7110–7123 *Nucleic Acids Research, 2009, Vol. 37, No. 21* *Published online 22 September 2009*
doi:10.1093/nar/gkp769

Phage T4 *mobE* promotes *trans* homing of the defunct homing endonuclease *I-TevIII*

Gavin W. Wilson and David R. Edgell*

Department of Biochemistry, Schulich School of Medicine & Dentistry, The University of Western Ontario, London, ON, N6A 5C1, Canada

Received July 24, 2009; Revised August 17, 2009; Accepted September 1, 2009

ABSTRACT

Homing endonucleases are site-specific DNA endonucleases that typically function as mobile genetic elements by introducing a double-strand break (DSB) in genomes that lack the endonuclease, resulting in a unidirectional gene conversion event that mobilizes the homing endonuclease gene and flanking DNA. Here, we characterize phage T4-encoded *mobE*, a predicted free-standing HNH family homing endonuclease. We show that *mobE* is promoterless and dependent on upstream transcription for expression, and that an internal intrinsic terminator regulates *mobE* transcript levels. Crucially, *in vivo* mapping experiments revealed a MobE-dependent, strand-specific nick in the non-coding strand of the *nrdB* gene of phage T2. An internal deletion of the predicted HNH catalytic motif of MobE abolishes nicking, and reduces high-frequency inheritance of *mobE*. Sequence polymorphisms of progeny phage that inherit *mobE* are consistent with DSB repair pathways. Significantly, we found that mobility of the neighboring *I-TevIII*, a defunct homing endonuclease encoded within a group I intron interrupting the *nrdB* gene of phage T4, was dependent on an intact *mobE* gene. Thus, our data indicate that the stagnant *nrdB* intron and *I-TevIII* are mobilized in *trans* as a consequence of a MobE-dependent gene conversion event, facilitating persistence of genetic elements that have no inherent means of promoting their own mobility.

INTRODUCTION

Bacteriophage genomes are littered with homing endonucleases, site-specific DNA endonucleases that function as mobile genetic elements by cleaving at cognate sites (the homing site) in genomes that lack the endonuclease (1). Homing endonucleases are a significant source of genomic variability in phage, as repair of DNA damage introduced by the endonuclease leads to inheritance of the endonuclease gene and flanking genetic markers, essentially shuffling large segments of DNA between genomes. Indeed, a number of homing endonucleases are responsible for a phenomenon called marker exclusion in T-even and HMU phage, whereby genetic markers of the donor (endonuclease-containing) phage predominate in the progeny of crosses with recipient (endonuclease-lacking) phage because of gene conversion events associated with repair of endonuclease-mediated double-strand breaks (DSBs) (2–4). For example, in progeny resulting from co-infection of *Escherichia coli* with phages T2 and T4, the strongest regions of exclusion of T2 markers correspond to the cleavage sites of T4-encoded homing endonucleases (2,5,6). Homing endonucleases are often encoded within group I introns and inteins but in phage genomes are commonly found as so-called free-standing endonucleases because the endonuclease is not obviously encoded within a self-splicing element. In particular, phage T4 encodes fifteen homing endonucleases, 12 of which are free-standing (7).

Phage T4 has served as a model system for studying homing endonuclease mobility with most efforts focused on I-TevI, encoded within a group I intron interrupting the thymidylate synthase (*td*) gene. Detailed analyses of I-TevI-mediated homing pathways have revealed that homing occurs within the context of recombination-dependent replication, and relies on the activities of a number of T4 recombination/repair and replication proteins (8–10). A crucial feature of the homing pathway is the introduction of a DSB by I-TevI that serves as a substrate for recombination-dependent repair and subsequent intron homing (11). Similar mobility pathways are likely used by free-standing endonucleases of the *seg* (*s*imilar to *e*ndonucleases of *g*roup I introns) family in T4 (2,6,12–14). *Seg* endonucleases, like I-TevI, are GIY−YIG homing endonucleases that introduce a DSB at their homing sites (2,6,12,13,15). In contrast, five of the 12 free-standing endonucleases in phage T4 belong to a group of related genes termed *mob* (*mob*ility)

*To whom correspondence should be addressed. Tel: +1 519 661 3133; Fax: +1 519 661 3175; Email: dedgell@uwo.ca

endonucleases (16). All Mob proteins share a conserved N-terminal HNH nuclease domain, whereas the C-terminal domain (presumably the DNA-binding domain) is specific to each protein. The presence of an HNH domain suggests that the *mob* enzymes are nicking endonucleases, as many characterized HNH-containing endonucleases nick their target substrates rather than introducing a DSB (17,18). Nonetheless, free-standing and intron-encoded HNH homing endonucleases in various phage systems promote homing to phage lacking the intron or endonuclease. In *Bacillus* HMU phage, mobility of the intron-encoded I-HmuI and I-HmuII endonucleases was associated with inheritance of flanking genetic markers, suggesting that intron mobility proceeded through a DSB-repair pathway (3,19). However, it is not known if the nick introduced by the HNH endonuclease is processed to a DSB by host or phage-encoded proteins to initiate a repair event, or if other mechanism(s) are used to promote mobility. Similarly, it is not known if the *mob* endonucleases of phage T4 are nicking enzymes, and if they are mobile endonucleases that promote mobility through DSB-repair pathways.

In contrast to the detailed studies on mobility pathways, comparatively little attention has focused on the regulation of endonuclease function within the context of the T4 infective cycle. The three intron-encoded endonucleases, I-TevI, I-TevII and I-TevIII, are all subject to translational regulation that limits expression to late times after infection (20), and I-TevI also functions as a transcriptional autorepressor by binding to an operator site that overlaps the late promoter that drives I-TevI expression (21). However, regulatory mechanisms that control expression of the twelve free-standing endonucleases in phage T4 are unknown. Because free-standing endonucleases are inserted in the intergenic region between genes, free-standing endonucleases have the potential to influence transcription of surrounding genes by displacing promoters or other regulatory elements of neighboring genes. Thus, a free-standing endonuclease must successfully integrate into the T4 transcriptional program to minimize any transcriptional impact on surrounding host genes.

To resolve the issue of whether the *mob* genes encode site-specific endonucleases that function as mobile elements, and to investigate how *mob* genes are integrated into the transcriptional program of T4, we have focused here on *mobE* that is encoded within the genomes of a number of T-even phage (7,22). In T4 and related phage, *mobE* is inserted between the genes for the large (*nrdA*) and small (*nrdB*) subunits of aerobic ribonucleotide reductase. The *nrd* genomic region of T-even phage is a hotspot of endonuclease insertion, as evidenced by numerous homing endonuclease insertions. One relevant insertion is that of I-TevIII, an HNH homing endonuclease encoded within a self-splicing group I intron that interrupts the *nrdB* gene of phage T4, but not the closely related T2 (23) (Figure 1). An internal deletion within I-TevIII renders the endonuclease non-functional and unable to promote homing of its host intron; however, a related phage, RB3, contains an intact *I-TevIII* gene that functions to promote intron mobility (24,25). Intriguingly, previous studies showed that *I-TevIII* and the group I intron is inherited at high frequency in the progeny of crosses between T4 and T2 (6). To account for this discrepancy between lack of I-TevIII activity and high-frequency inheritance, it was suggested that the neighboring endonuclease *mobE* (~700 bp upstream) promotes *I-TevIII* and intron mobility by cleaving within the T2 *nrdB* gene, facilitating inheritance in progeny phage by recombination of sequences flanking the cleavage site.

Here, we show that *mobE* is promoterless, and dependent on upstream promoters for expression. Moreover, an intrinsic terminator within the *mobE*-coding region limits *mobE* transcript abundance. We provide evidence that *mobE* is a mobile endonuclease with a target site in the 5′-end of the *nrdB* gene of phage T2. Crucially, we have shown that a functional *mobE* gene product is required for a strand-specific nick in the T2 *nrdB* gene, and that the nick is recombinogenic as sequencing of progeny phage from a mixed infection of T4 and T2 revealed co-conversion of sequence flanking the nick site. Significantly, high-frequency inheritance of the defunct I-TevIII was absolutely dependent on a functional *mobE*. Our results are consistent with a model whereby *I-TevIII* and its host intron are mobilized in *trans*, thus providing a mechanism by which the non-mobile intron can persist in phage populations despite having no inherent means of promoting mobility.

MATERIALS AND METHODS

Bacterial and phage strains

Bacteria and phage are listed in Table 1. *Escherichia coli* DH5α was used for plasmid manipulations, while *E. coli* sup^0 and *supF* were used for propagation of phage (6). *E. coli* strains were grown in LB medium supplemented with the appropriate antibiotics (100 μg/ml ampicillin or 50 μg/ml kanamycin).

Mutagenesis of phage T4

Mutagenesis was performed by plasmid-to-phage recombination. Plasmids containing the desired mutations were generated by SOEing PCR reactions (26). Two gene-specific segments were amplified and used as templates for a subsequent SOEing PCR reaction. Each construct was assembled such that ~500 bp of flanking sequence was included to enhance recombination. PCR reactions were performed with PWO polymerase (Roche) following the manufacturers protocol with 25 cycles and the templates and primers listed in Table S1. The SOEing products were phosphorylated using OptiKinase (USB Corporation) and ligated into dephosphorylated pCRZap using T4 DNA ligase (New England Biolabs) following the manufacturers protocol. The ligation mixtures were transformed into DH5α, and plasmids sequenced to confirm the mutation. Plasmids for mutagenesis were transformed into *E. coli* sup^0, grown to an OD_{600} of 0.4 and infected with T4 at an multiplicity of infection (m.o.i.) of 4. The cells were incubated for 10 min at 37°C then diluted 10^{-4} into

Table 1. Bacteria, phage and plasmids used in this study

Bacteria	Genotype	Source
E. coli DH5α	*fhuA2* Δ(argF-lacZ)U169 *phoA glnV44* Φ80 Δ(lacZ)M15 *gyrA*96 *recA1 relA1 endA1 thi-1 hsdR17*	Invitrogen
E. coli B40 *sup*^{O}	*F*$^-$, *argI140(am)*, *str*^{R}	J.D. Smith
E.coli B40 *supF*	*F*$^-$, *argI140(am)*, *str*^{R}, *sup*^{F}	J.D. Smith
Phage		
T4D	Wildtype	
T2L	Wildtype	
T4K10	*am38 am51 denA denB*	Selick *et al.* 1988
T4 Δ*mobE33-62*	Deletion of amino acids 33–62 of *mobE*, removing HNH motif	This study
Plasmids		
pCR2.1	TA cloning vector	Invitrogen
pCRZap	pCR2.1 with a cloned in StuI site (DE-369/DE-370) using EcoRI	This study

Table 2. Phage-to-phage crosses

Phage	Cross	*mobE*/input[a]	*I-TevIII*/input	*I-TevI*/input
T2 × T4	1	1.8	1.8	1.9
	2	2.0	2.0	2.0
	3	2.3	2.3	2.4
		2.1 ± 0.2[b]	2.0 ± 0.3	2.1 ± 0.3
T2 × T4Δ*mobE33-62*	1	1.4	1.2	1.9
	2	1.2	1.9	2.3
	3	1.8	ND	ND
		1.5 ± 0.3	1.6 ± 0.4	2.1 ± 0.3

[a]The number of progeny phage positive for each endonuclease gene is expressed as ratio of the observed number of progeny phage positive by hybridization with endonuclease-specific probes relative to the expected number of endonuclease-positive phage (the number of T4 input phage for each independent cross). See Supplementary Table S2 for raw data.
[b]Average and standard deviation calculated for each set of endonuclease/input ratios.

pre-warmed LB medium and grown for an additional 3 h, treated with chloroform and incubated at room temperature overnight. The lysates were plated on *sup*^{0} cells, and screened for the mutation of interest by Southern blot analysis using plaque lifts. Hybridizing plaques were cored, grown and sequenced.

Isolation of phage DNA

Approximately 1 ml of a 1×10^{10}-pfu/ml phage stock was precipitated as previously described (27), except that PEG6000 was used in place of PEG8000. Phage DNA was isolated using a Sigma gDNA kit (Sigma), following the gram-negative bacterial genomic DNA isolation protocol. Isolation of DNA from mixed phage infections was performed as above, except that 1 ml of cells was collected prior to phage infection, and at 10, 20 and 30 min post-infection. The culture was pelleted, flash frozen and stored at −80°C. Cell pellets were resuspended in 300 µl of lysis solution (100 mM NaCl, 50 mM Tris−HCl pH 7.2, 1% SDS, 10 mM EDTA, 5 mM $CaCl_2$ and 0.3 mg/ml proteinase K), and incubated at 65°C for 1 h followed by three phenol:chloroform extractions. The aqueous layer was treated with 1 µl RNase A/T1 (Ambion) for 10 min at room temperature, and DNA was precipitated with ethanol and resuspended in 100 µl TE buffer.

Mixed phage infections and mobility assays

Mobility assays were performed using mixed phage infections. *E. coli sup*^{0} or *supF* was grown to an OD_{600} of 0.4 and infected with equal amounts of T2 and T4 at a m.o.i. of 4. The infections were incubated at 37°C for 10 min, and subsequently diluted 1:10^4 into pre-warmed LB medium and grown at 37°C for 4 h. Chloroform was added to each culture, left overnight at room temperature, and then stored at 4°C. To screen for *mobE* and other endonucleases by oligonucleotide hybridization, progeny were plated on *sup*^{0} cells and transferred to positively charged nylon membranes (Millipore) by placing the membrane directly onto HA-top agarose LB plates for 3–5 min. The membrane was soaked in alkaline transfer buffer (1 M NaCl, 0.4 N NaOH) for 2 min, and then neutralization buffer (1 M NaCl, 0.5 M Tris−HCl pH 7.2) for 5 min. The membranes were air-dried, UV cross-linked, and prehybridized in 0.5 M sodium phosphate pH 7.2, 1 mM EDTA pH 8.0, 7% SDS, 1% bovine serum albumin for 2 h at 68°C. Fresh buffer was then added with the end-labeled oligonucleotide (1×10^6 c.p.m.). Oligonucleotides were end labeled with [γ^{32}P] ATP (PerkinElmer) using Optikinase following the manufacturer's directions (USB Corporation). The unincorporated nucleotide was removed using QIAGEN nucleotide removal columns. The membranes were hybridized overnight at 50°C for *mobE* probes, 47°C for *I-TevIII* probes, and 50°C for *I-TevI* probes. The membranes were then washed at the hybridization temperature in phosphate−SDS wash I buffer (40 mM sodium phosphate, 1 mM EDTA, 5% SDS, 0.5% BSA) for 5 min, and then with phosphate−SDS wash II buffer (40 mM sodium phosphate, 1 mM EDTA, 1% SDS) for 5 min, and rinsed with 0.1 × SSC. The membranes were dried briefly, and exposed to a phosphoimager screen (GE Healthcare).

Sequencing of progeny phage

Genomic DNA was isolated from plaques that were *mobE*$^+$/*I-TevIII*$^-$ or *mobE*$^+$/*I-TevIII*$^+$. The *nrdA* region was amplified using primers DE-416/DE-421, and the *nrdB* region was amplified using primers DE-422/DE-488. PCR products were directly sequenced using the same primers, and two additional primers (DE-423 and DE-424) were used to sequence the *I-TevIII*$^+$ phage. The sequences were aligned using ClustalW (28) and a user-generated program created (available upon request) to display the nucleotide polymorphism graph shown in Figure 5.

RNA extraction and purification

Approximately 8.2×10^7 cells/ml of *E. coli sup*^{0} was infected with T4 at an m.o.i of 8. Samples (6 ml) were immediately suspended in equal volumes of chilled RNA*later* and stored on ice (Ambion). RNA was extracted using a QIAGEN RNeasy minikit following the manufacturers protocol. The RNA was eluted in 80 µl of RNase-free water, and then treated with TURBO DNase (Ambion) following the manufacturers protocol. The treated RNA was ethanol precipitated and resuspended in 50 µl RNase-free water and stored at −80°C. The RNA samples were quantified using a spectrophotometer and the concentration was normalized to 2 µg/µl.

RT-PCR

RT-PCR was performed using 8 µg of total RNA and 20 pmol of DE-367. The RNA and primer mixture was heat denatured and snap cooled on ice. Reaction mixtures were incubated at 42°C for 1 h, and 75°C for 10 min with M-MuLV reverse transcriptase (New England Biolabs), with 0.5 µl RNAsin (Promega), 1.25 mM dNTPs, and the manufacturers supplied buffer. The RT products were purified using PCR cleanup columns (QIAGEN). PCR amplification was performed using 5 µl of the RT product as template, with primer pairs specific for *nrdA* or *mobE* (DE-365/DE-367 and DE-387/DE-367), and Taq DNA polymerase (New England Biolabs). Cycling conditions were 94°C for 30 s, 53.3°C for 30 s, and 72°C for 1 min, with 30 cycles. The amplified products were resolved on 1% TAE-agarose gels.

5′-RLM RACE

RNA-ligase mediated rapid amplification of 5′ cDNA ends was as described with minor modifications (29). The primers for reverse transcription were DE-415 for *nrdA* and DE-368 for *mobE*. The *nrdA*-specific nested primers were DE-415/DE-196 and DE-414/DE-197 with an annealing temperature of 56°C for 30 cycles. The *mobE*-specific nested primers were DE-368/DE-196 and DE-367/DE-197 with an annealing temperature of 53.3°C for 30 cycles

3′-RLM RACE

3′RACE was adapted from Gibb and Edgell (29). Primer DE-194 was used for reverse transcription and the first nested PCR amplification was performed with primers DE-366/DE-194. The second nested amplification used primers DE-387/DE-195 with annealing temperatures of 53.3°C for 30 cycles.

RNase protection assays

PCR products corresponding to a 226 bp region centered on the *mobE* transcriptional terminator were generated with DE-352 and DE-353. DE-352 included 13 nt of non-homologous sequence and a T7 RNA polymerase promoter, while DE-352 contained 13 nt of non-homologous sequence. The PCR product was gel purified and used as template for *in vitro* transcription with T7 RNA polymerase (New England Biolabs) and [α^{32}P] UTP (PerkinElmer). The RNA probe was gel-purified, and RNase protection assays were performed following the manufacturers protocol with a 1/50 dilution of RNAse A/T1 (Ambion). The digested products were electrophoresed on a 8%-denaturing polyacrylamide gel, and exposed to a phosphoimager screen (GE Healthcare).

Northern hybridization

Northern hybridzations were performed as previously described with the following adaptations (29). An end-labeled 100 bp ladder was loaded on each gel, and membranes were UV crosslinked prior to hybridization. The *mobE*-specific probe was generated by PCR with DE-365/DE-367 and the *nrdA* specific probe was generated using primers DE-363/DE-364. The PCR products were gel purified using a QIAGEN gel purification kit, and used as template for nick translation using the Invitrogen nick translation kit supplemented with [α^{32}P] dCTP (PerkinElmer). Each northern hybridization used 5×10^6 c.p.m. of the nick-translated probe for hybridization. The membranes were washed in 100 ml of 0.1 × SSC, 0.2% SDS for 10 min at 46°C and three times in 100 ml of 0.5 × SSC, 0.2% SDS for 10 min at 68°C. The membranes were then dried and exposed on a phosphoimager screen.

Primer extension assays

To map strand-specific nicks in phage DNA, 1/20th of the gDNA isolated from mixed phage infections was used as template in 20-µl primer extension reactions. The reactions contained Taq polymerase and the manufacturers supplied buffer (New England Biolads), 1.25 mM dNTPs, and 5 pmol of end-labeled oligonucleotide. The annealing temperatures used were 46.6°C for DE-465, and 60°C for DE-522. The reactions were cycled in a thermocycler with the following parameters: denaturation at 94°C for 30 s, annealing for 30 s at temperatures specific for each primer, and 30 s extension at 72°C for 30 (plasmid) or 40 (gDNA) cycles. For primer extension assays on plasmid templates, *E. coli sup*^{0} bearing pT2nrdB was grown to an OD_{600} of 0.4, and infected with T4K10 at an m.o.i. of 4. Samples (20 ml) samples were collected prior to infection, and at 5, 10 and 15 min post-infection. Cells were pelleted at 5000 × g for 5 min, flash frozen, and stored at −80°C. Pellets were thawed and plasmid DNA isolated using the Biobasic miniprep kit. To map the nicking site, a sequencing ladder was generated using a cycle sequencing kit (USB Corporation). The reactions contained 1.5 pmol end-labeled DE-465, pT2nrdB, and were cycled at 94°C for 30 s, 46.6°C for 30 s, and 72°C for 30 s for 51 cycles. The reactions were loaded on an 8% denaturing polyacrylamide gel.

Sequence comparison of T-even phage

The *nrdB* nucleotide sequences of T-even were divided into two groups, phage containing *mobE* and phage without *mobE*. The region corresponding to the *mobE*-nicking site was aligned in both groups using ClustalW (28), and a sequence logo generated (30). The accession

numbers or database sources for the sequences are as follows: T4 (NC_000866), T2 (AY262132.2/phage.bioc.tulane.edu), T6 (AY262134.2), RB2 (DQ178120.1), RB3 (AY262130.2), RB14 (AY262126.2), RB15 (DQ178121.1), RB27 (AY262128.2), RB32 (DQ904452.1), RB51 (phage.bioc.tulane.edu), RB61 (AY262131.2), RB69 (NC_004928.1), LZ7 (DQ178119.1), JS10 (NC_012741.1) and JS98 (NC_010105.1).

Oligonucleotides

Oligonucleotides are listed in Table S1.

RESULTS

mobE is co-transcribed with *nrdA*

As a first step to determining the function of *mobE*, we examined the transcriptional regulation of the endonuclease gene. Bioinformatic analysis of the *nrdA-nrdB* region of phage T4 failed to identify obvious T4 promoters immediately upstream of *mobE* (Figure 1). However, characterized early, middle and late promoters lie upstream of *nrdA* (31), suggesting that *mobE* may be part of a transcriptional unit that includes *nrdA*. To investigate the transcriptional regulation of *mobE*, we isolated total RNA at various time points post-infection for use in RT-PCR. Using primers specific to *mobE*, we amplified RT-PCR products at 3 min post-T4 infection and all times thereafter (Figure 2A and B, left panel). A separate RT-PCR experiment utilizing a set of primers spanning the *nrdA*/*mobE* junction revealed a similar pattern, with products amplified from RNA sampled at all times post-3 min infection (Figure 2B, right panel).

The RT-PCR data suggested that *mobE* and *nrdA* are co-transcribed. To confirm that *nrdA* and *mobE* are present on a common transcript, we performed northern blot analyses using *mobE* and *nrdA* specific probes that detected the presence of a ~4 kb hybridizing band, consistent with a polycistronic message that included *nrdA* and *mobE* (Figure 2C). The hybridizing band may also include transcripts that initiated upstream of the *td* gene and contain *nrdA* but not *mobE*, because we mapped a transcriptional terminator in the 5′-end of *mobE* (Figure 3). A smaller 3 kb fragment was also detected with both probes (Figure 2B and C), suggesting the presence of a transcript that terminated at the *mobE* terminator (Figure 3). To map potential transcription initiation sites, we used 5′ RLM-RACE with RNA pooled from early (3–7 min) or late (9–15 min) time points in separate reactions with primers specific for *nrdA* or *mobE*. With *nrdA*-specific primers, we detected a single amplification product of sufficient size to correspond to transcripts that initiated upstream of *nrdA*. We did not, however, detect products that corresponded to transcripts that initiated immediately upstream of *mobE* (data not shown). The 5′ RLM-RACE products were cloned and sequenced, and the 5′-ends mapped to a previously characterized middle promoter upstream of *nrdA* (31). Collectively, these results provide evidence that *mobE* lacks a promoter immediately upstream of its coding region and is dependent on promoters upstream of *nrdA*, and that the *nrdA* and *mobE* genes are part of a polycistronic message.

A hairpin internal to *mobE* acts as a transcriptional terminator

In the course of analyzing the *mobE* sequence for transcriptional regulatory elements, we noted the presence of a stem-loop structure 27-nt downstream of the T4 *mobE* AUG codon (Figure 3). This stem-loop structure has characteristics of Rho-independent

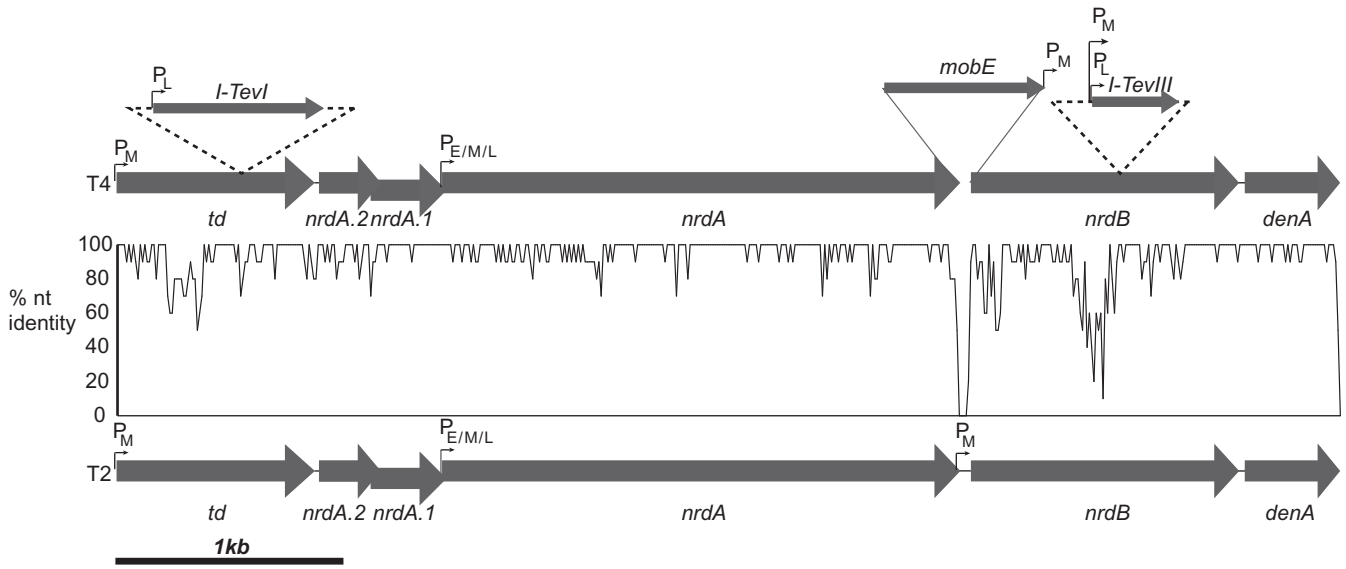

Figure 1. Overview of the genomic organization of the *nrd* region of phages T4 and T2. Shown is a plot of nucleotide identity over a 10-nt sliding window. Genes are indicated by rectangles with right-facing arrowheads, with italicized gene names indicated underneath. Dashed lines indicate self-splicing group I introns, while the solid line indicates the T4-specific *mobE* insertion. Early (E), middle (M) and late (L) promoters are indicated by right facing arrows.

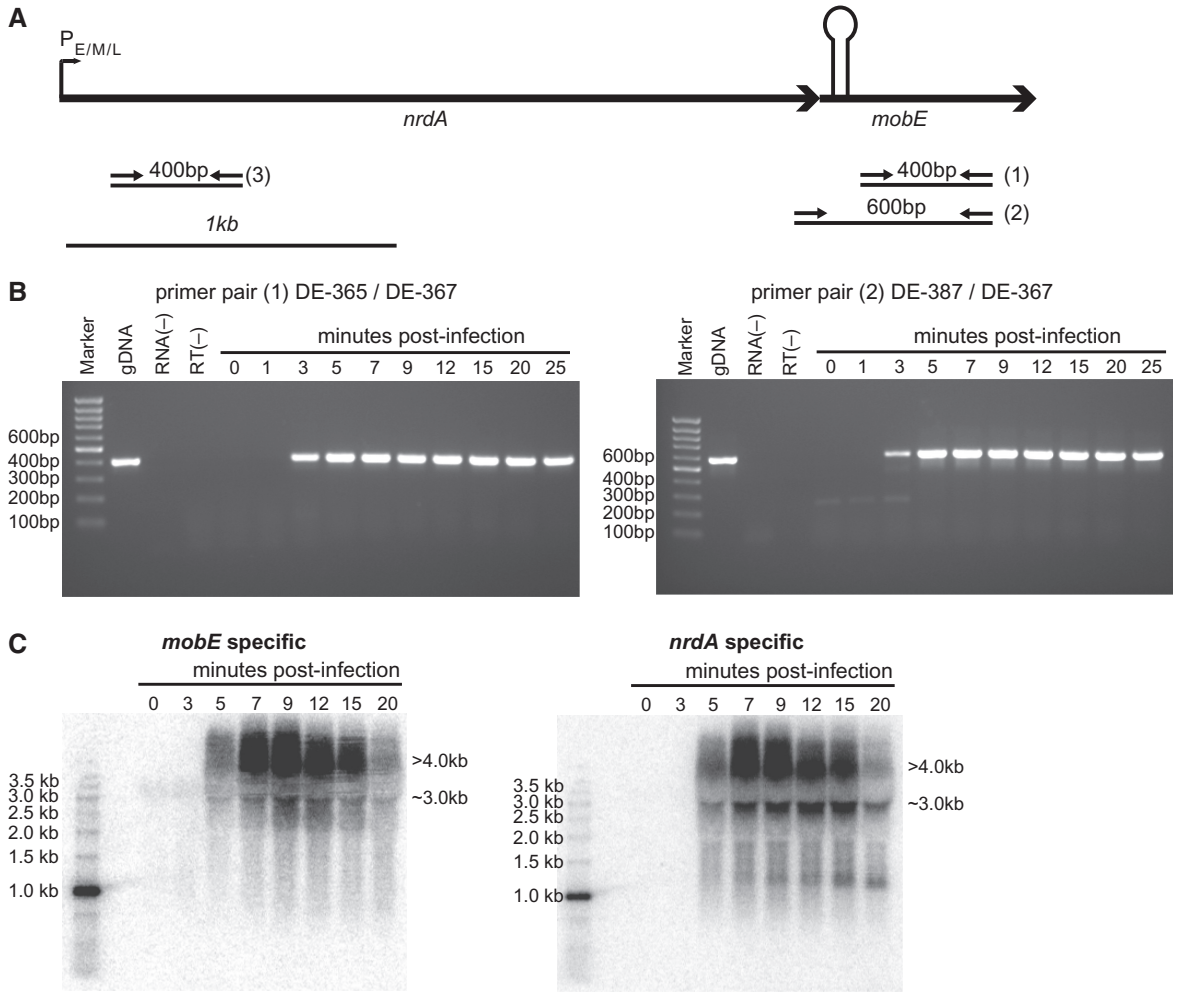

Figure 2. *mobE* and *nrdA* are co-transcribed. (**A**) Schematic of the *nrdA* and *mobE* genes of phage T4, with promoters and putative RNA hairpin sequences indicated. The position of primers used for RT-PCR or northern blot analyses are indicated by arrows, and the length of the products indicated. (**B**) RT-PCR analyses with primers specific to the *mobE* gene (left panel) or to the *nrdA*/*mobE* junction (right panel). gDNA, PCR with genomic DNA; RNA(-), RT-PCR reaction without added RNA; RT(-), reaction without added reverse transcriptase. (**C**) Northern blot analyses using probes specific to *mobE* (left panel) or to *nrdA* (right panel). Gene-specific probes were generated using primers (1) and (3) as indicated in panel (A) and listed in Table S1.

(intrinsic) transcriptional terminators, including a 9-nt stem, a 4-nt loop and a poly(U) tract following the stem (Figure 3B). Previous studies have noted the presence of a potential stem-loop structure, but did not test if it functioned as a terminator (31,32). To test if the terminator was functional, we used 3′ RLM-RACE to map the 3′-ends of transcripts. We sampled RNA at early (5 min) and late (15 min) time points, and found that the sequences of 7 of 11 3′ RACE clones mapped to a termination event at the poly(U) tract immediately following the hairpin (Figure 3B). The sequences of the remaining clones mapped to various positions downstream of the terminator, most likely representing transcriptional read through products.

To estimate the efficiency of the transcriptional terminator, we used a RNase protection assay with a 252-nt probe centered on the terminator (Figure 3A). In this assay, the probe contained 26 nt of non-homologous sequence that would not anneal to T4 transcripts, and thus be a substrate for digestion by RNases. This feature of the probe served as an internal control, allowing us to distinguish undigested full-length probe (252 nt) from transcriptional read through (226 nt) and termination products (~125 nt based on 3′RLM-RACE data). As seen in Figure 3C, the amount of termination at early time points was ~30%. At later time points, termination increased to 50% (Figure 3C, 15 and 20 min) reflecting an accumulation of transcripts over the course of phage infection. Thus, the predicted intrinsic terminator functions weakly, consistent with the short poly(U) tract following the hairpin. We also analyzed *mobE* sequences from additional T-even phage for the presence of the terminator, and found that two phages (LZ7 and RB15) possessed a number of nucleotide substitutions relative to other *mobE*-containing phage (Figure 4). Although the substitutions shorten the predicted stem and increase the loop length, the overall structure and poly(U) tract of the terminator is conserved, suggesting that

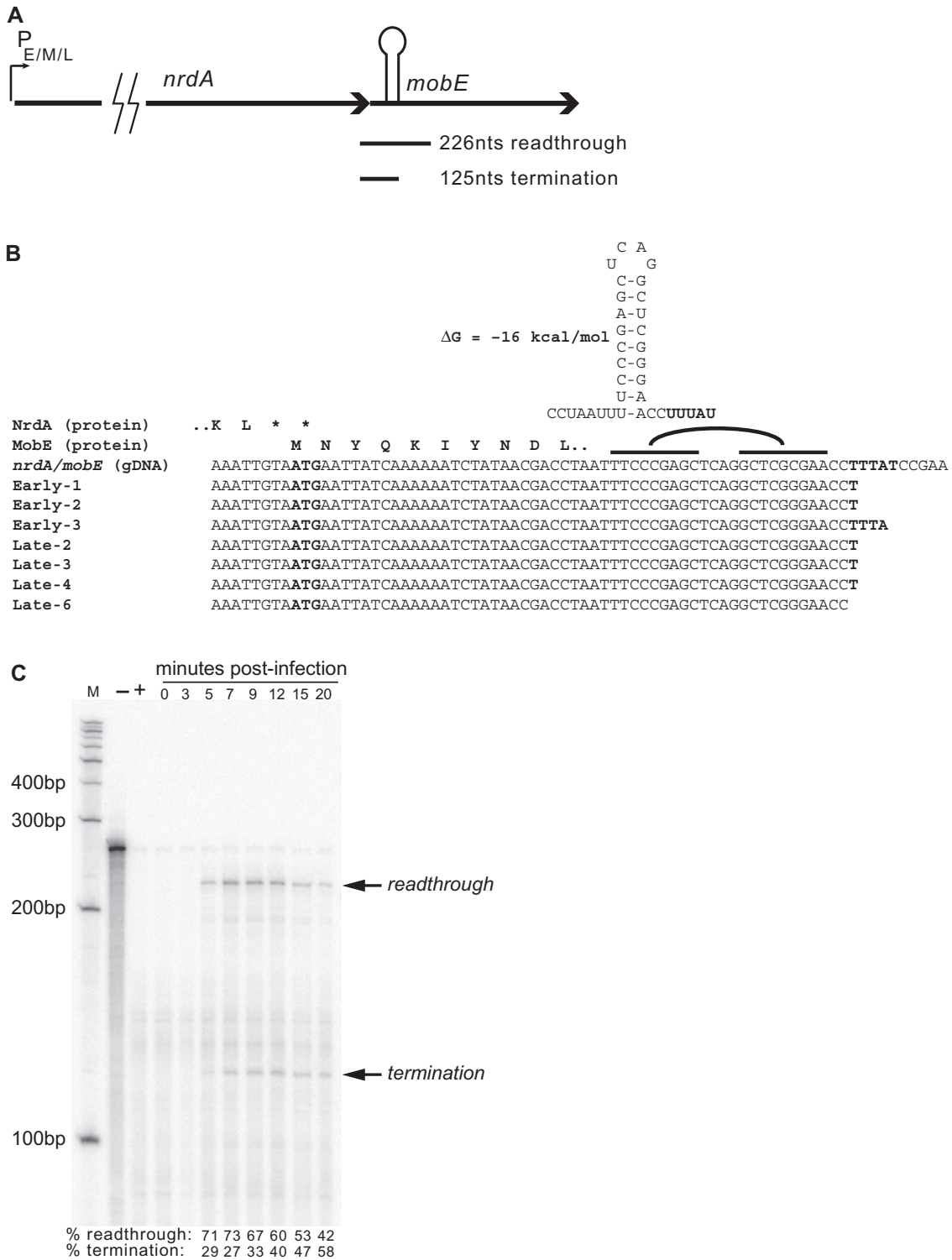

Figure 3. Transcriptional termination events in the 5′ region of *mobE*. (**A**) Schematic of the 5′ region of *mobE*, with the position of the Rho-independent termination indicated by a hairpin. The position of the probe and predicted sizes of the RNase digestion products are indicated below the *mobE* gene. (**B**) Sequences of 3′RACE clones from early or late RNA samples mapped onto the *mobE* genomic sequence. Nucleotides highlighted in bold are part of the poly(U) tract following the terminator structure, for which the sequence and structure is indicated above. (**C**) RNase protection assay to determine the functionality of the terminator. Shown is a representative protection experiment, with read through and termination products indicated. The extent of termination to read through product is indicated below the gel for each experimental time point. −, probe only; +, probe digested with RNases.

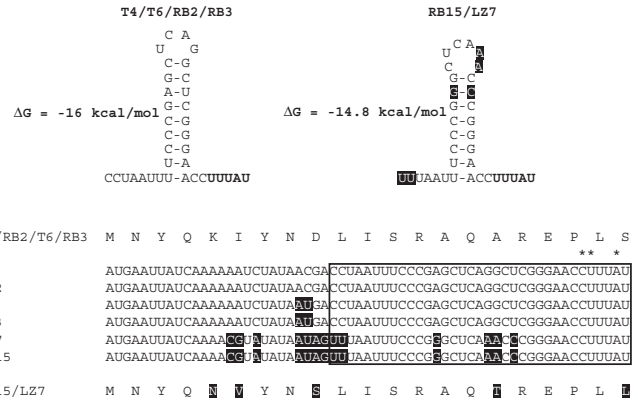

Figure 4. Conservation of the *mobE* terminator in related T-even phage. Shown are the nucleotide and amino acid sequences of the 5′ region of the *mobE* gene from various phages, with differences to the T4 sequence highlighted by black boxes. The predicted terminator sequence is indicated by a rectangle, and predicted secondary structures are shown above with nucleotide substitutions relative to T4 indicated by black boxes. Terminating nucleotides mapped to the T4 sequence by 3′RACE are indicated by asterisks.

transcriptional termination is an important regulatory mechanism in controlling gene expression in the *nrd* genomic region.

mobE is a mobile genetic element

To provide evidence that *mobE* is a mobile genetic element, we performed co-infections with T4 and T2 and screened progeny phage for the presence of *mobE* by Southern blots of plaque lifts using *mobE*-specific probes, as previously described (6). In this experiment, T2 serves as a recipient because it lacks *mobE*. We expressed the results of each phage cross as the ratio of progeny phage positive for *mobE* (and other endonuclease genes) relative to the input T4 phage, allowing direct comparison between experimental replicates that differed in the number of T4 input phage (Table 2 and Table S2). In crosses between wild-type T4 and T2, the inheritance of *mobE* in progeny was very high, with a ratio of 2.1 ± 0.2, suggesting a non-Mendelian inheritance characteristic of mobile endonucleases (2,6,12,13) (Table 2). Moreover, the inheritance of two other T4 homing endonucleases, *I-TevI* and *I-TevIII*, were also very high, with ratios of 2.0 ± 0.3 and 2.1 ± 0.3, respectively. To determine if the HNH motif of MobE is essential to promote high-frequency inheritance, we made an internal in-frame deletion that removed 29 amino acids of the predicted HNH motif, and used this T4*mobEΔ33-62* phage in co-infection experiments with T2. As shown in Table 2, the inheritance of *mobE* in crosses with T2 and T4*mobEΔ33-62* was reduced relative to inheritance in crosses with wild-type phage (1.5 ± 0.3 versus 2.1 ± 0.2), suggesting that deletion of the HNH motif abolishes mobility of MobE. The observation that *mobE* was present at a ratio of 1.5 ± 0.3 in progeny rather than the expected ratio of 1.0 based on the sequence similarity of the two phages is consistent with the fact that T4 generally excludes T2 genetic markers from progeny of a mixed infection (4). Moreover, the magnitude of reduction in *mobE* inheritance observed with the *mobEΔ33-62* mutant phage is similar to that observed with other T4 homing endonucleases (2,6,12,22). Collectively, the results of the phage co-infection experiments show that *mobE* is a mobile endonuclease, and that the HNH motif is critical for MobE to function as a mobile element.

Inheritance of the defunct T4 homing endonuclease I-TevIII is dependent on *mobE*

Previous experiments had suggested that *mobE* might also be responsible for the high-frequency inheritance of I-TevIII, a defunct homing endonuclease encoded within the *nrdB* group I intron. To test this hypothesis, we screened progeny from co-infection experiments with wild-type T4 and T2, and from co-infection experiments with T4 *mobE* mutants and T2, for the presence of *I-TevIII* (Table 2). With wild-type T4, the ratio of *I-TevIII* mirrored the ratio of *mobE* (2.0 ± 0.3). However, in mixed infections with T4 mutant phages and T2, the ratio of *I-TevIII* (1.6 ± 0.4) was reduced by the same amount as *mobE* inheritance. Thus, mobility of the defunct I-TevIII endonuclease is dependent on *mobE*.

Sequencing of progeny phage localizes the *mobE*-induced recombination point to the 5′-end of *nrdB*

In screening progeny from wild-type T4 and T2 co-infections, we noticed a number of phage did not inherit *I-TevIII* even though they did inherit *mobE* (*mobE*$^+$ *I-TevIII*$^-$). This observation suggested to us that the MobE recognition and cleavage site lies in the 5′ region of T2 *nrdB* because a recombination event that was initiated in the 5′ region of T2 *nrdB* would result in inheritance of both *mobE* and *I-TevIII* if exonucleolytic degradation extended in both directions from the cleavage site, into the 3′ regions of T2 *nrdB* and *nrdA*. If, however, exonucleolytic degradation was not extensive, and recombination events were initiated and resolved in the 5′ region of T2 *nrdB* near the MobE cleavage site, progeny phage would not inherit *I-TevIII*. To resolve this issue, we sequenced four *mobE*$^+$ *I-TevIII*$^-$ phage and found that three of the progeny were identical in sequence, allowing us to localize a recombination point (i.e. the transition site from T2 to T4 sequence) within the first 189 nt of T2 *nrdB* based on sequence polymorphisms of the parental T4 and T2 phages (Figure 5). The other *mobE*$^+$ *I-TevIII*$^-$ progeny phage had two recombination points, one in the 3′ region of *nrdA* near nucleotides 2237–2242 and another in the 5′ region of *nrdB*, near nucleotides 505–515 (Figure 5). Sequencing of the *nrd* genomic region from progeny phage that inherited both *mobE* and *I-TevIII* (*mobE*$^+$ *I-TevIII*$^+$) revealed T4 nucleotide polymorphisms throughout the 5′ region of *nrdB* and the 3′ region of *nrdA* (Figure 5). It is unlikely that the recombination events are influenced by co-conversion events initiated by cleavage of the upstream endonuclease *I-TevI*, which is ~3 kb distant from *mobE*, because *I-TevI*-mediated recombination events extend to a maximum of 2 kb downstream from the *I-TevI* cleavage site in the *td* gene (6).

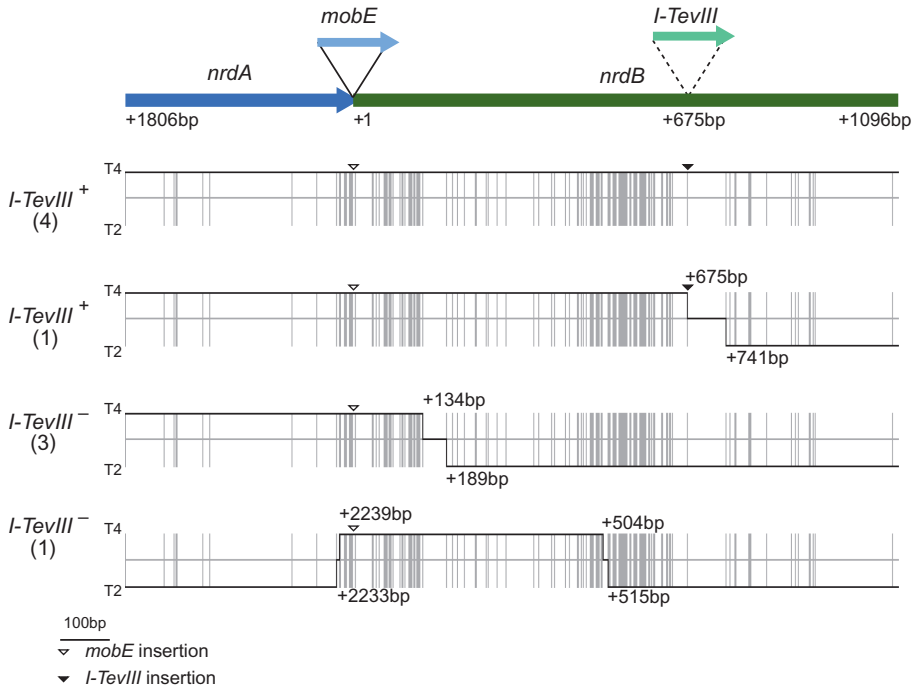

Figure 5. Sequence analysis of progeny phage. A schematic of the region sequenced is shown, with nucleotide positions relative to the AUG codon for each gene indicated. Nucleotide polymorphisms between T2 and T4 are indicated by vertical grey lines, and solid horizontal black lines indicate the nucleotide polymorphisms present in each progeny phage. All phages sequenced inherited *mobE* (*mobE*$^+$), and either inherited *I-TevIII* (*I-TevIII*$^+$) or did not inherit *I-TevIII* (*I-TevIII*$^-$). The number of progeny phage sequenced is indicated.

In vivo mapping of MobE-dependent nicks to the 5′ region of T2 *nrdB*

Sequencing of progeny phage revealed a recombination point in the 5′ region of the T2 *nrdB* gene, consistent with the hypothesis that *mobE* encodes a site-specific homing endonuclease with a recognition site in the *nrdB* gene. To provide evidence that *mobE* encodes a site-specific endonuclease, we mapped MobE-dependent breaks in the T2 *nrdB* gene *in vivo* over the course of a T4 and T2 co-infection (Figure 6). In this experiment, total DNA was isolated at various time points before and after co-infection, and used in primer extension experiments with a primer (DE-465) specific to the non-coding strand of the T2 *nrdA*/*nrdB* intergenic region, and with a primer (DE-522) that will anneal to the coding strand of the 3′ region of T2 and T4 *nrdB*. As a positive control for the primer extension reactions, genomic DNA was digested with DraI, which cleaves within the T2 *nrdB* gene.

As shown in Figure 6B, a nick was mapped to the non-coding strand of the T2 *nrdB* gene at 10 min post-T4 and post-T2 co-infection (Figure 6B, lanes 4–6), correlating with our *mobE* expression data (Figure 2). In contrast, no nicks were mapped to the coding strand of the *nrdB* gene (Figure 6C; an independent gel with longer exposure is shown in Figure S1). Control reactions using DraI-digested DNA-generated a product of the expected size on the coding strand (Figure 6C, lane 2 and Figure S1, lanes 1–3), suggesting that the lack of extension products on the coding strand is not due failure of the extension reaction. Furthermore, the nick in the non-coding strand was dependent on a functional *mobE* gene product, as no nicks were mapped to the same position on DNA isolated from T4*mobEΔ33-62* and T2 co-infections (Figure 6B, lanes 8–10). However, a band was evident at 10 min post-infection on the non-coding strand, but at a different position (Figure 6B, lane 8), likely representing a spurious primer extension product.

To validate that the 5′ region of the T2 *nrdB* gene was the sole determinant for MobE-dependent nicking, we amplified and cloned the T2 *nrdA*/*nrdB* intergenic region and the first 661 nucleotides of the T2 *nrdB* gene to create pT2nrdB, which was subsequently transformed into *E. coli sup*^{0}. The resultant strain was infected with T4K10, a T4 phage that cannot degrade host DNA upon infection in *sup*^{0} cells. Plasmid DNA was isolated at various times post-T4K10 infection, and used in primer extension reactions. Nicking of the non-coding strand of the *nrdB* gene in pT2nrdB was detected at all time points, but not in uninfected cells (Figure 7A, compare lane 3 with lanes 4–6). We mapped three prominent nicking sites to between nucleotides 122–126 of the T2 *nrdB* gene by electrophoresis of the primer extension reaction alongside a sequencing ladder generated from pT2nrdB with the same primer (Figure 7B and C). Moreover, the nick sites present *in vivo* on DNA isolated from T2 and T4 co-infections also mapped to the identical positions as the nicking sites on pT2nrdB (data not shown). It should be noted, however, that the multiple nicking sites could result from the nucleotide addition activity of Taq polymerase during primer extension reaction (33).

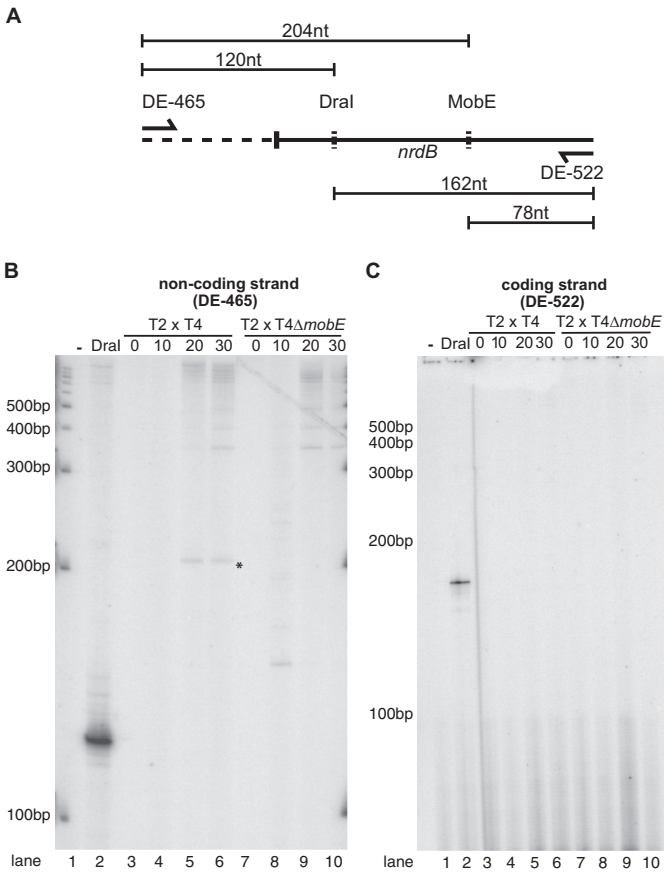

Figure 6. Mapping of MobE-generated nicks *in vivo*. (**A**) Schematic of the 5′ region of the *nrdB* gene showing the position of the primers used, the position of the DraI cleavage and MobE nicking sites, and the predicted lengths of the primer extension products. (**B**) Representative gel of primer extension reactions on the non-coding strand of DNA isolated at various time points (min) from co-infections with wild-type phages, or isolated from infections with T4 *mobE* deletion phage and T2. The sizes of a 100-bp ladder are labeled on the left edge of the gel, and the asterisk indicates the MobE nicking site present in wild-type but not mutant phage infections. −, no DNA added to primer extension; DraI, primer extension on DraI-digested gDNA. (**C**) Primer extension reactions on the coding strand, labeled as (B). For this representative gel, the DraI primer extension reaction was diluted 3-fold prior to electrophoresis. Note that the predicted size of the primer extension product is ∼78 nt.

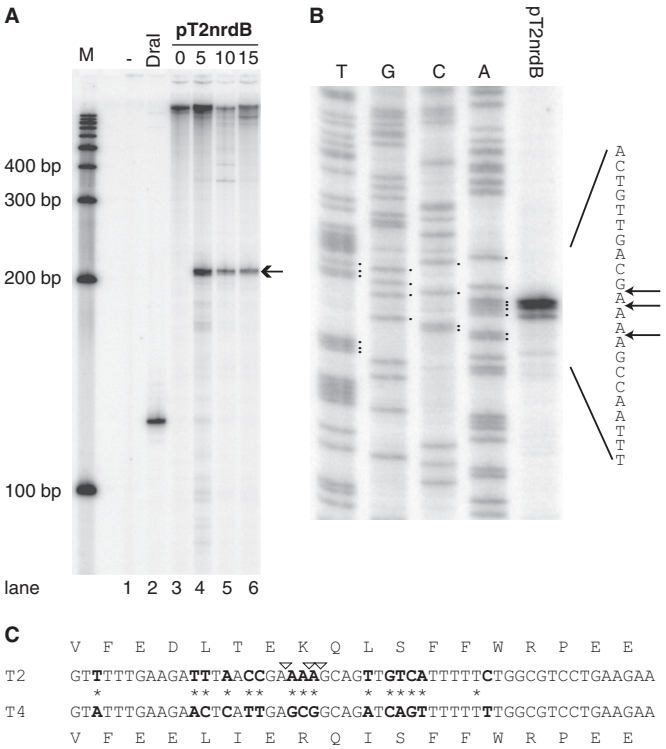

Figure 7. Primer extension reactions on plasmid substrates, and mapping of the MobE nicking sites. (**A**) Gel showing primer extension reactions on pT2nrdB plasmid DNA isolated at various time points (min) post infection with T4K10. The primer used in this experiment is the same primer used to maps nicks on the non-coding strand of the *nrdB* gene as shown in Figure 6. Nicks are indicated by a left-facing arrow. (**B**) High-resolution mapping of the MobE nicking sites by electrophoresis of the primer extension reaction alongside a sequencing ladder generated with the same primer. The sequence of the T2 *nrdB* gene is indicated to the right of the gel, with the nicking sites indicated by left-facing arrows. (**C**) Alignment of *nrdB* nucleotide and amino acid sequences from phages T2 and T4 in the MobE nicking-site region. Nucleotide differences are indicated by bold-type font and asterisks, and the nicking sites are indicated on the coding strand of the T2 sequence by triangles.

Collectively, these data are consistent with *mobE* encoding a site-specific nicking endonuclease with a target site in the 5′ region of the T2 *nrdB* gene.

DISCUSSION

Phage T4 is unique among T-even phage in that it encodes 15 homing endonucleases, accounting for ∼11% of the coding potential of the genome (7). Each of the 15 endonucleases are embedded within transcriptional units that include essential T4 gene products. In the case of *mobE*, the gene lies in the intergenic region separating the functionally critical *nrdA* and *nrdB* genes encoding the large and small subunits of aerobic ribonucleotide reductase, respectively (7). Multiple lines of evidence suggest that T4 and *mobE* have undergone extensive selection and adaptation to ensure that *mobE* is successfully integrated into the surrounding transcriptional program so as to limit any affect on function of the *nrd* genes. For instance, a middle promoter that drives expression of *nrdB* is located in the 3′ region of the *mobE* gene (32). In phage that lack *mobE* (Figure 8), an equivalent middle promoter is located in the extreme 3′-end of *nrdA*, or the *nrdA-nrdB* intergenic region. The location of the *nrdB* middle promoter in the *mobE* coding region, while critical for expression of *nrdB*, also selects against deletion of *mobE* from T4, as any deletion event that removed the middle promoter would be detrimental to the phage. Another function of the middle promoter in the 3′ region of *mobE* is likely to compensate for the presence of the Rho-independent transcriptional terminator in the 5′ region of *mobE*, ensuring sufficient *nrdB* transcript levels during phage infection. Intriguingly, in phage that lack *mobE*, no transcriptional terminator is located in the *nrdA−nrdB* intergenic region, whereas all phage

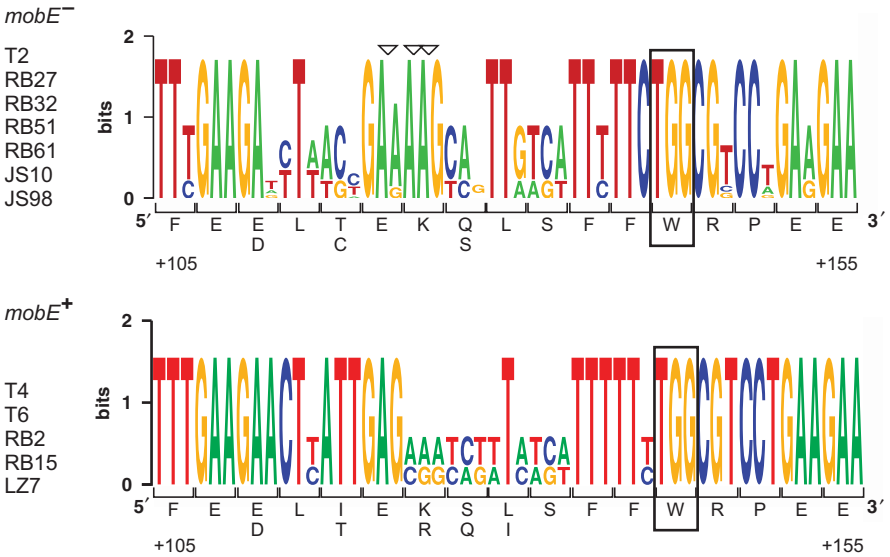

Figure 8. Logos representation of the *nrdB* nucleotide sequence up- and down-stream of the MobE-dependent nicking site for phage that possess or lack *mobE*. The amino acid sequence for each codon is indicated below the logos representation, as is the nucleotide position in the *nrdB* gene. The MobE nicking sites are indicated by triangles, and the conserved and functionally critical tryptophan residue in the NrdB protein is boxed by a rectangle. Note that the MobE-dependent nicking sites were mapped to the non-coding strand, but are indicated on the coding strand.

that encode *mobE* possess a terminator (Figure 5). This observation suggests that the terminator is an adaptation to down regulate expression of *mobE*, presumably because overexpression of MobE is in some way detrimental.

Previous studies on homing endonucleases in the *nrd* genomic region of T-even phage provided genetic evidence that *mobE* functioned as a mobile endonuclease, but MobE endonuclease activity could not be detected using *in vitro* synthesized protein and PCR-generated substrates (6,22) (G.W.Wilson and D.R.Edgell, unpublished data). Here, we have taken a different approach, and mapped a strand-specific nick *in vivo* to the T2 *nrdB* gene using genomic DNA isolated after co-infection with T2 and T4 (Figure 6). Furthermore, the nick was dependent on an intact *mobE* gene product, because an internal in-frame deletion that removed the predicted HNH catalytic motif abolished the appearance of the nick (Figure 6B). Thus, the available evidence suggests that MobE functions as a nicking HNH endonuclease to promote mobility, but we cannot rule out the possibility that MobE introduces a DSB, as does the phage RB3 HNH endonuclease, I-TevIII (24). Definitive evidence of MobE nicking or cleavage activity using purified protein and *in vitro* assays would resolve this issue. However, *in vitro* studies have been hindered by the fact that we cannot clone the *mobE* gene in a number of different expression vectors for purification. Finally, we acknowledge the possibility that MobE may not function as a homing endonuclease, but rather act as a regulator of another T4-encoded endonuclease that nicks or introduces a DSB in the T2 *nrdB* gene to promote *mobE* and *I-TevIII* mobility.

Interestingly, the MobE-dependent nicking sites map to a conserved region of the *nrdB* gene (Figures 7C and 8), as was found from studies of the recognition sites of other homing endonucleases that also map to conserved regions of their target genes (18,34,35). This region of *nrdB* contains nucleotide sequence that codes for a conserved and essential tryptophan residue that is involved in transfer of a radical to the NrdA subunit of ribonucleotide reductase to initiate reduction of NDPs to dNDPs (36,37) (Figures 7C and 8). Two prominent nicking sites are located in codon 42 of the T2 *nrdB* gene, which specifies a lysine (AAG) residue in T2 and all phages that do not encode *mobE*. The equivalent codon in the *nrdB* gene of *mobE*-containing phage is either a lysine (AAG, AAA) or arginine (CGG, CGA) (Figure 8). Based on these observations, we suggest that the potential MobE recognition site in the *nrdB* gene is bi-partite, mirroring the bipartite and modular structure of MobE itself; nucleotides encoding the conserved block of amino acids (FFWRPEE) would be a substrate for the DNA-binding module of MobE, whereas the N-terminal HNH catalytic domain would recognize the polymorphic region encompassing nucleotides 114–131. This mode of interaction of MobE with substrate would allow the endonuclease to recognize and nick its target site in related phage genome, but also prevent nicking of the host (*mobE*-containing) genome due to nucleotide polymorphisms that are suboptimal for nicking activity. Moreover, genetic and structural studies of other homing endonucleases have revealed similar modes of interaction with their substrates, including the HNH enzyme I-HmuI that possesses an extended two-domain structure consisting of a DNA-binding domain connected to a HNH catalytic domain by a linker (38,39).

In addition to promoting mobility of its own coding region between genomes, we have shown that an intact *mobE* gene is also responsible for high-frequency inheritance of the non-mobile *nrdB* intron and *I-TevIII*, resolving a number of experimental observations in the literature regarding the non-functional nature of

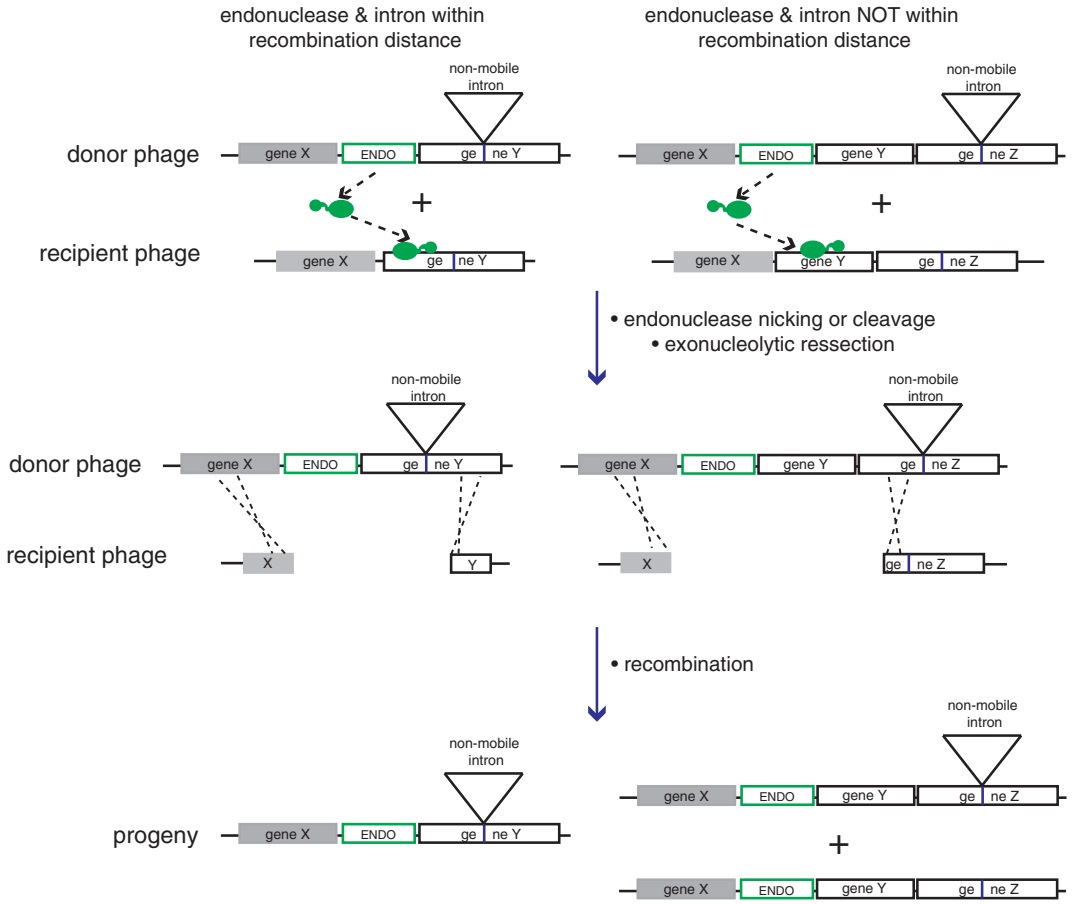

Figure 9. The *trans* homing pathway. Shown is a schematic of two potential gene arrangements in a donor phage genome, one where a functional endonuclease and non-mobile intron are encoded within recombination distance of each other (left), and one where they are not (right). Also shown is the recipient genome that lacks the endonuclease gene and intron. Nicking or cleavage of the recipient genome by the endonuclease initiates a recombination event that results in mobility of the endonuclease gene and intron (left), or only of the endonuclease gene (right). Note that co-conversion declines as a function of distance from the nicking or cleavage sites. Thus, the non-mobile intron will not be inherited in 100% of progeny if it is sufficiently distant from the nicking or cleavage site. Dashed lines between donor and recipient genomes indicate recombination points.

T4-encoded I-TevIII. Belfort and colleagues initially showed that cloned fragments of the *nrdB* intron were not mobile in plasmid-to-phage assays (40), an observation supported by the later finding that T4 I-TevIII possessed an inactivating internal deletion (25). Subsequent work, including data presented here, showed that the *nrdB* intron and *I-TevIII*, in spite of the inactivating deletion, were mobilized at high frequency in phage-to-phage crosses (6,22). These seemingly contradictory results can be rationalized by the fact that cloned fragments of T4 DNA used in the initial plasmid-to-phage assays contained only the C-terminal 67 residues of MobE, whereas *mobE* is fully functional in phage-to-phage assays and able to initiate a gene conversion event that results in high-frequency inheritance of the defunct *I-TevIII* and *nrdB* intron.

The mobilization of a defunct homing intron in the *nrdB* gene by the adjacently encoded MobE endonuclease is very similar to the recently described collaborative homing pathway involving a free-standing endonuclease mobilizing an endonuclease-lacking group I intron by cleaving at the intron insertion site of intronless alleles (41,42). One notable difference between the two pathways is the position of the endonuclease's target site; in the pathway described here, MobE does not nick at the intron insertion site in the intronless T2 *nrdB* gene, whereas in collaborative homing the free-standing endonuclease cleaves at the intron insertion site. Nonetheless, in both situations, cleavage or nicking by the free-standing endonuclease initiates a gene conversion event that ultimately mobilizes the endonuclease and intron. Thus, the *nrdB* intron can persist in the population in spite of having no inherent means of promoting its own mobility. This type of mobility is, in principle, similar to the mobilization of non-autonomous elements that lack the machinery to promote their own spread and instead are mobilized in *trans* by the machinery of functional full-length elements, which do not have to be in close proximity (43–45). However, the success of collaborative homing, and *trans* homing of the *nrdB* intron described here, is ultimately dependent on the close proximity of the free-standing endonuclease's cleavage site to the intron insertion site, because exonucleotyic processing of the DSB or nick in the recipient genome will extend a limited distance before recombination is initiated with the donor genome (Figure 9). Previous studies have shown

that endonuclease-mediated homing events in T4 result in co-conversion of flanking sequence to a maximum of 2 kb from the endonuclease's cleavage sites (6,10), suggesting that a free-standing endonuclease could mobilize a non-mobile intron only if they are both encoded within this critical distance. Such gene arrangements could occur by chance, but we propose that these arrangements may be common in small, endonuclease- and intron-rich genomes based on the finding that homing endonucleases and introns tend to target conserved sequences of functionally critical host genes (34,35,42). As these sequences occur infrequently, the probability that a free-standing endonuclease and group I intron are encoded within the critical recombination distance are relatively high. Moreover, a similar pathway could be used to mobilize adjacently encoded introns (or inteins), where only one intron encodes a functional endonuclease. Indeed, in bacterial and phage genomes, there are many examples of multiple insertions of endonuclease-containing and endonuclease-lacking introns (or inteins) in the same gene (46–50). Thus, the *trans* mobility of defunct or non-mobile genetic elements by an adjacently encoded functional homing endonuclease may be common in phage and bacterial genomes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank David Haniford and Ewan Gibb for critical reading of the manuscript, and helpful comments from the reviewers.

FUNDING

Canadian Institutes of Health Research (grant MOP77779 to D.R.E.); Early Researcher Award from the Government of Ontario (to D.R.E.). Funding for open access charge: Natural Sciences and Engineering Research Council of Canada

Conflict of interest statement. None declared.

REFERENCES

1. Belfort,M., Derbyshire,V., Cousineau,B. and Lambowitz,A. (2002) Chapter 31. In Craig,N., Craigie,R., Gellert,M. and Lambowitz,A. (eds), *Mobile DNA II*. ASM Press, New York, pp. 761–783.
2. Belle,A., Landthaler,M. and Shub,D.A. (2002) Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes Dev.*, **16**, 351–362.
3. Goodrich-Blair,H. and Shub,D.A. (1996) Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell*, **84**, 211–221.
4. Russell,R.L. and Huskey,R.J. (1974) Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. *Genetics*, **78**, 989–1014.
5. Edgell,D.R. (2002) Selfish DNA: new abode for homing endonucleases. *Curr. Biol.*, **12**, R276–R278.
6. Liu,Q., Belle,A., Shub,D.A., Belfort,M. and Edgell,D.R. (2003) SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. *J. Mol. Biol.*, **334**, 13–23.
7. Miller,E.S., Kutter,E., Mosig,G., Arisaka,F., Kunisawa,T. and Ruger,W. (2003) Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.*, **67**, 86–156.
8. Huang,Y.J., Parker,M.M. and Belfort,M. (1999) Role of exonucleolytic degradation in group I intron homing in phage T4. *Genetics*, **153**, 1501–1512.
9. Mueller,J.E., Clyman,J., Huang,Y.J., Parker,M.M. and Belfort,M. (1996) Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. *Genes Dev.*, **10**, 351–364.
10. Parker,M.M., Belisle,M. and Belfort,M. (1999) Intron homing with limited exon homology. Illegitimate double-strand-break repair in intron acquisition by phage T4. *Genetics*, **153**, 1513–1523.
11. Bell-Pedersen,D., Quirk,S.M., Aubrey,M. and Belfort,M. (1989) A site-specific endonuclease and co-conversion of flanking exons associated with the mobile *td* intron of phage T4. *Gene*, **82**, 119–126.
12. Brok-Volchanskaya,V.S., Kadyrov,F.A., Sivogrivov,D.E., Kolosov,P.M., Sokolov,A.S., Shlyapnikov,M.G., Kryukov,V.M. and Granovsky,I.E. (2008) Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage. *Nucleic Acids Res.*, **36**, 2094–2105.
13. Kadyrov,F.A., Shlyapnikov,M.G. and Kryukov,V.M. (1997) A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. *FEBS Lett.*, **415**, 75–80.
14. Sharma,M., Ellis,R.L. and Hinton,D.M. (1992) Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proc. Natl Acad. Sci. USA*, **89**, 6658–6662.
15. Sharma,M. and Hinton,D.M. (1994) Purification and characterization of the SegA protein of bacteriophage T4, an endonuclease related to proteins encoded by group I introns. *J. Bacteriol.*, **176**, 6439–6448.
16. Kutter,E., Gachechiladze,K., Poglazov,A., Marusich,E., Shneider,M., Aronsson,P., Napuli,A., Porter,D. and Mesyanzhinov,V. (1995) Evolution of T4-related phages. *Virus Genes*, **11**, 285–297.
17. Landthaler,M., Shen,B.W., Stoddard,B.L. and Shub,D.A. (2006) I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. *J. Mol. Biol.*, **358**, 1137–1151.
18. Landthaler,M. and Shub,D.A. (2003) The nicking homing endonuclease I-*Bas*I is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. *Nucleic Acids Res.*, **31**, 3071–3077.
19. Landthaler,M., Lau,N.C. and Shub,D.A. (2004) Group I intron homing in *Bacillus* phages SPO1 and SP82: a gene conversion event initiated by a nicking homing endonuclease. *J. Bacteriol.*, **186**, 4307–4314.
20. Gott,J.M., Zeeh,A., Bell-Pedersen,D., Ehrenman,K., Belfort,M. and Shub,D.A. (1988) Genes within genes: independent expression of phage T4 intron open reading frames and the genes in which they reside. *Genes Dev.*, **2**, 1791–1799.
21. Edgell,D.R., Derbyshire,V., Van Roey,P., LaBonne,S., Stanger,M.J., Li,Z., Boyd,T.M., Shub,D.A. and Belfort,M. (2004) Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat. Struct. Mol. Biol.*, **11**, 936–944.
22. Sandegren,L., Nord,D. and Sjöberg,B.-M. (2005) SegH and Hef: two novel homing endonucleases whose genes replace the *mobC* and *mobE* genes in several T4-related phages. *Nucleic Acids Res.*, **33**, 6203–6213.
23. Gott,J.M., Shub,D.A. and Belfort,M. (1986) Multiple self-splicing introns in bacteriophage T4: evidence from autocatalytic GTP labeling of RNA in vitro. *Cell*, **47**, 81–87.
24. Robbins,J.B., Stapleton,M., Stanger,M.J., Smith,D., Dansereau,J.T., Derbyshire,V. and Belfort,M. (2007) Homing endonuclease I-TevIII: dimerization as a means to a double-strand break. *Nucleic Acids Res.*, **35**, 1589–1600.

25. Eddy,S.R. and Gold,L. (1991) The phage T4 *nrdB* intron: a deletion mutant of a version found in the wild. *Genes Dev.*, **5**, 1032–1041.
26. Horton,R.M. (1997) In vitro recombination and mutagenesis of DNA. SOEing together tailor-made genes. *Methods Mol. Biol.*, **67**, 141–149.
27. Miller,E.S. and Carlson,K. (1994) Chapter 45.1. In Karam,J.D. (ed.), *Molecular Biology of Bacteriophage T4*. ASM Press, Washington, D.C., pp. 455–456.
28. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
29. Gibb,E.A. and Edgell,D.R. (2007) Multiple controls regulate the expression of *mobE*, an HNH homing endonuclease gene embedded within a ribonucleotide reductase gene of phage Aeh1. *J. Bacteriol.*, **189**, 4648–4661.
30. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
31. Tseng,M.J., Hilfinger,J.M., Walsh,A. and Greenberg,G.R. (1988) Total sequence, flanking regions, and transcripts of bacteriophage T4 *nrdA* gene, coding for alpha chain of ribonucleoside diphosphate reductase. *J. Biol. Chem.*, **263**, 16242–16251.
32. Tseng,M.J., He,P., Hilfinger,J.M. and Greenberg,G.R. (1990) Bacteriophage T4 *nrdA* and *nrdB* genes, encoding ribonucleotide reductase, are expressed both separately and coordinately: characterization of the nrdB promoter. *J. Bacteriol.*, **172**, 6323–6332.
33. Clark,J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677–9686.
34. Edgell,D.R., Stanger,M.J. and Belfort,M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J. Mol. Biol.*, **343**, 1231–1241.
35. Scalley-Kim,M., McConnell-Smith,A. and Stoddard,B.L. (2007) Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.*, **372**, 1305–1319.
36. Sahlin,M., Lassmann,G., Potsch,S., Slaby,A., Sjoberg,B.M. and Graslund,A. (1994) Tryptophan radicals formed by iron/oxygen reaction with Escherichia coli ribonucleotide reductase protein R2 mutant Y122F. *J. Biol. Chem.*, **269**, 11699–11702.
37. Nordlund,P. and Eklund,H. (1993) Structure and function of the *Escherichia coli* ribonucleotide reductase protein R2. *J. Mol. Biol.*, **232**, 123–164.
38. Derbyshire,V., Kowalski,J.C., Dansereau,J.T., Hauer,C.R. and Belfort,M. (1997) Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.*, **265**, 494–506.
39. Shen,B.W., Landthaler,M., Shub,D.A. and Stoddard,B.L. (2004) DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.*, **342**, 43–56.
40. Quirk,S.M., Bell-Pederesen,D. and Belfort,M. (1990) Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell*, **56**, 455–465.
41. Zeng,Q., Bonocora,R.P. and Shub,D.A. (2009) A free-standing homing endonuclease targets an intron insertion site in the *psbA* gene of cyanophages. *Curr. Biol.*, **19**, 218–222.
42. Bonocora,R.P. and Shub,D.A. (2009) A likely pathway for formation of mobile group I introns. *Curr. Biol.*, **19**, 223–228.
43. Yang,G., Zhang,F., Hancock,C.N. and Wessler,S.R. (2007) Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **104**, 10962–10967.
44. Dewannieux,M. and Heidmann,T. (2005) L1-mediated retrotransposition of murine B1 and B2 SINEs recapitulated in cultured cells. *J. Mol. Biol.*, **349**, 241–247.
45. Dewannieux,M., Esnault,C. and Heidmann,T. (2003) LINE-mediated retrotransposition of marked *Alu* sequences. *Nat. Genet.*, **35**, 41–48.
46. Landthaler,M. and Shub,D.A. (1999) Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. *Proc. Natl Acad. Sci. USA*, **96**, 7005–7010.
47. Liu,X.Q., Yang,J. and Meng,Q. (2003) Four inteins and three group II introns encoded in a bacterial ribonucleotide reductase gene. *J. Biol. Chem.*, **278**, 46826–46831.
48. Landthaler,M., Begley,U., Lau,N.C. and Shub,D.A. (2002) Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of *Staphylococcus aureus* phage Twort. *Nucleic Acids Res.*, **30**, 1935–1943.
49. Lazarevic,V., Soldo,B., Dusterhoft,A., Hilbert,H., Mauel,C. and Karamata,D. (1998) Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPβ. *Proc. Natl Acad. Sci. USA*, **95**, 1692–1697.
50. Nakayama,H., Morinaga,Y., Nomura,N., Nunoura,T., Sako,Y. and Uchida,A. (2003) An archaeal homing endonuclease I-PogI cleaves at the insertion site of the neighboring intron, which has no nested open reading frame. *FEBS Lett.*, **544**, 165–170.