, appears to have an accessory protein that stabilizes the hexamer from the larger 112 nm-diameter, $T = 16$ capsid (Johnson et al., 2015). The ΦM5 capsid contains a G-loop between α3 and α4 (residues 130–149) followed by a 30 amino acid insertion to the A domain between the first β-strand of its central five-stranded 15,423 β sheet and α5 (residues 165–185). A distinct bridge of density that could be filled by the A domain insertion reaches down from the β hinge of the A domain to the hairpin of the E loop within the same monomer (Fig. 3, lower inset). The A-domain insertion, G loop, and the E loop (residues 55–78) are three of the most variable regions of homologous capsid proteins. A host of examples of species-specific modifications work in a variety of ways to stabilize the HK97 central fold (Suhanovsky and Teschke, 2015). For example, in P22, a domain insertion called the I domain sits between the third and fourth strand of the A domain's central β-sheet (Hryc et al., 2017). In ΦM5, perhaps the A-domain bridge stabilizes the capsid by rigidifying the structurally important β-hinge after maturation (Teschke and Parent, 2010) through its interaction with the E loop, which then reaches into the neighboring member of the hexamer. This A-domain bridge may represent a novel method of capsid stabilization distinct from those observed in other $T = 7$ phages.

3.3. The ΦM5 virion proteome provides clues to the identity of several phage ORFs

The major capsid protein M5_11 is the most highly-represented protein in the ΦM5 proteome in total mass spectrum counts (Table 1). This, along with 14 other ORFs were detected in the phage particle proteome (Table 1). Three of these ORFs have > 25% identity to phage LUZ24 proteins: the major capsid protein; the portal (head-tail-connector) protein M5_04; and a scaffolding protein M5_09, which is required for proper capsid assembly in many phages (Aksyuk and Rossmann, 2011). A schematic of how these proteins would fit into a generic Podovirus is shown in Fig. 2C. There are an additional 3 LUZ24-like ORFs in the phage particle: M5_25; M5_16, tail tubular protein A; and M5_27 (Table 1, Fig. 2C). ΦM5 ORF M5_25 has moderate overall

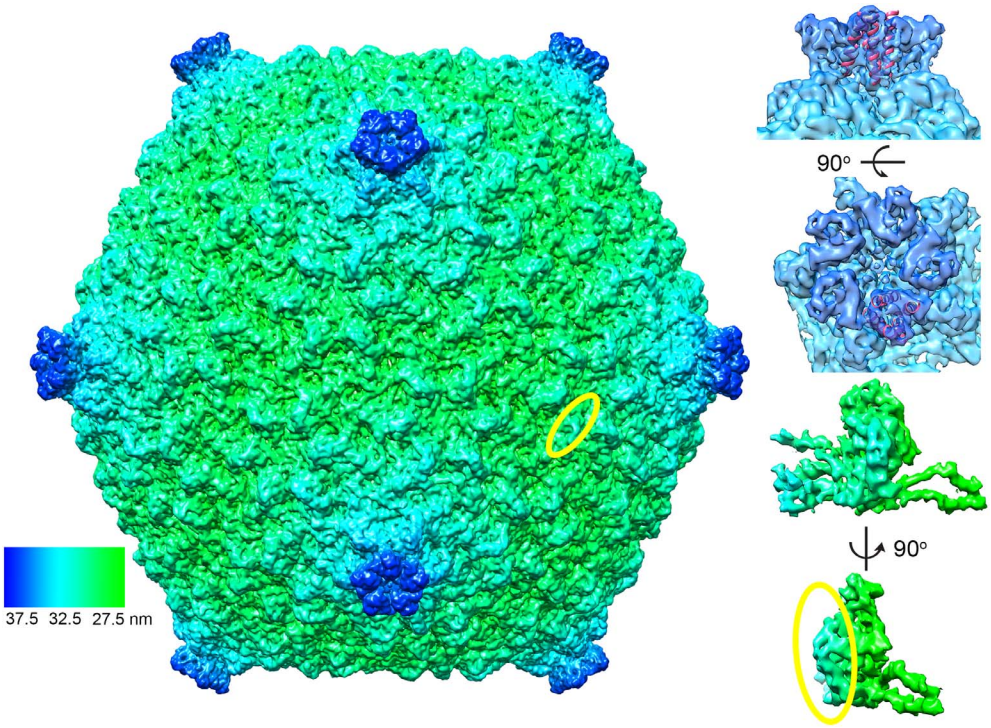


Fig. 3. cryo-EM structure of the ΦM5 capsid. The $T = 7$ capsid is about 75 nm in diameter, colored radially from the center. The capsid is smooth, with turrets at each pentamer that appear to be helical in nature (upper inset). A bridge of density, circled in yellow, joins the A domain to the E-loop (lower inset).

similarity to LUZ24 protein gp56 (Table 1), but the function of this protein is not known. Despite its location in the genome far from the capsid and portal proteins, M5_25 is one possible candidate for the pentamer decoration protein (Fig. 3). It has a short region similar to an immunoglobulin beta-sandwich fold (amino acids 316–391) (Kelley and Sternberg, 2009), a fold that has previously been observed in a smaller pentamer decoration protein (Lander et al., 2012). This fold partially overlaps with the region of homology M5_25 shares with LUZ24 protein gp56 (amino acids 212–501). However, M5_25 lacks the alpha-helical domains predicted in the pentamer decoration protein, and its position in the phage particle cannot be assigned with confidence. ΦM5 protein M5_16 is similar to LUZ24 gp60 and was matched by HH-PRED (Soding et al., 2005) to model 3j4b_A (Cuervo et al., 2013), T7 tail tubular protein A. This is also called the adapter or gatekeeper protein, which connects the portal protein to the "nozzle" or "hub" protein at the tip of a Podovirus tail (Fig. 2C) (Hardies et al., 2016). The identity of this nozzle protein in ΦM5 is not certain (see below). The side tail fibers of Podoviruses also attach to tail tubular protein A (Hardies et al., 2016). A candidate for these side fibers in ΦM5 is M5_27 (Tables 1 and 2, Figs. 2C and 5 A). M5_27 is not conserved in LUZ24 itself, but does share a short region of N-terminal homology with the related phage *Pseudomonas* phage tf (Fig. 4A) (Glukhov et al., 2012). At the C-terminal end of M5_27 is a region with similarity to *S. meliloti* phage ΦM9_134, a predicted tail fiber protein (Johnson et al., 2015), and to proteins from several *Brucella* Podoviruses (Hammerl et al., 2016; Tevdoradze et al., 2015). Like ΦM5, *S. meliloti* phage ΦM9 requires an intact LPS core on the outer membrane for host infection (Campbell et al., 2002; Campbell et al., 2003). The LPS of *S. meliloti* has also been demonstrated to be very similar to that of the closely-related bacterium *Brucella abortus* (Ferguson et al., 2004). This suggests a possible role for M5_27 in attachment to host cell LPS. In the middle of this conserved region in M5_27 is a peptidase S74/chaperone of endosialidase domain (amino acids 391–441), which was identified in *E. coli* phage K1F as a self-cleaving component of the tailspike (Stummeyer et al., 2006). HHPRED and PHYRE (Kelley and Sternberg, 2009; Soding et al., 2005) predictions also show similarity between the C-terminal 1/5th of M5_27 and the neck appendage protein of *Bacillus* phage ga-1 (3gud_A) (Schulz et al., 2010), the L-shaped tail fiber protein of phage T5 (4uw8_A) (Garcia-Doval et al., 2015), and the endosialidase of K1F (3gw6_A) (Schulz et al., 2010). However, M5_27 does not have the extensive beta-helix domain of most L-shaped tailspike proteins (Parent et al., 2014). Taken together, these predictions would be consistent with M5_27 serving as a side-tail fiber in which the C-terminus makes contact with the host cell and the N-terminus interacts with a protein in the proximal region of the tail.

The remaining 9 proteins in the ΦM5 proteome are not similar to LUZ24 proteins, but 3 of these (M5_17, M5_19 and M5_14) have regions of similarity to rhizophages (Table 2 and Fig. 4B–C). M5_17 has overall similarity to *Sinorhizobium* phage PBC5 protein 10 and is annotated as a minor tail protein in some Mycobacteriophages (Pope et al., 2015). M5_19 (Fig. 4B) does not have good matches to any conserved structural domains of phages, but it does have regions of similarity to ORFs in 4 different rhizophages (Fig. 4B and Supplemental Fig. 2). As with the protein M5_27, M5_19 has its best rhizophage match to a protein from the LPS-dependent *Sinorhizobium meliloti* phage ΦM9. This is consistent with M5_19 being an external virion protein that makes contact with the host cell surface. Based on its position in the genome relative to tail tubular protein A (M5_16), it is possible that M5_19 is the 'nozzle' or tail tubular protein B that caps the tip of the tail (Fig. 2C). M5_14 is the last of the proteins detected in the ΦM5 proteome that has significant homology with rhizophage proteins. By far the best match to a rhizophage is to the Podovirus *Mesorhizobium loti* phage vB_Mlo-P_Lo5R7ANS (Halmillawewa et al., 2014a) (Fig. 4C). The C-terminal end of M5_14 contains a GDSL/SGNH hydrolase domain (Kelley and Sternberg, 2009; Soding et al., 2005), which is highly-conserved (Akoh et al., 2004) in many bacterial and eukaryotic proteins, but is not often found in phage (Altschul et al., 1997). The function of M5_14 is unknown, but PHYRE predictions suggest that it has extensive alpha-helical regions (Kelley and Sternberg, 2009), and HHPRED (Soding et al., 2005) detects structural similarity to 4QNL, a tail fiber from *E. coli* phage G7C (Riccio et al., 2015). However, the tail of the N4-like phage G7C (Kulikov et al., 2012) has a double-ringed structure visible by TEM that does not resemble the tail of ΦM5. Alternatively, the alpha-helical character of M5_14 and its ORF position close to the ORF encoding the major capsid protein introduce the possibility that it could encode the pentamer decoration protein. Although the position of M5_14 in the

Table 1
ΦM5 predicted proteins detected in the sample proteome by ≥2 exclusive unique peptides. All phage proteins detected at a protein threshold of 1% FDR and a peptide threshold of 0.1% FDR are shown.

ΦM5 ORF name	Predicted function	Predicted MW	Length in amino acids	Exclusive unique peptides	Total spectrum counts assigned to protein	Percent protein coverage	Predicted peptides from trypsin digestion	Protein abundance index (exclusive unique peptides/ predicted peptides)	Protein content (weight% of identified phage proteins) (emPAIx MW/Σ(emPAI × MW) × 100	LUZ24 homolog % identity over the region of coverage	Homology to ORF in rhizophages (see Table 2)
phiM5_11	Major head subunit	35 kDa	323	26	5862	80.5%	80	**0.313**	13.7%	gp62 37%	–
phiM5_29	Hypothetical protein	39 kDa	363	20	1438	65.3%	82	**0.244**	10.8%	N/A	–
phiM5_28	Lysozyme/tail tape measure	51 kDa	514	22	727	58.8%	80	**0.238**	13.9%	N/A	–
phiM5_19	Predicted host-binding tail fiber protein	40 kDa	391	11	762	44.0%	51	**0.196**	8.4%	N/A	+
phiM5_04	Portal protein	80 kDa	716	36	1569	51.1%	167	**0.180**	15.4%	gp65 32%	–
phiM5_27	chaperone of endosialidase (possible tailspike protein)	49 kDa	468	15	659	45.5%	84	**0.167**	8.5%	phage tf gp54 43%	+
phiM5_31	Hypothetical protein	54 kDa	514	27	751	53.7%	144	**0.153**	8.4%	N/A	–
phiM5_17	Conserved in mycobacteria phages	13 kDa	126	2	21	15.9%	15	**0.133**	1.7%	N/A	+
phiM5_13	Hypothetical protein	9 kDa	91	2	152	27.5%	15	**0.133**	1.2%	N/A	–
phiM5_14	GDSL/SGNH hydrolase family protein	45 kDa	436	10	296	24.1%	68	**0.132**	6.0%	N/A	+
phiM5_25	Hypothetical protein	55 kDa	502	16	265	31.5%	110	**0.127**	7.0%	gp56 21%	–
phiM5_65	Hypothetical protein	8 kDa	65	2	44	30.8%	21	**0.095**	0.7	N/A	–
phiM5_16	Tail tubular A protein	24 kDa	205	6	117	30.7%	55	**0.073**	1.6%	gp60 24%	–
phiM5_09	Putative scaffolding protein	39 kDa	349	11	64	25.5%	91	**0.044**	1.5%	gp63 27%	–
phiM5_38	DNA polymerase A	77 kDa	688	5	34	6.7%	221	**0.018**	1.3%	N/A	–

Table 2
ΦM5 predicted proteins with homologs in other phages of rhizobia. See Supplemental Table 2C for sequence accessions.

	Host bacterium	*Sinorhizobium meliloti* 1021	*Sinorhizobium meliloti* 1021	*Sinorhizobium*	*Rhizobium leguminosarum* F1	*Rhizobium gallicum*	*Mesorhizobium loti*	*Rhizobium etli*
ΦM5 ORF	**predicted function or domains**	**ΦM9** (Myovirus)	**ΦM12** (Myovirus)	**PBC5** (Caudo-virus)	**vB_RleM_ PPF1** (Myovirus)	**vB_RglS_ P106B** (Siphovirus)	**vB_MloP_Lo5R7ANS** (Podovirus)	**RHEph_02** (Podovirus)
M5_02	Terminase large subunit	–	–	–	–	**24% ID/33% coverage** (P106B_01)	**30% ID/10% coverage** (Lo5R7ANS_63)	–
M5_14#	GDSL/SGNH hydrolase family protein	–	**28% ID/31% coverage** (M12_398)	–	–	**38% ID/17% coverage** (P106B_14)	**26% ID/69% coverage** (Lo5R7ANS_62)	–
M5_17#	Putative minor tail protein conserved in mycobacteriophages	–	–	**29% ID/98% coverage** (PBC5p10)	–	–	–	–
M5_19#	Predicted host-binding protein tail protein	**30% ID/66% coverage** (M9_136)	**38% ID/41% coverage** (M12_124)	**33% ID/22% coverage** (PBC5p14)	**63% ID/9% coverage** (PPF1_26)	**24% ID/18% coverage** (P106B_10)	–	**56% ID/6% coverage** (RHEph02_050)
M5_21	Predicted tail fiber assembly protein	**39% ID/82% coverage** (M9_137) **41% ID/67% coverage** (M9_138)	**26% ID/38% coverage** (M12_122)	**42% ID/81% coverage** (PBC5p15)	–	**25% ID/37% coverage** (P106B_39)	**42% ID/45% coverage** (Lo5R7ANS_58)	**36% ID/38% coverage** (RHEph02_051)
M5_22	Putative endolysin	–	–	–	–	–	–	**30% ID/21% coverage** (RHEph02_048)
M5_26	Hypothetical protein (possible acetyltransferase)	–	–	–	**38% ID/31% coverage** (PPF_15)	–	–	**23% ID/36% coverage** (RHEph02_058)
M5_27#	Chaperone of endosialidase; pfam13884; peptidase S74; putative tailspike	**47% ID/13% coverage** (M9_134)	–	–	–	–	–	–
M5_33	hypothetical protein (possible transcriptional regulator)	**33% ID/61% coverage** (M9_187)	–	–	–	–	–	–
M5_34	integrase	–	–	–	–	–	**26% ID/85% coverage** (Lo5R7ANS_12)	–
M5_35	DNA primase/polymerase	–	–	–	–	**24% ID/54% coverage** (P106B_71)	–	–
M5_36	Hypothetical protein	–	**29% ID/79% coverage** (M12_239)	–	–	–	–	–
M5_37	5′ nucleotidase, deoxy (Pyrimidine), cytosolic type C protein (NT5C)	**29% ID/44% coverage** (M9_176)	**28% ID/43% coverage** (M12_197)	–	–	–	–	**35% ID/30% coverage** (RHEph02_024)
M5_41	Hypothetical protein	–	**41% ID/11% coverage** (M12_470)	–	–	–	–	–
M5_42	HNH endonuclease	–	**36% ID/36% coverage** (M12_412)	–	–	–	–	–
M5_45	Predicted helicase	–	–	**31% ID/70% coverage** (PBC5p23)	–	–	–	–
M5_55	Hypothetical protein	**50% ID/40% coverage** (M9_205)	–	–	–	–	–	–
M5_62	Predicted Holliday junction resolvase	–	**28% ID/27% coverage**	–	–	–	–	–

(continued on next page)

Table 2 *(continued)*

	Host bacterium	*Sinorhizobium meliloti* 1021	*Sinorhizobium meliloti* 1021	*Sinorhizobium*	*Rhizobium leguminosarum* F1	*Rhizobium gallicum*	*Mesorhizobium loti*	*Rhizobium etli*
M5_63	Hypothetical protein	–	(M12_046) –	–	–	**24% ID/61% coverage** (P106B_22)	–	–
M5_64	Putative ATP-dependent Clp protease	–	–	–	–	**42% ID/65% coverage** (P106B_25)	–	–

#Denotes proteins which were detected in the ΦM5 proteome.

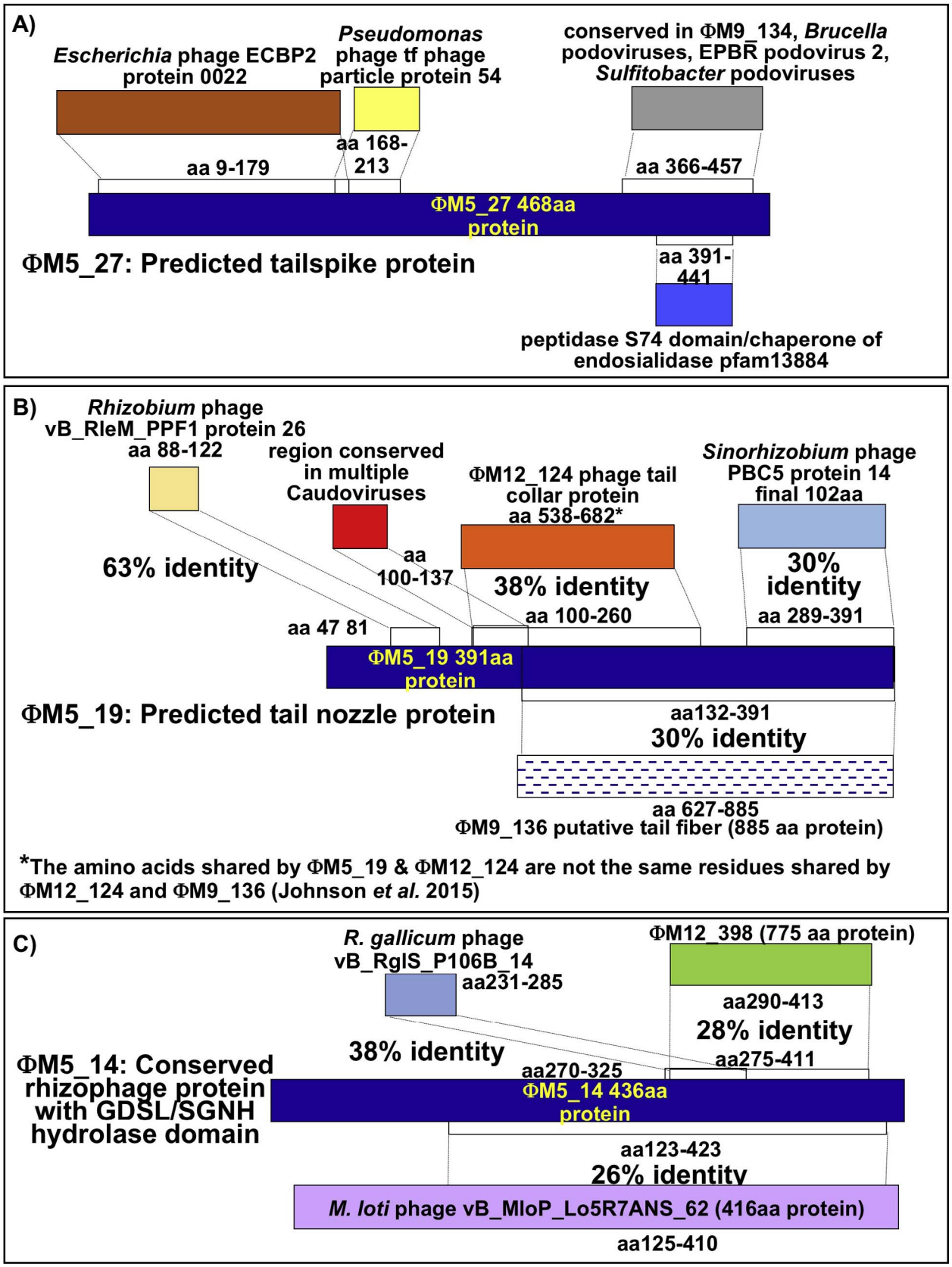
overall structure cannot be predicted, it, along with the other constituents of the ΦM5 proteome with rhizophage-like domains (M5_27, M5_17 and M5_19) are candidates for external phage particle proteins that make contact with the host *S. meliloti*.

Three of the remaining abundant proteins in the ΦM5 proteome are encoded by ORFs (M5_28, M5_29, and M5_31) adjacent to the predicted tailspike protein M5_27. These ORFs are similar to neither LUZ24 phages, nor to rhizophages, and do not have obvious homologs in any well-characterized phages (Altschul et al., 1997). M5_28 is a 514 amino acid protein that has a 4XP8 lysozyme domain at its N-terminus (Moak and Molineux, 2004), but no other conserved domains. M5_29 is similar to 5DZZ, an eukaryotic intermediate filament binding domain of desmoplakin (Biasini et al., 2014; Kang et al., 2016). M5_31 is also 514 amino acids and has a 2P4V GreB transcript cleavage domain at its C-terminus (Kelley and Sternberg, 2009; Soding et al., 2005; Vassylyeva et al., 2007). Based on their abundance in the proteome, the location of the ORFs next to the endosialidase-containing tailspike gene (Stummeyer et al., 2006), the presence of a lysozyme domain in one of the proteins (Moak and Molineux, 2004) and the somewhat large predicted size of the proteins (Hardies et al., 2016; Lavigne et al., 2006) it is possible that they are internal virion tail proteins that are injected into the host at the time of infection to form a transient tail tube (Hardies et al., 2016). The internal virion proteins that the short-tailed Podoviruses inject into host cells have highly diverged primary sequence and are difficult to recognize in phage genomes (Hardies et al., 2016). Some of these podoviral injected proteins with known functions include lysozyme, channel-forming proteins that allow the viral DNA to enter the host cell and effector proteins that can manipulate the host physiology (Hardies et al., 2016). Fig. 2B shows an electron micrograph of ΦM5 that has ejected its internal tail proteins in the absence of a host cell, and Fig. 2C show schematically how internal tail proteins are positioned in intact particles of some Podoviruses (Hu et al., 2013; Liu et al., 2010; Wu et al., 2016; Zhao et al., 2016).

Unusually, the ΦM5 phage structural gene region in the genome is interrupted by a set of ORFs that are not found in the phage particle proteome. These are M5_22, M5_23 and M5_24, which may be a host-cell lysis cassette, encoding respectively, an endolysin, a holin and a spanin. A phage holin is a membrane protein that begins the host lysis process by permeabilizing the inner membrane of a Gram negative host bacterium (Young, 2014). M5_23 is predicted by HH-PRED (Soding et al., 2005) to encode a colicin E1 (2i88_A), which can create holes in cell membranes (Elkins et al., 1997). M5_22 is a predicted N-acetylmuramidase (Altschul et al., 1997), which can function as a phage endolysin when a holin has provided access to the peptidoglycan cell wall within the Gram negative bacterial periplasm (Young, 2014). The identity of M5_24 is less obvious, but given its position next to an endolysin and holin, and the presence of a predicted transmembrane helix, it is a possible candidate for a spanin (Young, 2014). Spanins are responsible for membrane-fusion events that open the Gram-negative outer membrane, permitting release of progeny phage (Young, 2014). The C-terminal location of the transmembrane domain of M5_24 would be consistent with a u-type spanin, however some other features of the ORF are atypical of u-spanins (Young, 2013, 2014) making the function of this ORF uncertain.

3.4. The ΦM5 genome does not have overall synteny with characterized phages

Although many of the structural genes of ΦM5 are similar to those of the LUZ24 phages, other modules of the genome have no similarity to this phage genus, and the genome is not syntenic with any previously characterized phages. The ΦM5 genome sequence is 44,005 bp, of which 357 bp on either end are direct terminal repeats (DTRs) (Fig. 5A). In the initial Illumina read assemblies of two separate, plaque-purified isolates of ΦM5, the genome appeared to be circularly permuted with random breakpoints (data not shown). However, a



(caption on next page)

Fig. 4. ΦM5 structural proteins with extensive regions of homology with other rhizophages. A) M5_27, the predicted tailspike protein, has a C-terminal region of homology with the ΦM9_134 predicted tail fiber protein and with proteins from other phages of alphaproteobacteria. Within this region, it has a peptidase S74/chaperone of endosialidase domain similar to cleavage domains found in phage tailspike proteins. It also shares homology near the N-terminus with *Pseudomonas* phage tf, a LUZ24-like phage. This is consistent with a protein in which the C-terminus contacts a rhizobial host while the N-terminus interacts with other LUZ24-like proteins in the phage. B) M5_19 shares extensive homology with predicted tail fiber proteins of *S. meliloti* phages ΦM9 and ΦM12, and more limited regions of homology with proteins from additional rhizophages. (see Supplemental Fig. 2 for full alignment.) The most note-worthy matches to rhizophages in M5_19 are 260 amino acids that are similar to ΦM9 predicted tail fiber 136; 159 amino acids that are similar to the phage tail collar protein of ΦM12, and 103 amino acids that are similar to *Sinorhizobium* phage PBC5 protein 14. The similarity to tail fiber proteins of other rhizophages and the position of the M5_19 ORF relative tail tubular protein A (M5_16) in the ΦM5 genome, would be consistent with M5_19 serving as the 'nozzle' or tail tubular protein B that caps the tip of the tail. C) M5_14 shares extensive homology with protein 62 from the *M. loti* Podovirus vB_MloP_Lo5R7ANS. It also has regions with similarity to ΦM12_398 and a protein from the *R. gallicum* Siphovirus vB_RglS_P106B. M5_14 has a GDSL/SGNH hydrolase domain, but these proteins can have many different functions and the role of M5_14 in the phage particle is unclear.

region of higher Illumina read coverage was observed, which can be indicative of DTRs (Merrill et al., 2016). The true genome ends, including the 357 bp DTRs were mapped by restriction enzyme digestion (Supplemental Fig. 3) and by direct Sanger sequencing of the DNA from purified phage particles (Fig. 5B and Supplemental Fig. 3).

The genome contains 58 unique ORFs, with a copy of ORF 1 in each of the DTRs (ORFs 1.1 and 1.2) (Fig. 5A). Twenty-five of these ORFs are on the plus strand on the left arm of the genome (bases 1–21,823). Thirty-four are on the right arm (bases 21,854–44,005), with 31 of these on the minus strand and three on the plus strand (Fig. 5A). The only tRNA gene detected in the ΦM5 genome is a tRNA-Met from bases 7352 to 7426 on the plus strand.

ΦM5 does not have an ORF for its own RNA polymerase, so it is expected to be dependent upon the host transcription machinery. Searches for promoters predicted to be recognized by *S. meliloti* 1021 sigma factors were performed using the promoter sequence motif data from Schluter et al., 2013 and the PhiSite promoter hunter (http://www.phisite.org/main/index.php?nav=tools&nav_sel=hunter) (Klucar et al., 2010; Schlüter et al., 2013; Stano and Klucar, 2011). There are high scoring promoters spaced within 100 bp of start codons of only two ΦM5 ORFs. One of these is the plus-strand ORF M5_48 encoding a predicted HTH XRE family DNA binding protein. It has 2 overlapping sigma 70 and sigma H1 promoters spaced 21–45 bases upstream of its start codon. This protein is predicted by HH-PRED to share structural homology with the phage P22 c2 repressor protein, which is required for lysogeny of phage P22 (Watkins et al., 2008). The

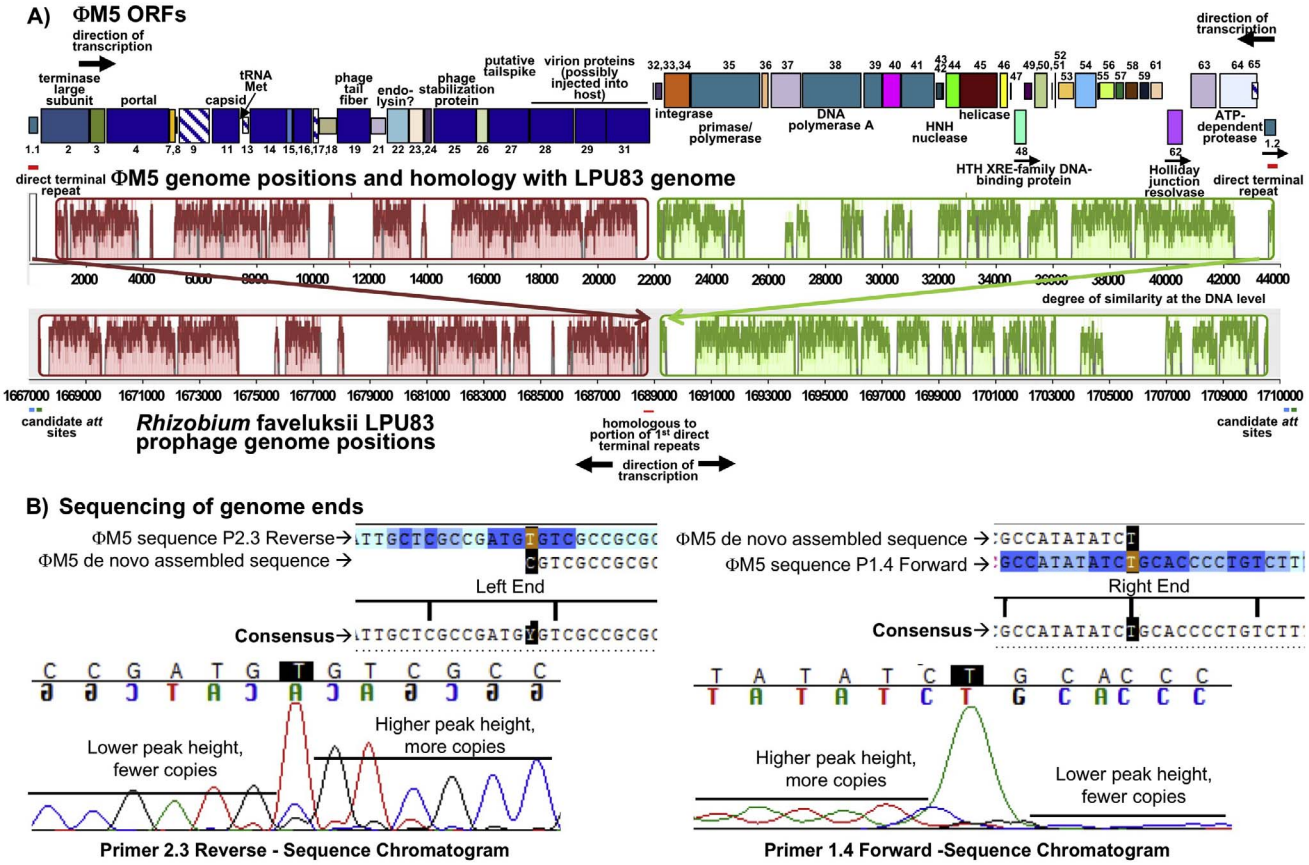


Fig. 5. Map of the features of ΦM5 genome. A) Top: The 44,005 bp genome of ΦM5, showing direction of transcription and ORF numbers with their genome position. The directions of transcription are shown in black arrows. ORF products detected in the ΦM5 proteome and most likely to be structural are shown in solid dark blue. ORF products that may be structural are shown in patterned dark blue. The boxes representing ORF products that share no regions of similarity between ΦM5 and LPU83 are drawn at half-height. Direct terminal repeats from ΦM5 genome positions 1–357 and 43,649–44,005 are shown with the repeated ORF 1.1 and 1.2. Bottom: Bases 1,667,000–1,710,000 of the chromosome of *Rhizobium favelukesii* LPU83, showing the two inverted blocks homologous to ΦM5, aligned in MAUVE. The left half of ΦM5, containing the terminase and the structural genes is 36.2% identical to LPU83 at the nucleotide level, while the right half, containing replicative functions and integrase is 33.4% identical. Candidate *att* sites for integration of the prophage into the LPU83 genome are CTGCTGGCGGAG at positions 1,667,176–1,667,188 and 1,710,460–1,710,472, and GCTACAAGCAGTTGAT at positions 1,667,238–1,667,253 and 1,710,952–1,710,967. There is partial homology of ΦM5 bases 318–357 of the first DTR to 1,688,823–1688783 of the LPU83 genome. B) The ends of the ΦM5 genome were determined by Sanger sequencing with multiple primers. The sequence chromatogram from one primer at each end of the genome is shown. The height of the chromatogram peaks is greater for reads from DNA that was packaged in the purified phage particles, and lower for reads from the residual prepackaging concatamer intermediate. The spiked T peaks at the ends of the phage DNA are from the sequencing of the terminal A base added to the 3′ end by Taq polymerase in the sequencing reaction.

other ORF with a closely-spaced, strong promoter is the minus-strand M5_38 ORF encoding a predicted DNA polymerase A, which has a predicted sigma 70 promoter 46–70 bases from the start codon. Other predicted strong promoters are located further from their potential target ORFs. A strong minus-strand sigma 70 promoter is predicted 256–331 bases from the start of ORF M5_61, which is predicted to encode a DUF3846 protein. Upstream of a predicted plus-strand Holliday-junction resolvase (RuvC, ORF M5_62) there are two predicted plus-strand promoters: a sigma 70 promoter at 157–183 bases upstream and a sigma E2 promoter at 128–149 bases upstream. In at least one phage, a RuvC resolvase is required for replication restart during theta-type DNA replication (Zecchi et al., 2012). Other predicted promoters in the ΦM5 genome have much lower PhiSite promoter hunter scores. The ORFs with predicted strong *S. meliloti* promoters are candidates for early genes involved in a lysis/lysogeny decision and in phage replication.

3.5. ΦM5 has strong overall homology to a prophage within the R. favelukesii LPU83 genome

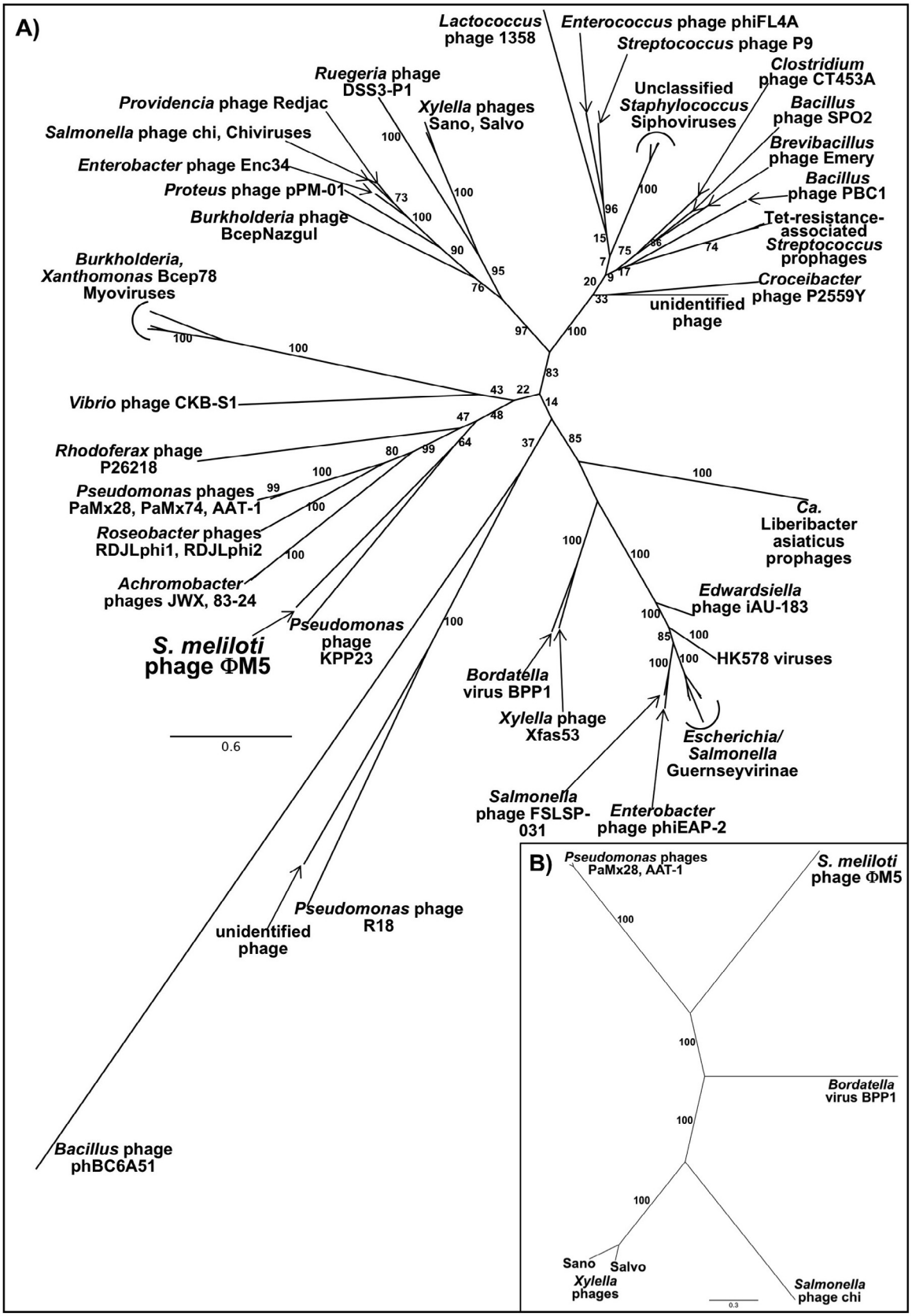
While ΦM5 is difficult to place phylogenetically among characterized free-living phages, it does have good homology over most of its genome to a prophage found in the chromosome of *Rhizobium favelukesii* LPU83 (Wibberg et al., 2014) at positions 1,667,000–1,710,000 bp (Altschul et al., 1990). The synteny of the ΦM5 genome with the *R. favelukesii* LPU83 prophage (Fig. 5A) suggests that a phage very similar to ΦM5 integrated into the LPU83 genome at some point in its history. Segments of the ΦM5 genome are also similar to other bacterial prophages and environmental metagenomic sequences (data not shown). The presence of a very similar prophage in LPU83, and ΦM5′s possession of an integrase gene and an ORF similar to the P22 c2 repressor suggested the possibility that it might form lysogenic infections. ΦM5 plaques are clear, rather than turbid, which is characteristic of lytic phages (Echols, 1972), but occasionally, *S. meliloti* colonies arise within plaques. To determine whether these colonies are lysogens or phage-resistant mutants, they were streaked multiple times to remove contaminating phage and then inoculated onto lawns of wild type *S. meliloti* 1021. The formation of plaques when a lysogenic strain is exposed to the naïve parent strain is indicative of spontaneous activation of prophages and infection of the indicator strain (Miller et al., 1998). No plaques arose on these lawns (data not shown). The lack of plaque formation suggests that either ΦM5 is not able to form lysogens on *S. meliloti* 1021, or that if lysogens are formed, they are quite stable.

Phage integrases are enzymes that catalyze the recombination of temperate phages into the host chromosome (Groth and Calos, 2004). The predicted integrase of ΦM5 (ORF M5_34) is a member of the XerC tyrosine recombinase INT_ICEBs1_C_like family, and, among characterized phages, is most similar to that of the Siphovirus *Mycobacterium* phage Giles (30% identity) (Morris et al., 2008). Giles is a highly mosaic phage and the close homologs of its integrase, like those of ΦM5, are more commonly observed in prophages than in characterized lytic phages (Morris et al., 2008). The ΦM5 integrase is shorter than that of Giles, lacking a 66N-terminal amino acid segment. ΦM5 does not have an ORF similar to the Giles excisionase/recombination directionality factor (RDF). RDFs are small, usually positively-charged proteins that are required for prophage excision in most phages (Lewis and Hatfull, 2001). There is no ORF with clear homology to known RDFs in the ΦM5 genome, but these proteins are notoriously difficult to identify (Lewis and Hatfull, 2001). Though its closest homolog is a prophage within a rhizobial genome, ΦM5 has not been observed to form lysogenic infections on *S. meliloti* 1021, and it is unclear if it has the capacity to do so.

3.6. The ΦM5 terminase large subunit is highly diverged from characterized terminases and may be a novel type

Phage terminases function in one of the last steps in a lytic phage infection: the cleavage of phage genomic DNA and the concomitant packaging of the DNA into phage capsids (Merrill et al., 2016). The terminase large subunit is the ATP-driven packaging motor, and the small subunit is the DNA recognition protein (Rao and Feiss, 2008). Although the ΦM5 terminase large subunit (ORF M5_02) is located close to the LUZ24-like structural genes in the genome, it appears to be from a different lineage. The packaging strategy of a phage (e.g. headful packaging, cohesive ends, short DTRs, etc.) can often be predicted based on the amino acid sequence of the terminase large subunit (Merrill et al., 2016; Rao and Feiss, 2008). However, the ΦM5 terminase large subunit is quite dissimilar from characterized phage terminases and this approach would not have predicted the presence of short DTRs at the genome ends in ΦM5. Among phages with a terminase large subunit similar to ΦM5 for which a packaging strategy has been determined, the headful-packaging Podovirus epsilon 15 is most similar to ΦM5 (21% identity) (Kropinski et al., 2007; Mcconnell et al., 1992). Some of the other characterized phages possessing a terminase with similarity to ΦM5 are *Clostridium* phage ΦCD27 (29% identity) (Mayer et al., 2008), *Pseudomonas* phage AF (25% identity) (Cornelissen et al., 2012), *Pseudomonas* phage vB_PaeP_Tr60_Ab31 (22% identity) (Latino et al., 2014), and *Xanthomonas citri* phage CP2 (22% identity) (Ahmad et al., 2014). Defined genome ends could not be detected for *Pseudomonas* phages AF and Ab31 (Latino et al., 2014) or *Clostridium* phage ΦCD27 or its close relatives (Mayer et al., 2008; Rashid et al., 2016) suggesting these phages are also headful packaging phages. (A phylogenetic tree showing the relationships between the terminase large subunit ORFs of these phages is shown in Supplemental Fig. 4, sequence information in Supplemental Table 2D). Better matches to the ΦM5 terminase large subunit are found in prophages within bacterial genomes and in environmental metagenomic sequences than among characterized phages (Supplemental Fig. 4). A phage terminase from the G18 group of deep water Mediterranean phages (41% identity) (Mizuno et al., 2013) and the *R. favelukesii* LPU83 terminase (68% identity) (Wibberg et al., 2014) have the greatest similarity to the ΦM5 terminase large subunit (Fig. 5A, Supplemental Fig. 4, Supplemental Table 2D). The *Candidatus* Liberibacter asiaticus prophages SC1 (24% identity) and SC2 (25% identity) (Zhang et al., 2011) also have a terminase large subunit ORF similar to that of ΦM5. These prophages are found within the genome of *Ca.* L. asiaticus, the Huanglongbing/citrus-greening disease bacterium that is related to the rhizobia (Zhang et al., 2011). SC1 has been detected as an excised, linear phage genome and has been observed by electron microscopy in infected periwinkle plants (Zhang et al., 2011). Deepening the mystery surrounding the ΦM5 terminase, the SC1 and SC2 prophages have cos sites and the cos sites are found at the ends of the excised linear form of SC1 (Zhang et al., 2011). Thus, terminases related to the ΦM5 terminase package DNA by a cohesive-end mechanism or a headful mechanism, but no others are known to form short DTRs.

A terminase small subunit ORF is usually located adjacent to the large subunit ORF, but the adjacent ORFs in ΦM5 have no similarity to terminase small subunit genes. Despite multiple attempts to identify a terminase small subunit by comparing ΦM5 ORFs with the predicted small subunit from phages and prophages with a similar large subunit, no ORF with similarity to a known terminase small subunit could be identified. Directly upstream of the ΦM5 terminase large subunit, in the DTR, is an ORF (M5_01.1) with structural similarity to the DNA-binding 1baz_A ARC repressor domain of P22 (Schildbach et al., 1999; Soding et al., 2005). One possibility is that this predicted DNA-binding protein functions in DNA recognition in packaging in ΦM5, but there is no evidence for this aside from proximity to the large subunit ORF. The fact that ΦM5 has direct terminal repeats, but its terminase large subunit is quite dissimilar from those of other phages known to employ this



(caption on next page)

Fig. 6. Phylogenetic trees based on DNA polymerase A and nearby conserved proteins. A) A conserved internal segment of ΦM5 DNA polymerase A (M5_38, amino acids 57–686) was aligned with orthologous sequences from other phages. Among characterized phages, the ΦM5 DNA polymerase A is most similar to that of the siphovirus *Pseudomonas* phage KPP23. The best tree places ΦM5 DNA polymerase A in a monophyletic group with DNA polymerase A proteins from very diverse phages (Siphoviruses, Podoviruses and Myoviruses). (See Supplemental Table 2E for protein sequence information). B) A tree was constructed in which the ΦM5 DNA polymerase A ORF and 3 nearby ORFs (DNA primase/polymerase, M5_35; DUF2815 protein, M5_39; and Cas4-like exonuclease, M5_41) were concatenated and aligned with orthologs from other phages that possess the same DNA polymerase A. Only 7 phages, ΦM5, *Pseudomonas* phages PaMx28 and AAT-1, *Bordetella* virus BPP1, *Salmonella* phage chi, and *Xylella* phages Sano and Salvo have orthologs for all 4 of these ORFs. The bootstrap percentage shown for each branch reflects the degree of confidence in the placement of that node in the phylogenetic reconstruction. The bar indicates branch distance. (See Supplemental Table 2F for protein sequence information.)

packaging strategy, suggests the possibility that the ΦM5 terminase is a novel type.

3.7. The DNA polymerase ORF of ΦM5 appears to derive from a different lineage than the virion structural ORFs, the integrase, or the terminase

The right arm of ΦM5 contains 34 ORFs, with the majority of these of unknown function. A few of these ORFs have clear similarity to replication proteins and other DNA-binding proteins (Altschul et al., 1997; Soding et al., 2005). DNA polymerase A, encoded by ORF M5_38, is located adjacent to a predicted, strong *S. meliloti* sigma 70 promoter (discussed above). This polymerase appears to be from a completely different phage lineage than the LUZ24-like structural ORFs, the Giles-like integrase, or the novel terminase. The phylogenetic tree shown in Fig. 6A shows that the ΦM5 DNA polymerase A is related to those of the Siphoviruses *Pseudomonas* phage KPP23 (33% identity) (Yamaguchi et al., 2014) and the *Pseudomonas* phage PaMx74/PaMx28 group (~35% identity) (Altschul et al., 1997; Sepulveda-Robles et al., 2012). The tree shows that this type of DNA polymerase A is found mostly in Siphoviruses, but also in a few lineages of Podoviruses and Myoviruses. A striking common link between the DNA polymerase A lineage (Fig. 6A) and the terminase large subunit lineage of ΦM5 (Supplemental Fig. 4) is made by the SC1 prophage of *Ca.* L. asiaticus (Zhang et al., 2011). The SC1 DNA polymerase A is 29% identical to that of ΦM5. The terminase large subunit and the DNA polymerase A ORFs are at nearly opposite ends of the genome, and it is difficult to speculate about the genomic history that might have derived these modules from a common lineage.

To understand how large a genome segment surrounding the DNA polymerase A ORF might come from a common lineage, ORFs across the right arm of the ΦM5 genome were compared to the ORFs from the phages shown in the tree in Fig. 6A (Altschul et al., 1997). Of the phages that have a DNA polymerase A similar to ΦM5, a small number also share the marker ORFs M5_35, a predicted primase/polymerase; M5_39, a predicted DUF2815 domain protein; and M5_41, a predicted Cas4-like exonuclease. The region of the ΦM5 genome containing these 4 ORFs extends for 8.5 kb across the right arm of the genome (positions 23,196–31,644). There does not appear to be a clear pattern of conservation of these ORFs among phages, but the DNA polymerase A and the Cas4-like exonuclease are most commonly found together, including in the *Ca.* L. asiaticus SC1 prophage (data not shown). Two of these ORFs, the primase/polymerase and the Cas4-like exonuclease, also have short regions of similarity to ORFs in rhizophages (Table 2). Only 7 phages have all 4 of these ORFs and a phylogenetic tree showing the relationships among them is shown in Fig. 6B. The similarity between the Siphoviruses *Salmonella* phage chi and *Xylella* phages Sano and Salvo has been noted previously (Ahern et al., 2014; Hendrix et al., 2015), but these ORFs are also shared with the Siphoviruses PaMx28 and AAT-1 and the Podovirus *Bordatella* phage BPP-1 (Liu et al., 2004). An additional ORF, M5_45, encoding a predicted helicase, is also shared by ΦM5, PaMx28 and AAT-1, extending this conserved region an additional 2.3 kb (to position 33,939). It appears that this genome segment has traveled as a module, but there is not strong pressure for co-conservation of individual ORFs.

The segment of the right arm of the ΦM5 genome between the predicted helicase (M5_45) and the Holliday junction resolvase (M5_62) contains 15 ORFs, 10 of which are not conserved in the closely-related prophage of *R. favelukesii* LPU83 or any of the phages with which ΦM5 shares other genome modules. Most of these ORFs appear to be either of bacterial origin or to have very poor matches to anything currently in the database (Altschul et al., 1997). This segment does not appear to form a coherent functional module like the structural genes or to come from an identifiable lineage like the DNA polymerase and its associated ORFs. It is possible that this is a genomic region that is permissive for the acquisition of new sequences. Such permissiveness would be consistent with the highly mosaic character of the ΦM5 genome.

4. Conclusions

Although the primary amino acid sequence of the ΦM5 major capsid protein is highly diverged from those of the distantly-related Podoviruses T7, P22, and epsilon 15, it has the $T = 7$ capsid geometry that is common to all of those phages. The most interesting feature of the ΦM5 capsid is the capsid protein's distinctive A-domain bridge that may stabilize the mature capsid by a novel interaction between its beta hinge and the E loop hairpin. The capsid-associated structural ORFs of ΦM5 appear to be from the same lineage as those of the LUZ24-like phages of *Pseudomonas*. Since this is the first-reported structure of a LUZ24-like phage, it is unknown if this novel stabilization interaction is common to other phages in this lineage. The other ΦM5 ORFs found in the proteome have extensive similarity to ORFs of previously-characterized rhizophages and are candidates for host-recognition proteins of the tail.

The genome of ΦM5 is highly mosaic with the structural genes, the integrase, the DNA polymerase and the terminase each deriving from a separate lineage. The integrase is similar to that of *Mycobacterium* phage Giles, which also has a highly mosaic genome. The DNA polymerase and an associated genome module appear to derive from the same lineage as that of the siphoviral phage PaMx74. The ΦM5 terminase is unlike previously-described terminases. The phages with the closest homologs of the terminase large subunit have < 30% identity with the terminase of ΦM5 and have headful packaging or cohesive ends rather than short DTRs. Thus, the ΦM5 terminase may define a new type of short-DTR-packaging terminase.

Nucleotide sequence and cryo-EM reconstruction accession numbers: The *S. meliloti* phage ΦM5 genome has been deposited in GenBank (http://www.ncbi.nlm.nih.gov/nuccore) with the accession number MF074189.1. The cryo-EM reconstruction has been deposited in the online Electron Microscopy Data Bank (http://www.ebi.ac.uk/pdbe/emdb/) under accession no. EMD-8689.

Acknowledgements

We thank Donald P. Breakwell for his generosity regarding ΦM5. This work was funded by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2014-67013-21579 to K.M.J., USDA NIFA SCRI Citrus Disease Research and Extension (CDRE) award 2016-70016-24844 to Dean Gabriel (University of Florida), subaward UFDSP00011165 to K.M.J. and NSF award MCB-1149763 to M.E.S.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the

online version, at http://dx.doi.org/10.1016/j.jsb.2017.08.005.

References

Adams, M.J., Lefkowitz, E.J., King, A.M.Q., Harrach, B., Harrison, R.L., Knowles, N.J., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Mushegian, A.R., Nibert, M., Sabanadzovic, S., Sanfacon, H., Siddell, S.G., Simmonds, P., Varsani, A., Zerbini, F.M., Gorbalenya, A.E., Davison, A.J., 2016. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016). Arch. Virol. 161, 2921–2949.

Adriaenssens, E.M., Krupovic, M., Knezevic, P., Ackermann, H.W., Barylski, J., Brister, J.R., Clokie, M.R.C., Duffy, S., Dutilh, B.E., Edwards, R.A., Enault, F., Bin Jang, H., Klumpp, J., Kropinski, A.M., Lavigne, R., Poranen, M.M., Prangishvili, D., Rumnieks, J., Sullivan, M.B., Wittmann, J., Oksanen, H.M., Gillis, A., Kuhn, J.H., 2017. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. Arch. Virol. 162, 1153–1157.

Ahern, S.J., Das, M., Bhowmick, T.S., Young, R., Gonzalez, C.F., 2014. Characterization of novel virulent broad-host-range phages of Xylella fastidiosa and Xanthomonas. J. Bacteriol. 196, 459–471.

Ahmad, A.A., Ogawa, M., Kawasaki, T., Fujie, M., Yamada, T., 2014. Characterization of bacteriophages Cp1 and Cp2, the strain-typing agents for Xanthomonas axonopodis pv. citri. Appl. Environ. Microbiol. 80, 77–85.

Akoh, C.C., Lee, G.C., Liaw, Y.C., Huang, T.H., Shaw, J.F., 2004. GDSL family of serine esterases/lipases. Prog. Lipid Res. 43, 534–552.

Aksyuk, A.A., Rossmann, M.G., 2011. Bacteriophage assembly. Viruses-Basel 3, 172–203.

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., Lipmann, D.J., 1990. A basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Ankrah, N.Y.D., Budinoff, C., Wilson, W.H., Wilhelm, S.W., Buchan, A., 2014. Genome sequences of two temperate phages, ΦCB2047-A and ΦCB2047-C, infecting *Sulfitobacter* sp. strain 2047. Genome Announc. 2, 00108–00114.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258.

Botstein, D., 1980. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. 354, 484–491.

Brewer, T.E., Elizabeth Stroupe, M., Jones, K.M., 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.

Brown, J.M., LaBarre, B.A., Hewson, I., 2013. Characterization of Trichodesmium-associated viral communities in the eastern Gulf of Mexico. FEMS Microbiol. Ecol. 84, 603–613.

Campbell, G.R., Reuhs, B.L., Walker, G.C., 2002. Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core. Proc. Natl. Acad. Sci. U.S.A. 99, 3938–3943.

Campbell, G.R., Sharypova, L.A., Scheidle, H., Jones, K.M., Niehaus, K., Becker, A., Walker, G.C., 2003. Striking complexity of lipopolysaccharide defects in a collection of *Sinorhizobium meliloti* mutants. J. Bacteriol. 185, 3853–3862.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R.A., Potter, C.S., Pulokas, J., Reilein, A., 2000. Leginon: an automated system for acquisition of images from vitreous ice specimens. J. Struct. Biol. 132, 33–45.

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

Chalkley, R.J., Baker, P.R., Huang, L., Hansen, K.C., Allen, N.P., Rexach, M., Burlingame, A.L., 2005. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. Mol. Cell. Proteomics 4, 1194–1204.

Cornelissen, A., Ceyssens, P.J., Krylov, V.N., Noben, J.P., Volckaert, G., Lavigne, R., 2012. Identification of EPS-degrading activity within the tail spikes of the novel Pseudomonas putida phage AF. Virology 434, 251–256.

Crockett, J.T., Hodson, T.S., Hyde, J.R., Schouten, J.T., Smith, T.A., Merrill, B.D., Crook, M.B., Griffitts, J.S., Burnett, S.H., Grose, J.H., Breakwell, D.P., 2015. *Sinorhizobium* phage phiM19, complete genome, KR052481. Microbiology and Molecular Biology, Brigham Young University, National Center for Biotechnology Information.

Crook, M.B., Draper, A.L., Guillory, R.J., Griffitts, J.S., 2013. The Sinorhizobium meliloti essential porin RopA1 is a target for numerous bacteriophages. J. Bacteriol.

Cuervo, A., Pulido-Cid, M., Chagoyen, M., Arranz, R., Gonzalez-Garcia, V.A., Garcia-Doval, C., Caston, J.R., Valpuesta, J.M., van Raaij, M.J., Martin-Benito, J., Carrascosa, J.L., 2013. Structural characterization of the bacteriophage T7 tail machinery. J. Biol. Chem. 288, 26290–26299.

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.

Deak, V., Lukacs, R., Buzas, Z., Palvolgyi, A., Papp, P.P., Orosz, L., Putnoky, P., 2010. Identification of tail genes in the temperate phage 16–3 of Sinorhizobium meliloti 41. J. Bacteriol. 192, 1617–1623.

Dekel-Bird, N.P., Avrani, S., Sabehi, G., Pekarsky, I., Marston, M.F., Kirzner, S., Lindell, D., 2013. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. Environ. Microbiol. 15, 1476–1491.

Desper, R., Gascuel, O., 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J. Comput. Biol. 9, 687–705.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Hayes, S. 2012. Geneious 5.6.6.

Dziewit, L., Oscik, K., Bartosik, D., Radlinska, M., 2014. Molecular characterization of a novel temperate sinorhizobium bacteriophage, phi LM21, encoding DNA methyltransferase with CcrM-like specificity. J. Virol. 88, 13111–13124.

Echols, H., 1972. Developmental pathways for temperate phage – lysis vs lysogeny. Annu. Rev. Genet. 6, 157–190.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Elkins, P., Bunker, A., Cramer, W.A., Stauffacher, C.V., 1997. A mechanism for toxin insertion into membranes is suggested by the crystal structure of the channel-forming domain of colicin E1. Structure 5, 443–458.

Endersen, L., Guinane, C.M., Johnston, C., Neve, H., Coffey, A., Ross, R.P., McAuliffe, O., O'Mahony, J., 2015. Genome analysis of Cronobacter phage vB_CsaP_Ss1 reveals an endolysin with potential for biocontrol of Gram-negative bacterial pathogens. J. Gen. Virol. 96, 463–477.

Felsenstein, J., 1985. Confidence-limits on phylogenies – an approach using the bootstrap. Evolution 39, 783–791.

Ferguson, G.P., Datta, A., Baumgartner, J., Roop 2nd, R.M., Carlson, R.W., Walker, G.C., 2004. Similarity to peroxisomal-membrane protein family reveals that *Sinorhizobium* and *Brucella* BacA affect lipid-A fatty acids. Proc. Natl. Acad. Sci. U.S.A. 101, 5012–5017.

Fernandez, J.J., Luque, D., Caston, J.R., Carrascosa, J.L., 2008. Sharpening high resolution information in single particle electron cryomicroscopy. J. Struct. Biol. 164, 170–175.

Finan, T.M., Hartweig, E., LeMieux, K., Bergman, K., Walker, G.C., Signer, E.R., 1984. General transduction in Rhizobium meliloti. J. Bacteriol. 159, 120–124.

Ganyu, A., Csiszovszki, Z., Ponyi, T., Kern, A., Buzas, Z., Orosz, L., Papp, P.P., 2005. Identification of cohesive ends and genes encoding the terminase of phage 16–3. J. Bacteriol. 187, 2526–2531.

Garcia-Doval, C., Caston, J.R., Luque, D., Granell, M., Otero, J.M., Llamas-Saiz, A.L., Renouard, M., Boulanger, P., van Raaij, M.J., 2015. Structure of the receptor-binding carboxy-terminal domain of the bacteriophage T5 L-shaped tail fibre with and without its intra-molecular chaperone. Viruses-Basel 7, 6424–6440.

Glazebrook, J., Walker, G.C., 1991. Genetic techniques in Rhizobium meliloti. Methods Enzymol. 204, 398–418.

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

Glukhov, A.S., Krutilina, A.I., Shlyapnikov, M.G., Severinov, K., Lavysh, D., Kochetkov, V.V., McGrath, J.W., de Leeuwe, C., Shaburova, O.V., Krylov, V.N., Akulenko, N.V., Kulakov, L.A., 2012. Genomic analysis of Pseudomonas putida phage tf with localized single-strand DNA interruptions. PLoS One 7.

Grigorieff, N., 2007. FREALIGN: high-resolution refinement of single particle structures. J. Struct. Biol. 157, 117–125.

Grose, J.H., Casjens, S.R., 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. Virology 468, 421–443.

Groth, A.C., Calos, M.P., 2004. Phage integrases: biology and applications. J. Mol. Biol. 335, 667–678.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Halmillawewa, A.P., Restrepo-Cordoba, M., Yost, C.K., Hynes, M.F., 2015. Genomic and phenotypic characterization of Rhizobium gallicum phage vB_RglS_P106B. Microbiology 161, 611–620.

Halmillawewa, A.P., Perry, B., Gavard, R., Yost, C.K., Hynes, M.F., 2014a. Genomic characterization of two T7-like *Mesorhizobium loti* phages vB_MloP_Lo5R7ANS and vB_MloP_Cp1R7ANS-C2. Department of Biological Sciences, University of Calgary, National Center for Biotechnology Information.

Halmillawewa, A.P., Restrepo-Cordoba, M., Perry, B., Yost, C.K., Hynes, M.F., 2014b. Characterization and complete genome sequence of the temperate phage vB_RleM_PPF1 capable of lysogenizing the *Rhizobium leguminosarum* strain F1. Department of Biological Sciences University of Calgary, National Center for Biotechnology Information.

Hammerl, J.A., Gollner, C., Al Dahouk, S., Nockler, K., Reetz, J., Hertwig, S., 2016. Analysis of the first temperate broad host range Brucellaphage (BiPBO1) isolated from B. inopinata. Front. Microbiol. 7.

Hardies, S.C., Thomas, J.A., Black, L., Weintraub, S.T., Hwang, C.Y., Cho, B.C., 2016. Identification of structural and morphogenesis genes of Pseudoalteromonas phage phi RIO-1 and placement within the evolutionary history of Podoviridae. Virology 489, 116–127.

Hendrix, R.W., Ko, C.C., Jacobs-Sera, D., Hatfull, G.F., Erhardt, M., Hughes, K.T., Casjens, S.R., 2015. Genome sequence of Salmonella phage chi. Genome Announc. 3.

Henn, M.R., Engelen, B., Levin, J., Malboeuf, C., Casali, M., Russ, C., Lennon, N., Chapman, S.B., Erlich, R., Young, S.K., Yandava, C., Zeng, Q., Alvarado, L., Anderson, S., Berlin, A., Chen, Z., Freedman, E., Gellesch, M., Goldberg, J., Green, L., Griggs, A., Gujja, S., Heilman, E.R., Heiman, D., Hollinger, A., Howarth, C., Larson, L., Mehta, T., Pearson, M., Roberts, A., Ryan, E., Saif, S., Shea, T., Shenoy, N., Sisk, P., Stolte, C., Sykes, S., White, J., Haas, B., Nusbaum, C., Birren, B., 2013a. Rhizobium phage RR1-A genomic sequence, pp. NC_021560. The Broad Institute Genome Sequencing Platform, National Center for Biotechnology Information.

Henn, M.R., Engelen, B., Levin, J., Malboeuf, C., Casali, M., Russ, C., Lennon, N., Chapman, S.B., Erlich, R., Young, S.K., Yandava, C., Zeng, Q., Alvarado, L., Anderson, S., Berlin, A., Chen, Z., Freedman, E., Gellesch, M., Goldberg, J., Green, L., Griggs, A., Gujja, S., Heilman, E.R., Heiman, D., Hollinger, A., Howarth, C., Larson, L., Mehta, T., Pearson, M., Roberts, A., Ryan, E., Saif, S., Shea, T., Shenoy, N., Sisk, P., Stolte, C.,

Sykes, S., White, J., Haas, B., Nusbaum, C., Birren, B., 2013b. Rhizobium phage RR1-B genomic sequence, pp. NC_021557. The Broad Institute Genome Sequencing Platform, National Center for Biotechnology Information.
Hodson, T.S., Hyde, J.R., Schouten, J.T., Crockett, J.T., Smith, T.A., Merrill, B.D., Crook, M.B., Griffitts, J.S., Burnett, S.H., Grose, J.H., Breakwell, D.P., 2015. *Sinorhizobium* phage phiN3, complete genome, KR052482. Microbiology and Molecular Biology, Brigham Young University, National Center for Biotechnology Information.
Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., Sullivan, M.B., 2017. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. 11. http://dx.doi.org/10.1038/ismej.2017.1016.
Hryc, C.F., Chen, D.H., Afonine, P.V., Jakana, J., Wang, Z., Haase-Pettingell, C., Jiang, W., Adams, P.D., King, J.A., Schmid, M.F., Chiu, W., 2017. Accurate model annotation of a near-atomic resolution cryo-EM map. Proc. Natl. Acad. Sci. U.S.A. 114, 3103–3108.
Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339, 576–579.
Huang, S.J., Zhang, S., Jiao, N.Z., Chen, F., 2015. Comparative genomic and phylogenomic analyses reveal a conserved core genome shared by estuarine and oceanic cyanopodoviruses. PLoS One 10.
Iranzo, J., Krupovic, M., Koonin, E.V., 2016. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. mBio 7.
Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M., 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics 4, 1265–1272.
Johnson, M.C., Tatum, K.B., Lynn, J.S., Brewer, T.E., Lu, S., Washburn, B.K., Stroupe, M.E., Jones, K.M., 2015. Sinorhizobium meliloti phage Phi M9 defines a new group of T4 superfamily phages with unusual genomic features but a common T = 16 capsid. J. Virol. 89, 10945–10958.
Kang, H., Weiss, T.M., Bang, I., Weis, W.I., Choi, H.J., 2016. Structure of the intermediate filament-binding region of desmoplakin. PLoS One 11.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kelley, L.A., Sternberg, M.J., 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363–371.
Klucar, L., Stano, M., Hajduk, M., 2010. phiSITE: database of gene regulation in bacteriophages. Nucleic Acids Res. 38, D366–D370.
Kropinski, A.M., Kovalyova, I.V., Billington, S.J., Patrick, A.N., Butts, B.D., Guichard, J.A., Pitcher, T.J., Guthrie, C.C., Sydlaske, A.D., Barnhill, L.M., Havens, K.A., Day, K.R., Falk, D.R., McConnell, M.R., 2007. The genome of epsilon 15, a serotype-converting, group E1 Salmonella enterica-specific bacteriophage. Virology 369, 234–244.
Krupovic, M., Prangishvili, D., Hendrix, R.W., Bamford, D.H., 2011. Genomics of bacterial and archaeal viruses: dynamics within the prokaryotic virosphere. Microbiol. Mol. Biol. Rev. 75, 610–635.
Kulikov, E., Kropinski, A.M., Golomidova, A., Lingohr, E., Govorun, V., Serebryakova, M., Prokhorov, N., Letarova, M., Manykin, A., Strotskaya, A., Letarov, A., 2012. Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB_EcoP_G7C. Virology 426, 93–99.
Lander, G.C., Baudoux, A.C., Azam, F., Potter, C.S., Carragher, B., Johnson, J.E., 2012. Capsomer dynamics and stabilization in the T = 12 marine bacteriophage SIO-2 and its procapsid studied by CryoEM. Structure 20, 498–503.
Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W., Lyumkis, D., Potter, C.S., Carragher, B., 2009. Appion: an integrated, database-driven pipeline to facilitate EM image processing. J. Struct. Biol. 166, 95–102.
Latino, L., Essoh, C., Blouin, Y., Thien, H.V., Pourcel, C., 2014. A novel Pseudomonas aeruginosa bacteriophage, Ab31, a chimera formed from temperate phage PAJU2 and P. putida lytic phage AF: characteristics and mechanism of bacterial resistance. PloS One 9.
Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.W., Kropinski, A.M., 2008. Unifying classical and molecular taxonomic classification: analysis of the Podoviridae using BLASTP-based tools. Res. Microbiol. 159, 406–414.
Lavigne, R., Noben, J.P., Hertveldt, K., Ceyssens, P.J., Briers, Y., Dumont, D., Roucourt, B., Krylov, V.N., Mesyanzhinov, V.V., Robben, J., Volckaert, G., 2006. The structural proteome of Pseudomonas aeruginosa bacteriophage phi KMV. Microbiol-Sgm 152, 529–534.
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.
Lawson, C.L., Patwardhan, A., Baker, M.L., Hryc, C., Garcia, E.S., Hudson, B.P., Lagerstedt, I., Ludtke, S.J., Pintilie, G., Sala, R., Westbrook, J.D., Berman, H.M., Kleywegt, G.J., Chiu, W., 2016. EMDataBank unified data resource for 3DEM. Nucleic Acids Res. 44, D396–D403.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Lefort, V., Heled, J., Guindon, S., Biomatters, 2012. Geneious 5.6.6, PhyML plugin.
Lewis, J.A., Hatfull, G.F., 2001. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. Nucleic Acids Res. 29, 2205–2216.
Liu, M., Gingery, M., Doulatov, S.R., Liu, Y., Hodes, A., Baker, S., Davis, P., Simmonds, M., Churcher, C., Mungall, K., Quail, M.A., Preston, A., Harvill, E.T., Maskell, D.J., Eiserling, F.A., Parkhill, J., Miller, J.F., 2004. Genomic and genetic analysis of Bordetella bacteriophages encoding reverse transcriptase-mediated tropism-switching cassettes. J. Bacteriol. 186, 1503–1517.
Liu, X., Zhang, Q., Murata, K., Baker, M.L., Sullivan, M.B., Fu, C., Dougherty, M.T., Schmid, M.F., Osburne, M.S., Chisholm, S.W., Chiu, W., 2010. Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. Nat. Struct. Mol. Biol. 17, 830–836.
Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.
Lukashin, A.V., Borodovsky, M., 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 26, 1107–1115.
Mallick, S.P., Carragher, B., Potter, C.S., Kriegman, D.J., 2005. ACE: automated CTF estimation. Ultramicroscopy 104, 8–29.
Mayer, M.J., Narbad, A., Gasson, M.J., 2008. Molecular characterization of a Clostridium difficile bacteriophage and its cloned biologically active endolysin. J. Bacteriol. 190, 6734–6740.
Mcconnell, M., Walker, B., Middleton, P., Chase, J., Owens, J., Hyatt, D., Gutierrez, H., Williams, M., Hambright, D., Barry, M., Sage, S., Fuller, G., Birdwell, M., Rydelski, M., Risley, S., Kat, B., 1992. Restriction endonuclease and genetic-mapping studies indicate that the vegetative genome of the temperate, Salmonella-specific bacteriophage, epsilon-15, is circularly-permuted. Arch. Virol. 123, 215–221.
Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.E., Ausubel, F.M., 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. J. Bacteriol. 149, 114–122.
Merrill, B.D., Ward, A.T., Grose, J.H., Hope, S., 2016. Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. BMC Genom. 17.
Miller, R.V., 1998. Methods for enumeration and characterization of bacteriophages from environmental samples. In: Burlage, R.S. (Ed.), Techniques in Microbial Ecology. Oxford University Press, New York, Oxford, pp. 218–235.
Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347.
Mirzaei, M.K., Eriksson, H., Kasuga, K., Haggard-Ljungquist, E., Nilsson, A.S., 2014. Genomic, proteomic, morphological, and phylogenetic analyses of vB_EcoP_SU10, a podoviridae phage with C3 morphology. PLoS One 9.
Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E., Ghai, R., 2013. Expanding the marine virosphere using metagenomics. PLoS Genetics 9.
Moak, M., Molineux, I.J., 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol. Microbiol. 51, 1169–1183.
Morris, P., Marinelli, L.J., Jacobs-Sera, D., Hendrix, R.W., Hatfull, G.F., 2008. Genomic characterization of mycobacteriophage giles: evidence for phage acquisition of host DNA by illegitimate recombination. J. Bacteriol. 190, 2172–2182.
Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R., 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646–4658.
Parent, K.N., Tang, J.H., Cardone, G., Gilcrease, E.B., Janssen, M.E., Olson, N.H., Casjens, S.R., Baker, T.S., 2014. Three-dimensional reconstructions of the bacteriophage CUS-3 virion reveal a conserved coat protein I-domain but a distinct tailspike receptor-binding domain. Virology 464, 55–66.
Peng, Y., Ding, Y.-J., Lin, H., 2013. Isolation, identification and lysis properties analysis of a *Vibrio parahaemolyticus* phage VPp1. Mar. Sci. Qingdao-Chin. Ed. 37, 101–111.
Pope, W.H., Carbonara, M.E., Cioffi, H.M., Cruz, T., Dang, B.Q., Doyle, A.N., Fan, O.H., Gallagher, M., Gentile, G.M., German, B.A., Farrell, M.E., Gerwig, M., Hunter, K.L., Lefever, V.E., Marfisi, N.A., McDonnell, J.E., Monga, J.K., Quiroz, K.G., Pong, A.C., Rimple, P.A., Situ, M., Sohnen, P.C., Stockinger, A.N., Thompson, P.K., Torchio, N.M., Toner, C.L., Ulbrich, M.C., Vohra, N.I., Zakir, A., Adkins, N.L., Brown, B.R., Churilla, B.M., Kramer, Z.J., Lapin, J.S., Montgomery, M.T., Prout, A.K., Grubb, S.R., Warner, M.H., Bowman, C.A., Russell, D.A., Hatfull, G.F., 2015. Genome sequences of mycobacteriophages AlanGrant, Baee, Corofin, OrangeOswald, and Vincenzo new members of cluster B. Genome Announc. 3.
Rao, V.B., Feiss, M., 2008. The Bacteriophage DNA Packaging Motor. Annu. Rev. Genet. 42, 647–681.
Rappsilber, J., Ryder, U., Lamond, A.I., Mann, M., 2002. Large-scale proteomic analysis of the human spliceosome. Genome Res. 12, 1231–1245.
Rashid, S.J., Barylski, J., Hargreaves, K.R., Millard, A.A., Vinner, G.K., Clokie, M.R.J., 2016. Two novel myoviruses from the north of Iraq reveal insights into clostridium difficile phage diversity and biology. Viruses-Basel 8.
Restrepo-Cordoba, M., Halmillawewa, A.P., Perry, B., Hynes, M.F., Yost, C.K., 2014. Isolation and characterization of *Rhizobium leguminosarum* phages from western Canadian soils and complete genome sequences of rhizobiophages vB_RleS_L338C and vB_RleM_P10VF. Department of Biological Sciences, University of Calgary, National Center for Biotechnology Information.
Riccio, C., Browning, C., Prokhorov, N., Letarov, A., Leiman, P., 2015. Crystal structure of tail fiber protein gp63.1 from *E. coli* phage G7C. RCSB PDB.
Sambrook, J., Russell, D.W., 2001. Molecular cloning: a laboratory manual, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
Santamaria, R.I., Bustos, P., Sepulveda-Robles, O., Lozano, L., Rodriguez, C., Fernandez, J.L., Juarez, S., Kameyama, L., Guarneros, G., Davila, G., Gonzalez, V., 2014. Narrow-host-range bacteriophages that infect Rhizobium etli associate with distinct genomic types. Appl. Environ. Microbiol. 80, 446–454.
Savalia, D., Westblade, L.F., Goe, M., Florens, L., Kemp, P., Akulenko, N., Pavlova, O., Padovan, J.C., Chait, B.T., Washburn, M.P., Ackermann, H.W., Mushegian, A., Gabisonia, T., Molineux, I., Severinov, K., 2008. Genomic and proteomic analysis of phEco32, a novel Escherichia coli bacteriophage. J. Mol. Biol. 377, 774–789.
Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., Wilbur, W.J., Yaschenko, E., Ye, J., 2011. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 39, D38–51.
Scheres, S.H., 2012. RELION: implementation of a Bayesian approach to cryo-EM

structure determination. J. Struct. Biol. 180, 519–530.
Schildbach, J.F., Karzai, A.W., Raumann, B.E., Sauer, R.T., 1999. Origins of DNA-binding specificity: role of protein contacts with the DNA backbone. Proc. Natl. Acad. Sci. U.S.A. 96, 811–817.
Schlüter, J.-P., Reinkensmeier, J., Barnett, M.J., Lang, C., Krol, E., Giegerich, R., Long, S.R., Becker, A., 2013. Global mapping of transcription start sites and promoter motifs in the symbiotic alpha-proteobacterium *Sinorhizobium meliloti* 1021. BMC Genom. 14, 156.
Schouten, J.T., Crockett, J.T., Hodson, T.S., Hyde, J.R., Smith, T.A., Merrill, B.D., Crook, M.B., Griffitts, J.S., Burnett, S.H., Grose, J.H., Breakwell, D.P., 2015. *Sinorhizobium* phage phiM7, complete genome, KR052480. Microbiology and Molecular Biology, Brigham Young University, National Center for Biotechnology Information.
Schulmeister, S.A., Krol, J.E., Vorhoelter, F.-J., Skorupska, A.M., Lotz, W., 2009. Sequence of the genome of Sinorhizobium meliloti bacteriophage PBC5, NC_003324.1. National Center for Biotechnology Information.
Schulz, E.C., Dickmanns, A., Urlaub, H., Schmitt, A., Muhlenhoff, M., Stummeyer, K., Schwarzer, D., Gerardy-Schahn, R., Ficner, R., 2010. Crystal structure of an intramolecular chaperone mediating triple-beta-helix folding. Nat. Struct. Mol. Biol. 17, 210–215.
Sepulveda-Robles, O., Kameyama, L., Guarneros, G., 2012. High Diversity and Novel Species of Pseudomonas aeruginosa Bacteriophages. Appl. Environ. Microbiol. 78, 4510–4515.
Shrum, D.C., Woodruff, B.W., Stagg, S.M., 2012. Creating an infrastructure for high-throughput high-resolution cryogenic electron microscopy. J. Struct. Biol. 180, 254–258.
Skennerton, C.T., Angly, F.E., Breitbart, M., Bragg, L., He, S.M., McMahon, K.D., Hugenholtz, P., Tyson, G.W., 2011. Phage encoded H-NS: a potential Achilles heel in the bacterial defence system. PLoS One 6.
Soding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248.
Stano, M., Klucar, L., 2011. phiGENOME: an integrative navigation throughout bacteriophage genomes. Genomics 98, 376–380.
Stroupe, M.E., Brewer, T.E., Sousa, D.R., Jones, K.M., 2014. The structure of Sinorhizobium meliloti phage ΦM12, which has a novel T = 19l triangulation number and is the founder of a new group of T4-superfamily phages. Virology 450–451, 205–212.
Stummeyer, K., Schwarzer, D., Claus, H., Vogel, U., Gerardy-Schahn, R., Muhlenhoff, M., 2006. Evolution of bacteriophages infecting encapsulated bacteria: lessons from Escherichia coli K1-specific phages. Mol. Microbiol. 60, 1123–1135.
Suhanovsky, M.M., Teschke, C.M., 2015. Natures favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. Virology 479–480C, 487–497.
Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new Leginon system. J. Struct. Biol. 151, 41–60.
Teschke, C.M., Parent, K.N., 2010. 'Let the phage do the work': Using the phage P22 coat protein structures as a framework to understand its folding and assembly mutants. Virology 401, 119–130.
Tevdoradze, E., Farlow, J., Kotorashvili, A., Skhirtladze, N., Antadze, I., Gunia, S., Balarjishvili, N., Kvachadze, L., Kutateladze, M., 2015. Whole genome sequence comparison of ten diagnostic brucellaphages propagated on two Brucella abortus hosts. Virol. J. 12.
Vassylyeva, M.N., Svetlov, V., Dearborn, A.D., Klyuyev, S., Artsimovitch, I., Vassylyev, D.G., 2007. The carboxy-terminal coiled-coil of the RNA polymerase beta'-subunit is the main binding site for Gre factors. EMBO Rep. 8, 1038–1043.
Veesler, D., Cambillau, C., 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol. Mol. Biol. Rev. 75, 423–433 first page of table of contents.
Watkins, D., Hsiao, C.L., Woods, K.K., Koudelka, G.B., Williams, L.D., 2008. P22 c2 repressor-operator complex: Mechanisms of direct and indirect readout. Biochemistry 47, 2325–2338.
Weigel, C., Seitz, H., 2006. Bacteriophage replication modules. Fems Microbiology Reviews 30, 321–381.
Wibberg, D., Tejerizo, G.T., Del Papa, M.F., Martini, C., Puhler, A., Lagares, A., Schluter, A., Pistorio, M., 2014. Genome sequence of the acid-tolerant strain Rhizobium sp LPU83. J. Biotechnol. 176, 40–41.
Wisniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nat. Methods 6, 359–362.
Wu, W.M., Leavitt, J.C., Cheng, N.Q., Gilcrease, E.B., Motwani, T., Teschke, C.M., Casjens, S.R., Steven, A.C., 2016. Localization of the Houdinisome (Ejection Proteins) inside the Bacteriophage P22 Virion by Bubblegram Imaging. mBio 7.
Yamaguchi, K., Miyata, R., Shigehisa, R., Uchiyama, J., Takemura-Uchiyama, I., Kato, S., Ujihara, T., Sakaguchi, Y., Daibata, M., Matsuzaki, S., 2014. Genome analysis of Pseudomonas aeruginosa bacteriophage KPP23 belonging to the family Siphoviridae. Genome Announc. 2.
Yamamoto, K.R., Alberts, B.M., Benzinge.R, Lawhorne, L., Treiber, G.,, 1970. Rapid bacteriophage sedimentation in presence of polyethylene glycol and its application to large-scale virus purification. Virology 40 734 000.
Young, R., 2013. Phage lysis: do we have the hole story yet? Curr. Opin. Microbiol. 16, 790–797.
Young, R., 2014. Phage lysis: three steps, three choices, one outcome. J. Microbiol. 52, 243–258.
Zecchi, L., Lo Piano, A., Suzuki, Y., Canas, C., Takeyasu, K., Ayora, S., 2012. Characterization of the holliday junction resolving enzyme encoded by the bacillus subtilis bacteriophage SPP1. PLoS One 7.
Zhan, Y.C., Huang, S.J., Voget, S., Simon, M., Chen, F., 2016. A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agents. Sci Rep-Uk 6.
Zhang, S., Flores-Cruz, Z., Zhou, L., Kang, B.H., Fleites, L.A., Gooch, M.D., Wulff, N.A., Davis, M.J., Duan, Y.P., Gabriel, D.W., 2011. 'Ca. Liberibacter asiaticus' carries an excision plasmid prophage and a chromosomally integrated prophage that becomes lytic in plant infections. Mol. Plant Microbe Interact. 24, 458–468.
Zhao, H.Y., Speir, J.A., Matsui, T., Lin, Z.H., Liang, L.F., Lynn, A.Y., Varnado, B., Weiss, T.M., Tang, L., 2016. Structure of a bacterial virus DNA-injection protein complex reveals a decameric assembly with a constricted molecular channel. PLoS One 11.