Available online at www.sciencedirect.com

SCIENCE DIRECT®

Virology 318 (2004) 245–266

VIROLOGY

www.elsevier.com/locate/yviro

The genome and proteome of coliphage T1

Mary D. Roberts,[a] Nancy L. Martin,[b,1] and Andrew M. Kropinski[b,*]

[a] *Biology Department, Radford University, Radford, VA 24142, USA*

[b] *Department of Microbiology and Immunology, Queen's University, Kingston, Ontario, Canada K7L 3N6*

Received 24 July 2003; returned to author for revision 18 September 2003; accepted 22 September 2003

Abstract

The genome of enterobacterial phage T1 has been sequenced, revealing that its 50.7-kb terminally redundant, circularly permuted sequence contains 48,836 bp of nonredundant nucleotides. Seventy-seven open reading frames (ORFs) were identified, with a high percentage of small genes located at the termini of the genomes displaying no homology to existing phage or prophage proteins. Of the genes showing homologs (47%), we identified those involved in host DNA degradation (three endonucleases) and T1 replication (DNA helicase, primase, and single-stranded DNA-binding proteins) and recombination (RecE and Erf homologs). While the tail genes showed homology to those from temperate coliphage N15, the capsid biosynthetic genes were unique. Phage proteins were resolved by 2D gel electrophoresis, and mass spectrometry was used to identify several of the spots including the major head, portal, and tail proteins, thus verifying the annotation.

Keywords: *Caudovirales*; *Siphoviridae*; Bacteriophage; T1; Genome; Evolution; Lytic phages; Complete sequence; Proteomics; 2D gel electrophoresis; Mass spectrometry; Portal protein

Introduction

Enterobacterial phage T1 (originally called phage alpha) is one of the original seven T phages named by Delbrück (Delbrück, 1945; Delbrück and Luria, 1942). T1 is not related to any of the other T phages and it remains the least studied of that group. T1 exhibits an average latent period of 13 min and produces a burst of about 100 progeny phage after infecting sensitive strains of *Escherichia coli* (Borchert and Drexler, 1980; Delbrück, 1945; Roberts and Drexler, 1981b). It also can infect some strains of *Shigella dysenteriae*. A member of the Family *Siphoviridae*, the T1 virion has a polyhedral head 55–60 nm in diameter with a flexible noncontractile tail 150 nm in length and 7 nm in diameter. Thirteen to 15 structural proteins are contained in the mature virion (Martin et al., 1976; Toni et al., 1976; Wagner et al., 1977). Previous studies determined the genome size to be approximately 48.5 kilobase pairs with a terminal redundancy of 2800 ± 530 bp (MacHattie et al., 1972). Using amber mutants of T1, it was demonstrated that this phage could transduce various bacterial markers and prophages from permissive donors to nonpermissive recipients (Bendig and Drexler, 1977; Drexler, 1970; Drexler and Christensen, 1979).

T1 initiates infection by reversibly binding to the FhuA outer membrane ferrichrome iron receptor of *E. coli*. Transfer of the phage through the outer membrane is facilitated by TonB and is coupled to the proton motive force of the cytoplasmic membrane (Hancock and Braun, 1976; Hantke and Braun, 1978). Shortly after T1 DNA enters infected cells, host DNA and protein synthesis are rapidly inhibited (Figurski and Christensen, 1974; Wagner et al., 1977). Phage specific mRNA synthesis using the host RNA polymerase begins shortly after infection and appears to be independent of phage DNA synthesis (Wagner et al., 1977). How T1 controls the expression of its genes is not precisely known. The approximately 31 T1 specific proteins detected in infected cells have been roughly classified based on their time of appearance as early, early–late (continuous), and late (Martin et al., 1976; Toni et al., 1976; Wagner et al., 1977).

Initiation of phage specific DNA synthesis occurs independent of the products of *E. coli* genes *dna*A, *dna*B, *dna*C, and *dna*T but T1 DNA elongation requires the presence of *E. coli* polymerase III as well as the *dna*G encoded primase (Bourque and Christensen, 1980). Genes *1* and *2* (according

* Corresponding author. Fax: +1-613-533-6796.
E-mail address: kropinsk@post.queensu.ca (A.M. Kropinski).
[1] Author responsible for the proteomic section of the manuscript.

0042-6822/$ - see front matter © 2003 Published by Elsevier Inc.
doi:10.1016/j.virol.2003.09.020

to the previous gene numbering system and hereafter referred to in this report as og *1* and *2* for original gene designation) of T1 are required for DNA synthesis. Mutants (designated DO) defective in these genes do not synthesize phage DNA (Figurski and Christensen, 1974; Walling and Christensen, 1981). About 60% of the nucleotides in T1 genomes are derived from the degradation of the *E. coli* chromosome (Christensen et al., 1981; Labaw, 1951). Mutations in host genome degradation were originally discovered because they enhance rates of transduction of host genes and they map to og *2.5* (Roberts and Drexler, 1981a,b). Degradation of the host chromosome is coupled to the ongoing synthesis of new T1 specific DNA and will not occur if T1 DNA production is blocked (Christensen et al., 1981). DA or DNA arrest mutants defective in og *3.5* and *4* begin DNA production normally but soon after halt the process and fail to produce progeny phage (Figurski and Christensen, 1974; Ritchie et al., 1980). These genes also encode a general recombination function Grn that is functionally similar to the λ Red and *E. coli* RecE systems. The products of og *3.5* and *4* (designated as Pog *3.5* and *4*) participate in recombining newly synthesized T1 genomes at their terminally redundant ends into multimeric concatemers from which monomers are packaged into phage heads (Pugh and Ritchie, 1984a; Ritchie and Joicey, 1978).

A partial model for T1 capsid formation has been reported by Ramsay and Ritchie (1984). They proposed the initial assembly of a prohead containing 250–300 copies of a 33-kDa molecular weight major capsid protein. Next, the products of up to six T1 genes appear to be required for the association of proheads with DNA concatemers. The completion of mature head formation requires the presence of two proteins. One of these is the 16-kDa molecular weight Pog *13.3*. The original gene designated *13.3* has been shown to encode a head protein necessary for headful cleavage (Ramsay and Ritchie, 1983). The second protein necessary to complete head assembly is Pog *12* and it appears to stabilize filled heads. The major component of the tail is a 26-kDa molecular weight protein. The details of T1 tail assembly have not been determined.

Packaging of T1 DNA into proheads occurs by a headful mechanism. It begins at the *pac* site located between og *1* and *2* and proceeds unidirectionally toward og *1*. The first headful is terminated and progressive packaging results in one to three additional headfuls of DNA, each with a full genome plus about 6.1% terminal redundancy (2.8 kb ± 530 bp) being packaged from the same concatemer (Gill and MacHattie, 1976; MacHattie and Gill, 1977). The *pip* mutation of T1, which maps between og *2.5* and *3*, reduces the efficiency of packaging initiation at *pac* and prevents processive packaging of the second and successive headfuls (Drexler and Christensen, 1986).

The first T1 genetic map was developed by Michalke (1967) who placed amber mutations into 19 separate genes with the map reading from gene *1* (og *1*) on the left to og *19* on the far right. He observed a high rate of recombination between markers at either end of the map, a phenomenon he could not fully explain. Figurski and Christensen (1974) later eliminated one of the genes and assigned functions to the remaining 18. As is common in bacteriophage genome organization, they found a clustering of functionally related genes with og *1*, *2*, and *4* governing the T1 early functions of DNA metabolism. Original genes *3* and *5–11* were found to encode tail functions and og *4* and *12–18* were found to govern the production of heads. Several additional genes have been found and given decimal numbers. They are: og *2.5* involved in DNA degradation (Roberts and Drexler, 1981a,b), the pip function mapping between og *2.5* and *3* and necessary for processive packaging of DNA (Drexler and Christensen, 1986), og *3.5* necessary for formation of concatemeric DNA (Ritchie et al., 1980), og *7–8* and *11.5* required for tail assembly, and og *13.3*, *13.7*, and *14.5* encoding head functions (Ritchie and Joicey, 1980). The *hr* mutation allows T1 to infect *ton*B cells and is located among tail genes in the vicinity of og *5* (Figurski and Christensen, 1974). Restriction mapping studies of T1 confirmed the circularly permuted nature of the genome and identified the *pac* site where DNA packaging is initiated (Ramsay and Ritchie, 1980). The genetic and physical maps of T1 were correlated using cloned restriction fragments to rescue amber mutations in all the known genes (Liebeschuetz and Ritchie, 1985).

To date, the sequences of only two T1 genes and their implied proteins have been published (Gassner et al., 1998; Scherzer et al., 1987; Schneider-Scherzer et al., 1990). One of these is the early gene (map location between og *2* and *3*) for a DNA *N*-6-adenine-methyltransferase. Partially overlapping this gene is a second open reading frame (ORF) coding for a putative protein with sequence homology to λ exonuclease. In this publication, we report the complete annotated sequence of the T1 genome and a partial analysis of the proteins of the T1 virion.

Results and discussion

Presentation of genetic map

If the genetic map is presented in the classical manner (Drexler, 1988; Roberts, 2001), the majority of the genes were on the complementary strand and the gene order was "Tail–Head–Terminase" rather than the more usual "Terminase–Head–Tail." The reverse complement of the sequence was trimmed of redundant sequence of this terminally redundant, circularly permuted genome. To define the "left end" of the genome, we initially used data from Ramsay and Ritchie (1980). Their restriction analysis suggested that the terminal *Bgl*I-**D** and *Bgl*II-**D** fragments were 5.95 and 3.41 MDa which correspond to 9015 and 5167 bp, respectively. DNA sequence was transposed from the 3′-end to the 5′-end to result in a sequence that corresponds to the Ramsay and Ritchie

restriction map. Subsequently, digestion with *Pst*I, *Sal*I, and *Eco*RI confirmed the location of the left end of the molecule.

DNA analysis

The T1 genome is terminally redundant and circularly permuted with a reported length of 46.9–49.5 kb (Drexler, 1988; MacHattie et al., 1972; Roberts, 2001) Digestion with *Pst*I suggests that the mass is 50.7 kb, and our finding that the unique sequence of T1 is 48,836 bp suggests that the terminal repeats at 1.9 kb are shorter than the 2.8 kb previously suggested.

The overall base composition of T1 DNA (45.6 mol% GC) is somewhat less than that of the host (51 mol% GC) and the predicted value calculated from the DNA melting temperature (48%, Drexler, 1988). The AT content profile exhibits two spikes: the first is located within the predicted left terminal repeat while the other is found at approximately 30 kb. The latter corresponds to the approximate position of the Cor homolog and will be discussed later (Table 1, Fig. 1).

In silico restriction analysis revealed that *Apa*I, *Bam*HI, *Bst*EII, *Kpn*I, *Sac*I, *Sac*II, *Sph*I, and *Xba*I do not cut the sequence. A characteristic of many of these sites is a central GC but this dinucleotide is not significantly underrepresented in the T1 genome. Other common restriction endonucleases that have a limited number of restriction sites include: *Nco*I and *Stu*I (1 site), *Eco*RI, *Nde*I, *Msc*I, *Pst*I, *Sma*I, and *Xho*I (two sites), *Hin*dIII and *Sal*I (three sites). In addition, the DNA does not contain *Eco*BI [TGA(N8) TGCT] or *Eco*KI [AAC(N6)GTGC] sites answering the old question about how this phage escapes the common type I restriction endonucleases present in its hosts. Other type I restriction endonucleases that only cut once are *Eco*AI (GAG(N7)GTCA) and *Eco*DI (TTA(N7)GTCY, while others such as *Eco*DXXI and *Eco*R124I theoretically cut frequently. It is noteworthy that while our calculated *Bgl*I/*Bgl*II map closely resembles the published T1 maps, the number of sites for *Eco*RI and *Pst*I is one less, and *Hin*dIII cuts once more than measured by Ramsay and Ritchie (1980). This suggests that the T1 in the Felix d'Herelle Reference Center for Bacterial Viruses may be a variant of the strain most commonly studied (Ramsay and Ritchie, 1982).

The T1 sequence contains a significant number of direct repeats (DRs) including tandem 80 bp DRs that map to *orf33* (*fibA*), which encodes a putative tail fiber. This protein contains two repeats of the following amino acid sequence TETEALSREIDQLKAQIGDDIQASLTDIRE at positions 927–956 and 960–989 which overlap the nucleotide repeat. It is not uncommon for tail fiber proteins to contain repeat sequences (Makhov et al., 1993). Using Radar (Rapid Automatic Detection and Alignment of Repeats (Heger and Holm, 2000) at its website http://www.ebi.ac.uk/Radar/index.html#), repeats were found in a number of proteins designated as tail fiber proteins in GenBank, including gpH from enterobacterial phage P2 (Table 1, Fig. 1).

Two other shorter repeats with the following consensus sequences AATAGCACNNNTTGNTAAAWC (20 occurrences) and TtWWNCAAAAAGT-GcTAt (five occurrences) were found but their significance remains unknown.

Genes

Since extensive phage genetics and proteomic analyses have already been carried out on coliphage T1, we are faced with the problem of the annotation of genes discovered during this research and their correlation with the previously numbered genes. While only two genes had been sequenced previously, 24 had been identified and mapped. The present study has identified 77 ORFs. It was decided to renumber the latter, but also to provide, where possible, a correlation with previously described gene loci.

Three criteria were used to define potential ORFs. They had to: (a) contain >30 codons, (b) be preceded by a sequence displaying similarity to the consensus ribosome-binding site [RBS, TAAGGAGGT; (Shine and Dalgarno, 1974, 1975)], and (c) employ ATG, GTG, or TTG as initiation codons. The majority of the 77 ORFs (Table 1) were identified using WebGeneMark.HMM (Lukashin and Borodovsky, 1998) while the remainder (16%) were recognized by visual inspection of the DNA sequence. As with other phage genomes, the genes of T1 were densely packed with many incidences of overlapping sequences. One unusual characteristic of the genes of this bacteriophage was the high percentage of small ORFs, which were particularly prevalent at the ends of the genome. Thirty-four percent of the T1 ORFs would result in proteins of less than 100 amino acid residues. ATG was the predominant initiation codon with only two ORFs beginning with GTG (*orf22* [*helA*], *orf43*) and a single one with TTG (*orf17*). The predominant termination codons were UAA (65%) and UGA (27%). While codon usage superficially resembled that of the host, notable exceptions occurred: The preferred T1 Leu codon is CUU, while it is CUG in the host, other examples include Pro CCU/CCG (T1/host), Thr ACU/ACC, Lys AAG/AAA, Ala GCU/GCG, and Gly (GGU/GGC) which is perhaps expected for a genome with a higher AT-content.

Forty-seven percent of the putative T1 genes resulted in BLASTP hits, but with few exceptions these showed a low degree of sequence similarity to their "homologs." The exceptions were the genes that we assume are involved in tail assembly which show >40% sequence identity usually to corresponding proteins from coliphage N15 (Ravin et al., 2000). The other gene showing a high degree of relatedness is *orf12* (*lys*) that encodes a lysozyme-like protein. The high number of proteins (7) showing transmembrane domains warrants further study. In the following sections, we will discuss the genetic organization of T1 from a modular perspective.

Table 1
Location of the genes of coliphage T1 and characterization of the proteins

Orf no.	Begin	End	Protein mass/ amino acids/p*I*	Motifs	Function	Homologs	% Identity
77	539	1114	21,072/191/5.8	–	–	–	
76	1127	1453	12,256/108/8.1				
75	1532	1762	8732/76/5.0	–	–	–	
74	1768	1995	8697/75/9.9	–	–	–	
73	1992	2102	4018/36/5.4	One transmembrane domain	–	–	
72	2099	2272	6048/57/4.5	Prosite: PS00013 PROKAR_LIPOPROTEIN (Prokaryotic membrane lipoprotein lipid attachment site)	–	–	
71	2344	2832	18,048/162/5.4	–	–	–	
70	2904	3389	17,932/161/7.4	–	–	–	
69	3389	3538	5661/49/8.5	–	–	–	
68	3619	3993	13,747/124/9.9	One transmembrane domain	–	–	
67	4047	4157	4165/36/4.0	One transmembrane domain	–	–	
66	4142	4360	8074/72/7.9	–	–	–	
65	4418	4864	17,232/148/8.4	–	–	–	
64 pseT	4948	5478	20,096/176/5.5	BLOCKS: IPB000150: Hypothet_cof	Putative exonuclease III or polynucleotide kinase/phosphatase	*Mycobacterium smegmatis* phage Bxz1 gene = *244* (NP_818294);	24.2
						Streptomyces avermitilis gene = *SAV732* (NP_828503);	26.7
						Escherichia coli phage T4 gene = *pseT* (NP_049834)	15.3
63A endC	5465	5959	18,790/164/10.2	pfam01844, HNH, HNH endonuclease	HNH endonuclease	*Xanthomonas oryzae* phage Xp10 gene = *57R* (NP_859005);	37.6
						Xanthomonas oryzae phage Xp10 gene = *17R* (NP_858964);	37.1
						Xanthomonas oryzae phage Xp10 gene = *49L* (NP_858997);	33.9
						Yersinia enterocolitica phage ϕYeO3-12 gene = *1.45* (NP_052076)	31.3
63B	5702	5959		Alternative start			
62 dnk	5962	6534	21,687/190/5.2	Blocks: IPB002891APS_kinase Prosite: PS00017 (ATP/GTP-binding site motif A [P-loop].	Putative kinase	*Escherichia coli* phage T5 gene = *dnk* (AAN17765)	22.5
61	6607	6816	7888/69/3.7	–	–	–	
60	6813	7157	13,090/114/6.4	–	–	–	
59	7157	7387	8832/76/9.3	–	–	–	
58	7387	7590	7433/67/5.0	–	–	–	
57	7755	7916	5959/53/9.5	–	–	–	
56	7939	8082	5213/47/5.8	–	–	–	
55	8084	8311	8155/75/9.7	One transmembrane domain	–	–	

Table 1 (*continued*)

Orf no.	Begin	End	Protein mass/ amino acids/p*I*	Motifs	Function	Homologs	% Identity
54 terS	8396	8920	19,272/174/5.0	–	Terminase (small subunit)	*Salmonella* phage P22 DNA gene = *3* (P04893);	28.6
						Salmonella phage LP-7 gene = *3* (P16937)	27.4
						Salmonella phage ST64T gene = *3* (NP_720325)	27.4
53 terL	8945	10,528	60,801/527/6.2	Tigrfam: TIGR01630 psiM2_ORF9	Terminase (large subunit)	*Xylella fastidiosa* gene = *XfasA1279* (ZP_00039342);	37.0
						Novosphingobium aromaticivorans gene = *Saro1993* (ZP_00094970);	30.4
						Actinobacillus actinomycetemcomitans phage Aaphi23 gene = *terL* (NP_852753)	26.2
52	10,583	11,866	48,014/427/4.7	Tigrfam: TIGR01555 phge_rel_HI1409 (phage-related protein, HI1409 family)	Putative portal protein	*Salmonella enterica* subsp. *enterica* serovar Typhi gene = *STY2039* (NP_456399);	22.4
						Shigella flexneri 2a str. 301 *SF0670* (NP_706598)	22.1
51	11,856	12,617	29,013/253/9.5	Prosite: PS00017 ATP_GTP_A [ATP/GTP-binding site motif A; Tigrfam: TIGR01641 (phageSPP1_gp7, phage putative head morphogenesis protein, SPP1 gp7 family) COG: COG2369, Uncharacterized protein, homolog of phage Mu protein gp30	Putative phage capsid morphogenesis protein or minor head protein	*Deinococcus radiodurans* gene = *DRA0097* (NP_285420)	25.3
50	12,620	13,732	40,100/370/5.0	–		*Mesorhizobium loti* gene = *mlr8006* (NP_108196);	29.3
						Listeria innocua gene = *lin1728* (NP_471064);	27.0
						Xylella fastidiosa gene = *XF1575* (NP_298864)	25.6
49	13,744	14,220	17,007/158/4.9	–	–	–	
48	14,282	15,049	26,588/255/4.3	pfam02368 Bacterial Ig-like domain (group 2)	–	*Yersinia pestis* gene = *YPO2116* (NP_405664)	22.0
47	15,140	16,099	35,290/319/5.7	–	Major head protein	*Xylella fastidiosa* gene = *XF1577* (NP_298866);	23.7
						Enterococcus faecalis gene = *EF1463* (NP_815184);	23.5
						Novosphingobium aromaticivorans gene = *Saro3167* (ZP_00096129)	22.6
46	16,150	16,437	10,588/95/4.7	–	–	–	

(continued on next page)

Table 1 (*continued*)

Orf no.	Begin	End	Protein mass/ amino acids/p*I*	Motifs	Function	Homologs	% Identity
45	16,482	16,892	15,435/136/8.0	Prosite: PS00029 Leucine zipper	–	–	
44	16,892	17,263	13,815/123/4.9	–	–	*Yersinia pestis* gene = *YPO2113* (NP_405661)	28.2
43	17,256	17,699	16,437/147/9.1	–	–	*Pseudomonas syringae* pv. *tomato* gene = *PSPTO2080* (NP_791903);	27.2
						Yersinia pestis gene = *YPO2114* (NP_405662);	20.5
						Pseudomonas syringae pv. *syringae* gene = *Psyr3313* (ZP_00127002)	17.0
42	17,689	18,087	15,169/132/8.6	–	–	–	
41	18,090	18,758	24,067/222/4.4	–	Putative major tail protein	*Ralstonia solanacearum* gene = *RSc1689* (NP_519810);	35.4
						Pseudomonas syringae pv. *syringae* B728a gene = *Psyr3312* (ZP_00127001);	32.0
						Yersinia pestis gene = *YPO2116* (NP_405664)	31.9
40	18,872	19,189	11,972/105/5.7	–	–	*Yersinia pestis* gene = *YPO2117* (NP_405665)	35.2
39	19,339	19,505	6320/55/4.7	–	–	*Yersinia pestis* gene = *YPO2118* (NP_405666)	28.8
38 ttm	19,547	22,420	103,663/957/5.6	Tigrfam: TIGR01541 (tape_meas_lam_C)	Tail tape measure protein	*Escherichia coli* phage N15 gene = *16* (NP_046911);	25.6
						Escherichia coli phage gene = *H* (NP_040595);	24.8
						Escherichia coli CFT073 gene = *c1585* (NP_753493)	24.7
37	22,423	22,776	12,982/117/5.1	–	Minor tail protein	*Escherichia coli* phage N15 gene = *17* (NP_046912);	34.7
						Escherichia coli phage ϕE125 Gp16 (NP_536372);	33.1
						Yersinia pestis gene = *YPO2120* (NP_405668)	31.9
36	22,856	23,638	29,027/260/7.1	Tigrfam: TIGR01600 Phage_tail_L	Minor tail protein L	*Escherichia coli* phage N15 gene = *18* (NP_046913);	44.4
						Escherichia coli phage HK97 gene = *18* (NP_037712);	43.9
						Escherichia coli phage HK022 gene = *18* (NP_037678)	41.6
35	23,635	24,369	28,317/244/6.0	pfam00877, NLPC_P60, (NLP/P60 family)	Minor tail protein	*Escherichia coli* phage HK022 gene = *19* (NP_037679);	45.9
						Escherichia coli phage HK97 gene = *19* (NP_037713);	45.5
						Escherichia coli phage N15 gene = *19* (NP_046914);	45.2

Table 1 (*continued*)

Orf no.	Begin	End	Protein mass/ amino acids/p*I*	Motifs	Function	Homologs	% Identity
34	24,366	24,964	20,896/199/8.9	two transmembrane domains	Putative tail assembly protein	*Yersinia pestis* gene = *YPO2129* (NP_405677);	46.9
						Escherichia coli phage HK022 gene = *21* (NP_037681);	41.5
						Escherichia coli phage N15 gene = *20* (NP_046915)	41.1
33 fibA	25,043	28,561	130,143/1172/4.7	pfam01576, Myosin_tail; smart00060, Fibronectin type 3 domain	Tail fiber	*Escherichia coli* phage HK022 gene = *24* (NP_037684);	40.6
						Escherichia coli phage N15 gene = *21* (NP_046916);	40.2
						Escherichia coli phage HK97 gene = *24* (NP_037717)	35.7
32	28,607	28,912	10,736/101/4.6	–	–	*Escherichia coli* phage N15 gene = *22* (NP_046917);	56.4
						Escherichia coli phage HK022 gene = *25* (NP_597886)	54.5
31	28,912	29,601	24,116/229/9.1	Prosite: PS00013 PROKAR_LIPOPROTEIN (Prokaryotic membrane lipoprotein lipid attachment site)	–	*Escherichia coli* phage N15 gene = *23* (NP_046918);	52.6
						Escherichia coli phage HK022 gene = *26* (NP_597887)	52.4
30 cor	29,623	29,850 C	8333/75/8.7	Prosite: PS00013 (Prokaryotic membrane lipoprotein lipid attachment site)	Cor homolog	*Escherichia coli* phage N15 gene = *cor* (NP_046919);	34.6
						Escherichia coli phage HK022 gene = *cor* (NP_037685)	34.6
29 recE	30,378	31,442	40,218/354/5.0	–	RecE; Exodeoxyribonuclease VIII (putative)	*Legionella pneumophila* gene = *orfB* (CAC33454);	23.7
						Mycobacterium smegmatis phage Che9c gene = *60* (NP_817737)	23.0
28 erf	31,484	32,164	25,222/226/8.5	pfam04404 Erf superfamily (DNA single-strand annealing proteins)	Recombination	*Lactobacillus delbrueckii* phage LL-H gene = *orf178B* (AAL77543);	26.8
						Lactobacillus delbrueckii phage mv4 gene = *orf244* (AAG31330);	26.4
						Streptococcus pyogenes phage ϕNIH1.1 gene = *phiNIH1.1_12* (NP_438125)	23.2
27 ssb	32,211	32,633	16,151/140/6.4	–	Single-stranded DNA binding protein	*Escherichia coli ssb* (NP_052656);	27.6
						Escherichia coli plasmid pO157 *ssb* (T42184)	26.0
26 fibB	32,695	34,881 C	77,957/728/4.9	–	Tail fiber	*Escherichia coli* phage HK022 gene = *28* (NP_597888);	16.0
						Escherichia coli phage HK97 gene = *stf* (NP_037718)	15.1
25	34,934	34,791 C	5598/47/8.2	–	–	–	

(continued of next page)

Table 1 (*continued*)

Orf no.	Begin	End	Protein mass/ amino acids/p*I*	Motifs	Function	Homologs	% Identity
24 priA	35,902	34,982 C	34,609/306/6.4	–	DNA primase	N.B.: Homology only at N terminus to: *Escherichia coli* O157:H7 gene = *ECs0303* (NP_308330);	14.3
						Escherichia coli prophage CP-933I gene = *Z0339* (NP_285990);	14.4
						Escherichia coli cryptic prophage phi-R73 DNA primase (C41830)	13.3
23	35,979	36,431 C	17,210/150/10.3	–	–	–	
22 helA	36,525	38,543	75,917/672/7.0	Prosite: PS00017—ATP/GTP-binding site motif A (P-loop); Pfam: PF00270 DEAD/DEAH box helicase; PF00271 Helicase conserved C-terminal domain	NTP-dependent helicase	*Escherichia coli CFT073* gene = *yejH* (NP_754607);	26.0
						Shigella flexneri gene = *SF2271* (NP_708083);	25.9
						Escherichia coli O157:H7 EDL933 gene = *yejH* (NP_288767)	25.9
21	38,540	38,956	15,805/138/8.6	–	–	–	
20 dam	39,024	39,737	27,006/237/6.2	Prosite: PS00092 (*N*-6 Adenine-specific DNA methylases signature)	Dam methylase	Phage T1 Dam (AAA87390);	100*
						Haemophilus influenzae phage HP1 gene = *orf13* (NP_043482);	28.0
						Haemophilus influenzae phage HP2 gene = *dam* (AAK37795)	27.6
19	39,734	39,985	9730/83/4.2	–	–	Phage T1 protein HP 83 (AAA87391)	100
18	40,052	40,261	7907/69/8.2	–	–	–	
17	40,456	40,743	10,714/95/4.9	–	–	–	
16	40,822	41,955	42,354/377/6.4	–	–	*Novosphingobium aromaticivorans* gene = *Saro1979* (ZP_00094956)	26.8
15 endA	42,028	42,510	18,941/160/9.1	pfam01844 HNH endonuclease	Endonuclease	*Xanthomonas oryzae* phage Xp10 gene = *17R* (NP_858964);	33.3
						Xanthomonas oryzae phage Xp10 gene = *57R* (NP_859005);	33.0
						Yersinia enterocolitica phage ϕYeO3-12 gene = *1.45* (NP_052076)	27.6
14	42,583	42,759	6600/58/5.7	–	–	–	
13 hol	42,878	43,093	7577/71/9.3	Single transmembrane domain	Holin (putative)	–	
12 lys	43,093	43,581	18,330/166/9.5	pfam00959, Phage_lysozyme; Blocks: IPB002196: Phage_lysozyme	Endolysin	*Escherichia coli* prophage CP-933K gene = *Z0960* (NP_286508);	43.3
						Salmonella typhimurium phage PS119 gene = *19* (LYCV_BPPS1);	40.5
						Salmonella typhimurium phage PS34 gene = *19* (LYCV_BPPS3)	39.9
11	43,581	43,982	14,175/133/8.6	Single transmembrane domain	–	–	
10	44,512	44,105 C	16,196/135/9.1	–	–	–	

Table 1 (*continued*)

Orf no.	Begin	End	Protein mass/ amino acids/p*I*	Motifs	Function	Homologs	% Identity
9	44,517	46,085 C	58,043/522/6.5	Merops: Unassigned peptidase M22	–	*Streptomyces laurentii* gene = *orf1542* (BAC21271);	22.0
						Nitrosomonas europaea gene = *Neur0222* (ZP_00002071)	21.1
8 endB	46,085	46,624 C	20,981/179/10.2	Smart: SM0507 HNH family of nucleases; pfam01844 HNH endonuclease	HNH endonuclease	*Xanthomonas oryzae* phage Xp10 gene = *57R* (NP_859005);	34.4
						Xanthomonas oryzae phage Xp10 gene = *17R* (NP_858964);	32.8
						Xanthomonas oryzae phage Xp10 gene = *49L* (NP_858997)	31.4
						Yersinia enterocolitica phage ϕYeO3-12 gene = *1.45* (NP_052076)	29.4
7	46,621	47,037 C	15,917/138/9.0	–	–	–	
6	47,118	47,327 C	8033/69/3.3	–	–	–	
5	47,331	47,555 C	8024/74/10.1	–	–	–	
4	47,633	47,776 C	5430/47/9.5	–	–	–	
3	47,773	48,093 C	11,889/106/9.7	–	–	–	
2	48,110	48,310 C	7606/66/9.3	–	–	–	
1	48,303	48,674 C	13,857/123/4.8	–	–	–	

Frameshifting

Programmed frameshifting is a phenomenon associated with translation, in which a stalled ribosome usually slips +1 or −1 and continues translating the message resulting in a protein with an altered carboxyl terminus. This requires a "slippery" sequence (XXY YYZ where X, Y, and Z can be the same nucleotide) and a downstream pseudoknot (Alam et al., 1999; Farabaugh, 1996; Harger et al., 2002). Coliphage lambda uses programmed ribosomal slippage at GGGAAAG to produce the gpG-T tail protein (Levin et al., 1993). While this motif occurs six times in T1, none of the occurrences would logically lead to a fusion protein. On the other hand, AAA AAA GAG (LysLysGlu) occurs in *orf18* and *orf11* and in both cases ribosomal slippage could generate fusion proteins involving +1 and −1 slippage, respectively. While both had downstream stem–loop structures, pknotsRG (Jens Reeder and Robert Giegerich; http://bibiserv.techfak.uni-bielefeld.de/pknotsrg/) revealed that only *orf11* has a downstream sequence which would form a pseudoknot. Translational slippage would result in slightly larger protein terminating in kheKKGGVSLLSYLYLLT rather than kheKKEA.

DNA replication and recombination

T1 encodes several proteins that play roles in DNA metabolism including its degradation, synthesis, and recombination. Our analysis of the T1 genome reveals several genes potentially involved in nucleotide metabolism including *orf64* which may be a polynucleotide kinase or phosphatase.

Initiation of DNA duplication involves the coordinated accumulation of replication-associated proteins at the origin of replication (Ori). In the case of *E. coli*, this entails the initial binding of the replisome organizer protein DnaA to tandem nucleotide motifs (DnaA-binding sites, consensus: TTATMCAMA; Brassinga et al., 2002) and the subsequent localized melting of the DNA duplex at adjacent, repetitive sequences called iterons (Carr and Kaguni, 2001, 2002). A domain located at the N terminus of DnaA permits association with DnaB (helicase) (Sutton et al., 1998) or a complex of DnaB with the helicase-loading protein DnaC. Other proteins added to the replication complex include clamp-loading protein (DnaX; PolC proteins τ and γ), clamp-binding protein (DnaN), DnaG (DNA primase), single-strand binding protein (SSB), and DNA polymerase III (PolC).

The work of Bourque and Christensen (1980) showed that host Pol C, and DnaX and DNA primase were required for T1 replication, while *dnaA*, *dnaB*, *dnaC*, and *dnaT* (Primasomal protein i) were not. The genomic sequence of T1 contains only a single copy of a potential DnaA-binding site, within *orf76*, suggesting why replication is independent of this protein. The replication of many other phages including P22 (Schanda-Mulfinger and Schmieger, 1980), P4 (Toccheti et al., 1999), and SPP1 (Pedre et al., 1994) is also known to be independent of host DnaA protein.

Orf24 (*priA*) encodes a 306-amino-acid protein with homology to prophage and bacteriophage primases. Based

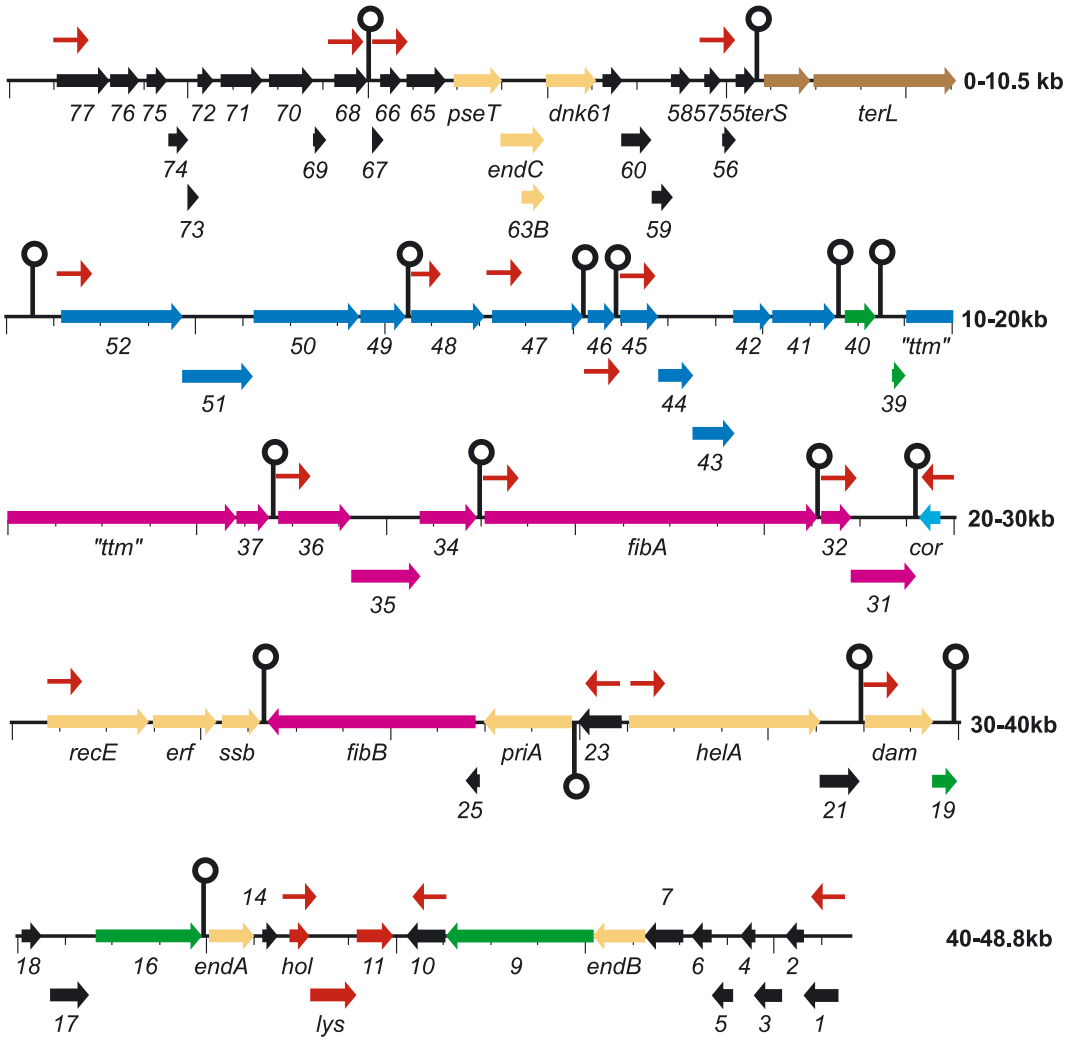

Fig. 1. Genetic map of the T1 genome. The arrows on or below the map line are colour-coded as follows: black (no homolog), gold (DNA breakdown, replication, or recombination), brown (packaging), blue (capsid biosynthesis), purple (tail synthesis), light blue (conversion), and green (homolog but no defined function). The vertical black rods topped with open circles indicate the presence of rho-independent terminators, while the slim red arrows above the map line indicate the location of putative promoters.

upon its mass (34.6 kDa), we assume that it is equivalent to the product of og *1* [38.0 kDa; (Wagner et al., 1977)]. This protein is significantly different from the host DnaG protein (581 aa) in that the latter contains several noteworthy motifs including the CHC2 zinc finger (Pfam: PF01807) and Toprim domains (Pfam: PF01751) while the former is merely defined as a member of COG4643 (“uncharacterized protein conserved in bacteria”).

Downstream from *orf24* and in the opposite orientation is a gene (*orf22*, *helA*) that clearly encodes an ATP-dependent helicase, which would explain why T1 grows on *dnaB*^{ts} hosts (Bourque and Christensen, 1980). This is probably equivalent to og *2* [65 kDa protein; (Ritchie et al., 1983)]. Interestingly, the early recombinational mapping studies indicated that these two genes were further apart. Helicases function to unwind double-stranded nucleotides in a 5′→3′ or 3′→5′direction and are classified into five superfamilies of which T1 gp*22* is a member of superfamily II (COG1061). Each of the proteins in these superfamilies contains up to seven identified motifs. In gp*22*, motif 1 ($G^{62}KT^{64}$) most probably corresponds to the Walker A box, which along with the Walker B box ($D^{148}ECH^{151}$, Motif II), is involved in interaction with MgATP. Motif VI could be $Q^{413}LLGRGMR^{420}$ which bears more than passing resemblance to QTIGRAAR from UvrB (Caruthers and McKay, 2002). If this is so, D^{148} may interact with Q^{413} and R^{417} which could interact with the gamma-phosphate of the bound ATP.

While these results indicate that T1 *gp22* encodes a helicase, explaining why host *dnaB*^{ts} mutants support phage replication, it is not obvious why it was possible to isolate T1 mutants deficient in this apparently “redundant” gene. Another apparently redundant gene encoding a single-stranded DNA binding protein maps to the replication region of the T1 genome. The presence of phage-encoded SSB homologs is quite common—being found in the

genomes of coliphages such as T3, T4, and T7. In the case of the two latter phages, the host SSB cannot substitute for the phage-encoded SSB (Kong and Richardson, 1998; Nakai and Richardson, 1988).

The growth of the phage on *dnaC*^{ts} strains of *E. coli* presumably reflects the fact that the primosome complex in T1 replication differs fundamentally from that of its host (Bourque and Christensen, 1980). Using Grigoriev DNA skew analysis, the inflection occurs at approximately position 36,540, suggesting that the origin of replication in T1 is located within the helicase gene as it is in *Salmonella* phage P22. Phage replication origins are frequently characterized by iterons such as are found with r1t (Zuniga et al., 2002), A2 (Moscoso and Suarez, 2000), BK5-T (Mahanivong et al., 2001), ϕ31 (Madsen et al., 2001) and TP901-1 (Ostergaard et al., 2001) and protein binding motifs. Examples of the latter include IHF and FIS interaction sites. There is no evidence for iterons within orf*22*, nor IHF-binding sites. On the other hand, there is a site (**ATTAAAAAACAGA**) that resembles the FIS-binding site found in the *E. coli* Ori (**ACTCAAAAACTGA**) (Hengen et al., 1997).

Phage T1 encodes a general recombination system termed "*grn*" containing two genes og *3.5*, which specifies a 20-kDa polypeptide and og *4*. We have no doubt, again based upon location, that these genes correspond to *orf28* (*erf*) and *orf29* (*recE*). Host *recABC* will not substitute for T1 grn while the host RecE recombinational pathway will (Pugh and Ritchie, 1984a). Of note is the fact that *orf29* specifies a homolog of host RecE. As with T7 replication, there is no evidence for circular intermediates in T1 replication (Pugh and Ritchie, 1984b; Ritchie and Joicey, 1978) and it is assumed that concatemeric molecules, which are the substrate for head-full packaging, are generated by end-to-end recombination. These are lacking in og *3.5*/og *4* mutants leading to failure to package DNA.

Lastly, the phage genome encodes a potent Dam methylase (gp*20*, *dam*) (Schneider-Scherzer et al., 1990). This and *orf19* were the only two T1 genes that could be found in GenBank, and an alignment of our data with that in GenBank (Accession No. AAA87390) reveals that the latter contains an in-frame deletion equivalent to 20 amino acids. The significance of this protein is unknown since *E. coli* also encodes Dam, but other phages specific for Dam-positive bacteria such as coliphages P1 (Coulby and Sternberg, 1988), T2, T4 and RB49 (Desplats et al., 2002), *Haemophilus* phage HP2 (Williams et al., 2002), and *Shigella* phage SfV (Allison et al., 2002) also encode a copy of this apparently redundant protein. It has been suggested that the activity of the host methylase may be insufficient to keep up with the replicating phage DNA (I.J. Molineux, personal communication).

DNA packaging

In all members of the *Cauloviridae* DNA, a complex of two proteins commonly referred to as terminase accomplish DNA encapsidation. We have identified two genes (*orf54* [*terS*] and *orf53* [*terL*]) that are probably involved in packaging. *Orf54* has sequence similarity to *gp3* (small subunit terminase) from members of the P22 group of viruses that also package by the headful mechanism. *Orf53* encodes a 527-amino-acid protein with homology to bacterial proteins and to *Methanobacterium* phage ψM2 and *Burkholderia cepacia* prophage Bcep781 terminases. Og *13.3* was also reported to be involved in T1 packaging but its functional equivalent in the sequence is unknown.

The *pac* site in the T1-like phage TLS has been located to a 60-bp region deficient in cytosine residues containing six repeats of GATT(T/r) [G. German, personal communication (German and Misra, submitted)]. The *pac* site of T1 was previously mapped 1 kb upstream of an *Eco*RI site which is to be found at 1324 (Ramsay and Ritchie, 1980). In T1, a 90-bp region between 220 and 309 lacks cytosine residues and contains five adjacent repeats of ATATA. If one includes the guanine at position 225, the first repeat (GATATA) superficially resembles the *pac* sites of TLS and P22 [AAGATTA (Casjens et al., 1987)].

Proteomics

Proteomic analysis of the T1 structural proteins was undertaken to aid in the genome annotation. An example of the 2D gel profile of the T1 structural proteins is shown in Fig. 2. Fifteen 15 proteins spots could be resolved with different relative mobilities but some of these exhibited several different p*I*'s. Of the 4 most abundant protein species on the gel, corresponding to MWs of approximately 17.7, 27.2, 30, and 32 kDa, none resolved into a single discreet spot suggestive of either multiple protein species and/or post-translational modifications. Peptide mass fingerprinting of these abundant proteins determined that they correspond to isoforms of gp's *49* (17.7 kDa), *48* (30 kDa), *47* (32 kDa), and *41* (27.2 kDa) (Table 2). It should be noted that at the mass tolerance levels used in this study (±25–50 ppm), the accuracy of matching a small number of peptides is considered sufficient for a positive identification by peptide mass fingerprinting (Clauser et al., 1999).

Gp41 is predicted to be a major tail fiber protein based on homology analysis and its position within the genome (Table 1). This protein probably corresponds to P10 based on Martin et al.'s (1976) estimate of its size (26 kDa) and abundance (16.3% of total protein). The spectra obtained included peptides at the predicted N terminus of this protein and did not include any major peptides unrelated to the tryptic peptide digest profile based on the amino acid sequence. The spectra obtained from the different isoforms showed no obvious indication of specific peptide modifications that would cause a significant shift in the p*I*, but this is not unusual since the peptide profile obtained covers only 52% of the protein.

Gp*49*, a protein with no apparent phage homologs, probably corresponds to P11 in previous studies (Martin

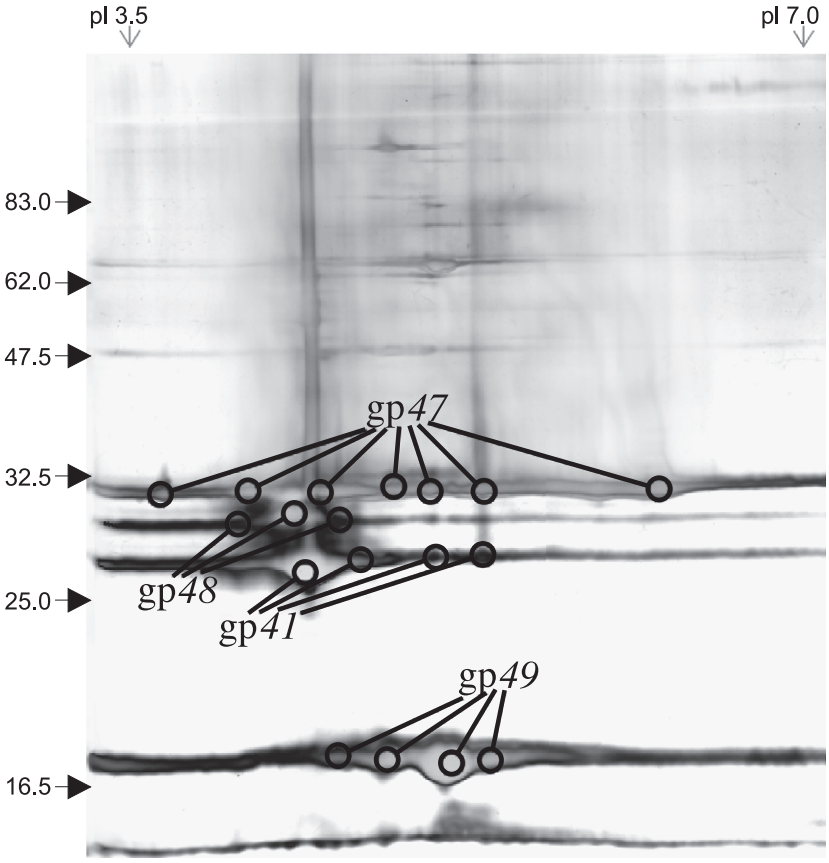

Fig. 2. 2D gel electrophoresis of the T1 structural proteins. The protein-containing portion (from p*I* 3.5–7) of a 2D gel of p*I* range 3–10 is shown. The sampling sites for mass spectrometric analysis for each of the major proteins are indicated by open circles. The positions of the molecular weight markers are indicated on the left of the gel in kilodaltons. The band at the bottom of the gel is the dye front.

et al., 1976). Tryptic peptide mass fingerprinting identified both the predicted N and C termini of this protein. Again, the peptide profiles of the most predominant isoforms showed no obvious differences, but as was the case for gp41, coverage was limited to 51%. Gp*49* is a major capsid component as determined by Ramsay and Ritchie (1984).

Gp*48* and *47* migrate in the gels at relative mobilities that differ only by 1–2 kDa although gp*48* forms a much tighter protein spot than gp*47*. A consideration of 1D SDS-PAGE results published previously (Martin et al., 1976; Wagner et al., 1977) suggests that their identification of the major protein in mature T1 particles [designated as either P7, 33 kDa (Martin) or P8, 36 kDa (Wagner)] corresponds to either gp*47* or a combination of gp*48* and gp*47*. Since this P7/P8 product was the most predominant protein band in 1D gels, it was concluded that this was the major head subunit protein. In the current 2D analysis, it appears that gp*47* is the most predominant protein and gp*48* is now visible as a separate protein because the numerous isoforms of gp*47* cause it to distribute across a large proportion of the gel (Fig. 2). Gp48 could also corresponds to a much less abundant product originally called P8 or P9 (Martin et al., 1976) or P9 (Wagner et al., 1977). Gp48 migrates at a larger than predicted MW, but there is currently no definitive evidence within the MS data to suggest why. It is common for proteins to migrate differently than predicted by their amino acid sequences due to retention of some secondary structure that influences SDS binding during electrophoresis. All of the detected isoforms of gp*47* migrate at a relative mobility corresponding to less than their predicted MW. Peaks within the mass spectra corresponding to the predicted C terminus of gp*47* were routinely detected although peptides corresponding to the first 40 amino acids were never detected. This may indicate that the N terminus of the protein has been post-translationally cleaved. Evidence for post-translational processing of gp47 was first reported by Martin et al. (1976) who showed that a radioactively labeled

Table 2
Mass spectroscopic analysis of T1 structural proteins

ORF	Predicted average mass	Experimental MW (kDa)	Predicted p*I*	Experimental p*I*	No. of matched peptides	% Coverage
49	17,008.21	17.7	5.01	4.6–5.1	7	51.9
48	26,588.83	30.0	4.47	4.3	7	35.7
47	35,290.52	32.0	5.61	4–6	17	63.0
41	24,068.17	27.2	4.48	4.5	13	51.4

40-kDa protein chased to the same position as the major 33-kDa band observed in mature T1 particles. The MS spectra from gp*47* also indicates at least two peptides have been modified, either post-translationally or during sample processing and further MS analysis is underway to determine the nature of these modifications.

Morphogenesis

Adsorption of T1 is a two-step process involving initial binding to the outer membrane ferric-siderophore uptake protein FhuA (TonA), and a subsequence energy-dependent step involving the inner membrane protein TonB, features it shares with phage ϕ80. T1 has two genes (*orf26*, *orf33*) specifying proteins with homology to tail fibers. *E. coli* phage K1-5 also has two tail fiber proteins that possess, respectively, K5 polysaccharide lyase and *N*-acetylneuraminidase-activity against K1 capsular polysaccharide (Scholl et al., 2001). In the case of T1, neither protein possesses identifiable enzymatic activity, and the situation may be analogous to that with coliphage T4 in which a number of proteins are required for tail fiber assembly, with gp*37* acting as the chemosensor (Miller et al., 2003).

Using the Pôle BioInformatique Lyonnais consensus secondary structure prediction program (Combet et al., 2000; Deleage et al., 1997) suggests that gp*26* (728 amino acids) has 47% random coil and 28% α-helical content the latter predominantly located between residues 86 and 261. The 1172-amino-acid gp*33* which has a similar degree of α-helicity and random coil (20% and 48%, respectively) with the α-helical portion being found between residues 851 and 1020. Since tail fibers are often composed of multiple copies of specific proteins, we investigated the presence of the coiled-coil motif in gp*26* and gp*33* using Paircoil (Berger et al., 1995) and MultiCoil (Wolf et al., 1997). The results indicate that gp*26* has two regions between residues 100 and 250 which display a strong probability of trimeric coiled-coil interactions.

Lysis

At the end of the latent period, bacteriophage release usually involves a two-gene lysis cassette composed of a holin and an endolysin (murein hydrolase). The holin creates pores in the inner or cytoplasmic membrane permitting the endolysin to access the peptidoglycan layer in the periplasm resulting in cell lysis and release of progeny viruses (Young, 1992; Young and Blasi, 1995). With the notable exceptions of members of the T4 and T7 coliphage groups, and *Staphylococcus aureus* phage 187, in all other phages studied, the endolysin gene is preceded or overlapped by a gene encoding a holin. The product of *orf12* (*lys*) is a polypeptide of 166 amino acids that displays sequence similarity to a variety of phage muramidases. Motif analysis indicates that it contains a member of the Pfam (pfam00959) defined as phage lysozyme, which cleaves the β-1,4-linkages between adjacent *N*-acetylmuramic acid and *N*-acetylglucosamine residues in cell wall peptidoglycan.

Holins are characterized by their relatively small size (67–145 amino acid residues), usually contain two to three membrane spanning helices, possess a charged C terminus and exhibit poor sequence identity to other members of this group of functionally similar proteins (Grundling et al., 2000; Young, 1992; Young and Blasi, 1995). Furthermore, they sometimes possess a dual-translation start regulatory motif, as do, for example, the holins of coliphages λ, HK97, and HK022 that have MetLysMet at the amino terminus. While one of the translated products is the holin, the other functions as an antiholin that contributes to the temporal regulation of the lysis event (Ramanculov and Young, 2001).

Orf13 (*hol*) specifies a 71-amino-acid single-transmembrane domain protein, which possesses a positively charged C terminus. Examples of single transmembrane domain holins include the *Borrelia burgdorferi* cp32 prophage holin protein BlyA (Damman et al., 2000), *Mycobacterium* phage Ms6 *hol* (Garcia et al., 2002), *Haemophilus influenzae* phage HP1 *hol* (Esposito et al., 1996), and a protein from *Lactococcus lactis* phage ϕU53 (http://tcdb.ucsd.edu/tcdb/background.php).

Downstream of the lysis cassette is *orf11* that specifies a protein containing 133 amino acids arranged into two transmembrane domains. It does not possess a dual start motif, but possesses charged N and C termini. Whether this functions as another holin or is similar to the poorly defined lysis proteins (Rz) in phages such as lambda and P22 is unknown at this time.

The appearance of the lysis cassette before the morphogenesis genes is characteristic of the situation with λ and P22 where the lysis genes precede the terminase genes.

Morons and homing endonucleases

Morons are sequences, flanked by promoter and transcriptional terminators, often found inserted within a group of cotranscribed genes (transcripton). Examination of the T1 genes (Fig. 1) reveals that *orf30* (*cor*) lies in the opposite orientation to its flanking genes and its orientation in coliphage N15. Furthermore, it is separated from *orf31* by a transcriptional terminator. This gene specified a protein called Cor involved in N15 lysogenic conversion, which is responsible for surface exclusion of T1, ϕ80, and N15 (Ravin et al., 2000). Homologs are also synthesized by temperate phage HK022 (Juhala et al., 2000) and ϕ80 (Matsumoto et al., 1985). Why a virulent phage such as T1 should specify such a gene is unknown.

The T1 sequence contains three ORFs *63* (*endC*), *15* (*endA*), and *8* (*endB*) which encode zinc-dependent HNH homing-endonucleases primarily related to those in *Xanthomonas oryzae* phage Xp10 (GenBank accession: AY299121) and *Yersinia enterocolitica* phage ϕYeO3-12 (GenBank: NC_001271). These site-specific endonucleases are found

Table 3
Signal sequences present in Coliphage T1

1. Transcriptional terminators

No.	Gene	Upstem	Loop	Downstem	Tail	ΔG (kcal/mol) at 37 °C	Midpoint distance (nt)
A. Top strand							
1	*76*	ggggua(u)cu	ugcuau	ag(u)uacccc		− 9.9	35
2	*72*	gggguugc	gauaua	guaacccc		− 13.3	36
3	*71*	ggggauu(ug)gcc	ugacu	ggu(au)aaucccc		− 13.8	31
4	*69*	ggccauu	ugauau	aguggcc	uuauu	− 9.5	35
5	*68*	ggggaaucg	acucugaa	cgguucccc	uuucuuuu	− 15.5	21
6	*66*	aggggaau	acgcuaug	auucccccu		− 11.4	22
7	*65*	gggcauuu	guua	gaaugccc	uucguu	− 10.6	38
8	*62*	gucggug(u)ggugauu	ugauau	aguuacc(u)caucgac		− 17.5	35
9	*55*	cccgc	uccg	gcggg	uuuuuuuau	− 8.6	18
10	*54*	gggcgc	gaga	gcgccc		− 14.4	15
11	*53*	gggcuguc	aauu	gacggccc	uuuuuauu	− 14.3	28
12	*49*	ggggacuaucc	uuuuuu	ggauaguccuu	uuuuuau	− 17.3	29
13	*47*	ggggcuaac	uua	guuaguccc	uuuuuuuauu	− 11.4	18
14	*46*	ggcgcu	caugu	agcgcc	uuuuuuuau	− 9.0	16
15	*41*	ggggg	cuugca	ccccc	uuauucauuu	− 8.9	21
16	*40*	ggcgguuu(au)cug	uuauau	cagaaaccgcc		− 14.4	20
17	*37*	gcccgc	cuugc	gcgggu	uuuuuugu	− 9.8	24
18	*34*	cccgc	cuugc	gcggg	uuuuuuuu	− 7.7	15
19	*33*	gggcauc	gaga	gaugccc	uuuucuuuu	− 14.2	16
20	*31*	gcccc	uugc	ggggc	uuuuauuuuau	− 9.1	8
21	*28*	gggcggc	uuag	gucgccc		− 12.6	16
22	*27*	gccccgc	auu	gcggggc	uuuuuauuacuu	− 13.4	15
23	*21*	ugccggg	auggaa	ucuggca	uuauuaucu	− 10.4	27
24	*19*	cccggc	gaaa	gucggg	uauaguuauuu	− 12.6	21
25	*18*	gggagc(a)cgca	aaggaa	ugcggc(a)ucuc		− 10.0	26
26	*16*	cccgc	gaaa	gcggg	uauuauuguuuu	− 12.0	32
B. Bottom strand							
1	*23*	gacgcguag(c)ggg	uuaacau	ccu(u)cuacgcguu	uuuuauuuu	− 17.2	24
2	*10*	gcucacuucg(a)ug	agggccaa	ua(a)cga(c)ggugagc		− 10.1	39
3	*6*	ggucauuu	gaua	aaguggcc	uuuauu	− 8.1	38
4	*4*	gggaguaa	gguaua	uuacuccc		− 11.2	39

N.B.: The "gene" column indicates that the potential terminator is downstream of the named gene; the presence of nucleotides in parentheses in the up- and downstems indicates that these nucleotides do not participate in bonding in the stem structure.

2. Promoters

Name	Location	Sequence	Spacer
ϕP77	167–201	**TTG**CCTAAATCGCAATCGTATAC**A**A**AAT**TACGCCAA**A**	16
ϕP68	3586–3626	**TTGA**AGCACGACAACCAACTGGAGG**TA**A**AAT**TTATGAAATT	19
ϕP67	3965–4004	C**TG**CT**A**TTTCGCTGCTTTCCATCA**T**G**TAAT**GCTAATGGG**G**	18
ϕP57	7587–7626	**TTGAC**CTAAATCCCTGAGCGGTGA**TA**GT**AT**TAATCCCGT**A**	18
ϕP52	10,515–10,549	**TGG**TTAAGCGGTAATAAATGAGA**TATAAT**TAGGGCTGTC	17
ϕP48	14,206–14,242	**TTG**CATCAACTGTAATTAGCG**TT****TAAT**ATGGGGACT**A**	15
ϕP47	15,062–15,099	**TTGAC**GCTTTAACAAAAAGTGC**TATTAT**TGAAGCCGT**G**	16
ϕP46	16,030–16,069	**TTG**CACCTCTAAGTGTACTGGTCT**TA**C**AAT**TTATCGCCC**G**	18
ϕP45	16,415–16,451	**TTGA**AGACGGCGGCGAATATTA**AT**C**AT**CAAGGCGCTC	15
ϕP36	22,799–22,836	**TTG**CGCGGGTTTTTTTGTAGCTG**TA**G**AAT**GGTTGCAG**G**	17
ϕP33	24,978–25,016	**TTG**CGCGGGTTTTTTTTCGCCTG**TATAAT**GAGTCCACC**G**	17
ϕP32	28,536–28,574	**TTGGCA**GGTTGACTGGCTCATGG**TAATAT**CAAGGGCATC	17
ϕP30	29,871–29,908	**TTGCAA**TAAAAGTAAATGCGACA**TACAAT**GCAATCAC**A**	17
ϕP29	30,319–30,356	**TTGAC**TAAACGCTGTTAATGGC**TATAAT**GGATTCATC**G**	16
ϕP23	36,473–36,511 (C)	**TTGA**ATCGTTGTAACGTTTACGC**TATAAT**ACACGGCAT**A**	17
ϕP22	36,420–36,458	TTGA**ATACTCATCTTGATTCCCC**TATA**GCACTTTTTGCT**	17
ϕP20	38,972–39,009	C**TG**C**C**GGGATGGAATCTGGCAT**TAT**T**AT**CTCACCAAA**A**	16
ϕP13	42,815–42,853	**TTG**TGTCAAATAAATTTGAAGGT**TA**A**AAT**CGACTCACTT	17
ϕP10	44,280–44,316	**TTGA**TCTTTCCCTTGCACCAG**TA**A**AAT**CCACCCTTGC	15

Table 3 (*continued*)

2. Promoters

Name	Location	Sequence	Spacer
ϕP1	48,739–48,778 (C)	**TTGA**TGGAATAATGAGCAATTAT**T**GG**AAT**AGCATTTTTT**G**	17

N.B.: In each case, the nucleotides showing the strongest relatedness to the consensus TTGACA (−35), TATAAT (−10), and A/G (+1) are indicated in bold and underlined.

3. IHF-binding sites

Number	Location	Sequence
1	41,500–41,520 (*orf16*)	TGCTTTAATATCAATAACTTG
2	35,325–35,345 (*orf24*)	TCATCAATAATCAAAATATTG
3	1127–1147 (*orf76*)	ATGTTTAATATCAAACCATTA
4	30,260–30,240 (*orf30–orf29* intergenic)	ATCAACAGAATCAATGGCTTA
5	32,479–22,459 (*orf27*)	GAATTGTAAATCATTCCGTTG
6	29,402–29,422 (*orf31*)	TTGGCAACTATCAAGTTGTTA
	Consensus:	ATCAA(N4)TTR

in group I and group II introns (Landthaler and Shub, 2003) or as independent genes in bacteriophages (Belle et al., 2002; Dalgaard et al., 1997). In the case of phage, these endonuclease genes could be considered analogous to Is or Tn elements in bacterial genomes.

Transcription

Early transcription in most coliphages belonging to the *Caudovirales* (Ackermann, 1999) involves host RNA polymerase recognition of promoter sites which are defined by the presence of the canonical hexamers (−35 TTGACA; −10 TATAAT) optimally separated by 15–19 bp (McLean et al., 1997). Transcription terminates by rho-dependent or rho-independent mechanisms and using Gester (Unniraman et al., 2002) we have discovered a considerable number of potential terminators, including rho-independent sites between genes (Fig. 1). Unlike the situation with coliphage T4 in which intrinsic transcriptional terminators frequently contained a UUCG or GNRA loop sequence (Miller et al., 2003), none of the putative T1 terminators contained these motifs (Table 3). The presence of transcriptional terminators leads us to speculate that downstream promoters must occur to direct transcription of subsequent genes. Obvious places where divergent promoters must exist include the intergenic regions between *orf29* and *orf30*, and between *orf22* and *orf23*. In both cases, one of the two putative promoters bore striking resemblance to the consensus sequence: ϕ**P29** (TTGACT (N16) TATAAT) and ϕ**P23** (TTGAAT (N17) TATAAT). Using BPROM and Martin Reese's Neural Network Promoter Prediction and restricting the possibilities to sequences with spacers of 15–19 bp lying predominantly in the non-coding upstream regions of genes, a number of potential promoters were discovered (Table 3). Further study will be required to verify that these are indeed promoters, and assess their relative strengths.

The binding of integration host factor (IHF) results in DNA bending, which may influence promoter activity (Sirko et al., 1998). A number of putative IHF-binding sites were discovered within T1 genes but potentially the most interesting was one located in the *orf29-30* intergenic region (Table 3).

General discussion

The complete sequence of phage T1 has revealed some of the secrets of this interesting virus, such as how it evades host restriction endonucleases, how the DNA replicates, is packaged and lysis occurs but a lot of information on the morphogenesis, regulation of gene expression, and evolution still remains obscure.

In T1 the majority of putative late transcription would, upon first glance, appear to occur in a single block from left to right resulting in expression of the genes involved in packaging and morphogenesis. This modality is found, with minor variations, in many other phages. Closer examination reveals that, particularly in the "late operon," T1 contains an unusually high content of rho-independent terminators. The presence of these leads us to speculate that either a transcriptional read-through system exists such as occurs as with lambdoid phage HK022 (Juhala et al., 2000) or that downstream promoters must exist to direct transcription of subsequent genes. We examined the proteins for homology to the lambdoid phage late gene positive regulator, Q, from a number of different phages including, HK022, N15, Stx1-converting phage, 21, and Aaphi23A using MEME (Bailey and Elkan, 1995). The only protein with a similar motif to HK022 Q (F**AHF**D**GC**R**SAAM**P**GKFW**RDC**HG**) was from T1 orf *9* (Y**AHF**I**GC**I**SAAM**L **GKFW**VQY**HG**).

Phage T1 possesses numerous 21 nt direct repeats located in the intergenic regions or overlapping the translational terminators of the preceding genes. While their high AT content is reminiscent of UP-elements in *E. coli* (Kolasa et al., 2002), their position (Fig. 3) suggests that they may function in a manner equivalent to eukaryotic enhancers. This suggests a far different approach to transcriptional regulation than has been seen with other phages such as

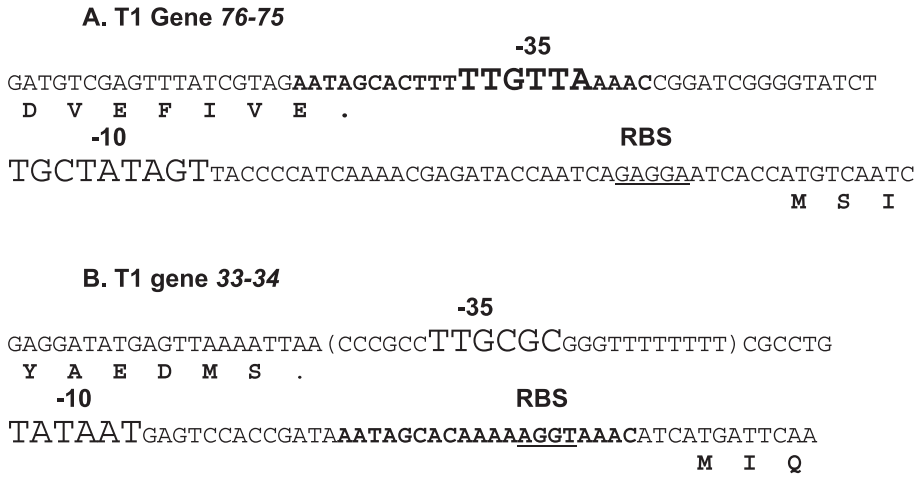

Fig. 3. Two examples of the location of the 21-nt direct repeat in T1. The sequence presented in bold is the direct repeat; putative RNA polymerase − 35 or − 10 regions (promoters) are indicated in larger type; and, the ribosome-binding site (RBS) is underlined. In the second case, the sequence contains a rho-independent terminator which is in parentheses.

the two common modalities exemplified by the lambdoid phages and T7. In the former case, a single promoter regulates transcription of the morphogenesis genes, while with the latter phages, multiple transcriptional start sites are located within the late region. T1 appears to divide the late region up into a series of transcriptional modules (transcriptons) flanked by rho-independent terminators and containing RpoD-dependent promoters and perhaps enhancers. This molecular approach may account for the short latent period of 13 min observed with coliphage T1 (Borchert and Drexler, 1980; Delbrück, 1945; Roberts and Drexler, 1981b).

T1 proteins for which homologs exist indicate a general truism about bacteriophage, the mosaic nature of their genomes (Casjens et al., 1992; Hendrix, 2002; Hendrix et al., 1999; Lawrence et al., 2002; Pedulla et al., 2003). In the case of T1, homologs occur in phages infecting bacterial phyla Actinobacteria (mycobacteriophage Bxz1), Firmicutes (lactococcus phage LL-H) and, as expected, members of the Proteobacteria. Two areas of particular interest are genes *38–30* which are related to coliphage N15, and, with the exception of gene *42*, genes *44–33* which are related to contiguous prophage sequences in *Yersinia pestis*.

One of the more interesting aspects of the genomics of T1 is the presence of four linked genes which have been implicated in tail assembly in a number of members of the *Siphoviridae* infecting, or carried by, members of the class Gammaproteobacteria orders *Enterobacteriales* (*Salmonella* and *Escherichia*) or *Alteromonadales* [*Shewanella*; (Heidelberg et al., 2002)]. Phage ϕE125 is carried by *Burkholderia thailandensis* (Woods et al., 2002), a member of the class Betaproteobacteria. An incomplete set of genes is also to be found in the genome of ϕ80. A further unifying feature is that all of the free-living phages (ϕE125, HK97, HK022, N15, and ϕ80) are classified as lambda-like viruses at NCBI. No homologs were discovered outside the phylum Proteobacteria. Taking a "total evidence approach" to the origin of T1, a phylogenetic tree was constructed by alignment of "polyproteins" composed of the gp*37-34* and their homologs. The results (Fig. 4) offer robust support for the existence of several lines of descent, which include lambda and its prophage relatives (Gifsy-1, Gifsy-2, and Fels-1), the N15-HK97-HK022 cluster, and three deeply rooted clades involving T1, ϕE125, and lambdaSo. The moron carrying Cor apparently was present in the phage genomes before the branching which gave rise to N15-HK022 and T1.

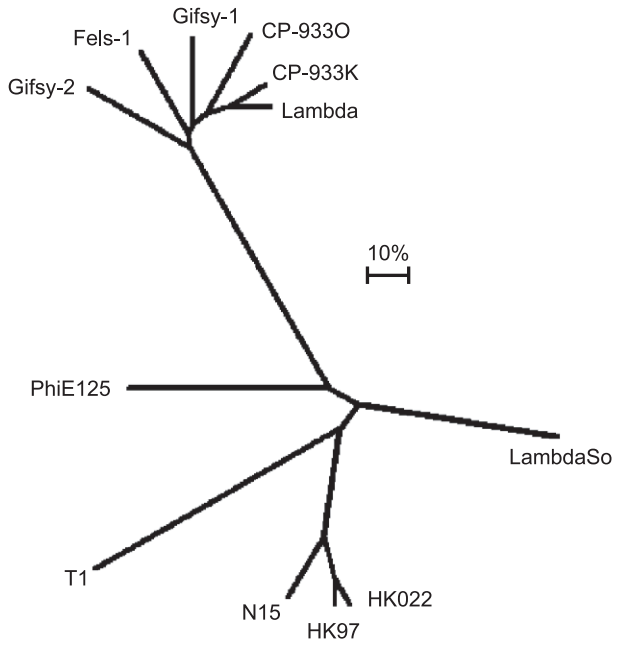

Fig. 4. An unrooted "total evidence approach" phylogenetic tree of phage and prophage tail assembly proteins (λgp*MLKI*–T1 gp*37-34*) derived from GenBank. Sequences were aligned using ClustalW and subsequently analyzed by neighbor joining with TreeCon.

Recently, Gregory German and Rajeev Misra (University of Arizona, Tempe) completed sequencing the genome of a proposed member of the ICTV T1-like phage genus (van Regenmortel et al., 2000). Phage TLS uses the TolC receptor (German and Misra, 2001) and contains many genes in common with T1. The notable exceptions are the presence of a gene encoding a cytosine methylase (*dcm*) within TLS, and the lack of *orf32-31-cor*. While the phylogenetic tree would suggest that speciation occurred in the distant past, the genomic layout suggests that these virulent phages evolved by speciation, and genetic adaptation to the new host via the accumulation of point mutations and not via the more usual recombinational exchange pathway observed with the temperate phages (Kovalyova and Kropinski, 2003). Presumably phages which induce rapid and massive host genome degradation are less likely to recombine with host prophages.

Materials and methods

Bacteria, bacteriophage, and media

Bacteriophage T1 and its host, *E. coli* HER1024, were obtained from Dr. Hans Ackermann (Felix d'Hérelle Reference Center for Bacterial Viruses, Université Laval, Quebec, Canada). The phage was originally purchased from the American Type Culture Collection (Manassas, VA; Accession No. 11303-B1) where it had been deposited by S.E. Luria. *E. coli* DH5α (F^- *deoR endA1 gyrA*96 *hsdR*17(rk^- mk^+) *recA*1 *relA*1 *supE*44 *thi*-1 Δ(*lacZYA-argFV169*) ϕ80 Δ*lacZ*ΔM15 Lambda$^-$) were used for recombinant DNA techniques. Bacteria were grown in Difco Luria-Bertani broth (LB; Fisher Scientific, Toronto, Ontario, Canada) or on LBA plates (LB with 1.5% [wt/vol] Difco agar). For phage titrations, 3 ml LB overlays containing 0.6% (wt/vol) agar were used. The titres of the phage lysates were determined using the agar overlay technique (Adams, 1959).

Phage purification and DNA isolation

Large volume phage lysates were treated with chloroform to enhance release of trapped phage and 1 μg/ml DNaseI (Boehringer Mannheim, Laval, Quebec, Canada) for 30 min. Subsequently, they were centrifuged at 10,000 × *g* for 15 min at 4 °C to pellet debris. The clarified lysate was placed on ice and the phage particles were precipitated with 10% [w/v] polyethylene glycol 8000 (BDH Chemicals, Toronto, Ontario, Canada) (Yamamoto et al., 1970), and the virus particles were subsequently purified by CsCl step and equilibration gradient centrifugations as previously described (Sambrook et al., 1989).

The purified phage preparation was dialysed in a 10K Slide-A-Lyzer cassette (Pierce Biotechnology Inc., Rockford, IL, USA) against TE buffer to remove CsCl. Proteinase K (Boehringer Mannheim) and SDS were added to 50 μg/ml and 0.5% [w/v], respectively, and the mixture was incubated at 56 °C for 1 h. After sequential extractions with phenol–chloroform–isopropanol mixture (25:24:1 [vol/vol/vol]; Fisher Scientific), and chloroform, phage DNA was precipitated with ethanol. After dissolving the DNA in TE, its concentration was analyzed spectrophotometrically (Sambrook et al., 1989).

Proteomics

An aliquot of the purified phage was dialyzed against 20 mM Tris–HCl (pH 7.6), 5 mM $CaCl_2$, 5 mM $MgCl_2$ using a 10K MW cutoff Slide-A-Lyzer cassette (Pierce) to remove CsCl. The dialyzed phages were then disrupted by sonication with five 5 s bursts and DNase I (Sigma, St. Louis, MO, USA) was added to a final concentration of 2 μg/ml and incubated overnight at 4 °C. The phage proteins were concentrated in Centricon YM-3 units (Pierce).

To prepare phage protein samples for isoelectric focusing (IEF), an aliquot of the concentrated phage solution was diluted 1:1 or 1:2 in rehydration buffer [7 M urea, 2 M thiourea, 4% CHAPS, 1% ASB-14, 50 mM DTT, and 0.2% (w/v) Bio-Lyte ampholytes (Bio-Rad Laboratories Inc., Hercules, CA, USA)]. IEF was carried out according to the manufacturer's protocol (Protein IEF cell, Bio-Rad) using IPG Ready Strips (7 cm, pH 3–10; Bio-Rad). Strips were actively rehydrated at 50 V for 12 h, then a rapid voltage ramping method was applied as follows: 100 V for 25 Vh, 500 V for 125 Vh, 4000 V for 90 kVh. The strips were maintained at 20 °C throughout the IEF period. Focused strips were incubated for 10 min in equilibration buffer (6 M urea, 0.375 M Tris pH 8.8, 2% SDS, 20% glycerol) containing 2% (w/v) DTT followed by a 10-min incubation in equilibration buffer containing 2.5% (w/v) iodoacetamide. The strips were then embedded in a 5% acrylamide stacking gel and the proteins resolved by 12% SDS-PAGE (Protean mini-gel, Bio-Rad) at 50 V for 10 min, followed by 150 V for 50 min.

Following electrophoresis, the gels were stained by the method of Shevchenko et al. (1996), which is compatible with mass spectrometry. Stained gels were scanned with a Powerlook II scanner (UMAX Data Systems, Remont, CA, USA) and spots were excised from the gel using a Proteomeworks Spotcutter (Bio-Rad). The gel pieces were destained, digested with trypsin, and extracted for peptide mass fingerprinting analysis using a MassPrep digestion robot (MicroMass). Briefly, destaining was done using 15 mM potassium ferricyanide and 50 mM sodium thiosulphate. The samples were then washed in 100 mM ammonium bicarbonate, followed by dehydration in acetonitrile. The trypsin digestion was carried out using 6 ng/μl trypsin (Promega) in 50 mM ammonium bicarbonate for 5 h at 37 °C. The tryptic peptides were extracted with 1% formic acid–2% acetonitrile followed by two additional extractions with 50% acetonitrile and, following in vacuo drying, they were reconstituted with 4 μl of 50% acetonitrile–0.1% trifluoroacetic acid (TFA). Two μl of matrix (5 mg/ml α-cyano-4-

hydroxy-*trans*-cinnamic acid in 50% acetonitrile–0.1% TFA–10 mM monobasic ammonium phosphate) was spotted onto a stainless steel 96-well MS target and allowed to dry, followed by the addition of 2 μl of the reconstituted sample. Samples were analyzed by MALDI-TOF using a MicroMass M@LDI mass spectrometer. Five spectra (50–100 laser shots/spectrum) were obtained for each sample. Peptide mass fingerprinting was conducted with the database search tool ProteinLynx Global Server 1.1 (Waters, Milford, MA, USA) searching against the T1 protein database generated from our annotation using the following restrictions: minimum three peptides to match, one missed cleavage and a mass tolerance of 50 ppm. Following the initial database search, the data was extensively analyzed to ensure the appropriate matches had been found and mass tolerances were adjusted to 25 ppm to increase accuracy. The spectra of at least four separate samples of each protein were analyzed and compared for accuracy.

Phage DNA cloning and DNA sequencing

We used two strategies to completely sequence the genome of T1. In the first case, T1 DNA (100 μg) was subject to partial digestion with a mixture of *Acc*II (Amersham Biosciences Canada Inc., Baie d'Urfé, Québec, Canada), *Alu*I, *Hae*III, and *Rsa*I (New England Biolabs, Pickering, Ontario, Canada) (80 units) in the presence of NEB2 buffer. Aliquots were removed at various times and the reaction was stopped by the addition of EDTA (pH 8) to 100 mM. The samples were subsequently heated at 70 °C for 15 min to inactivate the enzymes and then electrophoresed through 0.8% agarose gels. Bands of 1.5–3 kb were excised and the DNA recovered using the Prep-A-Gene matrix (Bio-Rad). The recovered fragments were ligated into pUC18-*Sma*I/BAP vector (Amersham Biosciences) and electroporated into DH5α cells using a Bio-Rad gene pulser. The cells were recovered in SOC medium (Sambrook et al., 1989) and after incubation for 1 h at 37 °C, aliquots were plated onto LB-Amp containing 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal) and ampicillin (100 μg/ml, Sigma).

Selected recombinant clones were grown overnight in LB containing ampicillin and plasmid DNA was isolated using the Concert Rapid Plasmid Miniprep System (Invitrogen Canada Inc., Burlington, Ontario). DNA sequencing was carried out at the Montreal Genome Centre (Montreal, Québec, Canada), while gap closure was achieved by directly sequencing off the phage DNA at the Robarts Research Institute (London, Ontario, Canada). The sequence data was stripped of poor-quality data and assembled into contigs using SeqManII (DNASTAR Inc., Madison, WI, USA).

Sequence analysis

A compendium of online tools (http://molbiol-tools.ca) was employed in the analysis of the putative genes. Open reading frames were identified using WebGeneMark.HMM at http://opal.biology.gatech.edu/ GeneMark/gmhmm2_prok.cgi (Lukashin and Borodovsky, 1998). In addition, the Find ORF feature of SeqEdit (DNASTAR) was employed to visually scan the sequence for potential genes. Proteins translated using the Translate DNA feature of EditSeq, which also provided their mass and isoelectric points, were scanned for homologs using BLASTP and Psi-BLAST (Altschul et al., 1990, 1997) at http://www.ncbi.nlm.nih.gov. Where homologs were identified, pairs of sequences were compared using Genestream's (Institute de Génétique Humaine) program ALIGN at its website (http://www2.igh.cnrs.fr/bin/align-guess.cgi). Proteins were also examined for conserved motifs using Prosite (Bairoch, 1992), Tigrfam, BLOCKS (Henikoff and Henikoff, 1996; Henikoff et al., 1999), PRINTS (Attwood et al., 1998, 2002), Pfam (Bateman et al., 2002), and Smart databases (Hofmann et al., 1999; Letunic et al., 2002; Schultz et al., 2000) at the PANAL site (http://mgd.ahc.umn.edu/panal/run_panal.html). To predict transmembrane domains, TMHMM (Sonnhammer et al., 1998) at the Center for Biological Sequence Analysis at The Technical University of Denmark (http://www.cbs.dtu.dk/services/TMHMM-2.0/) was employed.

For basic characterization of the DNA sequence, including codon usage, restriction sites, and motifs analysis, DNAMAN (Lynnon BioSoft, Vaudreuil, Québec, Canada) and Omiga (Oxford Molecular Group, Campbell, CA, USA) were employed. Grigoriev GC/AT skew analyses (Grigoriev, 1998, 1999, 2000) were carried out using the Genometrician's Scooter software (Saint-Sulpice, Switzerland). The DNA sequence was scanned for putative tRNA species using tRNAScan (Eddy and Durbin, 1994; Lowe and Eddy, 1997) at its website (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Potential transcriptional terminators (Brendel and Trifonov, 1984; Brendel et al., 1986) were assessed using the Microsoft Windows software program GeSTer (Unniraman et al., 2002). Putative promoters were screened for using Martin Reese's neural network prediction program at http://www.fruitfly.org/seq_tools/promoter.html, and BPROM (Softberry, Inc., Mount Kisco, NY, USA) at its website http://www.softberry. com/berry.phtml?topic=promoter-file.

To discover potential IHF-binding sites, a MEME search (Bailey and Elkan, 1995) was made with 62 IHF sites derived from the scientific literature, SEQSCAN [http://www.bmb.psu.edu/seqscan/matrixes/ihfmat.htm, (Goodrich et al., 1990)] and *E. coli* DNA-Binding Site Matrices (http://arep.med.harvard.edu/ecoli_matrices/ihf.html) (Robison et al., 1998). A 21-nt multilevel consensus WBHH WWWWWWTCAAYRADTTR was discovered which is quite consistent with previously designated consensus sequences—AATCAA(N4)TTA (Goodrich et al., 1990) or YAACTTNTTGATTW (Engelhorn et al., 1995). Unfortunately, the shorter motifs ignore the fact that IHF covers a 35-bp region of the DNA and interacts with bases 8 bp upstream

of the AATCAA region (Sirko et al., 1998). With this data, 21 and 14 nt consensus matrices were constructed at http://rsat.ulb.ac.be/rsat/ (van Helden et al., 2000a,b) and were used to screen the T1 sequence at BindGene [http://www.bioinf.man.ac.uk/~lockwood/BindGene.html, (Lockwood et al., 2003)].

A phylogenetic tree was generated using ClustalW alignment (http://www.ebi.ac.uk/clustalw/) with the data saved in Phylip format. This was subsequently analyzed using the default settings (Poisson correction, insertion/deletions not taken into consideration) of TreeCon (Van de Peer and De Wachter, 1997).

Nucleotide sequence accession number

The phage T1 sequence has been deposited with GenBank (accession number AY216660).

Acknowledgments

A.K. and N.L.M. acknowledge funding from the Natural Sciences and Engineering Research Council of Canada and Faculty of Health Science at Queen's University (A.K.). We are grateful to Greg German and Rajeev Misra for providing data on phage TLS, to D. MacLeod of the Protein Function Discovery Facility at Queen's University for his assistance in the preparation of samples for MS, and to Dr. Pierre Lepage (Montreal Genomics Institute) and the staff of the Robart's DNA sequencing Facility.

References

Ackermann, H.-W., 1999. Tailed bacteriophages: the order *Caudovirales*. Adv. Virus Res. 51, 135–201.

Adams, M.D., 1959. Bacteriophages. Interscience Publishers, New York.

Alam, S.L., Atkins, J.F., Gesteland, R.F., 1999. Programmed ribosomal frameshifting: much ado about knotting! Proc. Natl. Acad. Sci. U.S.A. 96, 14177–14179.

Allison, G.E., Angeles, D., Tran-Dinh, N., Verma, N.K., 2002. Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of *Shigella flexneri*. J. Bacteriol. 184, 1974–1987.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–4022.

Attwood, T.K., Beck, M.E., Flower, D.R., Scordis, P., Selley, J.N., 1998. The PRINTS protein fingerprint database in its fifth year. Nucleic Acids Res. 26, 304–388.

Attwood, T.K., Blythe, M.J., Flower, D.R., Gaulton, A., Mabey, J.E., Maudling, N., McGregor, L., Mitchell, A.L., Moulton, G., Paine, K., Scordis, P., 2002. PRINTS and PRINTS-S shed light on protein ancestry. Nucleic Acids Res. 30, 239–241.

Bailey, T.L., Elkan, C., 1995. The value of prior knowledge in discovering motifs with MEME. ISMB 3, 21–29.

Bairoch, A., 1992. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. 11, 2013–2088.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L., 2002. The Pfam protein families database. Nucleic Acids Res. 30, 276–280.

Belle, A., Landthaler, M., Shub, D.A., 2002. Intronless homing: site-specific endonuclease *Seg*F of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev. 16, 351–362.

Bendig, M.M., Drexler, H., 1977. Transduction of bacteriophage Mu by bacteriophage T1. J. Virol. 22, 640–645.

Berger, B., Wilson, D.B., Wolf, E., Tonchev, T., Milla, M., Kim, P.S., 1995. Predicting coiled coils by use of pairwise residue correlations. Proc. Natl. Acad. Sci. U.S.A. 92, 8259–8263.

Borchert, L.D., Drexler, H., 1980. T1 genes which affect transduction. J. Virol. 33, 1122–1128.

Bourque, L.W., Christensen, J.R., 1980. The synthesis of coliphage T1 DNA: requirement for host dna genes involved in elongation. Virology 102, 310–316.

Brassinga, A.K., Siam, R., McSween, W., Winkler, H., Wood, D., Marczynski, G.T., 2002. Conserved response regulator CtrA and IHF binding sites in the alpha-proteobacteria *Caulobacter crescentus* and *Rickettsia prowazekii* chromosomal replication origins. J. Bacteriol. 184, 5789–5799.

Brendel, V., Trifonov, E.N., 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. 12, 4411–4427.

Brendel, V., Hamm, G.H., Trifonov, E.N., 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomol. Struct. Dyn. 3, 705–723.

Carr, K.M., Kaguni, J.M., 2001. Stoichiometry of DnaA and DnaB protein in initiation at the *Escherichia coli* chromosomal origin. J. Biol. Chem. 276, 44919–44925.

Carr, K.M., Kaguni, J.M., 2002. *Escherichia coli* DnaA protein loads a single DnaB helicase at a DnaA box hairpin. J. Biol. Chem. 277, 39815–39822.

Caruthers, J.M., McKay, D.B., 2002. Helicase structure and mechanism. Curr. Opin. Struct. Biol. 12, 123–133.

Casjens, S., Huang, W.M., Hayden, M., Parr, R., 1987. Initiation of bacteriophage P22 DNA packaging series. Analysis of a mutant that alters the DNA target specificity of the packaging apparatus. J. Mol. Biol. 194, 411–422.

Casjens, S., Hatfull, G.F., Hendrix, R., 1992. Evolution of dsDNA tailed bacteriophage genomes. In: Koonin, E. (Ed.), Seminars in Virology. Academic Press, London, pp. 383–397.

Christensen, J.R., Figurski, D.H., Schreil, W.H., 1981. The synthesis of coliphage T1 DNA: degradation of the host chromosome. Virology 108, 373–380.

Clauser, K.R., Baker, P., Burlingame, A.L., 1999. Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal. Chem. 71, 2871–2882.

Combet, C., Blanchet, C., Geourjon, C., Deleage, G., 2000. NPS@: network protein sequence analysis. Trends Biochem. Sci. 25, 147–150.

Coulby, J.N., Sternberg, N.L., 1988. Characterization of the phage P1 dam gene. Gene 74, 191.

Dalgaard, J.Z., Klar, A.J., Moser, M.J., Holley, W.R., Chatterjee, A., Mian, I.S., 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res. 25, 4626–4638.

Damman, C.J., Eggers, C.H., Samuels, D.S., Oliver, D.B., 2000. Characterization of *Borrelia burgdorferi* BlyA and BlyB proteins: a prophage-encoded holin-like system. J. Bacteriol. 182, 6791–6797.

Delbrück, M., 1945. The burst size distribution in the growth of bacterial viruses. J. Bacteriol. 50, 131–135.

Delbrück, M., Luria, S.E., 1942. Interference between bacterial viruses. I. Interference between two bacterial viruses acting upon the same host, and the mechanism of virus growth. Arch. Biochem. 1, 111–114.

Deleage, G., Blanchet, C., Geourjon, C., 1997. Protein structure prediction. Implications for the biologist. Biochimie 79, 681–686.

Desplats, C., Dez, C., Tetart, F., Eleaume, H., Krisch, H.M., 2002. Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J. Bacteriol. 184, 2789–2804.

Drexler, H., 1970. Transduction by bacteriophage T1. Proc. Natl. Acad. Sci. U.S.A. 66, 1083–1088.

Drexler, H., 1988. Bacteriophage T1. In: Calendar, R. (Ed.), The Bacteriophages. Plenum, New York, NY, pp. 235–258.

Drexler, H., Christensen, J.R., 1979. Transduction of bacteriophage lambda by bacteriophage T1. J. Virol. 30, 543–550.

Drexler, H., Christensen, J.R., 1986. T1 pip: a mutant which affects packaging initiation and processive packaging of T1 DNA. Virology 150, 373–380.

Eddy, S.R., Durbin, R., 1994. RNA sequence analysis using covariance models. Nucleic Acids Res. 22, 2079–2088.

Engelhorn, M., Boccard, F., Murtin, C., Prentki, P., Geiselmann, J., 1995. In vivo interaction of the *Escherichia coli* integration host factor with its specific binding sites. Nucleic Acids Res. 23, 2959–2965.

Esposito, D., Fitzmaurice, W.P., Benjamin, R.C., Goodman, S.D., Scocca, J.J., 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. Nucleic Acids Res. 24, 2360–2368.

Farabaugh, P.J., 1996. Programmed translational frameshifting. Annu. Rev. Genet. 30, 507–528.

Figurski, D.H., Christensen, J.R., 1974. Functional characterization of the genes of bacteriophage T1. Virology 59, 397–407.

Garcia, M., Pimentel, M., Moniz-Pereira, J., 2002. Expression of Mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. J. Bacteriol. 184, 3034–3043.

Gassner, C., Schneider-Scherzer, E., Lottspeich, F., Schweiger, M., Auer, B., 1998. *Escherichia coli* bacteriophage T1 DNA methyltransferase appears to interact with *Escherichia coli* enolase. Biol. Chem. 379, 621–623.

German, G.J., Misra, R., 2001. The TolC protein of *Escherichia coli* serves as a cell-surface receptor for the newly characterized TLS bacteriophage. J. Mol. Biol. 308, 579–585.

German, G.J., Misra, R., 2003. The T1-like TolC- and lipopolysaccharide-specific (TLS) bacteriophage genome and the evolution of virulent phages. J. Mol. Biol. (submitted).

Gill, G.S., MacHattie, L.A., 1976. Limited permutations of the nucleotide sequence in bacteriophage T1 DNA. J. Mol. Biol. 104, 505–515.

Goodrich, J.A., Schwartz, M.L., McClure, W.R., 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. 18, 4993–5000.

Grigoriev, A., 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. 26, 2286–2290.

Grigoriev, A., 1999. Strand-specific compositional asymmetries in double-stranded DNA viruses. Virus Res. 60, 1–19.

Grigoriev, A., 2000. Graphical genome comparison: rearrangements and replication origin of *Helicobacter pylori*. Trends Genet. 16, 376–378.

Grundling, A., Blasi, U., Young, R., 2000. Biochemical and genetic evidence for three transmembrane domains in the class I holin, lambda S. J. Biol. Chem. 275, 769–776.

Hancock, R.W., Braun, V., 1976. Nature of the energy requirement for the irreversible adsorption of bacteriophages T1 and phi80 to *Escherichia coli*. J. Bacteriol. 125, 410–415.

Hantke, K., Braun, V., 1978. Functional interaction of the *ton*A/*ton*B receptor system in *Escherichia coli*. J. Bacteriol. 135, 190–197.

Harger, J.W., Meskauskas, A., Dinman, J.D., 2002. An "integrated model" of programmed ribosomal frameshifting. Trends Biochem. Sci. 27, 448–454.

Heger, A., Holm, L., 2000. Rapid automatic detection and alignment of repeats in protein sequences. Proteins 41, 224–237.

Heidelberg, J.F., Paulsen, I.T., Nelson, K.E., Gaidos, E.J., Nelson, W.C., Read, T.D., Eisen, J.A., Seshadri, R., Ward, N., Methe, B., Clayton, R.A., Meyer, T., Tsapin, A., Scott, J., Beanan, M., Brinkac, L., Daugherty, S., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Haft, D.H., Kolonay, J.F., Madupu, R., Peterson, J.D., Umayam, L.A., White, O., Wolf, A.M., Vamathevan, J., Weidman, J., Impraim, M., Lee, K., Berry, K., Lee, C., Mueller, J., Khouri, H., Gill, J., Utterback, T.R., McDonald, L.A., Feldblyum, T.V., Smith, H.O., Venter, J.C., Nealson, K.H., Fraser, C.M., 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. Nat. Biotechnol. 20, 1118–1123.

Hendrix, R.W., 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. 61, 471–480.

Hendrix, R., Smith, M.C., Burns, R.N., Ford, M.E., Hatfull, G.F., 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. U.S.A. 96, 2192–2197.

Hengen, P.N., Bartram, S.L., Stewart, L.E., Schneider, T.D., 1997. Information analysis of Fis binding sites. Nucleic Acids Res. 25, 4994–5002.

Henikoff, J.G., Henikoff, S., 1996. Blocks database and its applications. Methods Enzymol. 266, 88–105.

Henikoff, S., Henikoff, J.G., Pietrokovski, S., 1999. Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. Bioinformatics 15, 471–479.

Hofmann, K., Bucher, P., Falquet, L., Bairoch, A., 1999. The PROSITE database, its status in 1999. Nucleic Acids Res. 27, 215–219.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genetic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299, 27–51.

Kolasa, I.K., Lozinski, T., Wierzchowski, K.L., 2002. Effect of A_n tracts within the UP element proximal subsite of a model promoter on kinetics of open complex formation by *Escherichia coli* RNA polymerase. Acta Biochim. Pol. 49, 659–669.

Kong, D., Richardson, C.C., 1998. Role of the acidic carboxyl-terminal domain of the single-stranded DNA-binding protein of bacteriophage T7 in specific protein–protein interactions. J. Biol. Chem. 273, 6556–6564.

Kovalyova, I.V., Kropinski, A.M., 2003. The complete genomic sequence of lytic bacteriophage gh-1 infecting *Pseudomonas putida*—evidence for close relationship to the T7 group. Virology 311, 305–315.

Labaw, L.W., 1951. The origin of phosphorus in *Escherichia coli* bacteriophages. J. Bacteriol. 62, 169–173.

Landthaler, M., Shub, D.A., 2003. The nicking homing endonuclease I–BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. Nucleic Acids Res. 31, 3071–3077.

Lawrence, J.G., Hatfull, G.F., Hendrix, R.W., 2002. Imbroglios of viral taxonomy: genetic exchange and failings of phenetic approaches. J. Bacteriol. 184, 4891–4905.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., Bork, P., 2002. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res. 30, 242–244.

Levin, M.E., Hendrix, R.W., Casjens, S.R., 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. J. Mol. Biol. 234, 124–139.

Liebeschuetz, J., Ritchie, D.A., 1985. Correlation of the genetic and physical maps of phage T1. Virology 143, 175–184.

Lockwood, C.R., Bingham, C., Frayling, T.M., 2003. In silico searching of human and mouse genome data identifies known and unknown HNF1 binding sites upstream of beta-cell genes. Mol. Genet. Metab. 78, 145–151.

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., Borodovsky, M., 1998. GeneMark.hmm: a new solution for gene finding. Nucleic Acids Res. 26, 1107–1115.

MacHattie, L.A., Gill, G.S., 1977. DNA maturation by the "headful" mode in bacteriophage T1. J. Mol. Biol. 110, 441–465.

MacHattie, L.A., Rhoades, M., Thomas, C.A.J., 1972. Large repetition in the non-permuted nucleotide sequence of bacteriophage T1 DNA. J. Mol. Biol. 72, 645–656.

Madsen, S.M., Mills, D., Djordjevic, G., Israelsen, H., Klaenhammer, T.R., 2001. Analysis of the genetic switch and replication region of a P335-type bacteriophage with an obligate lytic lifestyle on *Lactococcus lactis*. Appl. Environ. Microbiol. 67, 1128–1139.

Mahanivong, C., Boyce, J.D., Davidson, B.E., Hillier, A.J., 2001. Sequence analysis and molecular characterization of the *Lactococcus lactis* temperate bacteriophage BK5-T. Appl. Environ. Microbiol. 67, 3564–3576.

Makhov, A.M., Trus, B.L., Conway, J.F., Simon, M.N., Zurabishvili, T.G., Mesyanzhinov, V.V., Steven, A.C., 1993. The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. Virology 194, 117–127.

Martin, D.T., Adair, C.A., Ritchie, D.A., 1976. Polypeptides specified by bacteriophage T1. J. Gen. Virol. 33, 309–319.

Matsumoto, M., Ichikawa, N., Tanaka, S., Morita, T., Matsushiro, A., 1985. Molecular cloning of phi80 adsorption-inhibiting cor gene. Jpn. J. Genet. 60, 483.

McLean, B.W., Wiseman, S.L., Kropinski, A.M., 1997. Functional analysis of sigma-70 consensus promoters in *Pseudomonas aeruginosa* and *Escherichia coli*. Can. J. Microbiol. 43, 981–985.

Michalke, W., 1967. Increased recombination frequency at the ends of T-1-chromosomes. Mol. Gen. Genet. 99, 12–33.

Miller, E.C., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Rüger, W., 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156.

Moscoso, M., Suarez, J.E., 2000. Characterization of the DNA replication module of bacteriophage A2 and use of its origin of replication as a defense against infection during milk fermentation by *Lactobacillus casei*. Virology 273, 101–111.

Nakai, H., Richardson, C.C., 1988. The effect of the T7 and *Escherichia coli* DNA-binding proteins at the replication fork of bacteriophage T7. J. Biol. Chem. 263, 9831–9839.

Ostergaard, S., Brondsted, L., Vogensen, F.K., 2001. Identification of a replication protein and repeats essential for DNA replication of the temperate lactococcal bacteriophage TP901-1. Appl. Environ. Microbiol. 67, 774–781.

Pedre, X., Weise, F., Chai, S., Luder, G., Alonso, J.C., 1994. Analysis of *cis* and *trans* acting elements required for the initiation of DNA replication in the *Bacillus subtilis* bacteriophage SPP1. J. Mol. Biol. 236, 1324–1340.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J.G., Jacobs Jr., W.R., Hendrix, R.W., Hatfull, G.F., 2003. Origins of highly mosaic mycobacteriophage genomes. Cell 113, 171–182.

Pugh, J.C., Ritchie, D.A., 1984a. Formation of phage T1 concatemers by the RecE recombination pathway of *Escherichia coli*. Virology 135, 200–206.

Pugh, J.C., Ritchie, D.A., 1984b. The structure of replicating bacteriophage T1 DNA: comparison between wild type and DNA-arrest mutant infections. Virology 135, 189–199.

Ramanculov, E., Young, R., 2001. An ancient player unmasked: T4 rI encodes a t-specific antiholin. Mol. Microbiol. 41, 575–583.

Ramsay, N., Ritchie, D.A., 1980. A physical map of the permuted genome of bacteriophage T1. Mol. Gen. Genet. 179, 669–675.

Ramsay, N., Ritchie, D.A., 1982. Restriction endonuclease analysis of variants of phage T1: correlation of the physical and genetic maps. Virology 121, 420–425.

Ramsay, N., Ritchie, D.A., 1983. Uncoupling of initiation site cleavage from subsequent headful cleavages in bacteriophage T1 DNA packaging. Nature 301, 264–269.

Ramsay, N., Ritchie, D.A., 1984. Phage head assembly in bacteriophage T1. Virology 132, 239–249.

Ravin, V., Ravin, N., Casjens, S., Ford, M.E., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. 299, 53–73.

Ritchie, D.A., Joicey, D.H., 1978. Formation of concatemeric DNA as an intermediate in the replication of bacteriophage T1 DNA molecules. J. Gen. Virol. 41, 609–622.

Ritchie, D.A., Joicey, D.H., 1980. Identification of some steps in the replication of bacteriophage T1 DNA. Virology 103, 191–198.

Ritchie, D.A., Christensen, J.R., Pugh, J.C., Bourque, L.W., 1980. Genes of coliphage T1 whose products promote general recombination. Virology 105, 371–378.

Ritchie, D.A., Joicey, D.H., Martin, D.T., 1983. Correlation of genetic loci and polypeptides specified by bacteriophage T1. J. Gen. Virol. 64, 1355–1363.

Roberts, M.D., 2001. T1-like viruses. In: Tidona, C.A., Darai, G. (Eds.), The Springer Index of Viruses. Springer-Verlag Heidelberg, Heidelberg, Germany, pp. 1–10.

Roberts, M.D., Drexler, H., 1981a. Isolation and genetic characterization of T1-transducing mutants with increased transduction frequency. Virology 112, 662–669.

Roberts, M.D., Drexler, H., 1981b. T1 mutants with increased transduction frequency are defective in host chromosome degradation. Virology 112, 670–677.

Robison, K., McGuire, A.M., Church, G.M., 1998. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. J. Mol. Biol. 284, 241–254.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Schanda-Mulfinger, U.E., Schmieger, H., 1980. Growth of *Salmonella* bacteriophage P22 in *Escherichia coli dna*(Ts) mutants. J. Bacteriol. 143, 1042–1045.

Scherzer, E., Auer, B., Schweiger, M., 1987. Identification, purification, and characterization of *Escherichia coli* virus T1 DNA methyltransferase. J. Biol. Chem. 262, 15225–15231.

Schneider-Scherzer, E., Auer, B., de Groot, E.J., Schweiger, M., 1990. Primary structure of a DNA (*N*6-adenine)-methyltransferase from *Escherichia coli* virus T1. DNA sequence, genomic organization, and comparative analysis. J. Biol. Chem. 265, 6086–6091.

Scholl, D., Rogers, S., Adhya, S., Merril, C.R., 2001. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. J. Virol. 75, 2509–2515.

Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., Bork, P., 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28, 231–234.

Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., Mann, M., 1996. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. U.S.A. 93, 14440–14445.

Shine, J., Dalgarno, L., 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. 71, 1342–1346.

Shine, J., Dalgarno, L., 1975. Terminal-sequence analysis of bacterial ribosomal RNA. Correlation between the 3′-terminal-polypyrimidine sequence of 16-S RNA and translational specificity of the ribosome. Eur. J. Biochem. 57, 221–230.

Sirko, A., Weglenska, A., Hulanicka, D., 1998. Integration host factor positively regulates *cysJIH* transcription. Mol. Gen. Genet. 258, 174–177.

Sonnhammer, E.L.L., von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow, J., Littlejohn, T., Major, F., Lathrop, R., Sankoff, D., Sensen, C. (Eds.), Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, CA, pp. 175–182.

Sutton, M.D., Carr, K.M., Vicente, M., Kaguni, J.M., 1998. *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. J. Biol. Chem. 273, 34255–34262.

Tocchetti, A., Galimberti, G., Deho, G., Ghisotti, D., 1999. Characterization of the oriI and oriII origins of replication in phage-plasmid P4. J. Virol. 73, 7308–7316.

Toni, M., Conti, G., Schito, G.C., 1976. Viral protein synthesis during replication of bacteriophage T1. Biochem. Biophys. Res. Commun. 68, 545–552.

Unniraman, S., Prakash, R., Nagaraja, V., 2002. Conserved economics of transcription termination in eubacteria. Nucleic Acids Res. 30, 675–684.

Van de Peer, Y., De Wachter, R., 1997. TREECON for Windows: a software package for the construction and drawing of evolutionary trees. Comput. Appl. Biosci. 10, 569–570.

van Helden, J., Andre, B., Collado-Vides, J., 2000a. A web site for the computational analysis of yeast regulatory sequences. Yeast 16, 177–187.

van Helden, J., Rios, A.F., Collado-Vides, J., 2000b. Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. Nucleic Acids Res. 28, 1808–1818.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., McGeoch, D.J., Pringle, C.R., Wickner, R.B., 2000. Virus Taxonomy: Classification and Nomenclature of Viruses—Seventh Report of the International Committee on the Taxonomy of Viruses. Academic Press, New York.

Wagner, E.F., Ponta, H., Schweiger, M., 1977. Development of *E. coli* virus T1: the pattern of gene expression. Mol. Gen. Genet. 150, 21–28.

Walling, L., Christensen, J.R., 1981. The synthesis of coliphage T1 DNA: studies on the roles of T1 genes 1, 2, and 4. Virology 114, 309–318.

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.

Wolf, E., Kim, P.S., Berger, B., 1997. MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci. 6, 1179–1189.

Woods, D.E., Jeddeloh, J.A., Fritz, D.L., DeShazer, D., 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. J. Bacteriol. 184, 4003–4017.

Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L., Treiber, G., 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. Virology 40, 734–744.

Young, R., 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. 56, 430–481.

Young, R., Blasi, U., 1995. Holins: form and function in bacteriophage lysis. FEMS Microbiol. Rev. 17, 191–205.

Zuniga, M., Franke-Fayard, B., Venema, G., Kok, J., Nauta, A., 2002. Characterization of the putative replisome organizer of the lactococcal bacteriophage r1t. J. Virol. 76, 10234–10244.