

# Complete nucleotide sequence of a P2 family lysogenic bacteriophage, $\phi$ MhaA1-PHL101, from *Mannheimia haemolytica* serotype A1

Sarah K. Highlander<sup>a,c,\*</sup>, Sarah Weissenberger<sup>a</sup>, Laura E. Alvarez<sup>a</sup>,  
George M. Weinstock<sup>a,b,c</sup>, Peter B. Berget<sup>d</sup>

<sup>a</sup> Department of Molecular Virology and Microbiology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

<sup>b</sup> Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

<sup>c</sup> Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

<sup>d</sup> Department of Biological Sciences, Carnegie Mellon University, 4400 5th Avenue, Pittsburgh, PA 15213, USA

Received 29 November 2005; returned to author for revision 11 March 2006; accepted 15 March 2006

Available online 2 May 2006

## Abstract

The 34,525 nucleotide sequence of a double-stranded DNA bacteriophage ( $\phi$ MhaA1-PHL101) from *Mannheimia haemolytica* serotype A1 has been determined. The phage encodes 50 open reading frames. Twenty-three of the proteins are similar to proteins of the P2 family of phages. Other protein sequences are most similar to possible prophage sequences from the draft genome of *Histophilus somni* 2336. Fourteen open reading frames encode proteins with no known homolog. The P2 orthologues are collinear in  $\phi$ MhaA1-PHL101, with the exception of the phage tail protein gene T, which maps in a unique location between the S and V genes. The phage ORFs can be arranged into 17 possible transcriptional units and many of the genes are predicted to be translationally coupled. Southern blot analysis revealed  $\phi$ MhaA1-PHL101 sequences in other A1 isolates as well as in serotype A5, A6, A9, and A12 strains of *M. haemolytica*, but not in the related organisms, *Mannheimia glucosida* or *Pasteurella trehalosi*.

© 2006 Elsevier Inc. All rights reserved.

**Keywords:** Bacteriophage; Genomic sequence; *Mannheimia haemolytica*

## Introduction

The family Pasteurellaceae includes the *Haemophilus*, *Actinobacillus*, *Pasteurella*, and *Mannheimia* genera of bacteria, which cause a variety of diseases in humans and animals. Bacteriophages have been reported in *Haemophilus influenzae* (Bendler and Goodgal, 1968; Harm and Rupert, 1963; Hendrix et al., 1990; Samuels and Clarke, 1969; Stuy, 1978; Williams et al., 2002), *Actinobacillus actinomycetemcomitans* (Loftus and Delisle, 1995; Stevens et al., 1982), *Pasteurella multocida* (Ackermann and Karaivanov, 1984; Gadberry and Miller, 1978; Pullinger et al., 2004), and *Mannheimia haemolytica* (Rifkind and Pickett, 1954; Saxena and Hoerlein, 1959) and bacteriophage-like sequences have been identified in *Histophilus somni*

(previously *Haemophilus somnus*) (Pontarollo et al., 1997). The genomes of two *H. influenzae* phages, HP1 and HP2, have been sequenced and both are members of the P2 family of temperate bacteriophage (Esposito et al., 1996; Williams et al., 2002). Recently, the complete genome of a lambdoid temperate bacteriophage from *A. actinomycetemcomitans* was reported (Resch et al., 2004). No other complete bacteriophage genome sequence has been reported for the Pasteurellaceae, though a Mu-like prophage,  $\phi$ flu, was identified within the genome of *H. influenzae* Rd (Fleischmann et al., 1995; Morgan et al., 2002) and HP1-like sequences were reported in *H. somni* (Pontarollo et al., 1997).

*M. haemolytica* is the primary bacterial pathogen in bovine respiratory disease complex (Whiteley et al., 1992), yet is also a commensal in ruminants including sheep and goats. Bacteriophage was first isolated from *M. haemolytica* in the 1950s (Rifkind and Pickett, 1954; Saxena and Hoerlein, 1959). In 1985, Richards, Renshaw and Sneed isolated an icosahedral bacteriophage from UV-treated isolates of *M. haemolytica*

\* Corresponding author. Department of Molecular Virology and Microbiology, Baylor College of Medicine, One Baylor Plaza, MS280 Houston, TX 77030-3411, USA. Fax: +1 713 798 7375.

E-mail address: [sarahh@bcm.tmc.edu](mailto:sarahh@bcm.tmc.edu) (S.K. Highlander).

biotype A, serotype 1 (A1) (Richards et al., 1985). The phage, called  $\phi$ PhaA1, was classified as a member of Bradley's group A (Bradley, 1967) and had a tail with a contractile sheath.  $\phi$ PhaA1 was unable to form plaques on any of the A1 strains tested leading the authors to speculate that all strains harbored the same lysogenic bacteriophage (Richards et al., 1985). The genome was not characterized but was assumed to be composed of double-stranded DNA. Froshauer et al. (1996) also showed that danofloxacin could induce a prophage in a serotype A1 isolate of *M. haemolytica*.

Here we report the complete DNA sequence of the double-stranded DNA bacteriophage,  $\phi$ MhaA1-PHL101, isolated from a bovine pneumonic isolate of *M. haemolytica*, A1, strain PHL101. This may be related to the bacteriophage originally isolated by Richards.

## Results and discussion

### Genome organization

The PHL101  $\phi$ MhaA1-PHL101 genome is 34,525 bp in length and has a 7 bp repeated sequence on each end. These repeats were identified by direct sequencing off the ends of the phage DNA. The terminal repeats form the *cos* or cohesive end site and have the sequence 5'-TGCGGGC. Fifty open reading frames were predicted from the sequence (Fig. 1 and Table 1). These were categorized as regulatory, structural, and morphogenic (assembly and packaging) or unknown (Fig. 1).

The genes encoding structural and assembly components of  $\phi$ MhaA1-PHL101 are clustered on the left end of DNA sequence (*orf1-orf30*). With the exception of *orf15*, this arrangement is identical to the organization of morphogenesis genes (*gpD* to *gpQ*) of bacteriophage P2 (Christie et al., 1998).

In  $\phi$ MhaA1-PHL101, the T-like gene lies between the hypothetical genes *orf14* and *orf15*. This is unlike the P2 genome where the T gene lies between the E and U genes. A *tblastx* comparison of the  $\phi$ MhaA1 and P2 genomes, visualized using the Artemis Comparison Tool (Carver et al., 2005), clearly shows the rearrangement of the T gene (Fig. 2). This organization was verified by repeated sequence reads across the region and by PCR performed using phage or chromosomal DNA and primers specific for the S and V genes (data not shown). The Artemis plot also illustrates protein similarities for the Q, P, O, N, M, L, R, S, V, J, I, H, FI, FII, E, U, D, and A proteins. Note that a gene encoding an ortholog of the tail fiber protein G, which should map between *orf22* and *orf25* (H and FI), is not present in  $\phi$ MhaA1-PHL101. The P2 tail protein gene can undergo a frameshift to extend the E protein to a longer E + E' protein with an identical amino terminus. This is thought to occur via translational frameshifting in a "slippery" hexa-T sequence upstream of the first stop codon (Christie et al., 2002). In  $\phi$ MhaA1-PHL101, a hexa-T sequence is not observed within the gene. However, using the Programmed Frameshift Finder (chainmail.bio.pitt.edu/~junxu/webshift.html; Xu et al., 2004a), a GGGAAAG slippery sequence was identified upstream of the stop codon. Addition of a base within this sequence would extend the E protein from 104 to 151 amino acids. Also unlike P2,  $\phi$ MhaA1-PHL101 encodes a P1-like signal-arrest-release (SAR) endolysin (Xu et al., 2004b) instead of the typical holin-endolysin pair. Similar SAR endolysins are found in HP1 and HP2 (Esposito et al., 1996; Williams et al., 2002) and in *H. somni* (Pontarollo et al., 1997). The adjacent predicted holin of  $\phi$ MhaA1-PHL101 is only weakly similar (e-05 over less than 80% of the protein) to a predicted holin protein in *H. somni*. Finally, the products of *orf1-3*, *orf5*, *orf6*, *orf8*, *orf9*, and *orf43* are also similar to corresponding

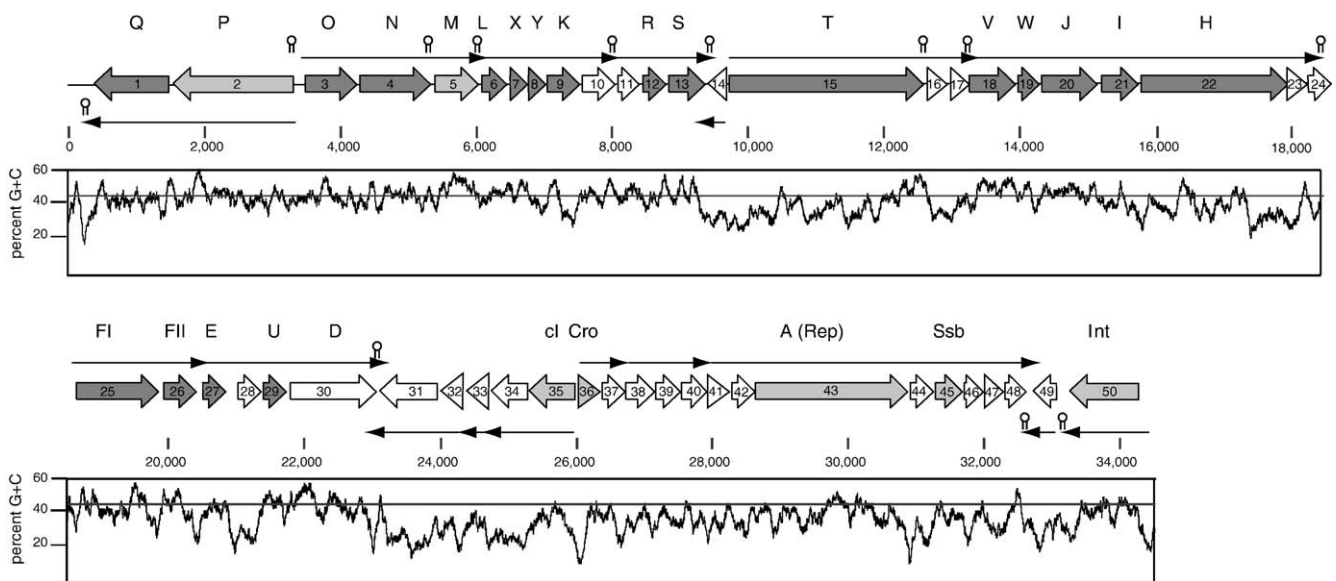


Fig. 1. Genome map of phage  $\phi$ MhaA1-PHL101. Genes are sequentially numbered from 1 to 50. Potential regulatory genes are indicated with light shading and morphogenesis genes are shown with dark shading. Unshaded arrows indicate hypothetical proteins. Protein designations are indicated by capital letters. Predicted transcriptional units are shown as arrows and terminators by the stem-loop structures. A %GC map, generated using a 100 nt sliding window, is shown below the gene map.

Table 1  
Bacteriophage  $\phi$ MhaA1-PHL101 ORF analysis

ORF	Range (bp)	Size (aa)	Predicted function	Top BLASTP match (organism)	Reference (gi number)	BLASTP e value	<i>H. somni</i> 2336 ortholog (gi number)	BLASTP e value
ORF1	1468–428	346	Capsid portal protein Q	Capsid portal protein ( <i>H. influenzae</i> R2866)	46133084	8e–117	53728326	2e–46
ORF2	3294–1477	605	Terminase, ATPase subunit	Terminase, ATPase subunit ( <i>H. influenzae</i> 86-028NP)	38058394	0e+00	53728325	1e–102
ORF3	3429–4256	275	Capsid scaffolding protein O	Chromosome segregation ATPase ( <i>H. influenzae</i> R2886)	42631022	1e–72	32029055	2e–70
ORF4	4270–5298	342	Major capsid protein N	gp5 ( <i>Salmonella</i> bacteriophage PSP3)	41057357	4e–101	32029056	3e–91
ORF5	5308–5997	229	Terminase, small subunit	Terminase, endonuclease subunit ( <i>H. influenzae</i> 86-028NP)	68058391	1e–45	32029057	2e–38
ORF6	6109–6624	171	Head completion protein L	orf4, phage P2-like head completion protein L ( <i>H. somni</i> )	915368	1e–46	53728618	3e–45
ORF7	6621–6833	70	Tail synthesis protein X	Phage P2-like tail protein X ( <i>H. somni</i> 2336)	32029059	8e–13	32029059	8e–13
ORF8	6839–7045	68	Possible holin	Possible holin ( <i>H. somni</i> )	915370 <sup>a</sup>	9e–06	32029060 <sup>a</sup>	9e–06
ORF9	7038–7604	188	Lysozyme	Lysozyme ( <i>H. somni</i> 2336)	53728269	9e–46	53728269	9e–46
ORF10	7601–8056	151	Conserved hypothetical phage protein	orf8 ( <i>H. somni</i> )	915373 <sup>a</sup>	2e–07	53728620 <sup>a</sup>	2e–07
ORF11	8205–8426	73	Conserved hypothetical protein	DnaK ( <i>H. influenzae</i> R2866)	42631014	4e–12		
ORF12	8423–8908	161	Tail completion protein R	Hypothetical protein ( <i>Azotobacter vinelandii</i> )	67159457	1e–24	32029063 <sup>a</sup>	5e–24
ORF13	8901–9359	152	Tail synthesis protein S	orf11 ( <i>H. somni</i> )	915376	8e–42	32029064	1e–41
ORF14	9688–9410	92	Conserved hypothetical protein	Hypothetical protein NTHI1870 ( <i>H. influenzae</i> 86-028NP)	68058378	6e–02	None	
ORF15	9727–12,639	970	Tail protein T	Tail protein ( <i>H. influenzae</i> R2866)	46133075	1e–80	53728622	2e–46
ORF16	12,704–12,880	58	Hypothetical protein	None			None	
ORF17	12,873–13,160	95	Conserved hypothetical protein	Plasmid stabilization protein ( <i>A. pleuropneumoniae</i> )	53728835	5e–10	None	
ORF18	13,289–13,894	201	Baseplate assembly protein V	Baseplate assembly protein V ( <i>H. influenzae</i> R2866)	46133073	3e–61	32029067	8e–41
ORF19	13,894–14,229	111	Baseplate assembly protein W	Baseplate assembly protein W ( <i>H. influenzae</i> 86-028NP)	68058370 <sup>a</sup>	1e–19	53728624 <sup>a</sup>	2e–19
ORF20	14,226–15,143	305	Baseplate assembly protein J	Baseplate assembly protein ( <i>H. somni</i> )	32029069	8e–102	32029069	8e–102
ORF21	15,130–15,762	210	Tail formation protein I	Tail formation protein ( <i>H. somni</i> 2336)	53728625 <sup>a</sup>	7e–52	53728625 <sup>a</sup>	7e–52
ORF22	15,765–18,044	759	Variable tail fiber protein H	Tail fiber protein ( <i>H. influenzae</i> R2866)	46133071 <sup>a</sup>	8e–40	53728626 <sup>a</sup>	1e–38
ORF23	18,045–18,287	80	Hypothetical protein	None			None	
ORF24	18,392–18,562	857	Pseudogene of conserved hypothetical protein	Hypothetical protein (bacteriophage Aaphi23)	45737848 <sup>a</sup>	1e–09	32028843 <sup>a</sup>	1e–06
ORF25	18,669–19,850	393	Tail sheath protein FI	Tail sheath protein FI ( <i>H. somni</i> 2336)	32030804	1e–103	32030804	1e–103
ORF26	19,859–20,365	168	Tail tube protein FII	Tail tube protein F11 ( <i>H. influenzae</i> R2866)	46133067	1e–52	32030805	2e–47
ORF27	20,444–20,758	104	Tail protein E	Hypothetical protein ( <i>Photobacterium luminescens</i> )	37524045 <sup>a</sup>	1e–11	32030806 <sup>a</sup>	4e–08
ORF28	20,962–21,309	115	Hypothetical protein	None			None	
ORF29	21,311–21,748	145	Tail protein U	Phage protein U ( <i>H. influenzae</i> R2866)	46133065	2e–41	46155983	7e–41
ORF30	21,748–22,986	412	Tail protein D	Phage protein D ( <i>H. influenzae</i> R2866)	46133064	1e–107	46155984	6e–105
ORF31	23,169–23,978	269	Conserved hypothetical phage protein	Hypothetical protein ( <i>Staph. haemolyticus</i> JCSC1435)	70726007 <sup>a</sup>	3e–14	None	
ORF32	24,010–24,270	86	Conserved hypothetical protein	Hypothetical protein ( <i>H. influenzae</i> R2866)	42630997 <sup>a</sup>	4e–10	None	
ORF33	24,370–24,699	111	Hypothetical protein	None			None	
ORF34	24,798–25,316	172	Conserved hypothetical protein	NAD-dependent DNA ligase ( <i>H. influenzae</i> R2866)	42630996	2e–34	None	
ORF35	25,320–26,006	228	CI family phage repressor	Transcriptional regulator ( <i>E. coli</i> F11)	75236053 <sup>a</sup>	3e–35	46155985 <sup>a</sup>	3e–34
ORF36	26,130–26,342	70	Cro family repressor	Cro ( <i>Salmonella typhimurium</i> bacteriophage ST104)	46358672	1e–13	46155986 <sup>a</sup>	1e–09
ORF37	26,326–26,610	94	Hypothetical protein	None				
ORF38	26,819–27,091	90	Conserved hypothetical protein	Hypothetical protein ( <i>H. somni</i> 2336)	46155987	6e–18	46155987	6e–18

(continued on next page)

Table 1 (continued)

ORF	Range (bp)	Size (aa)	Predicted function	Top BLASTP match (organism)	Reference (gi number)	BLASTP e value	<i>H. somni</i> 2336 ortholog (gi number)	BLASTP e value
ORF39	27,174–27,506	110	Conserved hypothetical protein	Chromosome segregation ATPase ( <i>H. somni</i> 2336)	46155988 <sup>a</sup>	7e–03	46155988 <sup>a</sup>	7e–03
ORF40	27,519–27,812	97	Hypothetical protein	None			None	
ORF41	27,965–28,207	80	Hypothetical protein	None			None	
ORF42	28,204–28,536	110	Hypothetical protein	None			None	
ORF43	28,533–30,893	786	Replication protein	Hypothetical protein ( <i>H. somni</i> 2336)	32030813	1e–101	32030813	1e–101
ORF44	30,906–31,238	110	Hypothetical protein	None			None	
ORF45	31,238–31,690	150	Single-stranded DNA binding protein	Single-stranded DNA-binding protein ( <i>H. somni</i> 129T)	23467571	2e–34	32029356	4e–334
ORF46	31,770–32,033	87	Conserved hypothetical phage protein	gp62 ( <i>Burkholderia cepacia</i> phage Bcep43)	41057713 <sup>a</sup>	2e–10	None	
ORF47	32,023–32,292	89	Hypothetical protein	None			None	
ORF48	32,267–32,557	96	Conserved hypothetical protein	ABC transport system, binding protein ( <i>H. somni</i> 129T)	23466938	1e–11	46155989	2e–11
ORF49	33,009–32,827	60	Conserved hypothetical protein	Hypothetical protein ( <i>H. somni</i> 2336)	32030618	6e–11	32030618	6e–11
ORF50	34,285–33,287	332	Integrase/recombinase	Integrase/recombinase ( <i>H. influenzae</i> )	1175903	5e–105	46155990 <sup>a</sup>	2e–69

<sup>a</sup> BLAST hit includes <90% of the query.

P2-like proteins in the *Haemophilus* phages, HP1 and HP2 (Esposito et al., 1996; Williams et al., 2002).

#### DNA replication

The phage contains two genes that are predicted to be involved in replication: *orf43*, which encodes the RepA protein and *orf45*, encoding a single-stranded DNA binding protein. The  $\phi$ MhaA1-PHL101 RepA protein (ca. 90 kDa) is similar to the 86.3 kDa bacteriophage P2 A protein (Liu et al., 1993), which nicks the DNA template at an origin of replication within the A gene coding sequence and initiates rolling-circle replication of the phage (Schnos and Inman, 1971). Though the  $\phi$ MhaA1-PHL101 RepA shares sequence similarity with the P2 A protein, it is most similar to a conserved hypothetical protein from *H. somni* and to the replication proteins encoded by the *H. influenzae* bacteriophages HP1 and HP2 (Esposito et al., 1996; Williams et al., 2002). All four Pasteurellaceae sequences contain a similar 24 bp sequence near the carboxy terminus of the protein (Supplementary Fig. 1). Based on this

conservation, its similarity to the P2 origin, and relative location within the coding region (Liu and Haggard-Ljungquist, 1996), we propose that this 24 bp sequence contains the origin of rolling circle replication for phages  $\phi$ MhaA1-PHL101, HP1, and HP2 (Supplementary Fig. 1a). Rolling circle replication proteins, such as the P2 A protein, contain three conserved amino acid domains, identified by Ilyina and Koonin (1992). The first domain is of unknown function, a second domain, with the motif HUHUU (where U is a hydrophobic residue), may be involved in binding of Mn<sup>2+</sup> or Mg<sup>2+</sup>, and a third domain contains two tyrosine residues, one of which is required for cleavage and ligation (Liu and Haggard-Ljungquist, 1996). The replication proteins of  $\phi$ MhaA1-PHL101, P2, HP1, HP2, and the *H. somni* sequence all contain these conserved motifs (Supplementary Fig. 1b), providing more evidence of their role in replication initiation. P2 phages do not encode a single-stranded binding protein, as does  $\phi$ MhaA1-PHL101, though it has been shown that the A protein binds to and nicks only single-stranded DNA (Liu and Haggard-Ljungquist, 1994). If the  $\phi$ MhaA1-PHL101 RepA has a similar requirement for

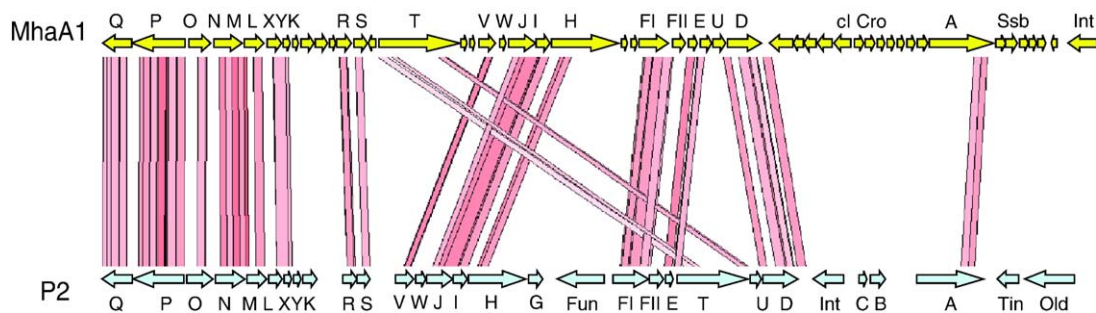


Fig. 2. Artemis Comparison Tool (ACT) plot (Carver et al., 2005) showing tblastx matches between  $\phi$ MhaA1-PHL101 (top) and P2 (bottom). Start and stop codons are indicated by the short vertical hash marks above and below the coordinates for each genome. The predicted open reading frames are shown as yellow ( $\phi$ MhaA1-PHL101) or blue (P2) arrows. Red to pink lines indicate regions of amino acid homology using a 200 bp window; darker colors indicate greater sequence identity. The locations of the T genes are marked by arrows.

single-stranded DNA at the origin, then it is possible that the single-stranded DNA binding protein encoded by this phage can stabilize an open structure for RepA action.

### Transcriptional organization

Based on the arrangement of open reading frames and using searches for potential *Escherichia coli*-like sigma 70-like promoter sequences and rho-independent terminator structures, we predict that the phage genome could be organized in up to 17 transcription units (Fig. 1 and Supplementary Figs. 1 and 2). This is probably an overestimate since some genes are likely to be transcribed by read-through of predicted terminators. Some of transcription units include only a single gene (e.g. *orf1*, *orf49*, and *orf50*), while others can include up to eight genes (e.g. *orf41* through *orf48*) within one operon. A sigma 70-like promoter could not be identified 5' to *orf16* so *orf16* and *orf17* may be transcribed by alternative sigma factors, may require accessory proteins, or may be transcribed by anti-termination of the upstream terminator. In bacteriophage P2, transcription of the late genes PQ, ONMLXYKRS, VJHG and FETUD requires the Ogr protein (Christie et al., 1986). In the related helper-dependent phage P4, the Delta protein is required for transcriptional activation. We did not identify homologs of Ogr or Delta in the genome of  $\phi$ MhaA1-PHL101, but we did observe a sequence similar to that bound by Delta on P2 (Julien and Calendar, 1995). Alignment of regions analogous to the late gene promoters of P2 (i.e. regions upstream of *orf3*, *orf18*, and *orf25*) revealed a consensus sequence, TGTNAAACCNTTT-NACA, centered 55 or 56 nt upstream of the predicted start-sites of the transcripts (based on the location of the predicted –10 sequences). The spacing and sequence upstream of these ORFs are similar to the P2 element of hyphenated dyad symmetry TGT-N<sub>12</sub>-ACA centered at position –55 with respect to the late promoters. Julien and Calendar (1995) have shown that this sequence is bound by the P4 Delta protein and Christie et al. (2003) have shown that it is important for transcriptional activation of the FETUD late promoter. The presence of such sequences on phage  $\phi$ MhaA1-PHL101 suggests that its late genes may also be regulated by transcriptional activation. One potential candidate for such an activator is ORF49. The predicted protein is small and basic, like Ogr and Delta, but it does not contain the zinc finger characteristic of them (Julien and Calendar, 1995). ORF49 does, however, contain a potential helix–turn–helix motif, so it may function as a DNA binding protein. A second candidate transcriptional activator is ORF11. This 73 amino acid protein has a region of limited sequence similarity to the DskA/TraR family of transcription factors (Marchler-Bauer et al., 2005) and contains six cysteines that could form a zinc finger.

Orthologs of the lambda repressors CI (ORF35) and Cro (ORF36) are encoded by  $\phi$ MhaA1-PHL101. The genes are divergently transcribed (Fig. 1) and four potential 8 nt operator-like sites (consensus TGGTTATA/TATAACCA) lie within the intergenic region (Supplementary Fig. 3). The presumptive binding sites, OR1 and OR3, overlap potential –35 and –10 sequences for the *cro* and *cI* genes, suggesting that the CI and

Cro proteins of  $\phi$ MhaA1-PHL101 function in a manner similar to that of phage lambda (Svenningsen et al., 2005). Thus, we postulate that CI can repress expression of the downstream replication genes during lysogeny, while Cro can inhibit transcription of *cI* during lytic growth. Two additional operator-like sequences map upstream of *orf38*; these sites are downstream of the predicted promoter sequence so repression could occur here as well.

Twelve potential rho-independent transcription terminators were identified within the phage genome (Fig. 1 and Supplementary Figs. 1 and 4). No terminator was identified between *orf36* and *orf48* or between *orf31* and *orf35*. These genes may be transcribed as single operons or may utilize factor-dependent terminators, such as a rho-dependent terminator or a phage-encoded terminator protein.

### Translational signals and coupling

Predicted ribosome binding sites and start codons for each ORF were identified (Supplementary Fig. 2). Only two genes, *orf23* and *orf46*, have a GTG start. A TTG start for *orf2* was chosen because it has a potential ribosome binding site (RBS) 5 nt upstream, while the upstream ATG start has no obvious RBS. All other predicted start codons are ATG. All genes, except the pseudogene (*orf24*), have purine-rich ribosome binding-like sequences that are close to the start codon. Only four genes (*orf3*, *orf20*, *orf25*, and *orf50*) have an exact match to the *E. coli* and *H. influenzae* consensus sequence GGAGG (Hayes and Borodovsky, 1998). With one exception (*orf31*), the spacing between the predicted RBSs and start codons is within the 3 to 15 nt range derived by Hayes and Borodovsky (1998).

Numerous opportunities for translational coupling between  $\phi$ MhaA1-PHL101 genes were observed. In nine cases, the stop codon and start codon are overlapping and in an additional four intergenic regions the start and stop codons lie within 3 nt. In such cases, the ribosome may slip to allow initiation of translation without release from the nascent mRNA (McCarthy and Brimacombe, 1994). This mechanism of initiation would seem likely where the downstream gene lacks a consensus RBS. For six genes, the upstream stop codon overlaps the predicted RBS and for an additional seven genes the stop codon of the upstream gene is embedded within the 5' end of the downstream gene. The potential translational linkages reinforce the assignment of genes to operons since it is assumed that the transcripts for these genes would be polycistronic. Specifically, overlap of translational initiation and termination signals suggest that *orfs* 8–10, 11–13, 18–23, 28–30, and 41–44 are within distinct transcriptional units. This is consistent with the assignments made based on predicted promoter and terminator sequences (Fig. 1 and Supplementary Fig. 2).

Using the MFold program of Zuker (2003), we identified three intergenic RNA structures that could sequester the RBSs and the start codons of genes (data not shown). These mapped within the translation initiation region of the tail formation protein I gene (*orf21*), a conserved hypothetical protein gene (*orf33*), and the replication protein gene (*orf43*). The structures are predicted to be involved in negative regulation of

translation. Note that all three ORFs appear to be translationally coupled to their upstream genes. Ribosome stalling in these regions could melt the hairpin structure to permit initiation of the downstream gene.

### Integration

Integration of P2-like bacteriophages into the host chromosome requires an integrase protein, the histone-like protein, IHF, and both phage and bacterial attachment sites (Frumerie et al., 2005). Integrase and IHF bind to the attachment sites to form a nucleoprotein complex called the intasome. This complex catalyzes synapsis, strand exchange, and ligation (Yu and Haggard-Ljungquist, 1993). *orf50* of phage  $\phi$ MhaA1-PHL101 encodes a protein similar to the tyrosine integrase family. The carboxy terminus of ORF50 contains sequences similar to the BoxA (contains a conserved arginine residue), BoxB (includes the H-X-X-R integrase signature), and BoxC (contains the catalytic tyrosine residue) clusters characteristic of such integrases (Esposito et al., 1996). The  $\phi$ MhaA1-PHL101 attachment site was identified by chromosomal DNA sequencing using primers directed outward from the phage:chromosome junctions or by sequencing PCR products generated using primers that flanked the junctions. The chromosomal attachment site (*attB*) in PHL101 has the sequence TGG TTC GAG TCC AGC TAG TCG CAC CAT. The phage attachment site was identified by first circularizing the phage sequence at the *cos* sites, followed by inspection to identify a phage sequence (*attP*) homologous to *attB*. The sequence of *attP* is TGG TTC GAG AGC TAG TCG CAC CAT. Note that the sequences do not align perfectly and that a three bp gap occurs in the phage sequence; this was not a sequencing error as the region was sequenced four times and in both directions. The phage integrates between open reading frames encoding bis (5'-nucleosyl)-tetraphosphatase and glyceraldehyde 3-phosphate dehydrogenase. More specifically, it integrates within the DNA sequence that encodes the T $\psi$ C loop of a valine tRNA gene. Integration does not interrupt the valyl-tRNA gene because phage sequence within *attP* (nt 1–14) restores the complete gene. tRNA genes are common sites for phage integration (Reiter et al., 1989): both P2 and HP1 integrate into a leucyl-tRNA gene (Hauser and Scocca, 1992; Pierson and Kahn, 1987; Reiter et al., 1989). Though it is not known if  $\phi$ MhaA1-PHL101 requires IHF for integration, a potential IHF binding site (TATCAAAAATGATG) maps near the *attL* site, between nt 34,385 and 34,398.

### G + C content and mosaicism and comparison with sequences in *H. somni* 2336

The G + C content across the  $\phi$ MhaA1-PHL101 sequence is illustrated in Fig. 1. The overall G + C content of the phage is 41.6% compared to 41% for the entire *M. haemolytica* genome, suggesting that the phage DNA was not recently acquired. Two regions of low G + C content are of note, however. First the region encompassing *orfs* 31 through 36 has an average G + C content of 34%. This region includes the phage repressor gene

(*orf35*), *cro* (*orf36*), and several genes that encode proteins of unknown function. Second, the region encompassing *orf15* (tail protein gene T) and *orf14* (hypothetical) has a G + C content of 38.4%. This is similar to that of *H. influenzae* (38.1%) and *H. somni* (37.4%, derived from the draft genome). Furthermore, the DNA sequence from *orf37* to *orf42* includes a number of hypotheticals and matches to *H. somni* cryptic prophage genes. Thus, it is possible that at least one of the low G + C regions was acquired from *H. somni* (Pontarollo et al., 1997). This is plausible because *M. haemolytica* and *H. somni* occupy the same niche in ruminants. Examination of the sequences flanking the low G + C regions did not reveal obvious sites of recombination, though short dyad repeats were observed within the regions.

Thirty-three predicted ORFs had orthologues in *H. somni* 2336 by blastp versus the nr database (Table 1). Each of the  $\phi$ MhaA1-PHL101 ORFs was queried against the most recent *H. somni* draft sequence by tblastn and a large cluster of orthologous genes was identified on contig 104. Contig 104 contained collinear orthologues of the following  $\phi$ MhaA1-PHL101 genes: *orf3–8*, *10*, *12*, *13*, *15*, *18–22*, *25–27*, *29*, *30*, *35*, *36*, *38*, *39*, *43*, *48*, *50*. A second contig (contig 102) contains genes corresponding to *orf1* and *orf2*; these were internal to the contig. These genes may be remnants of another similar phage or may indicate a *H. somni* assembly error. The attachment site for the putative *H. somni* 2336 phage is likely similar to that of  $\phi$ MhaA1-PHL101: an arginine-tRNA gene was found flanking the integrase gene on the contig. Despite the high level of similarity in gene organization and protein sequence, the DNA sequences of  $\phi$ MhaA1-PHL101 and the potential prophage region of contig 104 were not significantly similar (data not shown).

### Comparison with sequence of *M. haemolytica* ATCC BAA-410

We were interested in comparing the phage sequence from strain PHL101 to that of strain ATCC BAA-410, which has been sequenced to 12-fold coverage. The  $\phi$ MhaA1-PHL101 sequence mapped to a scaffold composed of five contigs in the draft assembly of BAA-410 (<http://www.hgsc.bcm.tmc.edu/projects/microbial/Mhaemolytica/>). PCR primers were designed to bridge the two gaps and PCR products were made using BAA-410 chromosomal DNA as a template. The PCR products were then sequenced and used to assemble a 99.9 kb chromosomal contig that spanned the entire prophage region. The prophage portion of this sequence has been deposited as GenBank accession number DQ426905. When the PHL101 phage sequence was compared to that of prophage  $\phi$ MhaA1-BAA-410, a high level of conservation was observed. Twenty-two single base substitutions and a 75 bp insertion were observed in the BAA-410 sequence as compared to PHL101. Of the single base substitutions, two fell within intergenic regions, one non-synonymous substitution occurred each in *orf17* and *orf28*, and 19 substitutions mapped to *orf22*, the predicted variable tail fiber gene. Seven of the 19 substitutions in *orf22* were non-synonymous and two changes within one codon (nt 16293 and 16295) resulted in a synonymous

substitution (Table 2). Finally, the 75 bp insertion observed in  $\phi$ MhaA1-BAA-410 encompassed a series of GCAA repeats found in both strains. In strain PHL101, 14 GCAA repeats were observed while 33 repeats were found in strain BAA-410. Such repeat regions have been found in the 5' regions of some virulence genes of *H. influenzae* (Hood et al., 1996) and *H. somni* (Wu et al., 2000), and in the *M. haemolytica* type I restriction modification locus (Highlander and Garza, 1996; Highlander and Hang, 1997). Expansion and contraction of these repeat sequences occur by slipped-strand mispairing within the region, causing frameshifts and truncation of the protein (Levinson and Gutman, 1987). It is a common mechanism for phase variation. In phage  $\phi$ MhaA1-PHL101, the tetrad repeats fall between *orf23* and *orf24* and only a short ORF (24 amino acids) initiating with a CTG codon spans this region, while in the  $\phi$ MhaA1-BAA-410 prophage, a 52 amino acid ORF with a TTG start and potential RBS is predicted. This longer ORF only matched other proteins with SKQA amino acid repeats such as the *H. influenzae* Lav autotransporter (Davis et al., 2001).

#### Distribution of $\phi$ MhaA1-PHL101 sequences in other *M. haemolytica* serotypes and in *Mannheimia glucosida* and *Pasteurella trehalosi*

*M. haemolytica* (previously designated *Pasteurella haemolytica*) was originally divided into 16 serological groups, based on capsular type. In 1999, Angen et al. reassessed this classification and broke the 16 groups into two genera and two species, namely *M. haemolytica* (serotypes A1, A2, A5, A6, A7, A8, A9, A12, A13, A14, A16), *M. glucosida* (serotype

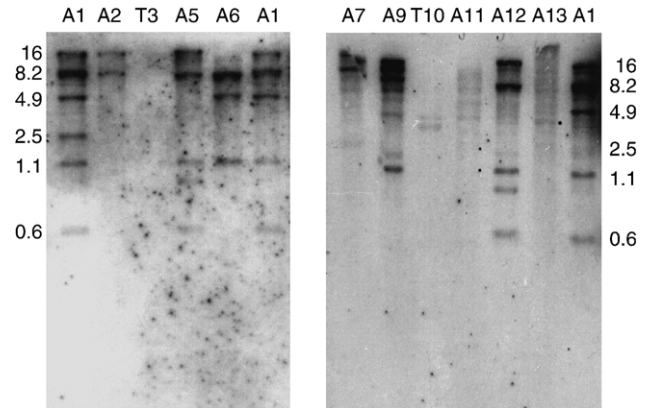


Fig. 3. Southern blots of *XmnI* digested chromosomal DNA from *M. haemolytica* (A1, A2, A5, A6, A7, A9, A12, A13), *M. glucosida* (A11), and *P. trehalosi* (T3, T10) strains. The blots were probed with  $\phi$ MhaA1-PHL101 DNA randomly labeled with  $^{32}\text{P}$ -dCTP.

A11), and *P. trehalosi* (serotypes T3, T4, T10 and T15) (Angen et al., 1999). We surveyed all 16 serotypes for sensitivity to mitomycin C treatment, sensitivity to  $\phi$ MhaA1-PHL101 by plaque assay, and for sequences homologous to  $\phi$ MhaA1-PHL101 by Southern blotting. Six of the strains (serotypes A1, A5, A6, A7, A8, and A13) were lysed following mitomycin C treatment, suggesting that they contain lysogenic phages. Phage  $\phi$ MhaA1-PHL101 did not form plaques on any of the other *M. haemolytica* strains (nor on *E. coli* or *Salmonella typhimurium* strains tested), yet a lysate from strain SH2052 (*M. haemolytica* serotype A8) formed zones of clearing on several of the strains, with the most pronounced clearing on serotypes A5, A8, and A9.

Southern blotting, using phage DNA as a probe, revealed the presence of specific  $\phi$ MhaA1-PHL101-like fragments in serotypes A1, A5, A6, A9, and A12 (Fig. 3). The six  $\phi$ MhaA1-PHL101 *XmnI* fragments were observed in all serotype A1 strains probed (data not shown). In *M. haemolytica* A2, A7, and A13 and *M. glucosida* (A11), high molecular weight DNA hybridized with the probe, but specific fragments characteristic of the A1 phage were not observed. No homologous DNA was observed in the *M. haemolytica* serotype A8, A14, and A16 strains (data not shown) nor in the *P. trehalosi* strains, T3, T4, T10, and T15 (Fig. 3 and data not shown). Based on these comparisons, we conclude that the A5, A6, A9, and A12 strains contain some sequences similar to  $\phi$ MhaA1-PHL101, though A9 and A12 are defective for induction by mitomycin C.

#### Conclusions

The bacteriophage  $\phi$ MhaA1-PHL101, isolated by SOS induction of a *M. haemolytica* serotype A1 strains, is a member of the P2 family of non-SOS inducible bacteriophages. Its predicted proteins are most similar to P2-like phage proteins in *H. influenzae* and predicted sequences of a prophage found in the draft genome of *H. somni*. The morphogenesis and packaging proteins of  $\phi$ MhaA1-PHL101 are most similar in sequence and genetic organization to those of P2, though the

Table 2  
Nucleotide differences between  $\phi$ MhaA1-PHL101 and the  $\phi$ MhaA1-BAA-410 prophages

Nucleotide	Gene	Codon change	Notes
16,283	<i>orf22</i>	TCT to TCG	
16,286	<i>orf22</i>	GTA to GTT	
16,289	<i>orf22</i>	GCA to GCG	
16,292	<i>orf22</i>	GAA to GAG	
16,293	<i>orf22</i>	CTC to TTA	16293 and 16295
16,295	<i>orf22</i>	CTC to TTA	In same codon
16,298	<i>orf22</i>	CGC to CGT	
16,304	<i>orf22</i>	ATT to ATC	
16,626	<i>orf22</i>	CAA to AAA	gln to lys
16,684	<i>orf22</i>	GAG to GGG	glu to gly
16,735	<i>orf22</i>	TGT to TAT	cys to tyr
16,940	<i>orf22</i>	GCT to GCG	
17,054	<i>orf22</i>	TTG to TTT	leu to phe
17,090	<i>orf22</i>	GTT to GTC	
17,158	<i>orf22</i>	TGC to TTC	cys to phe
17,162	<i>orf22</i>	AGC to AGA	ser to arg
17,207	<i>orf22</i>	AGC to AGT	
17,503	<i>orf22</i>	TGC to TTC	cys to phe
17,753	<i>orf22</i>	GGC to GGG	
18,298	<i>orf23-orf24</i>	75 nt insertion	
21,161	<i>orf28</i>	TAC to TGC	tyr to cys
24,317	<i>orf32-orf34</i>	None	In <i>orf32</i> promoter
26,120	<i>orf35-orf36</i>	None	Next to repressor binding site

Numbering is respect to the  $\phi$ MhaA1-PHL101.

regulatory proteins are more divergent and have different gene orders. Two important P2 regulatory functions were not assigned to  $\phi$ MhaA1-PHL101 proteins. First, P2 requires a protein, B, for second strand synthesis (Funnell and Inman, 1983). No such protein was identified on the *M. haemolytica* phage. Second, no excision (Cox) protein was identified. In P2, Cox functions as the repressor of the immunity repressor C and as the excisionase (Saha et al., 1987; Yu and Haggard-Ljungquist, 1993). In terms of transcription repression, the CI protein of  $\phi$ MhaA1-PHL101 may fulfill the first function, but its role as an excisionase cannot be predicted. Note that the Cox and CI proteins described here have absolutely no sequence similarity. CI may be responsible for the SOS-inducibility, however, since it contains an S24 peptidase motif in addition to a helix–turn–helix DNA binding motif. Proteins in the S24 family include the lambda phage repressor and LexA. Like other bacteriophage, the  $\phi$ MhaA1-PHL101 genome includes numerous hypothetical or conserved hypothetical genes. It is of interest to note that most of these cluster in the rightward part of the map and would be expected to be co-expressed with other regulatory genes. Some of these hypothetical proteins may then be involved in aspects of replication, recombination, and transcription control of  $\phi$ MhaA1-PHL101.

The most striking difference with respect to P2 is the location of the T tail protein gene. In P2, the gene maps between the E/E' and U genes. This is also the gene order in phages 186 from *E. coli* (NC\_001317), PSP3 (AY135486) and Fels2 from *Salmonella* (McClelland et al., 2001), and phage phi CTX from *Pseudomonas aeruginosa* (Nakayama et al., 1999). In contrast, the T gene of HP1 and HP2 and the phage K139 from *Vibrio cholerae* (Kapfhammer et al., 2002) is embedded with a group of conserved hypothetical proteins between the endolysin and tail fiber genes. In  $\phi$ MhaA1-PHL101 and in *H. somni*, the T gene maps between S and V genes (*orf13* and *orf18*). The %GC of the T gene of  $\phi$ MhaA1 is lower than that of *M. haemolytica* and other regions of the phage, suggesting that it may have been acquired from recombination with another phage in another organism. This is consistent with Botstein's theory of modular evolution of bacteriophages (Botstein, 1980). The  $\phi$ MhaA1-PHL101 T gene is flanked by two 12 base direct repeats: AAGCGGTCAATT. These are located between *orfs 13* and *14* and within the V gene. It is possible that this constitutes a type of mobile element that can recombine the T gene to different loci. No other example of the 12 bp repeat appears on the phage genome but a subsequence, AAGCGGTC, does occur, in tandem, between the E and U genes. This subsequence also occurs an additional seven times within the phage genome.

When we surveyed other *Mannheimia* strains for  $\phi$ MhaA1-PHL101-like sequences, we found similar sequences in *M. haemolytica* serotypes A1, A5, A6, A9, and A12, but not in *M. glucosida*, *P. trehalosi*, or other *M. haemolytica* serotypes. Thus,  $\phi$ MhaA1-PHL101 sequences are not uniformly distributed in the species. Indeed, no sequence similar to the phage was found in the recently published genome of *Mannheimia succiniciproducens* (Hong et al., 2004). Davies and Donachie (1996) have shown that serotypes A1, A5, A6, A9 and A12 have the same dominant LPS structure. This could explain the

distribution of phage DNA in these serotypes if LPS is the phage receptor.

## Materials and methods

### Bacterial strains and plasmids

Bacterial strains and plasmids are listed in Table 3. *M. haemolytica* strains were grown on sheep blood agar plates or in Bacto Brain Heart Infusion broth (Becton, Dickinson and Co., Sparks, MD) at 37 °C. For phage preparation, broth was supplemented with 10 mM MgSO<sub>4</sub>. *E. coli* strains were grown in Luria-Bertani medium containing 100 µg/ml ampicillin, when appropriate. Plasmid DNAs were introduced into *E. coli* cells by electroporation (Sambrook and Russell, 2001).

### Preparation of phage, phage DNA, and chromosomal DNA

DNA was isolated from bacteriophage produced by mitomycin induction of early log phase broth cultures of strain PHL101. The phage lysate was cleared by centrifugation at 12,000 × g for 15 min, then polyethylene glycol 8000 was added to 10% and the phage was precipitated on ice for 1 h. Phage particles were collected by centrifugation at 12,000 × g for 15 min. The pellet was resuspended in TMG buffer (10 mM Tris, pH 7.4, 5 mM MgSO<sub>4</sub>, 0.01% gelatin) then particles were purified by centrifugation for 60 min at 35,000 × g through a 5–40% glycerol gradient, using an SW41 rotor. Pelleted phage was incubated with DNase I (5 µg/ml) and RNase I (1 µg/ml) for 30 min at 37 °C, then EDTA was added to 20 mM. The sample was deproteinized by incubation for an hour at 65 °C in the presence of 0.1 mg/ml proteinase K, 1% sodium dodecyl sulfate, and 25 mM EDTA. DNA was extracted with TE (10 mM Tris, pH 8.0, 5 mM EDTA)-equilibrated phenol then with phenol-chloroform (1:1) followed by chloroform. The DNA was precipitated with ethanol. *M. haemolytica* chromosomal DNA was prepared by centrifugation of Triton lysates on ethidium bromide cesium chloride gradients (Sambrook and Russell, 2001). The chromosomal band was removed from the gradient, decolorized with isopropanol, and salt was removed by ethanol precipitation.

### DNA sequencing and assembly

Several fragments of phage  $\phi$ MhaA1-PHL101 (2.4 and 3.6 kb *Hind*III, 1.2, 4, 6.5, and 9 kb *Xmn*I) were cloned into the ColE1 vector pBS M13+ (Stratagene, La Jolla, CA) to create substrates for sequencing (see Table 3). These were sequenced using Big Dye chemistry (ABI, Foster City, CA) on an ABI Prism 377 sequencer using M13 forward and reverse primers or custom primers designed for template walking (IDT, Coralville, CA). We also attempted to create a shotgun library of  $\phi$ MhaA1-PHL101 DNA using nebulized DNA from CsCl purified particles but the library was non-random and included non-phage sequences. Gaps were closed and ends were sequenced by cycle sequencing of phage DNA using custom primers. To map the attachment site of phage  $\phi$ MhaA1-PHL101, PCR

Table 3  
Bacterial strains and plasmids

Strain or plasmid	Relevant genotype or properties	Reference or source
<i>Strains</i>		
<i>E. coli</i>		
DH5 $\alpha$	F' <i>recA1 endA1 hsdR17 supE44 thi-1 relA1</i> $\phi$ 80dlacZ $\Delta$ M15 $\Delta$ (lacZYA-argF)U169	(Hanahan, 1983)
<i>M. haemolytica</i>		
SH789 (ATCC BAA-410)	Serotype A1	G. Frank, NADC, Ames, IA
SH010 (PHL101)	Serotype A1	G. Frank, NADC, Ames, IA
SH2045 (I29)	Serotype A1	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2046 (J28)	Serotype A2	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2049 (G13)	Serotype A5	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2050 (A30)	Serotype A6	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2051 (H1)	Serotype A7	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2052 (BH21)	Serotype A8	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2053 (B1)	Serotype A9	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2056 (S209)	Serotype A12	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2057	Serotype A13	G. Frank, NADC, Ames, IA
SH2058	Serotype A14	G. Frank, NADC, Ames, IA
SH2060	Serotype A16	G. Frank, NADC, Ames, IA
<i>M. glucosida</i>		
SH2055 (KC282)	Serotype A11	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
<i>P. trehalosi</i>		
SH2047 (863)	Serotype T3	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2048 (S)	Serotype T4	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2054 (JF2)	Serotype T10	G. Frank, NADC, Ames, IA (Biberstein et al., 1960)
SH2059	Serotype T15	G. Frank, NADC, Ames, IA
<i>Plasmids</i>		
pBS M13+	Ap <sup>R</sup> , ColE1, fl(+)	Stratagene, La Jolla, CA
pSH2395	pBS <i>Hind</i> III: :2.4 kb <i>Hind</i> III of $\phi$ MhaA1-PHL101	This study
pSH2396	pBS <i>Hind</i> III: :3.6 kb <i>Hind</i> III of $\phi$ MhaA1-PHL101	This study
pSH2405	pBS <i>Hind</i> II: :9 kb <i>Xmn</i> I of $\phi$ MhaA1-PHL101	This study
pSH2408	pBS <i>Hind</i> II: :6.5 kb <i>Xmn</i> I of $\phi$ MhaA1-PHL101	This study
pSH2409	pBS <i>Hind</i> II: :1.2 kb <i>Xmn</i> I of $\phi$ MhaA1-PHL101	This study
pSH2419	pBS <i>Hind</i> II: :4 kb <i>Xmn</i> I of $\phi$ MhaA1-PHL101	This study

primers were designed (based on the current *M. haemolytica* genome sequence of strain BAA-410 at <http://www.hgsc.bcm.tmc.edu/projects/microbial/Mhaemolytica>) to flank the left and right chromosomal junctions. PCR was performed using PHL101 chromosomal DNA as a substrate and the PCR products were sequenced using the PCR and internal custom primers. Bases were called and DNA sequences were assembled using phred and phrap (Ewing and Green, 1998; Ewing et al., 1998). A total of 148,411 bases were included in the assembly, yielding approximately 4-fold coverage. Every base was sequenced at least twice and from both strands. The  $\phi$ MhaA1-PHL101 sequence has been deposited with GenBank as accession number DQ426904.

#### Genome annotation and analysis

GeneMark (Lukashin and Borodovsky, 1998) was used to predict open reading frames. Protein sequences were compared to the GenBank nr database (Wheeler et al., 2002). We followed the annotation conventions described in McLeod et al. (2004). Promoter sequences were predicted by manual inspection and by using BPROM (<http://www.softberry.com>). Transcription terminators were identified by visual inspection and using TransTerm (Ermolaeva et al., 2000). RNA structures were

created and free energies calculated using MFold (Zuker, 2003). Other basic analyses and manipulations were performed using the Lasergene suite of programs (DNASTAR, Inc., Madison, WI). Hypothetical and conserved hypothetical proteins were further analyzed by Pfam (Bateman et al., 2004), ScanProsite (Gattiker et al., 2002), BLOCKS (Henikoff and Henikoff, 1994), CDD (Marchler-Bauer et al., 2005), and HTH (Dodd and Egan, 1988), in an effort to assign a probable function.

#### Southern blotting

Ten micrograms of chromosomal DNA was digested with *Xmn*I and separated on a 0.8% agarose gel in TBE (89 mM Tris, 89 mM boric acid, 20 mM EDTA) at 30 V for 20 h. The DNA was depurinated by washing 15 min in 0.25 M HCl then the gel was equilibrated with transfer buffer (0.4 M NaOH, 0.6 M NaCl) and DNAs transferred to Nytran filters (Schleicher and Schuell, Keene, NH) in the same buffer, using a Turboblot apparatus (Schleicher and Schuell). Following transfer, the membrane was neutralized in 0.5 M Tris, pH 7.5, 1 M NaCl, preincubated in hybridization buffer (1 M NaCl, 50 mM Tris, pH 7.5, 1% SDS) containing 20  $\mu$ g/ml salmon sperm DNA (Sigma, St. Louis), then incubated overnight at 65 °C with 100 ng  $\phi$ MhaA1-PHL101 DNA that had been randomly labeled with <sup>32</sup>P- $\alpha$ -dCTP

using Klenow polymerase (Roche, Indianapolis, IN), in hybridization buffer containing 10% dextran sulfate (Amersham, Piscataway, NJ) plus 100 µg/ml salmon sperm DNA (Sigma, St. Louis, MO). Following overnight incubation, the blot was washed with 2× SSC (0.3 M NaCl, 30 mM sodium citrate), 0.5% SDS, then 1× SSC, 0.5% SDS, to remove unbound probe. The blot was then autoradiographed.

## Acknowledgments

We thank Glynn Frank for providing the *M. haemolytica* strains used in this work. We especially thank Roger Hendrix for critical reading of the manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2006.03.024.

## References

- Ackermann, H.W., Karaivanov, L., 1984. Morphology of *Pasteurella multocida* bacteriophages. *Can. J. Microbiol.* 30, 1141–1148.
- Angen, O., Mutters, R., Caugant, D.A., Olsen, J.E., Bisgaard, M., 1999. Taxonomic relationships of the [*Pasteurella*] haemolytica complex as evaluated by DNA–DNA hybridizations and 16S rRNA sequencing with proposal of *Mannheimia haemolytica* gen. nov., comb. nov., *Mannheimia granulomatis* comb. nov., *Mannheimia glucosida* sp. nov., *Mannheimia ruminalis* sp. nov. and *Mannheimia varigena* sp. nov. *Int. J. Syst. Bacteriol.* 49, 67–86.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., Eddy, S.R., 2004. The Pfam protein families database. *Nucleic Acids Res.* 32, D138–D141.
- Bendler, J.W., Goodgal, S.H., 1968. Prophage S2 mutants in *Haemophilus influenzae*: a technique for their production and isolation. *Science* 162, 464–465.
- Biberstein, E.L., Gills, M., Knight, H., 1960. Serological types of *Pasteurella haemolytica*. *Cornell Vet.* 50, 283–300.
- Botstein, D., 1980. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* 354, 484–490.
- Bradley, D.E., 1967. Ultrastructure of bacteriophages and bacteriocins. *Bacteriol. Rev.* 31, 230–314.
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., Parkhill, J., 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21, 3422–3423.
- Christie, G.E., Haggard-Ljungquist, E., Feiwell, R., Calendar, R., 1986. Regulation of bacteriophage P2 late-gene expression: the ogr gene. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3238–3242.
- Christie, G.E., Haggard-Ljungquist, E., Calendar, R., 1998. The Complete Genome of Bacteriophage P2. Genbank NC\_001895.
- Christie, G.E., Temple, L.M., Bartlett, B.A., Goodwin, T.S., 2002. Programmed translational frameshift in the bacteriophage P2 FETUD tail gene operon. *J. Bacteriol.* 184, 6522–6531.
- Christie, G.E., Anders, D.L., McAlister, V., Goodwin, T.S., Julien, B., Calendar, R., 2003. Identification of upstream sequences essential for activation of a bacteriophage P2 late promoter. *J. Bacteriol.* 185, 4609–4614.
- Davies, R.L., Donachie, W., 1996. Intra-specific diversity and host specificity within *Pasteurella haemolytica* based on variation of capsular polysaccharide, lipopolysaccharide and outer-membrane proteins. *Microbiology* 142, 1895–1907.
- Davis, J., Smith, A.L., Hughes, W.R., Golomb, M., 2001. Evolution of an autotransporter: domain shuffling and lateral transfer from pathogenic *Haemophilus* to *Neisseria*. *J. Bacteriol.* 183, 4626–4635.
- Dodd, I.B., Egan, J.B., 1988. The prediction of helix–turn–helix DNA-binding regions in proteins. *Protein Eng.* 2, 174–175.
- Ermolaeva, M.D., Khalak, H.G., White, W., Smith, H.O., Salzberg, S.L., 2000. Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.* 301, 27–33.
- Esposito, D., Fitzmaurice, W.P., Benjamin, R.C., Goodman, S.D., Waldman, A.S., Scocca, J.J., 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. *Nucleic Acids Res.* 24, 2360–2368.
- Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred: II. Error probabilities. *Genome Res.* 8, 186–194.
- Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred: I. Accuracy assessment. *Genome Res.* 8, 175–185.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHigh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-L., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., Venter, C.L., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.
- Froshauer, S., Silvia, A.M., Chidambaram, M., Sharma, B., Weinstock, G.M., 1996. Sensitization of bacteria to danofloxacin by temperate prophages. *Antimicrob. Agents Chemother.* 40, 1561–1563.
- Frumerie, C., Sylwan, L., Ahlgren-Berg, A., Haggård-Ljungquist, E., 2005. Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. *Virology* 332, 284–294.
- Funnell, B.E., Inman, R.B., 1983. Bacteriophage P2 DNA replication. Characterization of the requirement of the gene B protein in vivo. *J. Mol. Biol.* 167, 311–334.
- Gadberry, J.L., Miller, N.G., 1978. Characterization of a *Pasteurella multocida* bacteriophage. *Am. J. Vet. Res.* 39, 1565–1566.
- Gattiker, A., Gasteiger, E., Bairoch, A., 2002. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics* 1, 107–108.
- Hanahan, D., 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580.
- Harm, W., Rupert, C.S., 1963. Infection of transformable cells of *Haemophilus influenzae* by bacteriophage and bacteriophage DNA. *Z. Verbugsl.* 94, 336–348.
- Hauser, M.A., Scocca, J.J., 1992. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. *J. Biol. Chem.* 267, 6859–6864.
- Hayes, W.S., Borodovsky, M., 1998. Deriving ribosomal binding site (RBS) statistical models from unannotated DNA sequences and the use of the RBS model for N-terminal prediction. *Pac. Symp. Biocomput.* 3, 279–290.
- Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., Hatfull, G.F., 1990. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2192–2197.
- Henikoff, S., Henikoff, J.G., 1994. Protein family classification based on searching a database of blocks. *Genomics* 19, 97–107.
- Highlander, S.K., Garza, O., 1996. The restriction-modification system of *Pasteurella haemolytica* is a member of a new family of type I enzymes. *Gene* 178, 89–96.
- Highlander, S.K., Hang, V.T., 1997. A putative leucine zipper activator of *Pasteurella haemolytica* leukotoxin transcription and the potential for modulation of its synthesis by slipped-strand mispairing. *Infect. Immun.* 65, 3970–3975.
- Hong, S.H., Kim, J.S., Lee, S.Y., In, Y.H., Choi, S.S., Rih, J.K., Kim, C.H., Jeong, H., Hur, C.G., Kim, J.J., 2004. The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nat. Biotechnol.* 22, 1275–1281.
- Hood, D.W., Deadman, M.E., Jennings, M.P., Bisercic, M., Fleischmann, R.D., Venter, J.C., Moxon, E.R., 1996. DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Microbiology* 20, 11121–11125.

- Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archeobacteria. *Nucleic Acids Res.* 20, 3279–3285.
- Julien, B., Calendar, R., 1995. Purification and characterization of the bacteriophage P4 delta protein. *J. Bacteriol.* 177, 3743–3751.
- Kapfhammer, D., Blass, J., Evers, S., Reidl, J., 2002. *Vibrio cholerae* phage K139: complete genome sequence and comparative genomics of related phages. *J. Bacteriol.* 184, 6592–6601.
- Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.
- Liu, Y., Haggard-Ljungquist, E., 1994. Studies of bacteriophage P2 DNA replication: localization of the cleavage site of the A protein. *Nucleic Acids Res.* 22, 5204–5210.
- Liu, Y., Haggard-Ljungquist, E., 1996. Functional characterization of the P2 A initiator protein and its DNA cleavage site. *Virology* 216, 158–164.
- Liu, Y., Saha, S., Haggard-Ljungquist, E., 1993. Studies of bacteriophage P2 DNA replication. The DNA sequence of the *cis*-acting gene A and ori region and construction of a P2 mini-chromosome. *J. Mol. Biol.* 231, 361–374.
- Loftus, A., Delisle, A.L., 1995. Inducible bacteriophages of *Actinobacillus actinomycetemcomitans*. *Curr. Microbiol.* 30, 317–321.
- Lukashin, A., Borodovsky, M., 1998. GeneMark.hmm: new solutions for gene finding. *Nucl. Acids. Res.* 26, 1107–1115.
- Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H., 2005. CDD: a conserved domain database for protein classification. *Nucleic Acids Res.* 33, D192–D196.
- McCarthy, J.E.G., Brimacombe, R., 1994. Prokaryotic translation: the interactive pathway leading to initiation. *Trends Genet.* 10, 402–407.
- McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., Wilson, R.K., 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413, 852–856.
- McLeod, M.P., Qin, X., Karpathy, S.E., Gioia, J., Highlander, S.K., Fox, G.E., McNeill, T.Z., Jiang, H., Muzny, D., Jacob, L.S., Hawes, A.C., Sodergren, E., Gill, R., Hume, J., Morgan, M., Fan, G., Amin, A.G., Gibbs, R.A., Hong, C., Yu, X.J., Walker, D.H., Weinstock, G.M., 2004. Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J. Bacteriol.* 186, 5842–5855.
- Morgan, G.J., Hatfull, G.F., Casjens, S., Hendrix, R.W., 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria*, and *Deinococcus*. *J. Mol. Biol.* 317, 337–359.
- Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y., Hayashi, T., 1999. The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Mol. Microbiol.* 31, 399–419.
- Pierson III, L.S., Kahn, M.L., 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. *J. Mol. Biol.* 196, 487–496.
- Pontarollo, R.A., Rioux, C.R., Potter, A.A., 1997. Cloning and characterization of bacteriophage-like DNA from *Haemophilus somnus* homologous to phages P2 and HP1. *J. Bacteriol.* 179, 1872–1879.
- Pullinger, G.D., Bevir, T., Lax, A.J., 2004. The *Pasteurella multocida* toxin is encoded within a lysogenic bacteriophage. *Mol. Microbiol.* 51, 255–269.
- Reiter, W.D., Palm, P., Yeats, S., 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res.* 17, 1907–1914.
- Resch, G., Kulik, E.M., Dietrich, F.S., Meyer, J., 2004. Complete nucleotide sequence of the temperate bacteriophage Aa\*23 of *Actinobacillus actinomycetemcomitans*. *J. Bacteriol.* 186, 5523–5528.
- Richards, A.B., Renshaw, H.W., Sneed, L.W., 1985. *Pasteurella haemolytica* bacteriophage: identification, partial characterization, and relationship of temperate bacteriophages from isolates of *Pasteurella haemolytica* (biotype A, serotype 1). *Am. J. Vet. Res.* 46, 1215–1220.
- Rifkind, D., Pickett, M.J., 1954. Bacteriophage studies on the hemorrhagic septicemia Pasteurellae. *J. Bacteriol.* 67, 243–246.
- Saha, S., Haggard-Ljungquist, E., Nordstrom, K., 1987. The Cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *EMBO J.* 6, 3191–3199.
- Sambrook, J., Russell, D.W., 2001. *Molecular Cloning: A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Samuels, J., Clarke, J.K., 1969. New bacteriophage of *Haemophilus influenzae*. *J. Virol.* 4, 797–798.
- Saxena, S.P., Hoerlein, A.B., 1959. Lysogeny in *Pasteurella*: I. Isolation of bacteriophages from *Pasteurella* strains isolated from shipping fever and those from other infectious processes. *J. Vet. Res. Mhow.* 3, 53–66.
- Schnos, M., Inman, R.B., 1971. Starting point and direction of replication in P2 DNA. *J. Mol. Biol.* 55, 31–38.
- Stevens, R.H., B.F., H., Lai, C.H., 1982. Characterization of an inducible bacteriophage from a leukotoxic strain of *Actinobacillus actinomycetemcomitans*. *Infect. Immun.* 35, 343–349.
- Stuy, J.H., 1978. On the nature of nontypeable *Haemophilus influenzae*. *Antonie Van Leeuwenhoek* 44, 367–376.
- Svenningsen, S.L., Costantino, N., Court, D.L., Adhya, S., 2005. On the role of Cro in lambda prophage induction. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4465–4469.
- Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schrimi, L.M., Tatusova, T.A., Wagner, L., Rapp, B.A., 2002. Database resources of the national center for biotechnology information: 2002 update. *Nucleic Acids Res.* 30, 13–16.
- Whiteley, L.O., Maheswaran, S.K., Weiss, D.J., Ames, T.R., Kannan, M.S., 1992. *Pasteurella haemolytica* A1 and bovine respiratory disease: pathogenesis. *J. Vet. Intern. Med.* 6, 11–22.
- Williams, B.J., Golomb, M., Phillips, T., Brownlee, J., Olson, M.V., Smith, A.L., 2002. Bacteriophage HP2 of *Haemophilus influenzae*. *J. Bacteriol.* 184, 6893–6905.
- Wu, Y., McQuiston, J.H., Cox, A., Pack, T.D., Inzana, T.J., 2000. Molecular cloning and mutagenesis of a DNA locus involved in lipooligosaccharide biosynthesis in *Haemophilus somnus*. *Infect. Immun.* 68, 310–319.
- Xu, J., Hendrix, R.W., Duda, R.L., 2004a. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol. Cell* 15, 11–21.
- Xu, M., Struck, D.K., Deaton, J., Wang, I.-W., Young, R., 2004b. A signal-arrest-release sequence mediates export and control of phage P1 endolysin. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6415–6420.
- Yu, A., Haggard-Ljungquist, 1993. The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J. Bacteriol.* 175, 7848–7855.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.