Identification of a Gene Encoding a Functional Reverse Transcriptase within a Highly Variable Locus in the P2-Like Coliphages

Richard Odegrip,† Anders S. Nilsson, and Elisabeth Haggård-Ljungquist*

Department of Genetics, Microbiology and Toxicology, Stockholm University, S-106 91 Stockholm, Sweden

The P2-like coliphages are highly similar; the structural genes show at least 96% identity. However, at two loci they have genes believed to be horizontally transferred. We show that the genetic content at the second loci, the TO region, contains six completely different sequences with high AT contents and with different open reading frames. The product of one of them exhibits reverse transcriptase activity and blocks infection of phage T5.

Temperate coliphage P2 has three nonessential lysogenic conversion genes, *Z/fun*, *tin*, and *old*, conferring resistance to infections by phage T5, T-even phages, and lambda phages, respectively (1, 8, 11, 12, 15, 21). These genes have been shown to have a higher AT content than the rest of the genome and a codon usage that differs from that of the host, which suggests that they are horizontally transferred genes (3). P2-like prophages are common in *Escherichia coli*; about 30% of the ECOR collection (20) contain P2-like prophages. At the region equivalent to the P2 *Z/fun* gene, i.e., the Z region, which is located between the well-conserved tail genes *G* and *F₁*, these P2-like prophages have been shown to contain different gene cassettes surrounded by a highly similar inverted repeat, indicating a site-specific integration event. Similar inverted repeats, spacing other genes, can be found in genetically unstable regions in pathogenic enterobacteria (19). The *tin* and *old* genes are located at the right end of the P2 genetic map close to the *cos* site. The *old* gene encodes an exonuclease that blocks multiplication of lambdoid phages (16), and *tin* encodes a protein that inhibits T4 DNA synthesis by poisoning the T4 single-stranded binding protein (15). These lysogenic conversion genes have an unusual AT content and codon utilization and are believed to have been additions from foreign genomes. This is supported by the fact that genes for hypothetical proteins similar to Old are found in various bacterial genomes in the UniProt GenBank (Table 1), while the Tin protein so far is unique. Furthermore, P2-related phages found in other enterobacteria contain other genes at this locus (4–6, 14, 17, 18). To clarify the nature of this locus in the P2-like coliphages, we have sequenced and characterized the region equivalent to the P2 *tin* and *old* genes, i.e., the TO region, in seven of the P2-like prophages in the ECOR library.

Sequence variation between the *A* genes and the *cos* sites of the prophages. To sequence the region located between the *A* genes and the *cos* sites of the prophages, DNA was extracted from seven strains of the ECOR collection, which are known to contain P2-like prophages (P2-ECnb). A primer located at the

catalytic site of the *A* gene and a primer located to the left of the *cos* sequence were used for the PCR amplifications, and specific amplified DNA fragments of variable length were obtained. The fragments were either sequenced directly or first cloned and then sequenced. To obtain the complete region, plasmid primers and internal primers were used. All strains contained different DNA sequences in this region except P2-EC46 and P2-EC48, which were homologous to each other. The point of sequence divergence varies at the left end but is specific at the right end close to the *cos* site (Fig. 1). In most cases the point of divergence at the left end is within the *A* gene, generating a different C terminus of the A protein. The A proteins of P2-EC5, P2-EC58, and P2-EC67 are also slightly truncated, having five, three, and two fewer amino acids than P2, respectively.

Possible gene content in the variable region. The inserted sequences vary extensively in length (Fig. 2). P2 has the longest insert, about 3 kb, and P2-EC31 the shortest, about 1 kb. All inserts except that of P2-EC31 have a high AT content (62 to 67%), and a search for open reading frames (ORFs), using http:/www.ebi.ac.uk/emboss/transeq, and use of the bacterial translation table resulted in at least one ORF per insert when ORFs encoding proteins shorter than 60 amino acids (aa) were disregarded (Fig. 2). P2-EC31, which has an AT content of 51%, contains two open reading frames whose products show homologies to proteins of other P2-like coliphages and thus does not seem to have any horizontally transferred gene at this locus. Database searches, using the ψ-BLAST programs at http://www.ncbi.nlm.nih.gov/BLAST and FASTAUniProt at http://www.ebi.ac.uk/fasta33, for proteins similar to those encoded by the open reading frames in the other inserts showed that related putative proteins can be found in a variety of bacteria (Table 1). Interestingly, two open reading frames, orf570 of P2-EC30 and orf544 of P2-EC58, showed homology to genes encoding prokaryotic reverse transcriptases (RTs). The two genes show 69% identity with each other (with 100% identity in RT conserved regions), indicating either that they were inserted a very long time ago, leading to sequence divergence, or that they represent two independent integration events. The product of the other encoded open reading frame in P2-EC58 shows no significant identity to any protein in the UniProt GenBank, and a search for similar DNA sequences

* Corresponding author. Mailing address: Department of Genetics, Microbiology and Toxicology, Stockholm University, S-106 91 Stockholm, Sweden. Phone: 46 8 161270. Fax: 46 8 164315. E-mail: Elisabeth .Haggard@gmt.su.se.
† Present address: Isogenica Ltd., Babraham Hall, Babraham, Cambridge CB2 4AT, United Kingdom.

TABLE 1. Database search of ORFs in TO regions of P2 and P2-like prophages in the ECOR collection

Prophage insert	Insert size (bp)	% AT	ORF	FASTA UniProt protein	E score	Accession no.
P2	3,084	65	orf91 (109 aa)	Bacteriophage WΦ, Gg91, 93 aa	1.4E−29	Q7Y487
				Bacteriophage L-413C, 150 aa[a]	2.1E−29	Q858T3
				Escherichia coli, retron EC67, orfD, 100 aa[a]	6.4E−09	P21318
			tin (253 aa)	No significant hits		
			old (587 aa)	*Thermoanaerobacter tengcongensis*, TTE0672, 575 aa	7.2E−13	Q8RBY4
				Methanococcus maripaludis, MMP0332, 626 aa, putative ATP binding[a]	8.4E−08	Q6MOD9
				Vibrio vulnificus, VV10382, 530 aa	8.6E−08	Q8DF42
P2-EC5	3,033	67	orf555	*Nitrosomonas europaea*, NE1882, 429 aa[a]	2.7E−12	Q82TK5
				Listera innocua, Lin0834, 369 aa	1.4E−09	Q92DH9
				Vibrio cholerae, VCA0200, 374 aa[a]	3.9E−09	Q9KMW6
			orf333	No significant hits		
P2-EC30	1,806	66	orf570	*Synechocystis* sp. strain Sll1503, 508 aa, putative reverse transcriptase	6.9E−18	P72998
				Erwinia carotovora, ECA1057, 669 aa, putative reverse transcriptase[a]	1.4E−10	Q6D887
				Burkholderia pseudomallei, BPSL0582, 690 aa, putative reverse transcriptase[a]	2.6E−10	Q63XF5
P2-EC31	954	51	orf151	Bacteriophage L413C, orf91, 150 aa[a]	9.5E−59	Q858T3
				Bacteriophage WΦ, Gp91, 93 aa	5.1E−35	Q7Y487
				Bacteriophage P2, orf91, 109 aa[a]	3.5E−27	Q06426
			orf67	Bacteriophage PSP3, Gp37, 69 aa	1.9E−20	Q6K1F0
				Erwinia carotovora, ECA2626, 108 aa, putative ATP binding, Zn binding, protein metabolism	1.8E−05	Q6D3W7
			orf78	No significant hits		
P2-EC46	2,118	65	orf612	*Pseudomonas aeruginosa*, PA1370, 621 aa[a]	1.5E−63	Q913X4
				Brucella melitensis, BMEII1446, 631 aa[a]	2.3E−36	Q8YCT2
				Brucella abortus, BruAb2_0386, 631 aa[a]	5.9E−36	Q578W9
P2-EC58	2,203	62	orf86	No significant hits		
			orf544	*Synechocystis* sp. strain Sll1503, 508 aa, putative reverse transcriptase	8.2E−18	P72998
				Erwinia carotovora, ECA1057, 669 aa, putative reverse transcriptase[a]	4.2E−11	Q6D887
				Staphylococcus aureus, plasmid ETB, orf24, 590 aa, putative reverse transcriptase	9.5E−11	Q8VVS1
P2-EC67	2,498	62	orf73	*Rhodopseudomonas palustris*, RPA1189, 75 aa[a]	1.6E−04	Q6NAJ6
			orf226	*Escherichia coli* O6, c2409, 243 aa[a]	1.0E−25	Q8FGG8
			orf313	*Burkholderia cenocepacia*, Bcen2424, 351 aa[a]	1.3E−52	Q4LQV1

[a] Annotated as a hypothetical protein.

showed similarities to phages WΦ, PSP3, and 186, indicating that this region is of phage origin. This favors two independent integration events, but a later deletion in P2-EC30 cannot be excluded. Strain ECOR58 has previously been shown to produce multicopy single-stranded DNA (msDNA), and the reverse transcriptase promoting the synthesis of this branched DNA-RNA complex has been identified (13). However, this RT is only distantly related to the RT integrated into the P2-like prophage of strain ECOR58. Also, the unique features of the msDNAs, described by Inouye and Inouye (9), have not been found in P2-EC30 and P2-EC58.

orf570 encodes a functional reverse transcriptase. To determine whether the putative reverse transcriptase of P2-EC30 indeed displayed RT activity, orf570 was cloned in plasmid PCR2.1-TOPO (Invitrogen) so that it was under the control of the T7 promoter (pOTEC30). orf570 was expressed in vitro using the *E. coli* T7 S30 extract system for circular DNA (Pro-

mega) and tested in a Quan-T-RT system (GE Healthcare). The reaction mixture was added to the RT assay mixture containing [³H]TTP and RT DNA-RNA substrate coupled to scintillant. The homogenous Quan-T-RT assay makes use of the scintillation proximity assay principle. Only [³H]TTP nucleotides incorporated by a reverse transcriptase into a biotin-DNA-RNA primer-template linked to streptavidin fluomicrospheres (beads containing scintillant) are close enough to the scintillant to produce light. Unincorporated, tritiated nucleotides, free in solution, are unable to stimulate the scintillant and therefore produce no signal. As can be seen in Fig. 3 the in vitro-produced Orf570 clearly exhibits reverse transcriptase activity, comparable to the activity of 75 units of avian myeloblastosis virus reverse transcriptase in an *E. coli* S30 extract environment. However, its role in P2 phage biology was still unclear. Interestingly, Orf570 of P2-EC30 showed homology to AbiK of *Lactococcus lactis* (E score, 6.9E−3) which is a puta-

```
P2        AAA AAA TTC ACG TAT CCG GTC GCT GAT GAG TGG CCG GGA TTC TCA ACA GTA ATG GAG TGG ACA TGA TGGCAAAAATTCACGAGGTAAAGCT
P2-EC5    ... GG. ... ... ... ... ... ... ... ... ... ... ... .A. ..T GT. .T. TA. G.T AGC GCT .CA G.C..TA.TGC...GTTTT.TAT.GCA
P2-EC30   ... ... ... ... ... .T. ..A .A. ..G ... ... .T. .AG GAT AT. ..G .AT ..T A.. ... .TC .A..AT..TTTAA.T.CTC.G..TAGA
P2-EC31   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ............................
P2-EC46   ... ... ... ... ... AT. ... ... ... ... ... ... ... ... ... ... ... ... ... ... .A..............GC.ATATT..A.
P2-EC58   ... ... ... ... ... ... ... ... ... ... T.. ... ... ... AAC ..A T.A AAA TG. .T.AAATGGTTGA.TTT.CTTCC.CTTG
P2-EC67   ... ... ... ... ... ... ... ... ... ... .GG CAT CA. T.T ... .AA TG. .CTCAT.T.T.CAGT.TT.A.C.GTTT.
```

```
                    --------Primer used--------
P2        CGCCCGCTTCACGGGGCGGTTTTAATG   CACCTTAA  TAAAACTAATGAGCCATCAGTATTTCCAGTCAAAAGCC
P2-EC5                                  ..GT.GC.  ....CA..G.ACAAGCG.GCC.G.CAAGCCT.GCGCA.
P2-EC30   ...........T.A.A...........    ..GT.GC.  .T.TGTCCCGCTC.TCAGT.CTGCG.TCCATCCTGATT
P2-EC31                                  ..GA.GC.  CG..CACGCC.GAAGCG.GCC.GCA....CGGCGCT.A
P2-EC46                                  ..GG.GC.  AT.CCGCTT.TGAG.CA.GCCCG.C.TG.CGG.CGCA.
P2-EC58   .........T.................    ..GT.GC.  .G.CCACTCAACTAGCG.GCC.G.C.TGA.GTCGCCAG
P2-EC67   ...........T..A............    ..GT.GC.  ..G.TACT...GAT..G..CC.G.C.TGAC.GC.C..A
```
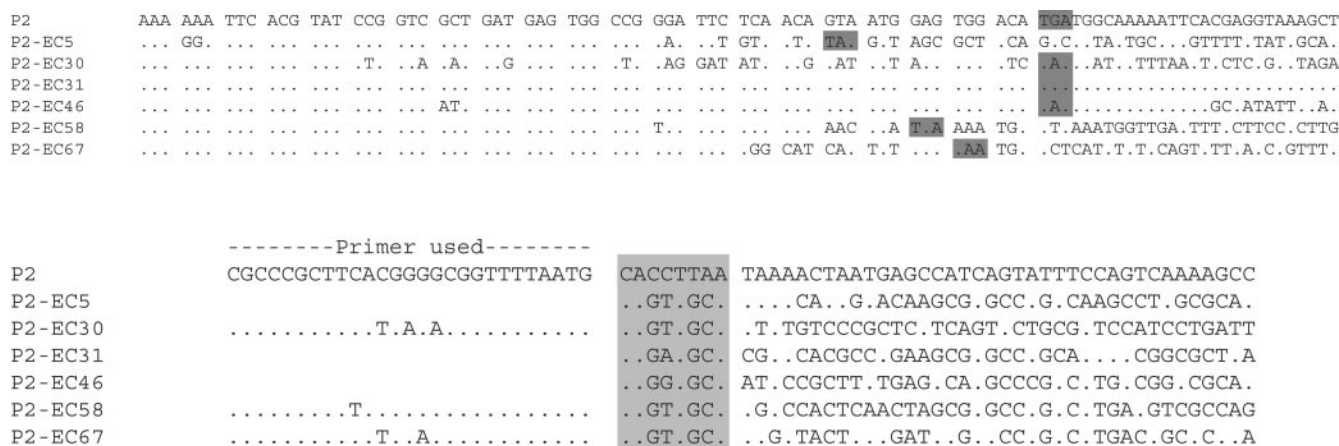
FIG. 1. Alignments of the DNA sequences at the right end of the variable region. Top panel, the sequence at the end of the *A* gene. The stop codon of the *A* gene is indicated by shading. Nucleotides identical to those in P2 are indicated by a dots. Bottom panel, the sequence at the right end, i.e., near *cos*. The point of sequence divergence at the right end is indicated by shading. Nucleotides identical to those in P2 are indicated by dots. For a map of this part of P2, see top line in Fig. 2.

tive reverse transcriptase with demonstrated antiphage activity (2, 7).

Orf570 excludes phage T5. To elucidate the possible antiphage activity of Orf570, the capacities of different phages, i.e., lambda, T2, T4, T5, and T6, all of which are known to utilize the same hosts as P2, to form plaques on lawns of *E. coli* strain BL21(DE3) (22) transformed with plasmid pOTEC30 expressing Orf570 were investigated. All phages except T5 formed plaques with equal efficiency on *E. coli* in the presence or absence of pOTEC30. The plating efficiency of T5 on cells expressing Orf570 was reduced more than 10^7-fold compared to that on cells lacking plasmid pOTEC30. Furthermore no exclusion effect of phage T5 was seen when expression of

Orf570 was down-regulated by transforming BL21(DE3) harboring pOTEC30 with plasmid pLysS (22), which reduces the amount of active T7 RNA polymerase in the cell, since the efficiency of plating was 1.1 compared to cells lacking pOTEC30. To further demonstrate Orf570 as the T5 exclusion agent, an internal deletion of orf570, including the RT conserved YRDD box, was constructed. When the orf570 mutant plasmid, pΔOTEC30, was transformed into BL21(DE3) and exposed to T5, the efficiency of plating was 1.6 compared to cells lacking pOTEC30. The fact that no exclusion of T5 was observed with pΔOTEC30 strongly suggests that Orf570 is responsible for the T5 exclusion activity, possibly through its reverse transcriptase activity.
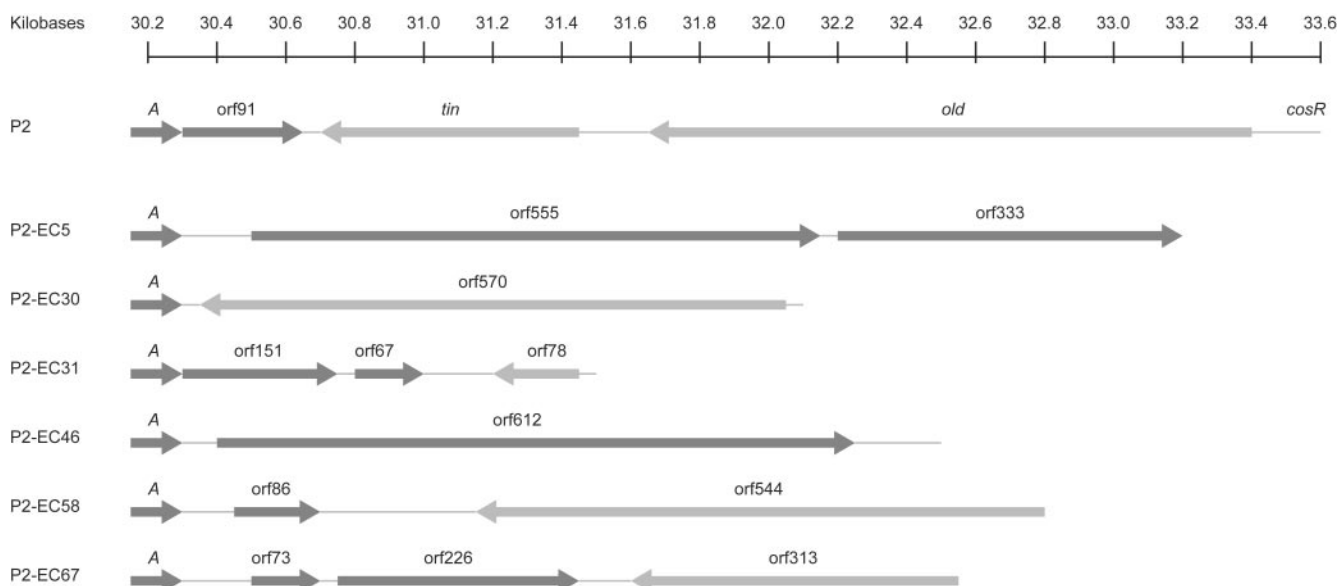

FIG. 2. Schematic drawing of the size and possible gene content of the TO region, equivalent to the P2 orf91-*tin-old* locus (from kb 30.2 to 33.6 in the P2 genome) in P2-like prophages in the ECOR collection. The locations, orientations, and lengths of the open reading frames are indicated by thick arrows. The accession numbers for the TO regions are as follows: P2-EC5, AM157364; P2-EC30, AM157365; P2-EC31, AM157367; P2-EC46, AM157366; P2-EC58, AM157368; and P2-EC67, AM157369.
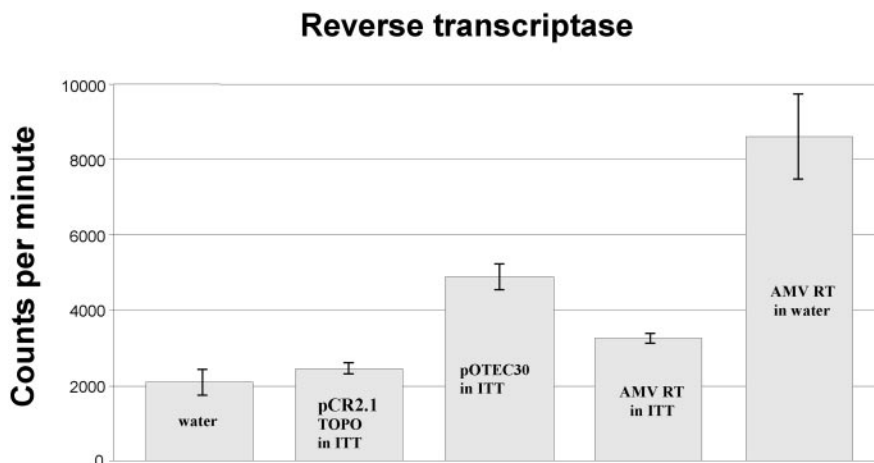
Reverse transcriptase

FIG. 3. Reverse transcriptase activity as measured by the Quan-T-RT assay system. Fifty microliters of in vitro-coupled transcription and translation (ITT) reaction mixture, containing 3 μg pOTEC30 or 3 μg pCR2.1-TOPO, was used per 200 μl of Quan-T-RT assay mixture. Negative controls were water only and ITT mix with pCR2.1-TOPO. Positive controls were 75 units of recombinant avian myeloblastosis virus (AMV) RT in ITT mix or water. The RT assay mixtures were incubated for 1 hour at 37°C. Results are means ± standard deviations ($n = 3$).

Although the results are based on a small sample, it seems likely that the region downstream of A is a very old location for nonessential and presumably horizontally transferred genes, since it is the only defined region of this kind that can be found at the same place in all P2-like phages irrespective of origin. The age of the region may explain why we have not been able to find any signature in the surrounding sequences that would imply an insertion mechanism. Most of these genes have not been characterized, but *Haemophilus influenzae* phages HP1 and HP2 carry an adenine methylase gene and coliphage WΦ carries a cytosine methylase gene (5, 17), and the nonessential genes *tin* and *old* lie at this position in phage P2. There are also defective P2-like coliphages in *E. coli* that carry a cassette containing a restriction endonuclease and a methyltransferase gene at this site (10). Based on the function of the genes that have been characterized and the similarity of genes in this region to genes encoding non-phage-associated hypothetical proteins in bacteria, it seems likely that the majority of genes are lysogenic conversion genes of key importance for bacteria. This is also evidently supported by the phage T5 exclusion activity of orf570 from P2-EC30. To our knowledge, this is the first report showing phage exclusion activity by a protein with demonstrated reverse transcriptase activity, although this has already been suggested by Fortier et al. (7), who showed that the presence of an RT motif in AbiK was critical for its antiphage activity. AbiK, which share 16% identity with Orf570, is an abortive infection system encoded by a lactococcal plasmid, and the molecular target of AbiK has been suggested to be phage DNA single-strand annealing proteins involved in recombination activities (2). Possibly, Orf570 utilizes a similar phage exclusion mechanism. Thus, the next challenge would be to pinpoint by what means Orf570 inhibits T5 growth.

Considering the genetic variation reported here, the variation at the Z region reported in a previous paper (19), and the variation of other characterized genes in P2-like coliphages, the role for P2-like coliphages in the evolution of the host seems to be to supply lysogenic conversion genes that exclude other phages. This increases the fitness for a P2-like prophage

and makes it possible to reside in the lysogenic state. Other genes, such as genes associated with virulence, will also increase the fitness of the lysogen, but the lysogen will still be vulnerable to attacks from lytic phages that would reduce or eradicate not only the population of bacteria but also any lysogenic phage that may be integrated in the bacterial genome. Prophages that protect the host by encoding factors that exclude all foreign DNA, such as restriction/modification systems, are expected to be found in most environments, but it seems reasonable to assume that specific phage exclusion genes are correlated to the most abundant phages in that particular environment.

REFERENCES

1. **Bertani, G.** 1958. Lysogeny. Adv. Virus Res. **5:**151–193.
2. **Bouchard, J. D., and S. Moineau.** 2004. Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. J. Bacteriol. **186:**3649–3652.
3. **Calendar, R., S. Yu, H. Myung, V. Barreiro, R. Odegrip, K. Carlson, L. Davenport, G. Mosig, G. E. Christie, and E. Haggård-Ljungquist.** 1998. The lysogenic conversion genes of coliphage P2 have unusually high AT content, p. 241–252. *In* M. Syvanen and C. I. Kado (ed.), Horizontal gene transfer. Chapman & Hall, London, United Kingdom.
4. **Elliott, J. M., A. A. Filippov, V. V. Kutyrev, A. G. Bobrov, O. A. Kirillina, V. L. Motin, P. S. Chain, and E. Garcia.** Unpublished data.
5. **Esposito, D., W. P. Fitzmaurice, R. C. Benjamin, S. D. Goodman, A. S. Waldman, and J. J. Scocca.** 1996. The complete nucleotide sequence of bacteriophage HP1 DNA. Nucleic Acids Res. **24:**2360–2368.
6. **Esposito, D., B. J. Schmidt, F. R. Bloom, and G. E. Christie.** Unpublished data.
7. **Fortier, L.-C., J. D. Bouchard, and S. Moineau.** 2005. Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. J. Bacteriol. **187:**3721–3730.
8. **Haggård-Ljungquist, E., V. Barreiro, R. Calendar, D. M. Kurnit, and H. Cheng.** 1989. The P2 phage *old* gene: sequence, transcription and translational control. Gene **85:**25–33.
9. **Inouye, S., and M. Inouye.** 1996. Structure, function, and evolution of bacterial reverse transcriptase. Virus Genes **11:**81–94.
10. **Kita, K., H. Kawakami, and M. Tanaka.** 2003. Evidence for horizontal transfer of the *Eco*T38I restriction-modification gene to chromosomal DNA

by the P2 phage and diversity of defective P2 prophages in *Escherichia coli* TH38 strains. J. Bacteriol. **185:**2296–2305.

11. **Lederberg, S.** 1957. Suppression of the multiplication of heterologous bacteriophages in lysogenic bacteria. Virology **3:**496–513.

12. **Lindahl, G., G. Sironi, H. Bialy, and R. Calendar.** 1970. Bacteriophage lambda: abortive infection of bacteria lysogenic for phage P2. Proc. Natl. Acad. Sci. USA **66:**587–594.

13. **Mao, J.-R., S. Inouye, and M. Inouye.** 1997. msDNA-Ec48, the smallest multicopy single-stranded DNA from *Escherichia coli*. J. Bacteriol. **179:** 7865–7868.

14. **McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson.** 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852–856.

15. **Mosig, G., S. Yu, H. Myung, E. Haggård-Ljungquist, L. Davenport, K. Carlsson, and R. Calendar.** 1997. A novel mechanism of virus-virus interactions: bacteriophage P2 Tin protein inhibits phage T4 DNA synthesis by poisoning the T4 single-stranded DNA binding protein, gp32. Virology **230:** 72–81.

16. **Myung, H., and R. Calendar.** 1995. The *old* nuclease of bacteriophage P2. J. Bacteriol. **177:**497–501.

17. **Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi.** 1999. The complete nucleotide sequence of φCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal transfer via bacteriophages. Mol. Microbiol. **31:**399–419.

18. **Nesper, J., J. Blass, M. Fountoulakis, and J. Reidl.** 1999. Characterization of the major control region of *Vibrio cholerae* bacteriophge K139: immunity, exclusion, and integration. J. Bacteriol. **181:**2902–2913.

19. **Nilsson, A. S., J. L. Karlsson, and E. Haggård-Ljungquist.** 2004. Site-specific recombination links the evolution of P2-like coliphages and pathogenic enterobacteria. Mol. Biol. Evol. **21:**1–13.

20. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.

21. **Sironi, G., H. Bialy, H. A. Lozeron, and R. Calendar.** 1971. Bacteriophage P2: interaction with phage lambda and with recombination-deficient bacteria. Virology **46:**387–396.

22. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.