Nucleic Acids Research, 2005, Vol. 33, No. 19 ***6203–6213***
doi:10.1093/nar/gki932

SegH and Hef: two novel homing endonucleases whose genes replace the *mobC* and *mobE* genes in several T4-related phages

Linus Sandegren, David Nord and Britt-Marie Sjöberg*

Department of Molecular Biology and Functional Genomics, Stockholm University, Svante Arrhenius väg 16, F3. SE-10691 Stockholm, Sweden

Received August 16, 2005; Revised and Accepted October 11, 2005 DDBJ/EMBL/GenBank accession nos+

ABSTRACT

T4 contains two groups of genes with similarity to homing endonucleases, the *seg*-genes (similarity to endonucleases encoded by group I introns) containing GIY-YIG motifs and the *mob*-genes (similarity to mobile endonucleases) containing H-N-H motifs. The four *seg*-genes characterized to date encode homing endonucleases with cleavage sites close to their respective gene loci while none of the *mob*-genes have been shown to cleave DNA. Of 18 phages screened, only T4 was found to have *mobC* while *mobE* genes were found in five additional phages. Interestingly, three phages encoded a *seg*-like gene (hereby called *segH*) with a GIY-YIG motif in place of *mobC.* An additional phage has an unrelated gene called *hef* (homing endonuclease-like function) in place of the *mobE* gene. The gene products of both novel genes displayed homing endonuclease activity with cleavage site specificity close to their respective genes. In contrast to intron encoded homing endonucleases, both SegH and Hef can cleave their own DNA as well as DNA from phages without the genes. Both *segH* and *mobE* (and most likely *hef*) can home between phages in mixed infections. We discuss why it might be a selective advantage for phage freestanding homing endonucleases to cleave both HEG-containing and HEG-less genomes.

INTRODUCTION

Homing endonucleases are mobile genetic elements that can promote their own horizontal transfer by a process called homing (1). This occurs if the homing endonuclease encounters a genome that does not contain the endonuclease gene (HEG) at the cognate site. The HEG-less genome is cleaved by the endonuclease and repaired by the cell machinery via double strand break repair using the HEG-containing allele as template thereby inserting the HEG at the cognate site in the cleaved genome [reviewed in (2)]. HEGs are found both intergenically (freestanding) and frequently inserted within intervening sequences (group I, or group II introns, or inteins) where they confer mobility both to themselves and the surrounding splicing element via homing (3–8). Insertion of a HEG into an intervening sequence is thought to be beneficial for the HEG because it increases the number of potential homing targets since the splicing capability of the intron allows an intron-associated HEG to be inserted also into coding sequences without disrupting their function (5,6). Four families of homing endonucleases have been described based on conserved sequence motifs within them, LAGLIDADG, GIY-YIG, H-N-H and His-Cys box [reviewed in (9)].

Bacteriophage T4 contains three intron-associated HEGs and at least 12 freestanding genes with similarity to intron-encoded homing endonucleases (10,11). Only two of the intron-associated HEGs encode active homing endonucleases (I-TevI and I-TevII) while the third (I-TevIII in the *nrdB* intron) is rendered inactive due to a large internal deletion (12). Seven of the freestanding HEGs have similarities to the GIY-YIG family of intron-encoded homing endonucleases (10,13,14) and the remaining five have similarity to the H-N-H family of homing endonucleases (11). The GIY-YIG genes are called *segA-G* (for similarity to endonucleases encoded by group I introns) and four of them have been shown experimentally to encode freestanding homing endonucleases (*segA* (10,15), *segE* (16), *segF* (formerly *gene 69*) (13) and *segG* (formerly *gene 32.1*) (14). The H-N-H genes are called *mobA-E* (for similarity to mobile endonucleases), but none of the *mob*-genes have so far been shown to possess

*To whom correspondence should be addressed. Tel: +46 8 164150; Fax: +46 8 166488; Email: britt-marie.sjoberg@molbio.su.se
Present address:
Linus Sandegren, Department of Medical Biochemistry and Microbiology, Uppsala, University, Sweden
+AY262125–AY262139, AY262142, AY262144, AY262145 and DQ178119-DQ178121

© The Author 2005. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

endonuclease activity or homing capability (17). The different *seg*-, and *mob*-genes are only scarcely represented in phages closely related to T4 (10,15,16,18). It is intriguing that T4 contains so many seemingly non-essential putative homing endonuclease genes while they are absent in many of its close relatives.

In a previous study of the distribution of group I introns among T4-related phages we PCR screened and sequenced the *td-nrdAB* and *nrdDG* regions from more than 20 phage strains (19). In T4 these regions contain the three well-characterized intron-encoded HEGs (I-TevI in *td*, I-TevII in *nrdD* and I-TevIII in *nrdB*) and also two freestanding, putative HEGs containing the H-N-H motif, *mobE* between *nrdA* and *nrdB*, and *mobC* between *nrdD* and *nrdG* (cf. Figure 5A). These two freestanding putative HEGs have previously only been found in T4 (10,15,16,18). Here we show that *mobE* genes are present in five additional T-even-like phages besides T4. No additional phages with *mobC* were found. Interestingly, some of the phages that lacked either *mobE* or *mobC* had completely unrelated genes inserted at the corresponding sites. One novel gene was found replacing *mobC* in three phages and shows similarity to the GIY-YIG *seg*-genes of T4 and a second gene was found replacing *mobE* in phage U5 and has no similarity to any known homing endonuclease. Both novel gene products displayed endonuclease activity with cleavage sites close to their gene loci in the phage genomes, a feature characteristic for homing endonucleases. We have therefore named the novel *seg*-gene *segH* and the U5 gene *hef* (for homing endonuclease-like function). Homing studies in mixed phage infections show that *segH* displays site-specific homing. In addition we have obtained evidence for homing of the *mobE* gene in phage T6.

MATERIALS AND METHODS

Bacteria and phages

Strains of the original T-evens, T2H, T4D and T6 were kindly provided by Elisabeth Haggård, Dept. of Genetics, Stockholm University. Strains RB2, RB3, RB27, RB32, RB49, RB69, LZ1, LZ7, TuIa and U5 were kindly provided by Karin Carlson, Dept. of Cell and Molecular Biology, Uppsala University. RB14, RB15, RB23 and RB51 are from our stocks, originally a gift from Sean Eddy and phage LZ2 was kindly provided for this study by Sean Eddy, Dept. of Genetics, Washington University of Medicine. *Escherichia coli* B_0 was used as host strain for all phage work and is from our stocks.

Screen for *mobC* and *mobE* genes

Amplifications were performed using 2.5 U of proofreading *Pfu* DNA polymerase (Stratagene) for 30 cycles according to the manufacturer's recommendations. Primers for PCR amplification were for *mobE*: A2 and BS20, and for *mobC*: D2 and G4 (Supplementary Table 1). PCR products were analysed on agarose gels and purified using the QIAquick PCR Purification kit (Qiagen) to remove buffers, nucleotides and primers before sequencing. PCR fragments were sequenced on both strands using DYEnamic ET Terminator Cycle Sequencing for MEGA Bace from Amersham Pharmacia Biotech. All sequences obtained in this study have been reported to GenBank™, either as updates to previously reported files AY262125–AY262139, AY262142 and AY262144–AY262145, or as novel accessions nos DQ178119–DQ178121.

In vitro expression of proteins

Targets for *in vitro* translations were amplified directly from phage using the following primer pairs with T7 promoter sequence and reticulocyte ribosome binding sequence: Retic T4 *mobE* T7 primer together with T4, T6, RB2, RB15 mobE end primer, Retic RB3/T6 *mobE* T7 primer together with T4, T6, RB2, RB15 mobE end primer or RB3 mobE end primer, Retic T4 *mobC* T7 primer together with G2, Retic *segH* T7 primer together with *segH* 3′ primer and Retic U5 *hef* T7 primer together with U5.1 (Supplementary Table 1). PCR products were purified using QIAquick PCR purification kit (Qiagen) and used as templates in *in vitro* translation reactions using the reticulocyte *in vitro* translation kit TNT® T7 Quick for PCR DNA (Promega) according to the manufacturers' recommendations. Radiolabelled [^{35}S]-Met was included in the reactions and products were separated on 12% polyacrylamide gels and analysed by PhosphoImager (FujiFilm FLA-3000).

In vivo expression and purification of SegH

A PCR fragment of *segH* from LZ2 was cloned into pET21(+) (Stratagene) and verified by DNA sequencing (MWG-biotech). Expression of C-terminally His-tagged *segH* was performed in *E.coli* strain BL21(DE3) codon plus, and cells were harvested after 2 h induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were disrupted in a X-press chamber (AB Biox) and extracted in sodium phosphate buffer (pH 7.4), containing 20 mM imidazol. The suspension was centrifuged and the supernatant was loaded onto a HisTrap™ HP column (Amersham Biosciences) and washed according to the manufacturer's recommendations. SegH was eluted at 116 mM imidazol using a stepwise increase of imidazol concentration.

In vitro endonuclease assays

All targets for cleavage assays were amplified from T2 by PCR and purified using QIAquick PCR purification kit (Qiagen). Primers for target amplifications shown in Figure 2B were: Pat11, G2(F), BS10 and A2(F) (Supplementary Table 1); (F) denotes primers that were Fluorescein-labelled in the 5′ end. *In vitro* translation products were used directly in cleavage reactions. Reaction conditions were as follows: up to 800 fmol target DNA, 5 µl *in vitro* translation reaction and 3 µg RNase A in a total volume of 50 µl were incubated at 37°C for 30 min in 66 mM K-acetate, 33 mM Tris-acetate (pH 7.9), 10 mM Mg-acetate, 0.1 mg/ml BSA (for MobC and SegH) and in 100 mM NaCl, 50 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 0.1 mg/ml BSA (for MobE and Hef). Incubations with target DNA and *in vitro* translation mix without primer DNA served as negative controls. Cleavage results were analysed on agarose gels by excitation at 473 nm and filtered at 520 nm (FujiFilm FLA-3000); ImageGauge v3.45 (FujiFilm) was used for the quantification of cleavage.

Primers for target amplifications shown in Figures 3A and B were: D3, D4, G4, H2, A2, U5.1 and 43B (Table 1,

Supplementary Data). *In vitro* translation products were used directly in cleavage reactions. Reaction conditions were as follows: up to 10 µl target DNA and 3 µl *in vitro* translation reaction were incubated for cleavage as above. For cleavage site mapping one ^{32}P-radiolabelled primer was used in the target PCR and cleavage products were separated on 8% polyacrylamide gels together with sequencing ladders produced with the fmol® DNA Cycle Sequencing System (Promega) using the same labelled primers: nrdB1, nrdB2, D4 and G2 (Supplementary Table 1).

Targets used in SegH competition assay were amplified from T2, using primers Pat11 and G2(F) and from T4 and LZ2 using Pat11 and G4(F) (Supplementary Table 1). Targets used in Hef competition assay were amplified from T2, T4 and U5 using primers BS10 and A2(F) (Supplementary Table 1). Competition assays with 0.4 and 0.8 µg of partially purified His-tagged SegH were performed with equimolar amounts of targets (700 fmol of T2, T4 and LZ2) in 20 µl reactions. The cleavage reaction was performed in 66 mM K-acetate, 33 mM Tris-acetate (pH 7.9), 10 mM Mg-acetate, 0.1 mg/ml BSA at 37°C, and samples were withdrawn for analyses as indicated. Competition assays with 5 and 10 µl of *in vitro* translated Hef were performed in 50 µl reactions with equimolar amounts of targets (700 fmol of T2, T4 and U5). The cleavage reaction

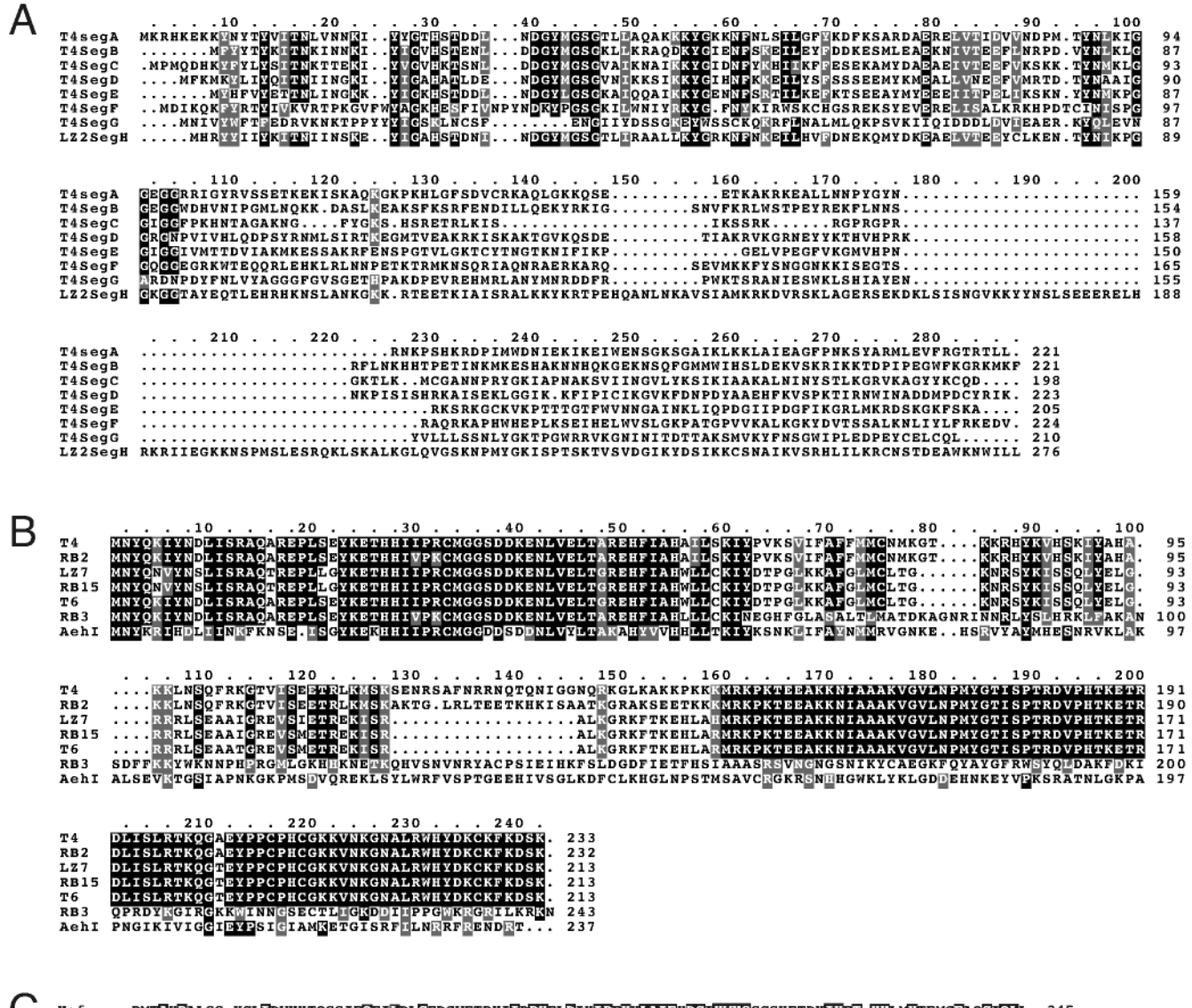


Figure 1. (**A**) Alignment of T4 Seg-proteins and LZ2 SegH. (**B**) Alignment of the different MobE proteins. The sequence of the MobE protein from Schizo-T-even phage AehI was obtained from the genome sequence, GenBank accession no. AY266303. Residues with black shading and white text are identical in 70% of the sequences, grey shading and black text denotes positions with similar residues in 70% of the sequences. (**C**) Alignment of a region of Hef (residues 251–345) with a region of a hypothetical protein from *Corynebacterium glutamicum* (residues 283–370 of 606, GenBank accession no. NP_600985) and a region from a representative ORF with unknown function that appears to be related to a diverse group of endonucleases (residues 9–102 of 144 from *Rhodopseudomonas palustris* hypothetical protein, GenBank accession no. NP_946397). Residues with black shading are conserved, and asterisks denote highly conserved residues among the diverse group of endonucleases.

was performed in 100 mM NaCl, 50 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 0.1 mg/ml BSA and 3 µg RNase A at 37°C, and samples were withdrawn for analyses as indicated. Samples were immediately purified using QIAquick PCR purification kit (Qiagen). Cleavage results were analysed on agarose gels by excitation at 473 nm and filtered at 520 nm (FujiFilm FLA-3000); ImageGauge v3.45 (FujiFilm) was used for the quantification of cleavage.

Mixed phage infections and PCR screening of homing

Equal amounts of two phage strains (2.5×10^9 p.f.u.) were mixed with 5×10^8 cells from a mid log-phase culture giving a multiplicity of infection of five of each phage. Incubation was continued for 90 min at 37°C with vigorous shaking. Cells were lysed with chloroform and progeny phage plated on new bacteria. For each cross 24 plaques were picked and resuspended in 20 µl of water. A 5 µl aliquot of the suspension was used as template in a 25 µl PCR with 1 U of *Taq* polymerase (Fermentas) according to the manufacturer's recommendations. Primers for screening of HEG presence were for *td* intron: *td*IVSA and *td*IVSB, for *nrdB* intron: *nrdB*IVSA and *nrdB*IVSB, for U5 *nrdD* intron: *nrdD*IVSA and *nrdD*IVSU5B, for *hef*: U5.1 and U5.2, for *segH*: segH1 and *segH* rev, for *mobE*: A2 and BS20, and for *mobC*: D2 and G2 (Supplementary Table 1).

Screening for markers flanking HEGs

Amplification of the *nrdD-nrdG* region from progeny phage was performed as described above using primers D8 and G4 and for the *nrdA-nrdB* region BS40 and 43B (Supplementary Table 1). Phage specific markers were screened by restriction cleavage at sites indicated in Figure 5C, for the *nrdG* gene of RB3/T2 crosses by direct sequencing using the G4 primer, and in Figure 5D by direct sequencing using primers BS40 and 43B.

RESULTS

Screen for *mobC* and *mobE* among T-even-like bacteriophages

The T4 *mobC* gene is inserted between genes *nrdD* and *nrdG* coding for the anaerobic ribonucleotide reductase and its activator protein, respectively. In a PCR screen of the region between *nrdD* and *nrdG* among 20 T-even-like phages we found that only T4 produced a fragment of *mobC* size. Phages T6, RB3 and LZ2 produced slightly larger PCR products and the remaining phages produced fragments indicating no additional gene inbetween *nrdD* and *nrdG* (data not shown). Surprisingly, sequencing of the PCR fragments from T6, RB3 and LZ2 showed that they all contain a completely different gene in this position with similarity to the T4 *seg*-genes (Figure 1A). We have named the new gene *segH*. The T6, RB3 and LZ2 *segH* sequences are very similar with the protein products only differing at five out of 276 amino acid residues (data not shown).

The putative H-N-H endonuclease gene *mobE* in T4 is located between the genes coding for the large (*nrdA*) and the small (*nrdB*) subunits of the aerobic ribonucleotide reductase. In a PCR screen of 18 T-even-like phages we found that phages T6, RB2, RB3, RB15 and LZ7 had insertions matching the length of T4 *mobE* in this region while phage U5 had a much longer insertion. Sequencing of the U5 region shows that it encodes a protein of 544 amino acids with no similarity to any known homing endonuclease, but containing a domain of unknown function (20) that appears to be related to a diverse group of endonucleases (Figure 1C). We have named the U5 gene *hef* for homing endonuclease-like function (see below). The remaining phages in the screen did not contain any gene inbetween *nrdA* and *nrdB*.

Sequencing of the *nrdA-mobE-nrdB* region in T4 showed that *mobE* differed from previously published sequences (21,22) by two frame shifts that brought two previously annotated open reading frames (ORF) into a continuous open reading frame from the stop codon of *nrdA* to the start codon of *nrdB* (19). Sequencing of T6, RB2, RB3, RB15 and LZ7 confirms that they also contain continuous *mobE* ORFs between *nrdA* and *nrdB* with somewhat varying sequences (Figure 1B). The genome sequence of the schizo-T-even phage Aeh1 also encodes a *mobE* gene of similar size in this region (Figure 1B). All in all this indicates that the original sequence of T4 *mobE* obtained from a cloned genome fragment (21,22) was of a non-functional gene (perhaps generated during the cloning procedure). Indeed, our attempts to clone T4 *mobE* have produced several frame shifted and mutated versions. However, we have now managed to clone T4 *mobE* and the gene product is highly toxic when expressed in *E.coli* (data not shown), a feature shared with numerous proven

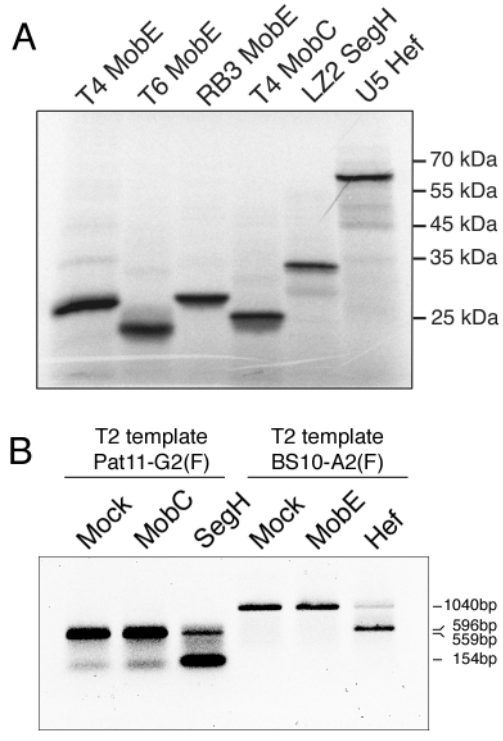


Figure 2. (**A**) PAGE gel 8% of the *in vitro* translation products. Proteins were labelled with [^{35}S]Methionine during *in vitro* translation and bands were visualized by phosphoimager analysis. (**B**) Agarose gel from cleavage assay with *in vitro* translated T4 MobC (*lane 2*), LZ2 SegH (*lane 3*), T4 MobE (*lane 5*) and U5 Hef (*lane 6*); control incubations with *in vitro* translation mix without template DNA (*lanes 1, 4*). DNA targets were labelled at one end with Flourescein-labelled primer thereby generating only one labelled band upon cleavage.

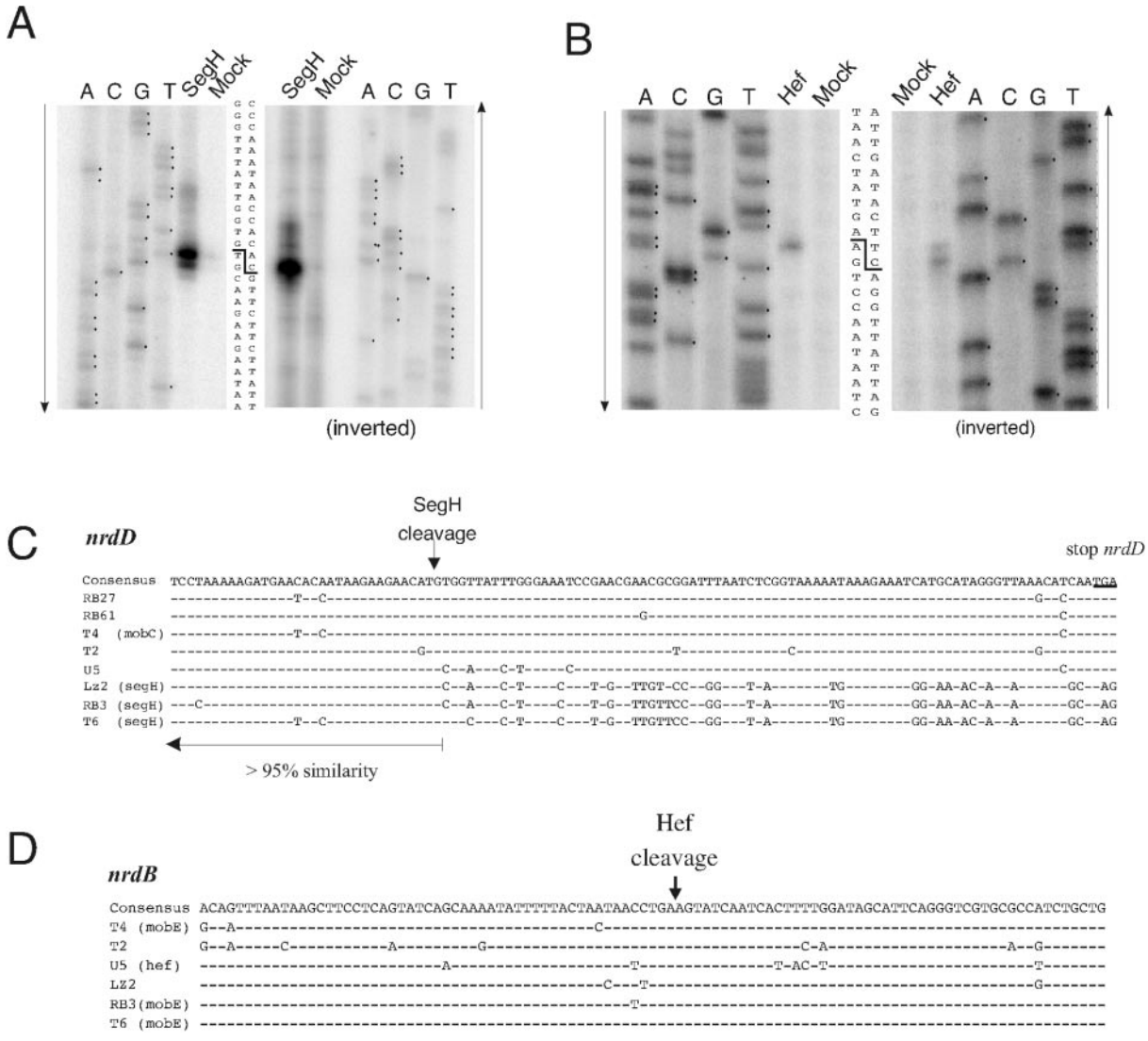


Figure 3. Sequencing reactions run alongside single strand labelled cleavage reactions to map (**A**) the SegH cleavage site, and (**B**) the Hef cleavage site. Note that the second gel is inverted to illustrate the cleavage sites on both strands. Sequence variation surrounding (**C**) the SegH cleavage site in some T-even-like phages with or without optional genes *mobC* or *segH*, and (**D**) the Hef cleavage site in T-even-like phages with and without optional genes *mobE* or *hef*.

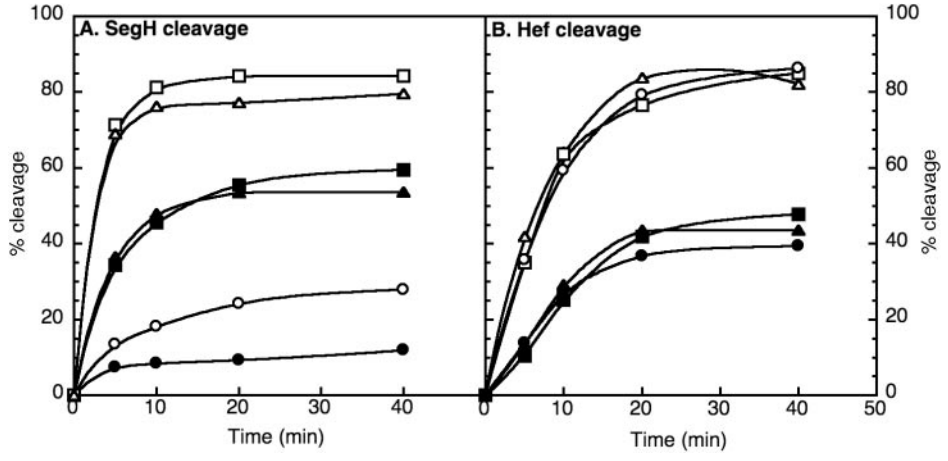
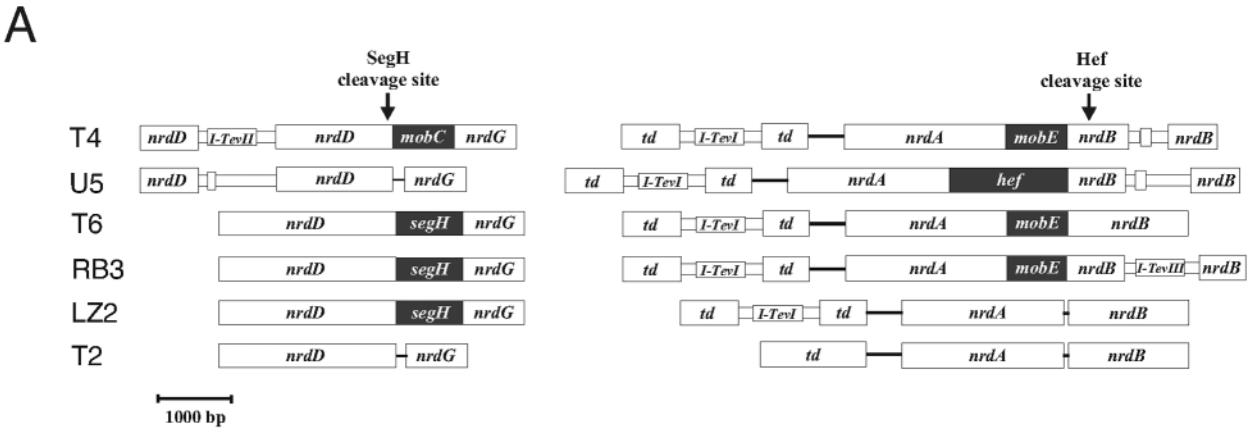
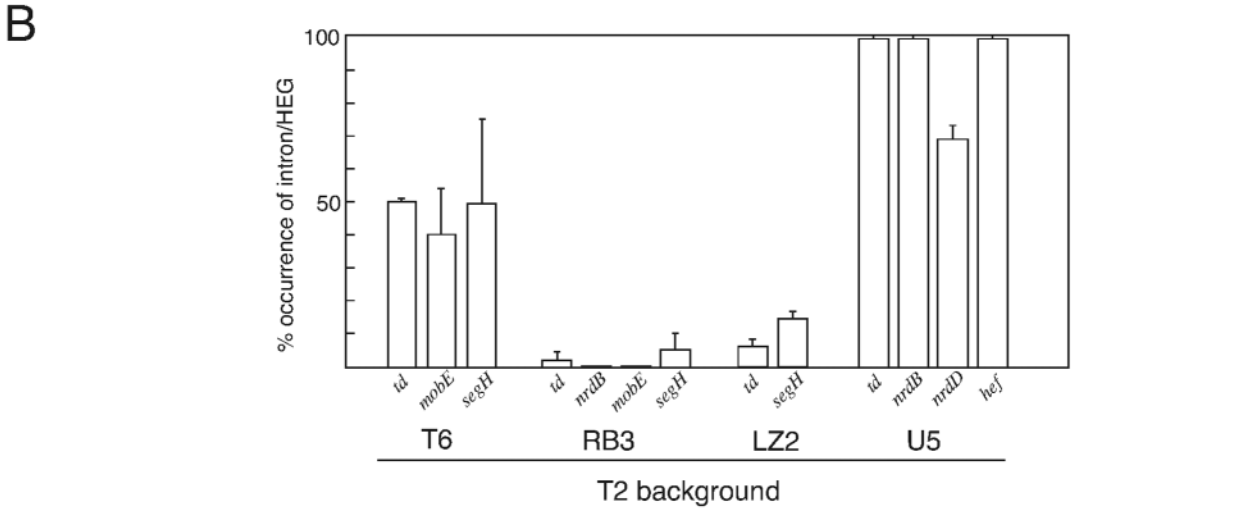
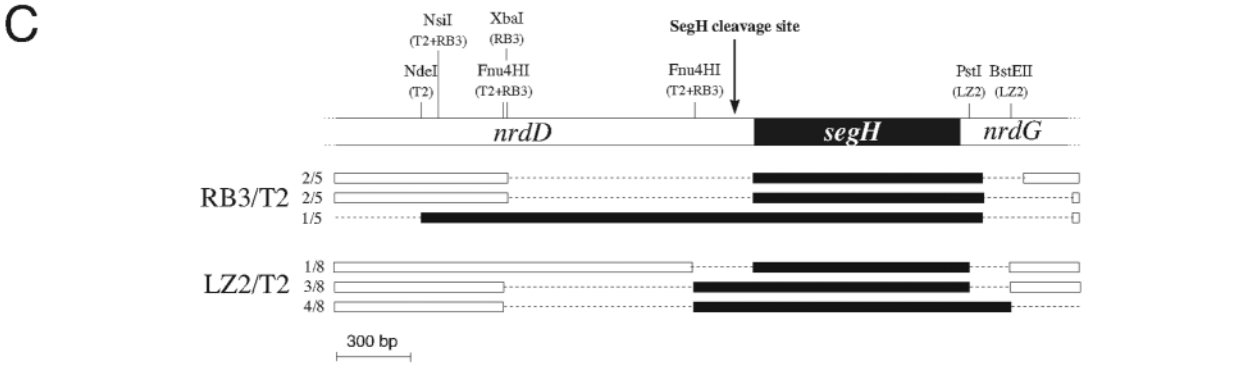
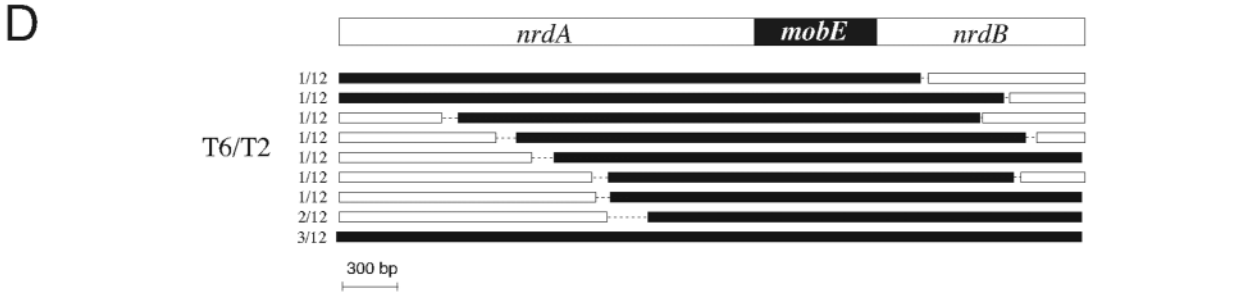


Figure 4. Competition between target sites for SegH and Hef endonucleases. (**A**) Purified SegH cleavage of target sites from T2 (triangles), T4 (squares) and LZ2 (circles). (**B**) Cleavage of T2 (triangles), T4 (squares) and U5 (circles) target sites by *in vitro* translated Hef. Open symbols denote incubations with twice as much endonuclease as in incubations denoted with closed symbols.


A
SegH cleavage site
Hef cleavage site
T4
U5
T6
RB3
LZ2
T2
nrdD
I-TevII
mobC
nrdG
td
I-TevI
nrdA
mobE
nrdB
hef
segH
I-TevIII
1000 bp


B
% occurrence of intron/HEG
100
50
td
mobE
segH
nrdB
nrdD
hef
T6
RB3
LZ2
U5
T2 background


C
NsiI (T2+RB3)
XbaI (RB3)
NdeI (T2)
Fnu4HI (T2+RB3)
SegH cleavage site
Fnu4HI (T2+RB3)
PstI (LZ2)
BstEII (LZ2)
nrdD
segH
nrdG
RB3/T2
2/5
2/5
1/5
LZ2/T2
1/8
3/8
4/8
300 bp


D
nrdA
mobE
nrdB
T6/T2
1/12
1/12
1/12
1/12
1/12
1/12
1/12
2/12
3/12
300 bp

homing endonuclease genes (23). The toxicity to *E.coli*, its scattered distribution in phages closely related to T4 and the strong exclusion of T2 markers around *mobE* in mixed infections with T4 (14,19) suggest that MobE is an endonuclease with homing properties.

SegH and Hef: two novel sequence specific endonucleases

To test if any of these putative homing endonucleases are functional we *in vitro* expressed the gene products of three of the *mobE* genes (from T4, T6 and RB3, respectively), T4 *mobC*, LZ2 *segH* and U5 *hef* from PCR fragments with a T7 promoter sequence inserted before the start codon of the respective gene. All *in vitro* reactions yielded proteins of the expected molecular masses (Figure 2A).

After degradation of RNA the *in vitro* translated proteins were used directly for endonuclease assays. As can be seen in the left part of Figure 2B incubation of LZ2 SegH translation extract with a PCR fragment fluorescently labelled at one end and spanning the *nrdD-nrdG* region from T2 generated a specific double strand cleavage product while no such activity was detected with a T4 MobC translation or with a mock translation reaction where no T7 translation template was added. SegH could also cleave DNA of the same genetic region from T4 and LZ2, and from the lengths of the cleavage products generated with DNA targets amplified with different primer combinations we were able to locate the SegH cleavage site to the end of the *nrdD* gene for all phages (data not shown).

Double strand cleavage was also shown to occur when *in vitro* translated U5 Hef was incubated with a fluorescently labelled PCR fragment spanning the *nrdA-nrdB* region from T2 (Figure 2B, right part). However, *in vitro* translated T4 MobE or a mock translation reaction without added T7 template had no endonuclease activity. Further analysis showed that Hef could also cleave T4, T6 and RB3 DNA at the same position around 250 bp into the *nrdB* gene (data not shown).

We further tested the T4, T6 and RB3 variants of MobE for cleavage of T2, T6 and RB3 DNA ranging from the intergenic region upstream of *nrdA* to the end of the *nrdB* gene (~2800 bp for T2, 3800 bp for T6 and 4000 bp for RB3) with seven different buffer conditions. No cleavage products were detected with any templates or cleavage conditions tested. The same negative results were obtained with T4 MobC on T2, T4 and LZ2 DNA spanning 1500–2000 bp including the whole *nrdG* gene and half of the *nrdD* gene (data not shown). Since several proteins with the H-N-H motif recently have been shown to nick only one strand of the DNA instead of introducing double strand breaks (24,25) we also tested T4, T6 and RB3 MobE and T4 MobC for cleavage with PCR fragments that were isotopically labelled on one strand or the other but still without detecting any cleavage (data not shown). We conclude that MobE and MobC do not display any endonuclease activity *in vitro* with the templates and under conditions tested.

Mapping of SegH and Hef cleavage sites and competition between different target sites

The freestanding T4 homing endonucleases characterized so far differ from intron-encoded homing endonucleases in that their cleavage sites are located several 100 bp away from the HEG insertion site (10,13–16). As mentioned above we roughly mapped the SegH cleavage site to the end of *nrdD* and the Hef site to within the first 300 bp of *nrdB*. By isotopic labelling of only one DNA strand at a time we were able to determine the exact cleavage sites on each strand (Figure 3A and B). Both SegH and Hef introduce double strand breaks with 2 nt 3′ extensions like many characterized phage homing endonucleases (6,13,14,26).

Cleavage by SegH occurs 82 and 84 nt from the end of the *nrdD* gene, on the coding and template strand respectively (Figure 3C). Interestingly, the SegH cleavage site coincides exactly with the position of a major shift in sequence similarity between the *segH* containing phages T6, RB3 and LZ2 and the T-even-like phages that do not contain *segH* (Figure 3C). Upstream of the cleavage site there is >95% sequence identity throughout the *nrdD* gene among all phages sequenced (19) while the last part of the *nrdD* gene, downstream of the cleavage site in the *segH* containing phages, shows only about 60% identity to *segH*-less phages. This shift in sequence similarity fits well with the co-conversion of flanking markers between the endonuclease cleavage site and the inserted HEG that is generated upon homing (27–29) and suggests that *segH* originates from a phage whose sequence has diverged substantially from the rest of the T-even-like phages.

Cleavage by U5 Hef occurs 229 and 231 nt downstream of the start of *nrdB,* on the template and coding strands respectively (Figure 3D). This region of *nrdB* is quite well conserved at the DNA level among the T-even-like phages. Therefore it is not surprising that Hef cleaves DNA from T2, T4, T6 and RB3 as well as from U5. The position of the cleavage site in the middle of the first U5 *nrdB* exon fits well with our previous prediction that it is in fact the homing activity of Hef that mobilizes the U5 *nrdB* intron (that contains a non-functional I-TevIII HEG, cf. Figure 5A) (19).

For most intron-encoded homing endonucleases, insertion of the intron disrupts the recognition site. However, for freestanding homing endonucleases the separation of the cleavage site and the insertion site frequently results in both HEG-containing as well as HEG-less alleles being cut (13–16). Extensive sequence variation which limits the frequency of

Figure 5. (**A**) Distribution of optional genes and gene order of flanking genes for the *nrdD-nrdG*, and the *td-nrdA-nrdB* regions in phages used as donors and recipients in the homing experiments. Genes are indicated by large boxes, introns with thin boxes and intergenic regions with black lines. Optional genes are dark grey with white text. Note that frame-shifted HEGs are indicated by short/interrupted ORF boxes. The cleavage positions of SegH and Hef are shown. (**B**) Frequency of screened introns and HEGs in progeny of mixed infections. Error bars show the standard error of the mean. Homing proficient genes are generally over represented in progeny from mixed infections (i.e. U5 *td* IVS) while non-homing genes are expected to have around 50% occurrence (i.e. U5 *nrdD* IVS). General exclusion by the recipient phage can reduce the effect of homing and therefore some homing proficient genes are present in less than 50% of the progeny. Such progeny were screened for recombination points between donor and recipient sequences close to the gene of interest, indicative of homing. (**C**) Screen for co-conversion of flanking markers around the *segH* gene in RB3/T2 and LZ2/T2 crosses. (**D**) Screen for co-conversion of flanking markers around the *mobE* in T6/T2 crosses. Gene order and position of restriction sites used to determine sequence specific markers are shown at the top. White boxes denote recipient alleles, black boxes denote donor alleles and dashed lines denote regions where recombination has occurred. Frequencies of the different chimerical sequences in the progeny are indicated to the left.

recombination events between the cut site and the HEG insertion site has been suggested to increase the homing efficiency of the HEG by expanding the region to be replaced (13,28). If efficient recombination were possible in the region between the cut and the HEG insertion site, repair would proceed without transfer of the HEG in the majority of repair events. If SegH recognition includes sequences downstream of the cleavage site the extensive sequence differences between HEG-containing and HEG-less alleles in this region could make *segH*-containing alleles resistant to cleavage.

To test if SegH and Hef show any cleavage specificity for different phage DNA sequences, competition assays were performed. Partially purified SegH shows a higher cleavage activity towards T2 (without any HEG in the *nrdDG* region) and T4 (with *mobC* in the *nrdDG* region) compared to the LZ2 *nrdDG* region in a mix of equimolar concentrations of PCR amplified targets (Figure 4A). If this is due to differences in target binding or in cleavage activity remains to be determined (work in progress). Competition assays with *in vitro* translated U5 Hef protein showed no preferential cleavage activity towards a specific target in a mixture of equimolar concentrations of PCR amplified *nrdAB* regions from T2 (without any HEG in the *nrdAB* region), T4 (with *mobE* in the *nrdAB* region) and U5 (Figure 4B) in accordance with the high degree of sequence similarity around the cleavage site between the phages.

Homing of *segH* and *mobE* in mixed phage infections

To assess if *segH* and *hef* can home to HEG-less phages and to further investigate the reported unidirectional inheritance of *mobE* (19), we performed mixed infections (two different phages infecting the same bacterium) using donor phages containing various combinations of *mobE*, *segH* and *hef* (Figure 5). If only neutral recombination occurs, the theoretical occurrence of an allele from either phage in the progeny of a mixed infection is 50%. However, homing endonucleases have been shown to cause a phenomenon termed localized marker exclusion where genetic markers from the HEG-less recipient phage are excluded in the progeny of the cross while the HEG is strongly over represented (13,14,30). In addition, some phages generally exclude other phage alleles to varying extent by still largely unknown mechanisms (30). We have shown previously that T2 generally excludes T6 and RB3 markers but that genes with homing properties (i.e. the *td* intron) can partly counteract this exclusion (19). The use of partially excluded phages as donors is a good set-up for homing studies since the donor alleles found in the progeny are most likely from homing events.

To mid-log phase bacteria we added equal amounts of donor and recipient phages in excess to minimise single infection events, and the relative frequency of occurrence of HEGs in the progeny was determined. In those phages where introns known to be homing proficient were present the frequency of these in the progeny was also determined for comparison. To assess *mobE* and *segH* homing we used either T6, RB3 or LZ2 as donor phages while U5 was used as donor for homing studies of *hef* (Figure 5A). T2 was used as the recipient in all crosses. Figure 5B shows that all RB3 and LZ2 markers tested were strongly excluded by the T2 recipient. Interestingly, both RB3 and LZ2 *segH*, and in one cross also T6 *segH,* were more frequent than the respective *td* introns that are homing proficient. This indicates that also *segH* is homing proficient and can propagate itself in spite of exclusion. Notably, also *mobE* was found at a similar frequency as *td* in the T6/T2 crosses suggesting that *mobE* is homing proficient. Progeny from the RB3/T2 crosses contained no *mobE* genes or *nrdB* introns in accordance with previous reports that this region of RB3 is very strongly excluded by T2 (12,19,30). In the U5/T2 crosses *hef* was strongly over-represented in the progeny (Figure 5B), and so were also the nearby *td* and *nrdB* introns, whereas, the *nrdD* intron with its non-functional HEG was found at a lower frequency.

To analyse further if homing of the *segH*, *mobE* and *hef* genes has taken place we analysed the origin of the regions flanking these genes in all progeny that contained the HEG, from the T2 crosses (Figures 5C and D). Recombination points between donor and recipient alleles were found close to the *segH* gene in all progeny screened from both RB3/T2 and LZ2/T2 crosses (Figure 5C). This is strong evidence that homing has taken place for this gene and corroborates our earlier data with the *td* intron (19) showing that homing proficient genes can overcome exclusion. Similarly, we found that 9 of the 12 progeny from the T6/T2 crosses had at least one recombination point close to the *mobE* gene strongly suggesting that T6 *mobE* is homing proficient (Figure 5D). In a previous study we observed that the homing deficient T4 *nrdB* intron was transmitted with the same frequency as the homing proficient *td* and *nrdD* introns in T4, and speculated that T4 *mobE* mobilized the *nrdB* intron as well as its own gene (19). Similar observations were made by Liu *et al*. (14). Our current results from the T6/T2 crosses strongly support this hypothesis and suggest that T4 and T6 MobE may have a cleavage site early in the T2 *nrdB* gene. A similar screen of the progeny from the U5/T2 cross did not locate any T2 markers throughout the *nrdA* and *nrdB* genes (data not shown) making it impossible to determine if *hef* has been transferred via homing. Generally, U5 seems to exclude most phages over large genetic regions (T2, T4, RB23, LZ1, our unpublished data) making the homing results in these regions uninformative as there appears to be little mixing of alleles between U5 and other phages. The same difficulty is apparent with T4 that strongly excludes most T-even-like phages over large regions (30,31). Further analysis of homing of U5 *hef* as well as T4 *mobC* therefore has to await the construction of exact deletion variants of these strains for unbiased homing analyses (work in progress).

DISCUSSION

We have found that, like the T4 introns, the putative H-N-H endonuclease genes *mobC* and *mobE* of bacteriophage T4 are infrequently represented in closely related phages and that several phages have an unrelated gene at either of these positions. Phages T6, RB3 and LZ2 have a gene homologous to the T4 *seg*-genes replacing T4 *mobC* between their *nrdD* and *nrdG* genes. This gene is the eighth *seg*-like gene found among the T-even-like phages and we therefore name it *segH* (10,13,14). We show here that *segH*, like several of its relatives in T4, encodes an endonuclease with a cleavage site close to its gene locus, a feature common to homing endonucleases. Analysis of the co-conversion region between the *segH* cut site and insertion site indicates that *segH* has been transferred to the T-even-like phages from a genetically quite distant source.

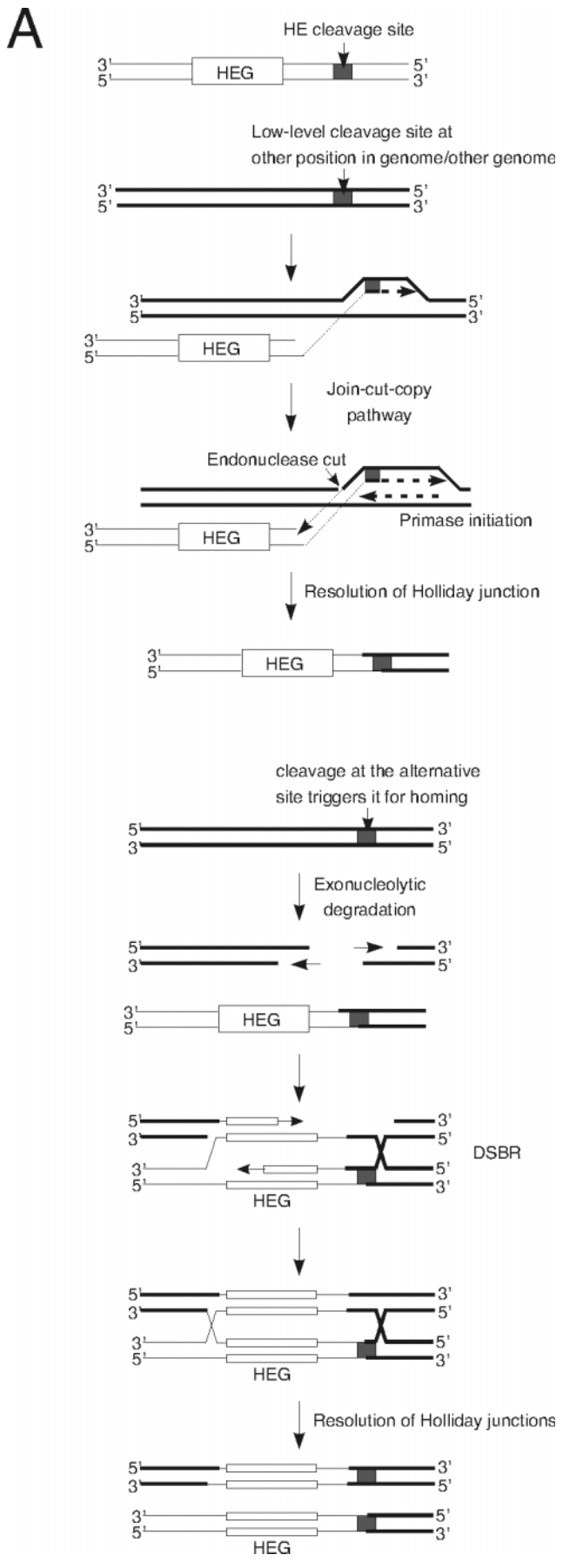


Figure 6. (**A**) Proposed pathway for transposition of a HEG via cleavage of its own genome. Occasional cleavage of the own genome by the homing endonuclease at the cleavage site gives it the opportunity to invade sites with sequence similarity at other positions in the genome, or in another genome during a mixed infection event. Single strand invasion followed by Join-cut-copy replication (38) and resolution of the Holliday junction will form a chimeric chromosome where the HEG has been joined to the new site. Short regions of initial pairing will be stabilized by DNA replication from the 3′ end of the invading strand. In those cases where the homing endonuclease can cleave the alternative site the chimeric chromosome may subsequently be used as template for double strand break repair (DSBR) using the identical part (black lines) and any other short sequence similarity on the opposite side of the HEG for initiation of repair. This will result in the HEG being inserted at the new position in the genome. (**B**) Scheme of horizontal transfer of HEG to a cognate site within a phage population and invasion and spread to new sites within the same genome or an unrelated genome. The scheme is an adaptation of the cycle of intron gain and loss from reference (19). Heavy arrows indicate preferred transfer, and light arrows less frequent events.

A novel gene was also found in phage U5 between *nrdA* and *nrdB* (in place of the *mobE* gene in T4). For *in vitro* cleavage assays the product of this gene generates a double strand cut in the downstream *nrdB* gene, and we name this gene *hef*

(for homing endonuclease-like function). Even though *hef* encodes a domain related to a diverse group of endonucleases it does not show any sequence similarity to the other known HEG families. In line with the nomenclature proposed for homing endonucleases (9,32) we suggest that the endonuclease products of *segH* and *hef* are called F-TevVIII and F-TevIX, respectively.

Our results show that *segH* from LZ2 and RB3 can induce homing in mixed infections with T2 as recipient phage and that T6 *mobE* in mixed infections with T2 also promotes its own propagation via recombination although we have still not been able to demonstrate *mobE* cleavage *in vitro*. Since T-even phages use hydroxymethylcytosine instead of cytosine and also glucosylate the HMC residues *in vivo*, with the glucosylation varying in pattern and amount between phages (33), it is not unlikely that our *in vitro* experiments with unmodified DNA do not produce the conditions needed for recognition and cleavage by these enzymes.

Both SegH and Hef produce double strand cuts generating 2 nt 3′ extensions and, at least *in vitro*, they cleave both HEG-containing and HEG-less alleles as seen in previously tested T4 Seg-endonucleases (13–16). This is in sharp contrast to the majority of intron-encoded homing endonucleases that do not cleave their own intron-containing alleles. Cleavage of the HEG-containing allele is likely to reduce the efficiency of homing. Why then do so many of the freestanding homing endonucleases cleave their own DNA? We propose that modest cleavage of the own genome does not reduce phage viability and that it may increase the frequency of transposition of the HEG to new sites within the same genome or between genomes (Figure 6). Frequent transposition to new sites followed by new spread between phages via homing is vital for the survival of a HEG in a population (19,34,35).

T4 utilizes its very efficient recombination machinery to ensure replication of the ends of its linear chromosome and at the same time increases the number of replication forks on the phage DNA (36,37). Single stranded 3′ ends of T4 chromosomes are highly recombinogenic and invade cognate sites on sister chromosomes followed by replication with the invaded DNA as template [for mechanism see (38)]. We propose that it is beneficial for the spreading of a homing endonuclease that the DNA is cut near the HEG region since this will generate recombinogenic DNA ends close to the HEG that can invade a second genome-copy. Invasion will in the majority of cases, due to sequence identity, occur at a cognate site in an uncleaved genome-copy resulting in increased genome replication from that position and will therefore not have a negative effect on phage viability. However, in the event of invasion at a secondary site within the same genome or the genome of a second infecting phage, strand invasion will lead to relocation of sequences from the cleavage site and onwards including the HEG (Figure 6A). This can result in genome shuffling within a single genome or in mixing of genomes between phages infecting the same bacterial cell (39,40). Due to the requirement of sequence similarity for strand invasion, such sites may be secondary low-level cleavage sites for the homing endonuclease. After transposition to the new site, HEGs that can cleave HEG-less alleles at the new site will be able to initiate homing to that site and will spread in the population (Figure 6B). However, extensive cleavage of the HEG-containing genome is likely to compromise the efficiency of homing and therefore mutations (either in the recognition site of the HEG-containing genome or in the HEG itself) that lower self-cleavage will produce HEGs with a higher homing frequency that will promote their own super-Mendelian inheritance in the phage population. In agreement with this, many of the characterized T4 *seg*-genes have relaxed cleavage site specificity with several alternative cleavage sites in other parts of the genome and most of them also cleave their own DNA although they have a preference for cleaving HEG-less alleles (13–16). In addition variations in *in vivo* DNA modification patterns between phages may alter the cleavage specificity further. Our results show that SegH from LZ2 has a preference for site-specific cleavage of T2 and T4 over LZ2 *in vitro*. T2 and T4 both lack the *segH* but T4 has the *mobC* gene instead at the equivalent site. The sequence variation surrounding the SegH cleavage site is less between T2 and T4 but quite high downstream of the cleavage site between LZ2 on the one hand and T2 and T4 on the other. This high variation downstream of the site, differentiating the *segH*-containing phages LZ2, RB3 and T6 from T2 and T4 may contribute to the difference in target preference.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Marcus Boman for help with the *mobC* screen, Kajsa Eriksson for help with the purification of SegH, and Patrick Young for helpful discussion and constructive criticism on the manuscript. This work was supported by the Swedish Science Research Council. Funding to pay the Open Access publication charges for this article was provided by the Swedish Science Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Dujon,B. (1989) Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene*, **82**, 91–114.
2. Belfort,M. and Perlman,P.S. (1995) Mechanisms of intron mobility. *J. Biol. Chem.*, **270**, 30237–30240.
3. Perlman,P.S. and Butow,R.A. (1989) Mobile introns and intron-encoded proteins. *Science*, **246**, 1106–1109.
4. Lambowitz,A.M. (1989) Infectious introns. *Cell*, **56**, 323–326.
5. Belfort,M. (1990) Phage T4 introns: self-splicing and mobility. *Annu. Rev. Genet.*, **24**, 363–385.
6. Bell-Pedersen,D., Quirk,S., Clyman,J. and Belfort,M. (1990) Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Res.*, **18**, 3763–3770.
7. Belfort,M. (1991) Self-splicing introns in prokaryotes: migrant fossils? *Cell*, **64**, 9–11.
8. Eddy,S.R. and Gold,L. (1992) Artificial mobile DNA element constructed from the EcoRI endonuclease gene. *Proc. Natl Acad. Sci. USA*, **89**, 1544–1547.
9. Belfort,M. and Roberts,R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
10. 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.

11. 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.
12. 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.
13. 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.
14. 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.
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. 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.
17. 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.
18. Repoila,F., Tetart,F., Bouet,J.Y. and Krisch,H.M. (1994) Genomic polymorphism in the T-even bacteriophages. *EMBO J.*, **13**, 4181–4192.
19. Sandegren,L. and Sjöberg,B.-M. (2004) Distribution, sequence homology and homing of group I introns among T-even like bacteriophages. Evidence for recent transfer of old introns. *J. Biol. Chem.*, **279**, 22218–22227.
20. Marchler-Bauer,A., Anderson,J.B., Cherukuri,P.F., DeWeese-Scott,C., Geer,L.Y., Gwadz,M., He,S., Hurwitz,D.I., Jackson,J.D., Ke,Z. *et al.* (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.*, **33**, D192–D196.
21. Sjöberg,B.M., Hahne,S., Mathews,C.Z., Mathews,C.K., Rand,K.N. and Gait,M.J. (1986) The bacteriophage T4 gene for the small subunit of ribonucleotide reductase contains an intron. *EMBO J.*, **5**, 2031–2036.
22. 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.
23. Kowalski,J.C. and Derbyshire,V. (2002) Characterization of homing endonucleases. *Methods*, **28**, 365–373.
24. 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.
25. Landthaler,M. and 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.
26. Edgell,D.R. and Shub,D.A. (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc. Natl Acad. Sci. USA*, **98**, 7898–7903.
27. 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.
28. 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.
29. Mueller,J.E., Smith,D. and Belfort,M. (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Dev.*, **10**, 2158–2166.
30. Russell,R.L. and Huskey,R.J. (1974) Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. *Genetics*, **78**, 989–1014.
31. Russell,R.L. (1967) Speciation among the T-even bacteriophages. California Institute of Technology, Pasadena, CaliforniaPhD Thesis.
32. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
33. Carlson,K., Raleigh,E.A. and Hallmann,S. (1994) Restriction and modification. In Karam,J.D. (ed.), *Molecular Biology of bacteriophage T4*. American Society for Microbiology, Washington, DC, pp. 369–381.
34. Goddard,M.R. and Burt,A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl Acad. Sci. USA*, **96**, 13880–13885.
35. Haugen,P., Wikmark,O.G., Vader,A., Coucheron,D.H., Sjøttem,E. and Johansen,S.D. (2005) The recent transfer of a homing endonuclease gene. *Nucleic Acids Res.*, **33**, 2734–2741.
36. Dannenberg,R. and Mosig,G. (1981) Semiconservative DNA replication is initiated at a single site in recombination-deficient gene 32 mutants of bacteriophage T4. *J. Virol.*, **40**, 890–900.
37. Luder,A. and Mosig,G. (1982) Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. *Proc. Natl Acad. Sci. USA*, **79**, 1101–1105.
38. Mosig,G. (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu. Rev. Genet.*, **32**, 379–413.
39. Hendrix,R.W., Smith,M.C., Burns,R.N., Ford,M.E. and Hatfull,G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl Acad. Sci. USA*, **96**, 2192–2197.
40. Breitbart,M. and Rohwer,F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends Microbiol*, **13**, 278–284.