Virology 450-451 (2014) 84–97

Contents lists available at ScienceDirect

Virology

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

The genome, proteome and phylogenetic analysis of *Sinorhizobium meliloti* phage ΦM12, the founder of a new group of T4-superfamily phages

Tess E. Brewer [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 Biophysics, Florida State University, 91 Chieftan Way Tallahassee, FL 32306-4380 United States

ARTICLE INFO

Available online 25 December 2013

Keywords:
Bacteriophage
ΦM12
PhiM12
T4-like phage
Sinorhizobium meliloti
Alphaproteobacteria
Proteome
gp23
gp20
Class II ribonucleotide reductase

ABSTRACT

Phage ΦM12 is an important transducing phage of the nitrogen-fixing rhizobial bacterium *Sinorhizobium meliloti*. Here we report the genome, phylogenetic analysis, and proteome of ΦM12, the first report of the genome and proteome of a rhizobium-infecting T4-superfamily phage. The structural genes of ΦM12 are most similar to T4-superfamily phages of cyanobacteria. ΦM12 is the first reported T4-superfamily phage to lack genes encoding class I ribonucleotide reductase (RNR) and exonuclease *dexA*, and to possess a class II coenzyme B_{12}-dependent RNR. ΦM12's novel collection of genes establishes it as the founder of a new group of T4-superfamily phages, fusing features of cyanophages and phages of enteric bacteria.

Introduction

It is an understatement to say that the advent of next-generation sequencing technology has reminded us of how little we know about phage diversity. In fact, the term "dark matter of the biosphere" has been used to describe the breadth and depth of phages yet to be characterized (Filee et al., 2005). Phage T4 and its close relatives are among the phages that have been characterized in the greatest detail at the molecular level (for reviews see Comeau and Krisch, 2008; Leiman et al., 2010). It has become clear in recent years with the discovery of new lineages of T4-superfamily phages, that T4 itself and other T4-like phages of enteric bacteria are only one branch of this complex family tree (Sullivan et al., 2010; Chen et al., 2007; Adriaenssens et al., 2012). The "T4-like cyanophage" sub-group of T4-like phages is another branch of this family that has been surveyed in depth. The genomes of many of these cyanophages have been sequenced and annotated (Sullivan et al., 2010, 2005; Weigele et al., 2007; Dreher et al., 2011; Mann et al., 2005), and a structure for the capsid of one of these, Syn9, has been determined using cryo-electron microscopy (Weigele et al., 2007).

T4 is the prototypical myovirus, the family of dsDNA-genome bacteriophages with long, contractile tails (Hatfull and Hendrix, 2011). T4-like phages make up a substantial portion of the known myoviruses, which, along with podoviruses and siphoviruses compose the *Caudovirales* order of dsDNA-genome bacteriophages (Hatfull and Hendrix, 2011). The 195-kb genome sequence of ΦM12, described here, is the third complete genome of a T4-superfamily phage that infects an alphaproteobacterial host to be deposited in the National Center for Biotechnology Information (NCBI) GenBank (Table 1). The ΦM12 genome is also one of only a few completely sequenced genomes of the many phages that infect rhizobial bacteria (Table 1).

ΦM12 is one of two phages routinely used to perform generalized transduction in *Sinorhizobium meliloti* 1021 (Finan et al., 1984; Martin and Long, 1984), and is a standard tool used for genetic manipulation of this bacterium (Glazebrook and Walker, 1991). *S. meliloti* 1021 is a rhizobial strain used as a model organism for the study of symbiotic nitrogen fixation with its legume plant hosts *Medicago sativa* (alfalfa), *Medicago truncatula* (barrel medic), and other plants of the *Medicago* and *Melilotus* genera (Gibson et al., 2008; Jones et al., 2007b). ΦM12 was originally isolated from a commercial rhizobial seed inoculant prepared for field crop use on alfalfa (Finan et al., 1984). The presence of lytic phage within bioinoculant preparations of beneficial bacteria raises the

* Corresponding author. Tel.: +1 850 645 8743.
E-mail address: kmjones@bio.fsu.edu (K.M. Jones).

0042-6822
http://dx.doi.org/10.1016/j.virol.2013.11.027

Table 1
Phage genomes. Phages that infect rhizobial and other alphaproteobacterial hosts for which the complete genome has been reported.

Phage name	Phage type	GenBank Accession number	Host bacterium	References
T4-like phages that infect alphaproteobacterial hosts				
Sphingomonas phage PAU	myovirus	NC_019521	*Sphingomonas paucimobilis*	Pope et al. (2012)
Pelagibacter phage HTVC008M	myovirus	NC_020484	*Pelagibacter ubique*	Zhao et al. (2013)
Phages that infect rhizobial hosts				
Rhizobium phage 16-3	siphovirus	NC_011103	*Sinorhizobium meliloti* Rm41	Ganyu et al. (2005)
S. meliloti phage PBC5	unclassified Caudovirus	NC_003324	*S. meliloti* 1021	Schulmeister et al. (2009)
Rhizobium phages RHEph01-RHEph10	unclassified phages	accessions JX483873-JX483881	*Rhizobium etli*	Santamaria et al. (2013)
Rhizobium phages RR1-A and RR1-B	unclassified phages	accessions NC_021560 and JF974314	*Rhizobium radiobacter* P007	Henn et al. (2013a,2013b)

possibility that these bacterial populations may experience attrition due to phage predation. Conversely, co-inoculation of legume crops with both phage directed against native, inefficient-nitrogen-fixing rhizobial populations along with the desired, efficient nitrogen-fixing strain of rhizobia has been proposed as a method of maximizing crop yields (Basit et al., 1992; Jones et al., 2007a).

ΦM12 is also interesting in that its initial interaction with *S. meliloti* host cells is different from that of many other phages that infect this bacterium. Mutants of *S. meliloti* that have a defective, truncated lipopolysaccharide (LPS) core oligosaccharide are resistant to many *S. meliloti* lytic phages, but remain sensitive to ΦM12 (Campbell et al., 2002, 2003). The *S. meliloti* outer membrane protein encoded by *ropA1* was recently determined to be the site of adsorption of ΦM12 on the *S. meliloti* cell surface (Crook et al., 2013). The sequence, annotation, and comparative analysis of the genome (Accession KF381361.1), and the proteome of ΦM12, described here, along with the structural characterization of ΦM12 in the accompanying manuscript (Stroupe et al., in press) are important steps in the further characterization of the biology of this phage and its interaction with host *S. meliloti* cells.

Results and discussion

Phylogeny places ΦM12 with uncultured myoviruses of the T4-superfamily

Initial BLAST searches (Altschul et al., 1997) with predicted ORFs from the ΦM12 genome indicated that its structural genes and replication machinery are more similar to those of T4-superfamily phages than to any other phage group. In order to place ΦM12 in a phylogenetic context with other members of the T4 superfamily, we used the amino acid sequence of the gp20 portal protein to construct a phylogenetic tree. The amino acid sequence of gp20 has become a commonly-used marker for constructing phylogenetic trees of T4-like phages and other myoviruses (Fuller et al., 1998; Zhong et al., 2002; Sullivan et al., 2008).

The ΦM12 gp20 full-length amino acid sequence was used in a BLAST (Altschul et al., 1997) search to find similar sequences. The most similar 149 sequences are all from the gp20 protein of T4-superfamily phages, some of which have never been cultured and are known only from PCR of gene 20 from environmental samples. A conserved internal amino acid sequence from each of these 149 proteins (listed in Supplemental Table 1A) was aligned with the orthologous 181 amino acid fragment from ΦM12 using the MUSCLE program (Edgar, 2004). This alignment was then used as the basis for generating an unrooted tree using the phyML algorithm (Guindon and Gascuel, 2003; Lefort et al., 2012) (Fig. 1A). This tree places the ΦM12 gp20 protein in a clade with 36 other gp20 sequences isolated from uncultured phages in environmental samples (Wilhelm et al., 2006; Wang et al., 2011, 2010; Zhong et al., 2002; Short and Suttle, 2005; Jameson et al., 2011) (Fig. 1). The gp20 sequences in these studies are from diverse environments, ranging from agricultural fields (Wang et al., 2011, 2010) to freshwater (Short and Suttle, 2005; Wilhelm et al., 2006) and marine (Jameson et al., 2011; Short and Suttle, 2005; Zhong et al., 2002) environments, demonstrating that sequences related to ΦM12 are found in a wide variety of settings. However, many of the nodes in the subtree containing ΦM12 gp20 are poorly supported by bootstrap analysis (close-up view in Fig. 1B), suggesting that ΦM12 is quite distant from many of its relatives in these environmental samples. This also suggests that the closest relatives of ΦM12 have not yet been identified and characterized. A phylogenetic analysis performed with the ΦM12 gp23 major capsid protein, another phage protein that has frequently been used as a phylogenetic marker (Comeau and Krisch, 2008; Edgar, 2004; Drummond et al., 2012), also groups ΦM12 with sequences from uncultured phages (data not shown).

Overall, the phyML analyses of the gp20 and gp23 amino acid sequences suggest that the sequence information currently available in public databases is not yet sufficient to permit a clear determination of the phylogenetic position of ΦM12 among the T4-superfamily phages. However, it appears that ΦM12 is somewhat more closely related to T4-like cyanophages than to T4-like phages that infect enteric bacteria.

ΦM12 has features of both T4-like cyanophages and T4-like phages of enteric bacteria

Additional features of the ΦM12 sequence further support the observation that ΦM12 represents a new lineage of T4-superfamily phages distinct from those that have been previously characterized. ΦM12 does contain most of the proteins that make up the "core" set common to all T4-like phages (Table 2); however, it is the first-reported, sequenced member of this group that lacks the genes encoding class I ribonucleotide reductase (RNR), *nrdA* and *nrdB*, and the exonuclease *dexA* (Sullivan et al., 2010; Filee et al., 2006) (see Supplemental Table 1B for additional references). (Note that ΦM12 does have a *nrdJ* gene, encoding a class II RNR, discussed below.) Beyond the T4-like phage core gene set, however, ΦM12 shares very few proteins with either the T4-like cyanophages or the T4-like phages of enteric bacteria.

To further elucidate the relationships between ΦM12 and other T4-like phages, based on the core gene set, we compared each of the ΦM12 T4-like phage "universal core proteins" and nearly-universal core proteins (Sullivan et al., 2010) with the orthologous proteins from 10 cyanophages and 7 non-cyanophage T4-like phages chosen from different sub-classes of these phage groups (Petrov et al., 2010; Sullivan et al., 2010) (Table 2, Supplemental Table 1B.) Each of the 42 full-length ΦM12 T4-like phage core proteins in Table 2 was aligned with its orthologs in MUSCLE (Edgar, 2004; Drummond et al., 2012) and the percent identity

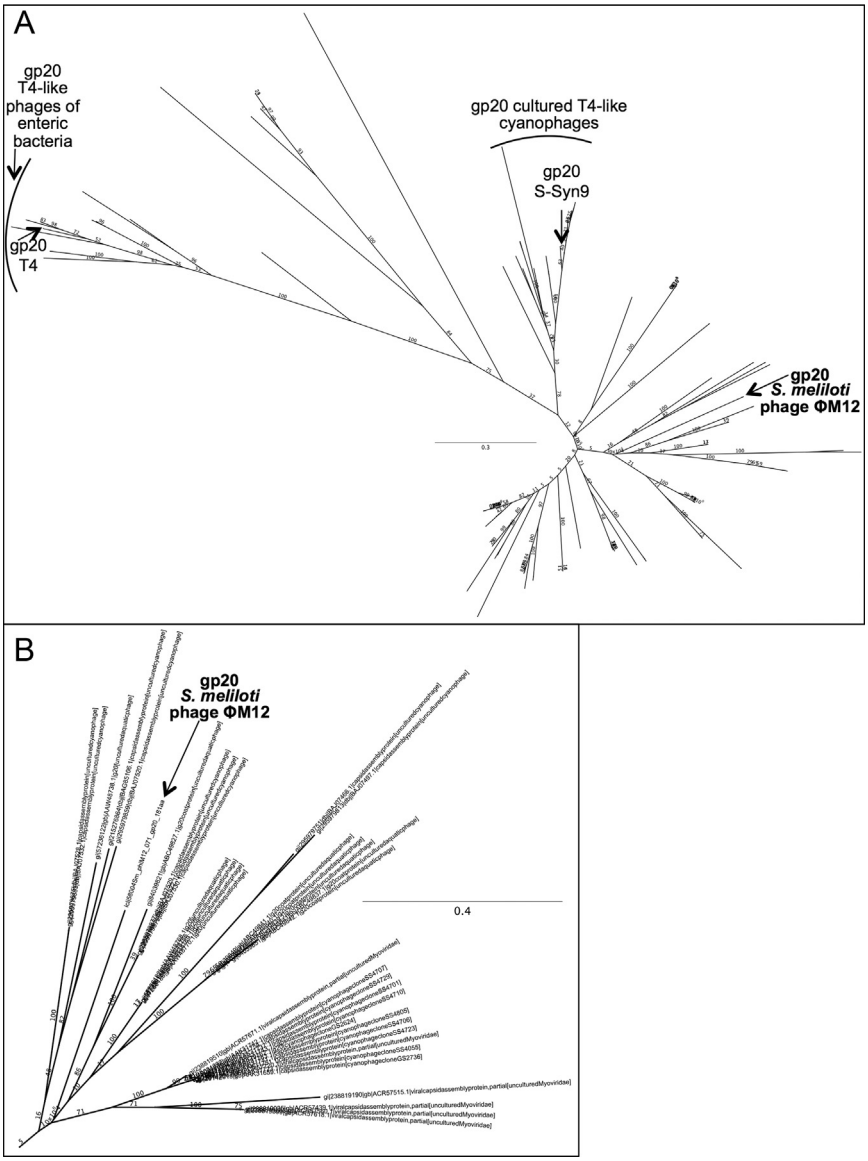

Fig. 1. (A) An unrooted gp20 (portal vertex protein) tree generated by phyML from a MUSCLE alignment of a 181 amino acid internal sequence from 149 sequences (listed in Supplemental Table 1A). The bootstrap percentage value for each branch is shown on the tree and the bar is a marker of branch distance length. gp20 of T4 and gp20 of Syn9 are highlighted because both the genome (Weigele et al., 2007; Miller et al., 2003) and the structure (Weigele et al., 2007; Fokine et al., 2013, 2004) of these phages have been characterized. (B) A close-up view of the sub-tree containing ΦM12 gp20 protein. The sequences in this subtree are denoted with an asterisk (∗) in Supplemental Table 1A.

with each ortholog was determined. (Note that unlike most phages, ΦM12 has two copies of gp19 tail protein (similar to the Campylomyoviruses (Javed et al., 2013)) and a truncated second copy of gp17 terminase.) The core proteins are presented in Table 2 in the order in which they are found in the genome starting with gp41 (see Fig. 2, and Supplemental Fig. 1, for gene order).

With the exception of UvsX, NrdC (glutaredoxin), Td (thymidylate synthase) and some components of the phage particle baseplate, ΦM12 core proteins are more similar to their orthologs from cyanophages and *Pelagibacter* phage HTVC008M than to their orthologs from T4-like phages of enteric bacteria. However, even the most highly conserved protein, the major capsid protein gp23, has less than 60% identity with the major capsid protein from any cyanophage, suggesting that ΦM12 is widely diverged. Almost all of the predicted structural proteins in Table 2 were detected in the ΦM12 lysate proteome (see below, Table 3 and the accompanying manuscript (Stroupe et al., in press).

Genomic synteny of ΦM12 with other T4-superfamily phages

As in other T4-like phages, the ΦM12 genome is circularly permuted (Miller et al., 2003), and ORFs with related functions are organized in clusters (Table 2, Fig. 2, Supplemental Fig. 1). When compared to the genome organization of the most closely-related T4-like cyanophages, the synteny in gene clusters 1 and 2 is highly conserved (Fig. 2). However, there are some differences in gene spacing between ΦM12 and the cyanophages. The large (10 kb) gap in ΦM12 between gp19 and gp18 that defines the boundary between gene clusters 1 and 2 has been observed in very few (Javed et al., 2013) T4-superfamily phages (Fig. 2, Supplemental Fig. 1). (This gap has been confirmed by PCR and Sanger-sequencing, as described in Experimental Procedures.) Also, ΦM12 does not have the gap between gp17 and gp16 that is present in some cyanophages (Fig. 2). Interestingly, ΦM12 has two sequential copies of gp17, one of which is more similar to that of cyanophage

Table 2

MUSCLE multiple alignment percent identity score of full-length amino acid sequences of ΦM12 T4-like phage universal core and nearly universal core ORFs with those of 17 other T4-like phages. See Supplemental Table 1B for sequence accessions and references.

Best match to ΦM12 protein → Poorest match

		←Cyanophages→											|--Phages of enteric bacteria→					
gene cluster 1	gene product	function	S-SM2	P-SSM4	P-Syn1	HTVC008M	S-Syn9	S-Syn19	S-ShM2	P-Syn33	S-SM1	P-HM1	S-CRM01	T4	Aeh	44RR	KVP40	RB43	ΦW-14
	gp41	DNA primase-helicase	52.1	51.0	46.9	47.4	49.5	48.6	47.1	49.0	49.0	49.6	43.5	35.9	37.9	38.3	36.6	36.3	31.0
	UvsX	recombination protein	25.9	27.4	25.6	27.8	26.8	25.1	24.2	24.5	25.6	27.4	26.3	46.9	47.8	---	46.4	46.0	33.3
	gp43	DNA polymerase	42.1	40.9	39.8	38.7	40.6	39.4	41.5	39.6	40.1	40.5	40.0	30.0	31.9	31.3**	31.1	30.5	29.6
	RegA	translational re-pressor/early genes	45.3	53.0	48.9	49.3	51.5	52.2	49.6	52.2	52.2	46.7	52.3	49.6	49.2	46.7	48.4	48.0	28.5
	gp62	clamp loader subunit	36.4	33.3	35.9	39.4	34.8	34.1	31.3	33.3	34.8	37.2	32.6	17.0	18.7	20.4	28.9	22.4	24.4
	gp44	clamp loader subunit	46.3	43.3	40.6	43.7	45.4	47.3	45.7	46.7	44.4	44.8	42.0	31.3	32.9	31.6	35.4	31.4	32.1
	gp45	sliding clamp DNA polym. accessory	33.0	31.5	31.1	31.6	24.2	31.4	28.2	32.9	33.8	31.7	28.7	24.9	26.1	27.2	23.7	28.8	19.9
	gp46	recombination endo-nuclease subunit	43.6	42.0	42.9	40.1	42.0	44.2	43.5	42.5	42.8	39.9	35.9	35.2	24.6	29.7	26.6	31.5	23.7
	gp47	recombination endo-nuclease subunit	34.0	33.4	37.1	36.2	39.1	38.0	36.1	37.4	37.7	35.9	36.2	25.4	23.4	25.1	30.8	26.6	21.6
	gp55	sigma factor for late transcription	36.2	40.2	41.0	34.4	40.4	38.4	39.1	36.0	39.5	44.0	40.8	25.7	19.4	22.0	21.7	20.3	20.7
	UvsW	helicase	41.0	43.7	42.1	39.9	41.9	42.5	42.3	41.6	42.7	42.6 **	37.2	34.2	34.6	32.5	37.3	32.6	28.9
	UvsY	recombination,repair and ssDNA binding	34.2	34.7	26.1	32.4	31.4	32.8	36.6	28.2	28.3	27.6	28.5	24.1	23.4	---	20.7	22.1	15.2
	gp3	head proximal tip of tail tube	28.6	26.7	21.8	20.2	23.7	27.7	28.0	25.1	22.8	27.6	25.1	21.6	21.6	21.6	21.8	19.5	16.2
	gp23	precursor of major head subunit	54.8	57.0	50.8	56.8	58.8	57.7	55.5	57.8	56.4	53.6	48.5	32.8	34.4	33.6	34.9	33.7	41.3
	gp22	scaffold prohead core protein	23.9	28.6	26.0	25.1	27.5	27.4	26.3	27.8	27.5	24.2	19.3	23.4	24.8	21.2	18.2	23.3	14.5
	gp21	prohead core scaffold/protease	42.0	43.3	44.6	44.5	43.8	43.8	45.1	41.5	42.0	40.4	40.2	24.4	29.1	28.1	30.9	25.2	31.1
	gp20	portal vertex protein	46.3	46.7	43.2	41.8	47.1	45.4	44.5	46.5	45.4	44.0	42.0	31.8	30.5	32.3	33.0	33.3	29.0
	gp19	tail tube monomer	33.3	30.3	29.9	31.8 18.5	30.9	34.8	25.8	31.9	33.0	24.7	34.4	29.8	30.0	32.9	29.1 15.4	33.1	31.2 15.3

	gene product	function	S-SM2	P-SSM4	P-Syn1	HTVC008M	S-Syn9	S-Syn19	S-ShM2	P-Syn33	S-SM1	P-HM1	S-CRM01	T4	Aeh	44RR	KVP40	RB43	ΦW-14
gene cluster 2	gp18	tail sheath monomer	29.8	32.9	33.9	38.8	32.5	33.3	29.7 **	33.0	33.5	32.7	29.8	28.0	28.2	27.7	29.8	28.5	19.6
	gp17.1	terminase, large sub	42.4	41.2	41.8	43.3	39.7	43.5	45.1	43.7	42.6	42.3	39.4	36.1	34.6	36.5	35.9	37.1	29.2
	gp17.2	terminase, (truncated 2nd copy)	27.3	25.5	27.3	26.0	25.1	23.8	25.1	23.4	23.8	25.1	27.3	21.6	22.0	20.8	21.4	23.0	17.6
	gp16	terminase, small subunit	25.5	24.8	23.8	17.9	22.9	22.4	23.1	23.6	23.6	28.6	24.7	18.7	13.3	14.4	13.2	15.0	12.5
	gp15	proximal tail sheath stabilization	26.0	28.5	32.7	33.2	31.5	26.2	31.0	34.2	33.1	28.4	25.9	22.7	22.8	21.3	17.0	20.1	20.6
	gp14	neck protein	15.9	24.9	28.8	25.9	19.9	24.1	19.4	21.3	26.4	16.5	16.8	22.7	20.5	19.5	23.0	20.6	24.4
	gp13	neck protein	33.3	28.9	33.1	36.5	29.7	30.5	31.6	28.9	31.2	30.4	23.4	17.9	19.0	18.3	19.7	14.1	22.0
gene cluster 3	gp8	baseplate wedge	18.2	21.0	16.6	18.8	21.7	20.9	21.4	21.2	21.4	19.7	17.5	7.8	8.9	8.8	8.1	8.2	---
	gp7	baseplate wedge initiator	4.1	3.0	5.2	8.6	3.7	3.4	3.2	3.5	3.7	3.6	7.4	6.1	5.5	6.3	5.5	6.1	---
	gp6	baseplate wedge	29.0	26.1	28.6	31.5	29.4	27.0	24.4	25.2	26.7	28.8	28.2	17.7	20.6	20.2	20.2	20.7	21.7
	gp25	baseplate wedge	28.3	26.6	32.6	31.3	32.6	22.3	23.6	28.8	27.3	26.3	25.6	25.2	26.4	27.8	24.1	29.1	29.4
	gp5	baseplate hub + tail lysozyme	10.3	9.4	8.0	19.3	9.7	9.7	9.4	9.4	9.2	8.9	11.1	15.3	14.7	15.6	22.8	15.1	9.7
	gp53	baseplate wedge	17.6	16.0	20.3	19.3	19.9	20.1	17.5	22.7	19.9	20.6	22.1	16.8	21.1	16.0	18.2	19.7	22.3
	gp48	baseplate tail tube cap	13.2	11.0	13.4	14.1	13.4	13.6	13.4	15.5	10.4	10.7	18.0	11.7	11.0	9.6	9.7	9.8	15.3
gene cluster 4	gp4	head completion protein	45.8	49.3	32.7	37.8	49.3	48.0	47.3	48.0	48.7	43.7	25.5	36.4	39.6	41.6	40.6	40.8	32.8
	gp19.2	tail tube monomer (2nd copy)	14.9	18.0	19.7	17.9 15.0	18.8	15.4	16.0	17.1	18.0	14.9	16.0	12.9	13.3	12.5	14.9 14.3	14.6	16.3 14.8
	gp26	baseplate hub	34.6	27.4	32.2	34.0	32.7	32.2	32.1	32.0	33.6	35.1	30.1	20.8	17.7	14.3	15.8	12.0	20.8
	gp51	baseplate hub assembly catalyst	39.6	28.3	40.0	26.2	29.3	27.5	37.7	27.5	34.0	36.0	22.6	27.3	---	25.5	35.3	15.7	20.8
cluster 5	gp33	late promoter transcription factor	21.3	33.3	22.7	18.1	25.3	29.3	25.3	29.3	33.3	28.0	21.3	13.8	18.7	8.8	13.0	12.3	20.0
	gp59	DNA helicase loader	34.7	34.2	---	33.0	28.9	31.1	28.1	34.2	31.7	32.5	36.4	24.5	22.1	21.7	23.7	22.3	---
	gp32	ssDNA binding protein	47.0	47.0	41.7	46.3	46.6	47.4	43.0	47.2	46.5	45.1	46.5	33.0	27.7	33.3	37.0	32.1	25.5
cluster 6	NrdC*	glutaredoxin	25.3	22.8	24.4	24.4	15.5	26.6	25.6	26.9	26.9	21.8	24.1	34.5	28.9	---	27.2	27.8	---
	Td	thymidylate synthase	6.6	8.2	9.8	---	10.0	7.3	8.4	8.7	9.7	8.6	7.6	46.8	46.0	50.8	39.4	47.1	13.1
	gp61	DNA primase	39.3	38.0	41.7	37.9	40.5	39.7	42.0	40.1	40.9	40.1	37.6	33.1	33.5	32.5	29.6	32.5	29.6
	NrdA	ribonucleotide reductase subunit A	absent in phiM12																
	NrdB	ribonucleotide reductase subunit B	absent in phiM12																
	DexA	exonuclease A	absent in phiM12																

*Only close NrdC orthologs are included in this comparison; **denotes sequences for which two ORFs separated by an intron have been unified.

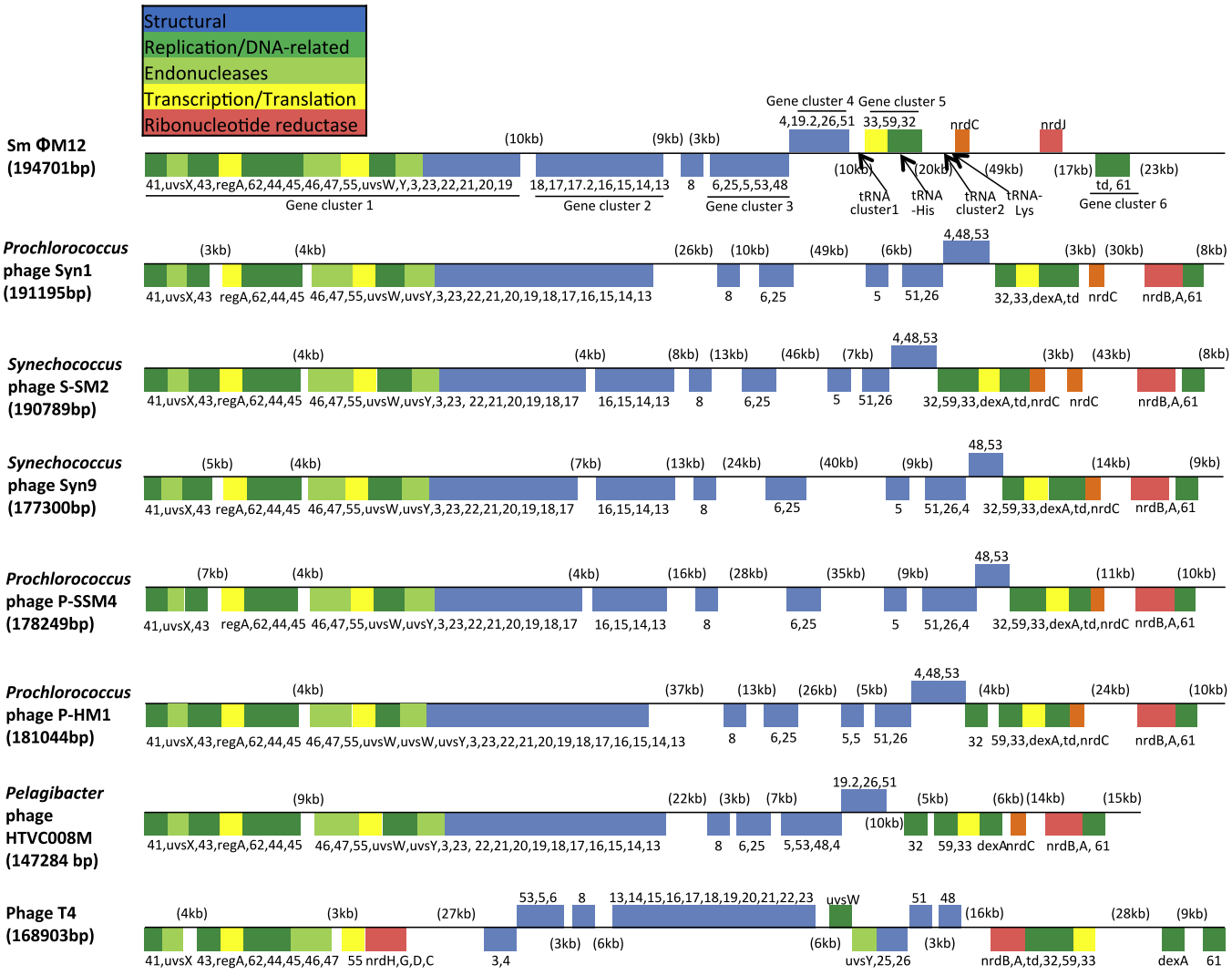

Fig. 2. Gene order of T4 core genes and nearly universal core genes in ΦM12 compared with the most closely-related cyanophages (based on the number of core protein sequence best matches) Syn1; S-SM2; Syn9; P-SSM4; P-HM1; and with *Pelagibacter* phage HTVC008M. The gene order of phage T4 itself is also shown. The location of the tRNA gene clusters is shown for ΦM12, but not for the other phages.

S-ShM2, while the second, truncated copy is more similar to those of P-Syn1 and S-CRM01 (Table 2). It is possible that the partial duplication of ΦM12 gp17 is due to past genetic rearrangements at this locus. The order of the next ΦM12 core genes, gp8, gp6 and gp25 is the same in ΦM12 and most cyanophages, however their spacing suggests that genetic material has been inserted or deleted (Fig. 2, Supplemental Fig. 1).

After gp25, synteny breaks down. In ΦM12, gp6, 25, 5, 53 and 48 are organized in cluster 3, which is more similar to the organization in *Pelagibacter* phage HTVC008M (Fig. 2) and *Aeromonas* phage Aeh1 (Supplemental Fig. 1) than in the cyanophages. Also, the organization of the ΦM12 *nrdC* (glutaredoxin), *td* (thymidylate synthase) and gp61 (DNA primase) is completely different from the cyanophages. Interestingly, this portion of the ΦM12 genome appears to be a mosaic. ΦM12 NrdC and Td proteins are more similar to those of the T4-like phages of enteric bacteria than those of the cyanophages (Table 2). In the cyanophage genomes, the aerobic RNR genes *nrdA* and *nrdB* are located between *nrdC* and gp61. In contrast, in ΦM12, *nrdA* and *nrdB* are absent, and *nrdJ* (class II, B_{12}-dependent RNR) is located between *nrdC* and *td* (Fig. 2). Our analysis suggests that extensive genomic rearrangements have occurred since the ΦM12 lineage separated from the last common ancestor it shares with T4-like cyanophages.

All 10 predicted ΦM12 tRNA genes and pseudogenes are in the region between gp48 and gp61 that appears to have undergone genetic rearrangements (tRNA genes and pseudogenes are summarized in Supplemental Table 2). Five of the 7 ORFs predicted to encode homing endonucleases are also in these less-conserved regions of the genome, and two of them are closely associated with tRNA gene clusters (Supplemental Table 2). It is possible that the tRNA genes and the homing endonucleases of ΦM12 have played a role in the extensive rearrangements in this part of the genome.

This region of the ΦM12 genome also contains a predicted ochre suppressor tRNA (codon TAA) (Supplemental Table 2). If this tRNA is transcribed during host infection and the resulting tRNA can be charged with an amino acid, this suppressor tRNA would be expected to have a greater impact on the translation of ΦM12 proteins than on those of the host *S. meliloti*. Nearly half (42.9%) of the 366 predicted ORFs in the ΦM12 genome have a TAA stop codon, in contrast to only 16.7% of the ORFs encoded on the chromosome of this phage's host *S. meliloti* 1021 (Capela et al., 2001). Several predicted virion structural proteins have a TAA stop

Table 3
ΦM12 predicted proteins detected in the sample proteome by > 3 exclusive unique peptides. (1) All phage structural proteins present at a protein content weight % > 0.30; (2) other phage proteins present at a protein content weight % > 2.5%. (Additional proteins with a protein content weight % > 0.30% are shown in Supplemental Table 3.).

ΦM12 ORF name	T4 ortholog gene product	Predicted MW	Length in amino acids	Exclu-sive 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) x100
1) Structural proteins									
phiM12_065	**gp23, precursor of major head subunit**	46 kDa (40 kDa after cleavage between A57// A58)	434(377 after cleav-age)	37 (35 of peptides between A57-L434)	260 (258 are between A57-L434)	76% of 434 aa (88% after cleavage)	76 (64 after cleav-age)	0.487 (0.547 after cleavage)	7.84% (after cleavage)
phiM12_134	**gp8, baseplate wedge subunit**	65 kDa	594	26	75	41%	94	0.277	4.44%
phiM12_137	**gp6, baseplate wedge subunit**	66 kDa	594	29	69	50%	147	0.197	2.92%
phiM12_135	**gp7, baseplate wedge subunit**	108 kDa	1007	14	25	26%	132	0.106	2.30%
phiM12_066	**gp22, scaffold prohead core**	33 kDa	303	18	56	71%	75	0.240	1.87%
phiM12_071	**gp20, portal vertex protein of head**	62 kDa	540	22	53	40%	156	0.141	1.85%
phiM12_074	**gp19, tail tube protein**	19 kDa	171	10	42	57%	32	0.313	1.57%
phiM12_107	**gp18, tail sheath monomer**	66 kDa	625	13	27	30%	122	0.107	1.41%
phiM12_146	**gp5, baseplate hub**	58 kDa	537	9	14	26%	107	0.084	0.96%
phiM12_161	**gp26, baseplate hub subunit**	27 kDa	234	11	22	52%	75	0.147	0.83%
phiM12_054	**gp3, proximal tip of tail tube**	19 kDa	165	4	6	36%	25	0.160	0.64%
phiM12_142	**gp25, baseplate wedge subunit**	15 kDa	134	7	12	71%	38	0.184	0.63%
phiM12_152	**gp53, baseplate wedge subunit**	23 kDa	190	5	9	35%	45	0.111	0.51%
phiM12_153	**gp48, baseplate tail tube cap**	30 kDa	275	3	6	20%	51	0.059	0.34%
phiM12_117	**gp15, proximal tail sheath stabilization**	31 kDa	268	3	8	15%	57	0.053	0.31%
2) Other abundant phage proteins									
phiM12_129	**VrlC** (see Table 4)	197 kDa	1802	61	184	42%	317	0.192	8.48%
phiM12_124	**gp12-like, short tail fiber** phage tail-collar domain-associated protein (see Table 4)	92 kDa	901	20	67	37%	63	0.317	7.66%
phiM12_398	**conserved hypothetical protein** (contains SGNH-hydrolase domain; has similarity to Rhizo-bium phage RHEph08 protein)	81 kDa	775	27	72	45%	101	0.267	5.34%
phiM12_182	**putative glycanase/laminarinase**	45 kDa	405	23	93	69%	76	0.303	3.51%
phiM12_446	**thymidylate synthase (td)**	35 kDa	306	21	86	68%	67	0.313	2.84%
phiM12_034	**dUTPase**	20 kDa	180	16	82	66%	36	0.444	2.74%
phiM12_229	**hypothetical protein**	12 kDa	105	7	15	81%	12	0.583	2.60%

Table 4

MUSCLE multiple alignment percent identity score of full-length amino acid sequences of ΦM12 "T4-like cyanophage core" proteins, and "T4-like non-cyanophage core" proteins with those of 16 other T4-like phages. ΦM12 has only 4 out of 25 T4-like, cyanophage core proteins (and VrlC), and only 2 out of 30 non-cyanophage core proteins. (See Supplemental Table 1B for sequence accession numbers and references.).

Best match to ΦM12 protein → Poorest match

		←Cyanophages→											--Phages of enteric bacteria→					
gene product	**function**	**S-SM2**	**P-SSM4**	**P-Syn1**	**HTVC008M**	**S-Syn9**	**S-Syn19**	**S-ShM2**	**P-Syn33**	**S-SM1**	**P-HM1**	**S-CRM01**	**T4**	**Aeh**	**44RR**	**KVP40**	**RB43**	**ΦW-14**
Cyanophage core proteins																		
PhoH T4-GC 322	P-starvation inducible protein Sm_phiM12_266	30.0	29.5	31.3	30.0	27.2	29.4	28.8	30.4	29.4	32.7	30.9	---	---	---	18.7	---	17.1
T4-GC 313	hypothetical protein Sm_phiM12_262	24.0	29.1	26.3	26.7	25.6	29.1	29.8	27.9	30.2	21.6	19.0	---	---	---	---	---	---
T4-GC 15 gp27?	virion structural protein Sm_phiM12_150	16.6	15.1	13.8	16.4	17.8	19.2	16.4	18.0	15.3	15.5	---	---	17.6	---	15.4	---	14.5
T4-GC 321	hypothetical protein Sm_phiM12_265	15.9	13.4	16.0	20.7	18.3	19.5	12.2	17.1	14.6	9.8	14.6	---	---	---	---	---	---
VrlC	Predicted structural protein Sm_phiM12_129	16.7	18.0	12.8	21.3	17.5	17.6	17.1	18.8	18.2	18.4	13.9	---	---	---	---	---	19.2
These 21 cyanophage core proteins are not found in ΦM12		PsbA/T4-GC 280; MazG pyrophosphatase/T4-GC 184; Hsp20 heat shock protein/T4-GC170; Hli03 high-light inducible protein/T4-GC 267; CobS porphyrin biosynthetic protein/T4-GC 150; T4-GC 4 hypothetical (hyp) protein; T4-GC 146 hyp. protein; T4-GC 190 hyp. protein; T4-GC 139 hyp. protein; T4-GC101/155 hyp. protein protein; T4-GC 312/443/1092/1149 hyp. protein; T4-GC 176 hyp. protein; T4-GC 201 hyp. protein; T4-GC 49 hyp. protein; T4-GC 71 hyp. protein; T4-GC 112/1330 hyp. protein; T4-GC 142 hyp. protein; T4-GC 152 hyp. protein; T4-GC 250 hyp. protein; T4-GC 198 hyp. protein; T4-GC 43 hyp. protein														
		S-SM2	**P-SSM4**	**P-Syn1**	**HTVC008M**	**S-Syn9**	**S-Syn19**	**S-ShM2**	**P-Syn33**	**S-SM1**	**P-HM1**	**S-CRM01**	**T4**	**Aeh**	**44RR**	**KVP40**	**RB43**	**ΦW-14**
Non-cyanophage core proteins																		
NrdC.11	conserved hypothetical protein Sm_phiM12_318	---	---	---	---	---	---	---	---	---	---	---	26.3	32.3	27.7	27.0	32.3	---
gp12	short tail fiber Sm_phiM12_124	---	---	---	10.8	---	---	---	---	---	---	15.1 12.3	12.7	8.6	9.1	8.9	12.9	---
These 28 non-cyanophage core proteins are not found in ΦM12		dCMP deaminase; Dda DNA helicase; DsbA dsDNA binding protein; gp1 dNMP kinase; gp9; gp10; gp11; gp24; gp30; gp31; gp34; gp49; gp52; gp54; gp60; T4-GC 1491 hyp. protein; T4-GC 1584 hyp. protein; T4-GC 1630 hyp. protein; NrdD; NrdH; PseT; RIIA-RIIB membrane-associated protein; RNaseH; RnlA RNA ligase; Tk thymidine kinase; T4-GC 1557 hyp. protein; T4-GC 1559 hyp. protein; Wac fibritin neck whiskers														

codon. One possibility is that read-through of the TAA stop codon by the ΦM12 ochre suppressor tRNA permits the production of a fraction of these proteins as functional variants.

Four "accessory core" proteins characteristic of T4-like cyanophages are predicted from the ΦM12 genome sequence

Beyond the T4-like phage core gene set, ΦM12 shares very few proteins with either the T4-like cyanophages or the T4-like phages of enteric bacteria. An "accessory core" set of 25 "gene clusters" ("T4-GC"s) that are present in all sequenced T4-like cyanophages, but not in T4-like phages of enteric bacteria, has been defined by Sullivan et al. (2010). ORFs similar to 4 of these gene products are found in the ΦM12 genome (Table 4). These are a truncated and likely nonfunctional version the phosphate-starvation-inducible ATPase PhoH (Kim et al., 1993); a predicted virion protein with structural similarity (Soding, 2005) to the gp27 T4 baseplate hub subunit (Veesler and Cambillau, 2011); and Sm_phiM12_262 and Sm_phiM12_265, ORFs of unknown function. The gp27-like protein and Sm_phiM12_262 are present in the ΦM12 lysate proteome (Supplemental Table 3), while Sm_phiM12_265 and PhoH are not.

Sm_phiM12_129, which encodes a predicted 1802 amino acid phage structural protein VrlC, may also be considered part of the T4-like cyanophage accessory core gene set, because it is in all of the T4-like cyanophages. This ΦM12 ORF is located 10 bp from the ORF encoding the gp8 baseplate wedge protein (Sm_phiM12_134). Homologs of VrlC are not present in T4 or the T4-like phages of enteric bacteria, but are found in *Pelagibacter* phage HTVC008M (Zhao et al., 2013) and in *Deftia* phage ΦW-14 (Kropinski et al., 2010) (Table 4). The proximity of the genes encoding VrlC and gp8 raise the intriguing possibility that VrlC could be part of the baseplate. VrlC is one of the more abundant proteins in the

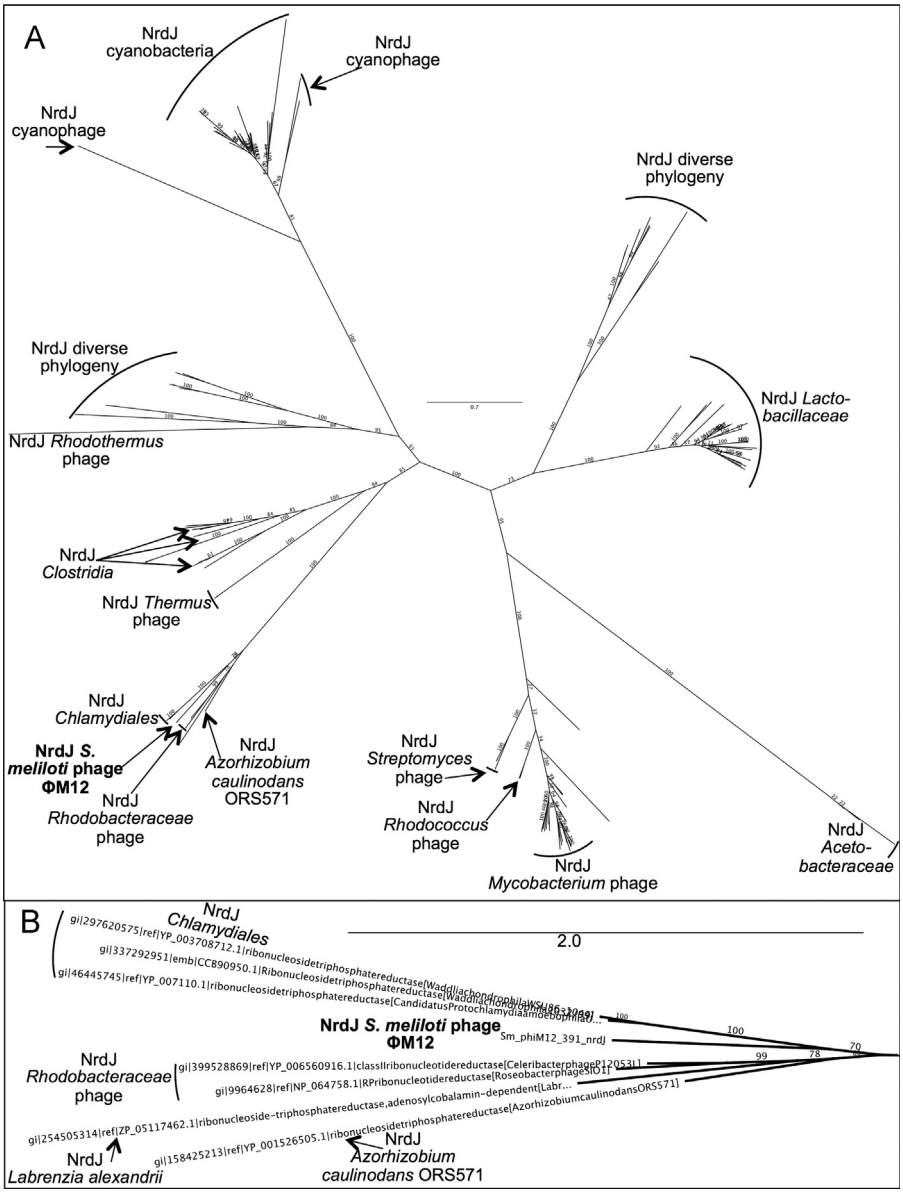

Fig. 3. (A) An unrooted tree generated by phyML from a MUSCLE alignment of the full-length NrdJ RNR protein. The bootstrap percentage value for each branch is shown on the tree and the bar is a marker of branch distance length. (B) A close-up view of the sub-tree containing ΦM12 NrdJ, which is closely related to those from several species of primitive *Chlamydia* (Horn and Wagner, 2004, Bertelli et al., 2010), the dinoflagellate-associated alphaproteobacterium *Labrenzia alexandrii* (Biebl et al., 2007), the rhizobial species *Azorhizobium caulinodans* ORS571 (Lee et al., 2008), the podovirus *Roseobacter* phage SiO1 (Rohwer et al., 2000), and the unclassified dsDNA phage *Celeribacter* phage P12053L (Kang et al., 2012). These sequences are marked with an asterisk (*) in Supplemental Table 1C. Accessions and references are shown in Supplemental Table 1C.

ΦM12 proteome (Table 3). The homolog of this protein in cyanophage Syn9, known as gp105, is also found in purified virions (Weigele et al., 2007).

The other 21 cyanophage accessory core ORFs that are found in all T4-like cyanophages are not present in the ΦM12 genome (Table 4). The absence of these genes from the ΦM12 genome lends additional weight to the hypothesis that this phage is the first sequenced example of a novel lineage of T4-superfamily phages.

Two ΦM12 "accessory core" proteins characteristic of T4-like phages of enteric bacteria.

An accessory core gene set has been described that is present in the T4-like phages of enteric bacteria, but is absent from the T4-like cyanophages (Comeau et al., 2007; Sullivan et al., 2010). The ΦM12 genome has homologs for only two (Table 4). These are NrdC.11 (a protein of unknown function, encoded by Sm_phiM12_318, that is distinct from the NrdC glutaredoxin), and Sm_phiM12_124, an ORF with some similarity to the T4 short tail fiber protein gp12 (Comeau et al., 2007; Sullivan et al., 2010; Kells and Haselkorn, 1974). Both of these proteins are found in the ΦM12 proteome, with the gp12-like protein being one of the more abundant proteins (Table 3).

Unique features of Φm12

Ribonucleotide reductase NrdJ

ΦM12 has a class II coenzyme B_{12}-dependent RNR rather than an aerobic class I or anaerobic class III RNR (Booker and Stubbe, 1993). This ΦM12 gene, *nrdJ*, is predicted to encode a monomeric B_{12}-dependent class II RNR of the TIGR02505 model (Booker and Stubbe, 1993), and this protein is detected in the ΦM12 lysate proteome (Supplemental Table 3). An advantage for organisms carrying a class II RNR is that unlike class I, which requires oxygen, and class III, which is strictly anaerobic, a B_{12}-dependent RNR is insensitive to oxygen levels and can function in a wider range of environments (Lundin et al., 2010). However, phages that have a class II RNR would be expected to be restricted to a host organism that either synthesizes B_{12} (adenosylcobalamin) or can readily acquire this scarce cofactor from the environment. *S. meliloti*, the ΦM12 host bacterium, can synthesize B_{12} (Taga and Walker, 2010; Taga et al., 2007).

Monomeric class II RNR is found in many bacterial lineages, (Sintchak et al., 2002; Lundin et al., 2010; Booker and Stubbe, 1993), and in phages that infect mycobacteria, cyanobacteria, and *Streptomyces* (Lundin et al., 2010). An unrooted phyML tree comparing the amino acid sequence of the ΦM12 monomeric class II RNR with a representative collection of other class II RNRs is shown in Fig. 3. The ΦM12 RNR is closely related to those from several species of primitive *Chlamydia* (Horn and Wagner, 2004; Bertelli et al., 2010), the dinoflagellate-associated alphaproteobacterium *Labrenzia alexandrii* (Biebl et al., 2007), the rhizobial species *Azorhizobium caulinodans* ORS571 (Lee et al., 2008), the podovirus *Roseobacter* phage SiO1 (Rohwer et al., 2000), and the unclassified dsDNA phage *Celeribacter* phage P12053L (Kang et al., 2012). This tree is well-supported by bootstrap analysis (Fig. 3B). The presence of monomeric class II RNR in such disparate bacterial lineages and phages suggests that phage have served a critical role in horizontal transfer of this class of RNR (Lundin et al., 2010). The ΦM12 RNR is not related to the dimeric class II RNR (Taga and Walker, 2010) found in its host bacterium *S. meliloti*.

Other predicted ΦM12 ORFs with similarity to proteins in NCBI databases

In addition to the 42 T4-like phage core proteins shown in Table 2, the amino acid sequences of another 112 ORFs in the ΦM12 genome have BLAST matches (Altschul et al., 1997) to ORFs in the NCBI database with expect values (*E*-values) <0.009. Supplemental Table 4 shows the best match among cellular organisms and phages/viruses of each of these 112 ORFs. (Thirty-six of these 112 ORFs are in the ΦM12 lysate proteome and are denoted with asterisks (*) in Supplemental Table 4.) Ten of these 112 ORFs are found only in alphaproteobacteria (the class that includes ΦM12's host *S. meliloti*) or in alphaproteobacteria and phages.

The ΦM12 genome contains ORFs that may function in interactions with host cell polysaccharides and cell envelope. Sm_phiM12_182 is predicted to encode a glycanase/laminarinase with 21% amino acid sequence identity to the *S. meliloti* ExsH enzyme, which cleaves the exopolysaccharide succinoglycan (York and Walker, 1998). The putative glycanase encoded by Sm_phiM12_182 is one of the most abundant proteins in the phage sample proteome (Table 3). Succinoglycan secreted by *S. meliloti* can form a very thick, gelatinous matrix around these cells. It is possible that a glycanase enzyme released along with phage from lysed host cells would help the phage to penetrate this dense matrix and access the surface of more host cells. This also raises the intriguing possibility that phages have served as an agent of change for the rhizobial polysaccharide repertoire as carriers of glycanase genes.

Another ORF is Sm_phiM12_402, which is predicted to encode a lysozyme. This protein was detected at a low abundance in the proteome (Supplemental Table 5). This is a candidate for an "endolysin", which phage use to attack the cell walls of their host bacteria from within (Oliveira et al., 2013). However, there is not an obvious candidate for a 'holin' ORF, which permeabilizes the inner bacterial membrane and allows the endolysin access to the cell wall (Wang et al., 2000).

Proteome of the Φm12 phage lysate

In order to determine which of the predicted proteins of ΦM12 are expressed during host infection and lysis, we used trypsin proteolysis followed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) to assay the total proteome of the ΦM12 lysate. Of the 97 phage proteins identified in the ΦM12 lysate shotgun mass spectrometry proteome by at least 3 unique, exclusive peptides, the predicted gp23 major capsid protein (Sm_phiM12_065) was confirmed as the most abundant phage protein by protein content weight percent (Table 3). This protein was also identified by N-terminal sequencing of a protein band that runs at ~42 kDa on a denaturing gel of phage sample (Fig. 4). The N-terminal sequence of this band (APTTNTGNIATY) is found at positions 57–68 in the predicted amino acid sequence of gp23. In phage T4, the first 65 amino acids are cleaved off gp23 during capsid maturation (Mesyanzhinov et al., 2004), and it appears that ΦM12 gp23 is processed in a similar fashion.

The ΦM12 homologs of T4 gp20 (portal protein), gp15 (proximal tail sheath stabilization protein), and gp3 (proximal tip of tail tube protein) are found in the ΦM12 proteome (Table 3). Also present in the proteome are the ΦM12 homologs of the T4 tail proteins gp19 (tail tube protein) and gp18 (tail sheath protein), and the T4 baseplate proteins gp8, gp7, gp6, gp25, gp5, gp48, gp53, and gp26 (Table 3).

The ΦM12 protein encoded by Sm_phiM12_124 is abundant in the proteome (Table 3) and is 12.7% identical to the T4 short tail

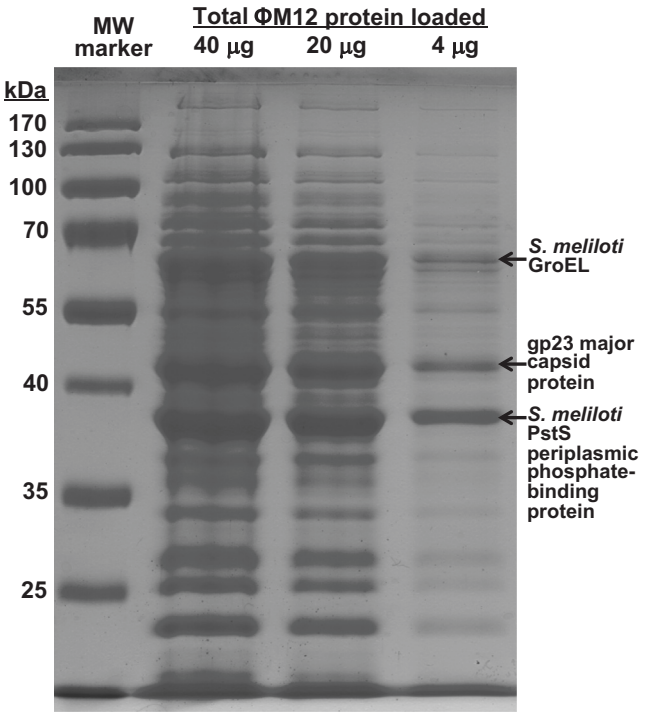

Fig. 4. ΦM12 proteins separated by SDS-PAGE and identified by N-terminal sequencing. Total ΦM12 phage sample was run on a 12% SDS-PAGE Tris-glycine gel, stained with Coomassie Brilliant Blue and imaged. The N-terminal sequence of the proteins in the indicated bands was determined.

fiber protein gp12. In ΦM12, this gp12-like ORF is located near the ORF encoding the gp13 neck protein, as it is in T4 (Miller et al., 2003).

ΦM12 lacks most of the genes that are predicted to encode long tail fibers in the cyanophages (Weigele et al., 2007), or in the T4-like phages of enteric bacteria (Leiman et al., 2010). However one protein, ΦM12 ORF Sm_phiM12_135 has some similarity to BSV9_gp101 (YP_717768.1), one of 6 *Synechococcus* phage Syn9 proteins predicted to be part of the tail fiber (Weigele et al., 2007). Sm_phiM12_135 has 6.1% identity to the T4 gp7-like base-plate wedge initiator protein (and has been labeled as a gp7 protein in Table 2). Sm_phiM12_135 is abundant in the proteome of ΦM12 (Table 3), but it has not yet been determined if it is localized to the baseplate, the tail fiber or another location.

Conclusions

This is the first report of the genome and proteome of a T4-superfamily phage that infects a rhizobial bacterium and only the third report of a T4-superfamily phage that infects an alphaproteobacterium. The genomic and proteomic data reported here for *S. meliloti* phage ΦM12 demonstrate that it is the founder of a novel group of T4-superfamily phages that has features of both T4-like cyanophages and T4-like phages of enteric bacteria, yet is distinct from both of those groups. The ΦM12 genome has a gene order and organization that is unique among T4-superfamily phages. This organization suggests that part of the genome is a mosaic, which has undergone rearrangements and lost the conserved T4-like phage ORFs encoding DexA, NrdA, and NrdB, and a portion of the cyanophage ORF encoding PhoH. It is also the only sequenced T4-superfamily phage reported that has a B_{12}-dependent class II ribonucleotide reductase (NrdJ), which is most closely related to NrdJ sequences from primitive species of *Chlamydia*. ΦM12 also encodes and expresses a predicted glycanase that is similar to the ExsH exopolysaccharide glycanase produced by its host *S. meliloti*, suggesting the intriguing possibility that ΦM12 modulates the extracellular polysaccharide matrix surrounding its host during lytic attack.

Experimental procedures

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 (TY) medium (0.5% tryptone, 0.3% yeast extract, 10 mM $CaCl_2$) supplemented with 500 μg/mL streptomycin. Optimal production of ΦM12 virions was obtained by inoculating 10 μL of phage into 25 mLs of *S. meliloti* 1021 at a density of $OD_{600}=0.2$. The infected culture was incubated at 30 °C overnight or until lysis was apparent, at which point it was centrifuged at 3.8 k × *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 a 1/5 volume of chloroform at 4 °C until further purification. Phage titer was monitored by plaque assays (Finan et al., 1984).

Phage purification and genomic DNA isolation

Chloroform-extracted ΦM12 phage was pelleted by centrifugation at 45 krpm for 6 h in Beckman Ultra-Clear 25 mm × 89 mm tubes in a Ti70 rotor/Beckman Coulter Optima XL-100 K ultracentrifuge (Beckman Coulter, Indianapolis, IN). Phage pellets were resuspended in phage buffer (10 mM sodium phosphate buffer, pH 7; 1 mM $MgSO_4$) plus 0.001% SDS. The resuspended phage were concentrated and washed in a 50 kDa molecular weight cutoff (MWCO) Amicon concentrator (Millipore, Billerica, MA). Concentrated phage were suspended in 10 mL Buffer EX from the Large Construct Kit (Qiagen, Valencia, CA) and treated twice with DNase using 1 unit (80 μg) of ATP-dependent exonuclease (Qiagen) to remove *S. meliloti* genomic DNA. After digestion, particulate matter was removed by passage through a 40 μm polypropylene mesh strainer (BD Biosciences, Franklin Lakes, NJ). Prior to capsid lysis, the ATP was removed to inactivate the exonuclease by washing in a 50 kDa MWCO Amicon concentrator. 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 RNase A (Qiagen).

Phage DNA was digested with HindIII (New England Biolabs, Ipswich, MA) and the DNA-banding pattern was compared to that obtained by Finan et al. (1984). HindIII-digested fragments were cloned into pBluescript KSII+ (Agilent Technologies, Santa Clara, CA), and plasmids containing insert were isolated and submitted for Sanger sequencing using an ABI 3730 Genetic Analyzer (Applied Biosystems, Carlsbad, CA).

SOLiD sequencing of the ΦM12 genome, contig assembly and Sanger re-sequencing

Phage DNA sequencing library preparation was performed with a 5500 SOLiD Fragment Library Kit (Applied Biosystems) at the University of Chicago Comprehensive Cancer Center Genomics Core Facility. A single-end 75 bp sequencing run was performed on one lane of a 5500xl SOLiD System (Applied Biosystems). Initial assembly of nodes was performed with the Velvet *de novo* assembler (Zerbino, 2010), and Velvet nodes and Sanger sequences of pBluescript clones were assembled into contigs with

Sequencher version 4.7 (Gene Codes Corporation, Ann Arbor, MI). Ambiguous sequences and gaps between contigs were resolved by PCR with either Taq or Q5 polymerase (New England Biolabs) anchored in regions of known sequence, followed by Sanger sequencing of the products. The genome sequence between positions 77301 and 194701 has a low density of highly conserved T4-like phage core protein ORFs. Because of this low density of conserved sequences, 54% of the SOLiD genome sequence between positions 77301–194701 was confirmed by PCR amplification and Sanger sequencing of PCR products. There is good agreement between the HindIII, EcoRI, ClaI, and AccI ΦM12 DNA restriction patterns shown in Finan et al. (1984) and the predicted restriction sites in the completed ΦM12 genome.

ORF prediction and analysis

ORFs were predicted using the NCBI ORF Finder (Sayers et al., 2011) and GeneMark.hmm for Prokaryotes (Version 2.8) (Lukashin and Borodovsky, 1998) and Prodigal (Hyatt et al., 2010). tRNAs were predicted using tRNAScan-SE (Lowe and Eddy, 1997). Additional adjustments to ORFs were made based on MyRast (Aziz et al., 2008) and Kodon (Applied Maths, Inc., Austin, TX, USA) (Andrew Kropinski, PhD, personal communication).

Construction of phylogenetic trees based on amino acid sequence

MUSCLE multiple sequence alignments were performed on the sequences in Supplemental Table 1A–C using Geneious (Edgar, 2004, Drummond et al., 2012). The maximum number of iterations selected was 8, with the anchor optimization option. The trees from iterations 1 and 2 were not retained. The distance measure for iteration1 was kmer6_6, and for subsequent iterations was pctid_kimura. The clustering method was UPGMB for all iterations. Unrooted phyML trees were constructed from the MUSCLE alignments using the PhyML plugin within Geneious (Guindon and Gascuel, 2003, Lefort et al., 2012). PhyML was performed with LG amino acid substitution matrix (Le and Gascuel, 2008) with the proportion of invariable sites fixed and 4 substitution rate categories. The fast Nearest-Neighbor Interchange (NNI) tree topology search (Desper and Gascuel, 2002) was used and 100 bootstraps were performed.

Proteomic analysis of phage lysate

ΦM12 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). Phage proteins were separated on 12% SDS-PAGE gels in 1X Tris-glycine buffer (Sambrook and Russell, 2001), stained 60 min with 0.25% Coomassie Brilliant Blue, 10% acetic acid/90% methanol (Sambrook and Russell, 2001), destained 8–16 h and scanned. For N-terminal sequencing of proteins, phage protein gels were electroblotted to Immobilon-P PVDF membrane (Millipore) in a Bio-Rad Mini-Trans-Blot cell (Bio-Rad). Blots were stained for 10 min. with 0.25% Coomassie Brilliant Blue, 10% acetic acid/90% methanol, destained overnight, equilibrated in Milli-Q water, air-dried and photographed. Protein bands were cut from the stained membrane and N-terminal sequence analysis was performed in the Florida State University Department of Biological Science Analytical Laboratory on a Procise 492 CLC using standard sequencing manufacturer protocols (Applied Biosystems, Foster City, CA).

For shotgun proteomics of ΦM12, phage equivalent to 100 μg of total protein was trypsin-digested 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), with some of the modifications in the unpublished FASP protocol for viruses from the M. Sullivan laboratory (Univ. of Arizona) (http://www.eebweb.arizona.edu/faculty/mbsulli/protocols/FASP_Protocol_Viruses.pdf). Frozen phage sample was resuspended in ~7 M urea/Tris-HCl solution (provided with the FASP kit) and reduced with 10 mM DTT (Sigma-Aldrich) for 30 min at 25 °C. The sample was loaded on a 30 kDa MWCO filter (Wisniewski et al., 2011) (FASP kit) and the DTT was washed out with urea/Tris-HCl solution. The cysteines were alkylated for 20 min with iodoacetimide (FASP kit) applied to the filter. The iodoacetimide and urea were washed out and exchanged for 50 mM ammonium bicarbonate buffer, (ABC buffer) (FASP kit). Trypsin from porcine pancreas (Sigma-Aldrich), was resuspended in ABC buffer and applied to the filter at a ratio of 1:20 trypsin:protein sample. After digestion overnight at 37 °C, tryptic peptides were eluted twice with ABC buffer and once with 0.5 M NaCl. The shotgun proteome analysis was performed at the Florida State University Translational Science Laboratory. There the sample was acidified, desalted (ZipTip, Millipore), and analyzed using an externally calibrated Thermo LTQ Orbitrap Velos nLC-ESI-LTQ-Orbitrap (high-resolution electrospray tandem mass spectrometer) operated in data-dependent mode with the following parameters: A 2 cm, 100 μm internal diameter (i.d) trap column (SC001 Easy Column, Thermo-Scientific, San Jose, CA, USA) was followed by a 10 cm analytical column of 75 μm i.d. (SC200 Easy Column, Thermo-Scientific). Both columns had C18-AQ packaging. Separation was performed using Easy nanoLC II (Thermo-Scientific) with a continuous, vented column configuration. A 5 μL (~500 ng) sample was loaded onto the trap from a 20 μL sample loop. The flow rate was 300 nL/min for separation on the analytical column. Mobile phase A was 99.9H_2O (EMD Omni Solvent), with 0.1% formic acid and mobile phase B was 99.9% ACN, with 0.1% formic acid. A 90 min linear gradient from 0% to 45% B was performed. The LC eluent was directly nano-sprayed into an LTQ Orbitrap Velos mass spectrometer (Thermo-Scientific). The MS data were acquired using the following parameters: 10 data-dependent collisional-induced-dissociation (CID) MS/MS scans per full scan. All measurements were performed at room temperature and three technical replicates of each sample were run.

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 KJonesPhage.fasta (6654 entries) (a combined list of predicted ΦM12 ORFs from this study and annotated ORFs from the composite *S. meliloti* 1021 genome (Galibert et al., 2001)).

Scaffold (version Scaffold_4.0.5, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications with greater than 95.0% probability established by the Scaffold Local FDR algorithm were accepted. Protein identifications with greater than 99.9% probability and containing at least 3 identified peptides (FDR 3.7%) were accepted. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii et al., 2003).

Calculation of protein abundance index (PAI) and protein content weight percent

The 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 predicted ΦM12 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 modifications: Oxidation (M) and Carbamidomethyl (C); variable modifications: Phospho (STY); and report multiple charges. The calculation of protein content weight percent for each identified phage protein was emPAI x MW/Σ(emPAI x 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).

Acknowledgments

We thank William J. Buikema at the University of Chicago Comprehensive Cancer Center Genomics Core Facility for help with initial assembly of SOLiD sequences, Margaret Seavy in the Florida State University Department of Biological Science Analytical Laboratory for N-terminal sequencing of proteins, Steve Miller for Sanger sequencing of plasmids and PCR products, and Rakesh Singh and Roger Mercer in the Florida State University Translational Science Laboratory for mass spectrometry/shotgun proteomic analysis. We are very grateful to Andrew Kropinski for his recommendations on ORF predictions. We also thank Peter Beerli, Darin R. Rokyta, Kenneth H. Roux and Brian K. Washburn for critical reading of the manuscript. This work was funded by USDA National Institute of Food and Agriculture, Agriculture and Food Research Initiative award 2010-65108-20582 to K.M.J, and American Heart Association award 10IRG4300065 to M.E.S.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2013.11.027.

References

Adriaenssens, E.M., Van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceyssens, P.J., De Proft, M., Kropinski, A.M., Noben, J.P., Maes, M., Lavigne, R., 2012. T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by 'Dickeya solani'. PLoS One 7, e33227.

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.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Basit, H.A., Angle, J.S., Salem, S., Gewaily, E.M., 1992. Phage coating of soybean seed reduces nodulation by indigenous soil Bradyrhizobia. Can. J. Microbiol. 38, 1264–1269.

Bertelli, C., Collyn, F., Croxatto, A., Ruckert, C., Polkinghorne, A., Kebbi-Beghdadi, C., Goesmann, A., Vaughan, L., Greub, G., 2010. The Waddlia genome: a window into chlamydial biology. PLoS One 5, e10890.

Biebl, H., Pukall, R., Lunsdorf, H., Schulz, S., Allgaier, M., Tindall, B.J., Wagner-Dobler, I., 2007. Description of *Labrenzia alexandrii* gen. nov., sp. nov., a novel alphaproteobacterium containing bacteriochlorophyll a, and a proposal for reclassification of *Stappia aggregata* as *Labrenzia aggregata* comb. nov., of *Stappia marina* as *Labrenzia marina* comb. nov. and of *Stappia alba* as *Labrenzia alba* comb. nov., and emended descriptions of the genera Pannonibacter, Stappia and Roseibium, and of the species *Roseibium denhamense* and *Roseibium hamelinense*. Int. J. Syst. Evol. Microbiol. 57, 1095–1107.

Booker, S., Stubbe, J., 1993. Cloning, sequencing, and expression of the adenosylcobalamin-dependent ribonucleotide reductase from *Lactobacillus leichmannii*. Proc. Nat. Acad. Sci. U.S.A. 90, 8352–8356.

Campbell, G.R., Reuhs, B.L., Walker, G.C., 2002. Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core. Proc. Nat. 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.

Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E., Dreano, S., Gloux, S., Godrie, T., Goffeau, A., Kahn, D., Kiss, E., Lelaure, V., Masuy, D., Pohl, T., Portetelle, D., Pühler, A., Purnelle, F., Ramsperger, U., Renard, C., Thebault, P., Vandenbol, M., Weidner, S., Galibert, F., 2001. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. Proc. Nat. Acad. Sci. U.S.A. 98, 9877–9882.

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; , pp. 1194–1204.

Chen, C.R., Lin, C.H., Lin, J.W., Chang, C.I., Tseng, Y.H., Weng, S.F., 2007. Characterization of a novel T4-type *Stenotrophomonas maltophilia* virulent phage Smp14. Arch. Microbiol. 188, 191–197.

Comeau, A.M., Bertrand, C., Letarov, A., Tetart, F., Krisch, H.M., 2007. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. Virology 362, 384–396.

Comeau, A.M., Krisch, H.M., 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol. Biol. Evol. 25, 1321–1332.

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

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

Dreher, T.W., Brown, N., Bozarth, C.S., Schwartz, A.D., Riscoe, E., Thrash, C., Bennett, S.E., Tzeng, S.C., Maier, C.S., 2011. A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. Environ. Microbiol. 13, 1858–1874.

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

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

Filee, J., Bapteste, E., Susko, E., Krisch, H.M., 2006. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol. Biol. Evol. 23, 1688–1696.

Filee, J., Tetart, F., Suttle, C.A., Krisch, H.M., 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc. Nat. Acad. Sci. U.S.A. 102, 12471–12476.

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.

Fokine, A., Chipman, P.R., Leiman, P.G., Mesyanzhinov, V.V., Rao, V.B., Rossmann, M. G., 2004. Molecular architecture of the prolate head of bacteriophage T4. Proc. Nat. Acad. Sci. U.S.A. 101, 6003–6008.

Fokine, A., Zhang, Z., Kanamaru, S., Bowman, V.D., Aksyuk, A.A., Arisaka, F., Rao, V.B., Rossmann, M.G., 2013. The molecular architecture of the bacteriophage T4 neck. J. Mol. Biol. 425, 1731–1744.

Fuller, N.J., Wilson, W.H., Joint, I.R., Mann, N.H., 1998. Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. Appl. Environ. Microbiol. 64, 2051–2060.

Galibert, F., Finan, T.M., Long, S.R., Pühler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R.W., Dreano, S., Federspiel, N.A., Fisher, R.F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R.W., Jones, T., Kahn, D., Kahn, M.L., Kalman, S., Keating, D.H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M.C., Pohl, T.M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorhölter, F.J., Weidner, S., Wells, D.H., Wong, K., Yeh, K.C., Batut, J., 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science 293, 668–672.

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.

Gibson, K.E., Kobayashi, H., Walker, G.C., 2008. Molecular determinants of a symbiotic chronic infection. Annu. Rev. Genet. 42, 413–441.

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

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

Hatfull, G.F., Hendrix, R.W., 2011. Bacteriophages and their genomes. Curr. Opin. Virol. 1, 298–303.

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

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. In.:. National Center for

Biotechnology Information: The Broad Institute Genome Sequencing Platform. (NC_021557).
Horn, M., Wagner, M., 2004. Bacterial endosymbionts of free-living amoebae. J. Eukaryotic Microbiol. 51, 509–514.
Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. 11, 119.
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.
Jameson, E., Mann, N.H., Joint, I., Sambles, C., Muhling, M., 2011. The diversity of cyanomyovirus populations along a North-South Atlantic Ocean transect. ISME J. 5, 1713–1721.
Javed, M.A., Ackermann, H.W., Azeredo, J., Carvalho, C.M., Connerton, I., Evoy, S., Hammerl, J.A., Hertwig, S., Lavigne, R., Singh, A., Szymanski, C.M., Timms, A., Kropinski, A.M., 2013. A suggested classification for two groups of *Campylobacter myoviruses*. Arch. Virol..
Jones, J.B., Jackson, L.E., Balogh, B., Obradovic, A., Iriarte, F.B., Momol, M.T., 2007a. Bacteriophages for plant disease control. Annu. Rev. Phytopathol. 45, 245–262.
Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E., Walker, G.C., 2007b. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. Nat. Rev. Microbiol. 5, 619–633.
Kang, I., Jang, H., Oh, H.M., Cho, J.C., 2012. Complete genome sequence of Celeribacter bacteriophage P12053L. J. Virol. 86, 8339–8340.
Kells, S.S., Haselkorn, R., 1974. Bacteriophage T4 short tail fibers are the product of gene 12. J. Mol. Biol. 83, 473–485.
Kim, S.K., Makino, K., Amemura, M., Shinagawa, H., Nakata, A., 1993. Molecular analysis of the phoH gene, belonging to the phosphate regulon in *Escherichia coli*. J. Bacteriol. 175, 1316–1324.
Kropinski, A.M., A. Villegas, E.J. Lingohr, (2010) Deftia Phage phiW-14, Complete Genome, NC_013697.1. In: National Center for Biotechnology Information, pp.
Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25, 1307–1320.
Lee, K.B., De Backer, P., Aono, T., Liu, C.T., Suzuki, S., Suzuki, T., Kaneko, T., Yamada, M., Tabata, S., Kupfer, D.M., Najar, F.Z., Wiley, G.B., Roe, B., Binnewies, T.T., Ussery, D.W., D'Haeze, W., Herder, J.D., Gevers, D., Vereecke, D., Holsters, M., Oyaizu, H., 2008. The genome of the versatile nitrogen fixer *Azorhizobium caulinodans* ORS571. BMC Genomics 9, 271.
Lefort, V., J. Heled, S. Guindon, Biomatters, (2012) Geneious 5.6.6, PhyML plugin.
Leiman, P.G., Arisaka, F., van Raaij, M.J., Kostyuchenko, V.A., Aksyuk, A.A., Kanamaru, S., Rossmann, M.G., 2010. Morphogenesis of the T4 tail and tail fibers. Virol. J. 7, 355.
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.
Lundin, D., Gribaldo, S., Torrents, E., Sjoberg, B.M., Poole, A.M., 2010. Ribonucleotide reduction—horizontal transfer of a required function spans all three domains. BMC Evol. Biol. 10, 383.
Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A., Krisch, H.M., 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine Synechococcus strains. J. Bacteriol. 187, 3188–3200.
Martin, M.O., Long, S.R., 1984. Generalized transduction in *Rhizobium meliloti*. J. Bacteriol. 159, 125–129.
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.
Mesyanzhinov, V.V., Leiman, P.G., Kostyuchenko, V.A., Kurochkina, L.P., Miroshnikov, K.A., Sykilinda, N.N., Shneider, M.M., 2004. Molecular architecture of bacteriophage T4. Biochemistry (Mosc) 69, 1190–1202.
Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156. (table of contents).
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.
Oliveira, H., Melo, L.D., Santos, S.B., Nobrega, F.L., Ferreira, E.C., Cerca, N., Azeredo, J., Kluskens, L.D., 2013. Molecular aspects and comparative genomics of bacteriophage endolysins. J. Virol. 87, 4558–4570.
Petrov, V.M., Ratnayaka, S., Nolan, J.M., Miller, E.S., Karam, J.D., 2010. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. Virol. J. 7, 292.
Pope, W.H., J. Hua, G.F. Hatfull,R.W. Hendrix, (2012) Sequence of the Genome of Sphingomonas Phage PAU. In: National Center for Biotechnology Information, pp.
Rappsilber, J., Ryder, U., Lamond, A.I., Mann, M., 2002. Large-scale proteomic analysis of the human spliceosome. Genome Res. 12, 1231–1245.
Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F., Azam, F., 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. Limnol. Oceanogr. 45, 408–418.
Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual, third ed. 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., 2013. Narrow host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. Appl. Environ. Microbiol.
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–D51.
Schulmeister, S.A., J.E. Krol, F.-J. Vorhoelter, A.M. Skorupska,W. Lotz, (2009) Sequence of the Genome of *Sinorhizobium meliloti* Bacteriophage PBC5, NC_003324.1. In: National Center for Biotechnology Information, pp.
Short, C.M., Suttle, C.A., 2005. Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Appl. Environ. Microbiol. 71, 480–486.
Sintchak, M.D., Arjara, G., Kellogg, B.A., Stubbe, J., Drennan, C.L., 2002. The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. Nat. Struct. Biol. 9, 293–300.
Soding, J., 2005. Protein homology detection by HMM-HMM comparison. Bioinformatics 21, 951–960.
Stroupe, M.E., T. E. Brewer, D. Sousa, K. M. Jones, The structure of Sinorhizobium meliloti phage ΦM12, which has a novel T=19ltriangulation number and is the founder of a new group of T4-superfamily phages, Virology http://dx.doi.org/10.1016/j.virol.2013.11.027, in press.
Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., Defrancesco, A.S., Tan, G., Fu, R., Lee, J.A., Waterbury, J.B., Bielawski, J.P., Chisholm, S.W., 2008. Portal protein diversity and phage ecology. Environ. Microbiol. 10, 2810–2823.
Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., Chisholm, S.W., 2005. Three *Prochlorococcus cyanophage* genomes: signature features and ecological interpretations. PLoS Biol. 3, e144.
Sullivan, M.B., Huang, K.H., Ignacio-Espinoza, J.C., Berlin, A.M., Kelly, L., Weigele, P. R., DeFrancesco, A.S., Kern, S.E., Thompson, L.R., Young, S., Yandava, C., Fu, R., Krastins, B., Chase, M., Sarracino, D., Osburne, M.S., Henn, M.R., Chisholm, S.W., 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ. Microbiol. 12, 3035–3056.
Taga, M.E., Larsen, N.A., Howard-Jones, A.R., Walsh, C.T., Walker, G.C., 2007. BluB cannibalizes flavin to form the lower ligand of vitamin B12. Nature 446, 449–453.
Taga, M.E., Walker, G.C., 2010. *Sinorhizobium meliloti* requires a cobalamin-dependent ribonucleotide reductase for symbiosis with its plant host. Mol. Plant Microbe Interact. 23, 1643–1654.
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).
Wang, G.H., Asakawa, S., Kimura, M., 2011. Spatial and temporal changes of cyanophage communities in paddy field soils as revealed by the capsid assembly protein gene g20. Fems Microbiol. Ecol. 76, 352–359.
Wang, G.H., Murase, J., Asakawa, S., Kimura, M., 2010. Unique viral capsid assembly protein gene (g20) of cyanophages in the floodwater of a Japanese paddy field. Biol. Fert. Soils 46, 93–102.
Wang, I.N., Smith, D.L., Young, R., 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. 54, 799–825.
Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., King, J., Hatfull, G.F., Lawrence, J.G., Hendrix, R.W., 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine Prochlorococcus and Synechococcus. Environ. Microbiol. 9, 1675–1695.
Wilhelm, S.W., Carberry, M.J., Eldridge, M.L., Poorvin, L., Saxton, M.A., Doblin, M.A., 2006. Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of g20 genes. Appl. Environ. Microbiol. 72, 4957–4963.
Wisniewski, J.R., Zielinska, D.F., Mann, M., 2011. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. Anal. Biochem. 410, 307–309.
Wisniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nat. Methods 6, 359–362.
York, G.M., Walker, G.C., 1998. The *Rhizobium meliloti* ExoK and ExsH glycanases specifically depolymerize nascent succinoglycan chains. Proc. Nat. Acad. Sci. U. S.A. 95, 4912–4917.
Zerbino, D.R., (2010) Using the Velvet de Novo Assembler for Short-read Sequencing Technologies. Current Protocols in Bioinformatics/Editoral Board, Andreas D. Baxevanis… [et al.] Chapter 11: Unit 11 15.
Zhao, Y., Temperton, B., Thrash, J.C., Schwalbach, M.S., Vergin, K.L., Landry, Z.C., Ellisman, M., Deerinck, T., Sullivan, M.B., Giovannoni, S.J., 2013. Abundant SAR11 viruses in the ocean. Nature 494, 357–360.
Zhong, Y., Chen, F., Wilhelm, S.W., Poorvin, L., Hodson, R.E., 2002. Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. Appl. Environ. Microbiol. 68, 1576–1584.