Western University
Scholarship@Western

Digitized Theses

Digitized Special Collections

2009

PHAGE T4 mobE IS A REGULATED HOMING ENDONUCLEASE GENE THAT IS REQUIRED FOR THE trans HOMING OF THE DEFUNCT HOMING ENDONUCLEASE I- TevIII

Gavin W. Wilson
Western University

Follow this and additional works at: https://ir.lib.uwo.ca/digitizedtheses

Recommended Citation

Wilson, Gavin W., "PHAGE T4 mobE IS A REGULATED HOMING ENDONUCLEASE GENE THAT IS REQUIRED FOR THE trans HOMING OF THE DEFUNCT HOMING ENDONUCLEASE I- TevIII" (2009). *Digitized Theses*. 3869.
https://ir.lib.uwo.ca/digitizedtheses/3869

This Thesis is brought to you for free and open access by the Digitized Special Collections at Scholarship@Western. It has been accepted for inclusion in Digitized Theses by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

PHAGE T4 *mobE* IS A REGULATED HOMING ENDONUCLEASE GENE THAT IS REQUIRED FOR THE *trans* HOMING OF THE DEFUNCT HOMING ENDONUCLEASE I-TevIII

(Spine title: Characterization of T4 *mobE* and *I-TevIII trans* homing)

(Thesis format: Monograph)

by

Gavin W. Wilson

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master Of Science

School of Graduate Studies
The University of Western Ontario
London, Ontario, Canada

© Gavin W. Wilson 2009

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor

Dr. David Edgell

Supervisory Committee

Dr. Greg Gloor

Dr. David Haniford

Examiners

Dr. Greg Gloor

Dr. Caroline Schild-Poulter

Dr. Priti Krishna

The thesis by

Gavin W Wilson

entitled:

Phage T4 *mobE* is a Regulated Homing Endonuclease Gene That Is Required for the *trans* Homing of the Defunct Homing Endonuclease I-TevIII

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

Date________________________

Chair of the Thesis Examination Board

Abstract

The *E. coli* bacteriophage T4 has fifteen homing endonuclease gene insertions, five of which are predicted to be free-standing HNH homing endonucleases, named *mobA* to *mobE.* The focus of my studies is the *mobE* homing endonuclease gene found inserted between the conserved aerobic ribonucleotide reductase large (*nrdA*) and small subunits (*nrdB*). I showed that *mobE* is expressed during T4 infection, and a conserved Rho-independent transcriptional terminator regulates *mobE* expression. I demonstrated that *mobE* is a functional homing endonuclease gene that introduces a nick on the non-coding strand of the T2 *nrdB* gene. In T2xT4 mixed infections, *mobE* is inherited in ~91.3% of the progeny and the inheritance was reduced to ~72.8% in infections with T4 bearing a *mobE* internal deletion of the predicted HNH motif. Consequently, MobE activity also causes *trans* homing of the defunct *I-TevIII* homing endonuclease gene encoded within an intron interrupting the T2 *nrdB* gene.

Keywords: HNH homing endonuclease; strand-specific nicking; *I-TevIII*; *mobE*; phage T4; phage T2; T-even phage; trans homing pathway; post-transcriptional regulation; homing endonuclease evolution

Acknowledgements

I would like to extend my most sincere thanks to Dr. David Edgell, without his expertise on phage genetics and homing endonucleases, the work presented in this thesis would not have been possible. Furthermore, I would like to thank Dave for his insightful advice regarding this work, and for always being available for consultation.

I would also like to thank my committee; Dr. Greg Gloor, and Dr. David Haniford for their expertise and helpful advice. Furthermore, I would like to thank Dr. Gloor for his help with the statistical analysis of my phage-to-phage cross data.

Many thanks to Ewan Gibb, the "RNA-mastermind" for his advice regarding the RNA experimentation presented in this work. Furthermore, I would like to thank Ewan for teaching me the basics of wet-lab work while I was a summer student in the Edgell lab.

Finally, I would like to thank the Edgell lab graduate students, Jon Carter, Ben Kleinstiver, and Connor Moffatt, as well as the Edgell lab technician Nancy Friedrich for their helpful advice regarding my master's thesis.

Table of Contents

CERTIFICATE OF EXAMINATION ii

ABSTRACT iii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES viii

LIST OF ABBREVIATIONS ix

CHAPTER ONE – INTRODUCTION 1

1.1. Mobile genetic elements 1
1.2. Homing endonucleases – general properties 2
1.3. Homing endonuclease families and structure 2
1.4. HNH endonucleases 5
1.5. Homing endonuclease gene mobility pathways 9
1.6. The regulation of homing endonucleases 13
1.7. Evolutionary considerations regarding free-standing homing endonucleases . 16
    (i) Specificity versus Toxicity 17
    (ii) Host resistance to homing endonuclease cleavage 18
    (iii) Target site location 19
    (iv) Impact on the transcription of flanking genes 20
1.8. The *Escherichia coli* phage T4 20
1.9. Phage T4 as a model system for homing endonucleases 24
1.10. Marker exclusion in the progeny from T2 and T4 mixed infections 28
1.11. The putative homing endonuclease gene *mobE* 28

CHAPTER TWO – MATERIALS AND METHODS 31

2.1 Bacterial and phage strains 31
2.2 Mutagenesis of phage T4 / T2 31
2.3 Creation of pT2nrdB^R 37
2.4 Isolation of phage DNA 37
2.5 Mixed phage infections and mobility assays 38
2.6 Sequencing of progeny phage 39
2.7 RNA extraction and purification 40
2.8 RT-PCR 40

2.9 5' RLM RACE 41
2.10 3' RLM RACE 41
2.11 RNase protection assays 41
2.12 Northern blots 42
2.13 Primer extension assays 42
2.14 Southern blots 43
2.15 Sequence comparison of T-even phage 44

CHAPTER THREE – RESULTS 45

3.1 Computational analysis of the *nrd operon* in phages T4 and T2 45
3.2 Profiling *mobE* expression during T4 infection 48
3.3 The expression of *mobE* is regulated by a transcriptional terminator 52
3.4 *mobE* encodes a functional and mobile endonuclease 55
3.5 Inheritance of *I-TevIII* is dependent on *mobE* function 63
3.6 Sequencing of T2xT4 progeny localizes the MobE target site to T2 *nrdB* 64
3.7 Creation of a MobE resistant T2 phage 67
3.8 Primer Extension Mapping of MobE-induced nicks in T2 *nrdB* 67
3.9 Southern Blot mapping of double strand breaks *in vivo* 73

CHAPTER FOUR – DISCUSSION 77
4.1 The *mobE* insertion has successfully integrated into the *nrd* operon 77
4.2 An intact *mobE* gene is required for *mobE* mobility 81
4.3 MobE promotes the inheritance of *I-TevIII* 82
4.4 The location of *mobE*-dependent nicks map to conserved regions of *nrdB* 87
4.5 Stand-specific nicks and the homing pathway 91
4.6 Future directions and experiments 94

REFERENCES 96

CURRICULUM VITAE 106

List of Figures

Figure 1. Endonuclease mediated homing pathways 4

Figure 2. The structure of the HNH homing endonuclease I-HmuI 8

Figure 3. The homing pathway initiated by a DSB in the recipient genome 12

Figure 4. Schematic of the *nrd* operon in phages Aeh1 and T4. 15

Figure 5. The temporal regulation of transcription in the phage T4 23

Figure 6. HEG insertions in the phage T4 genome 26

Figure 7. Overview of the genomic organization of the T4 and T2 *nrd* operon 47

Figure 8. *mobE* and *nrdA* are co-transcribed 51

Figure 9. Transcriptional termination in the 5' end of *mobE* 54

Figure 10. Conservation of the *mobE* transcriptional terminator in T-even phage 57

Figure 11. T2 x T4 mixed infections 59

Figure 12. Sequence analysis of progeny phage 66

Figure 13. Mapping of MobE-dependent nicks *in vivo* using primer extension 69

Figure 14. Mapping of MobE-dependent nicks on plasmid substrates 72

Figure 15. AseI Southern blot of phage mixed infection genomic DNA 76

Figure 16. Conservation of the promoters upstream of *nrdB* in T-even phage 80

Figure 17. *trans* homing versus collaborative homing 86

Figure 18. Logos representation of the *mobE*-dependent nicking sites in T-even phage 90

Figure 19. Model for the conversion of a strand-specific nick to a DSB 93

List of Tables

Table I. Bacterial strains, Phage Strains, and Plasmids 32

Table II. Oligonucleotides used in this study 33

Table III. Phage-to-phage mixed infections 60

Table IV. Normalized phage-to-phage mixed infections 61

List of Abbreviations

bp	base-pairs
cDNA	complementary DNA
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTPs	deoxyribonucleotide triphosphates
DSB	double-strand break
gp	essential gene product
HEG	homing endonuclease gene
kb	kilobases
NTP	ribonucleotide triphosphate
nts	nucleotide(s)
ORF	open reading frame
PCR	polymerase chain reaction
phage	bacteriophage
RNAP	RNA polymerase
RLM-RACE	RNA ligase mediated rapid amplification of cDNA ends
RNase	ribonuclease
RBS	ribosomal binding site
RLM-RACE	RNA ligase mediated rapid amplification of cDNA ends
RNA	ribonucleic acid
RPA	RNase protection assays
wt	wild-type

Introduction

1.1 Mobile genetic elements

Mobile genetic elements are sequences of DNA that are capable of movement within an organism's genome or between one organism's genome to another (1). Typically, the mobile genetic element encodes the enzymes necessary for the mobility and the maintenance of the element within the host genome. The movement of mobile DNA significantly contributes to the genome dynamics of the host organisms and the horizontal transfer of genetic information. Mobile genetic elements occupy a large proportion of an organism's genome; for example, ~45% of the human genome is composed of these elements (2). The movement of these elements can cause various gene rearrangements and mutations in an organism's genome, including gene conversion, duplication, deletion, and disruption events (1). The effect of the aforementioned mutations can have a deleterious effect on an organism's viability, for example mobile genetic elements have been implicated in cell transformation in humans (1,3). Moreover, the study of these elements is essential to understand the function and role of the different sequences present in an organism's genome.

Mobile genetic elements are predominantly considered selfish conferring no selective advantage to their hosts; the sole purpose of their encoded genes is the propagation and spread of themselves (4-6). Notable selfish genetic elements include retrotransposons such as the Alu sequence in eukaryotes, DNA transposons such as insertion sequence elements in prokaryotes, and homing endonuclease genes (HEGs) that are found in a range of organisms (7-9). Nonetheless, some mobile genetic elements have collected gene insertions that are beneficial to their host for example, antibiotic resistance genes, virulence genes, and iron uptake genes (10). One interesting exception to the trend of mobile genetic elements being selfish is the staphylococcal cassette chromosome (11). The cassette chromosome encodes the methicillin resistance gene, which is the primary cause of methicillin resistant *Staphylococcus aureus*. Collectively, the effects caused by the movement of mobile genetic elements whether neutral, deleterious or beneficial highlight the importance of their molecular biology and evolution.

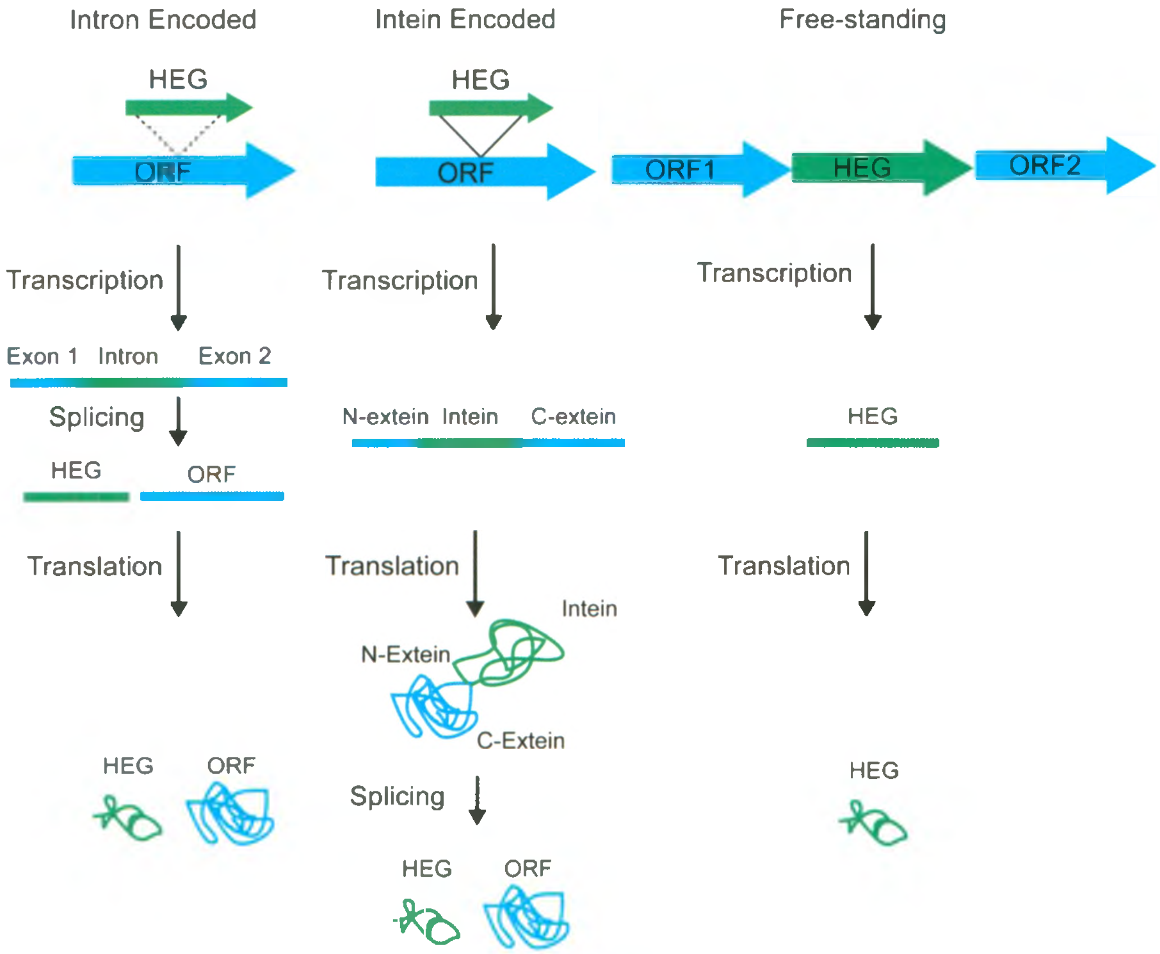
1.2 Homing Endonucleases – general properties

HEGs are selfish mobile genetic elements that are found in unicellular eukaryotes, bacteria, and bacteriophage (9,12). HEGs are often found encoded in an intron or intein that is inserted within a conserved gene, or as a free-standing ORF between two conserved genes (**Figure 1**). The HEG encodes a site-specific DNA endonuclease that catalyzes a sequence-specific break in the DNA of a genome lacking the HEG insertion. These breaks promote mobility of the HEG through the homing pathway, leading to dominant non-mendelian inheritance of the HEG and its flanking sequence (Discussed in section 1.5). The movement of HEGs causes DNA to be shuffled between related genomes, implicating them as a major source of genetic variability.

1.3 Homing endonuclease families and structure

Homing endonucleases are categorized into different families based on conserved amino acids in their catalytic motifs. The known families are, HNH, GIY-YIG, LAGLIDADG, His-Cys-Box, and the recently discovered PD-(D/E)-XK and VSR-like

Figure 1. Endonuclease mediated homing pathways. The three types of homing endonuclease insertions are shown. Intron encoded homing endonucleases are spliced out at the transcript level, which results in an intact host gene mRNA and HEG mRNA. Intein encoded homing endonucleases are spliced out at the protein level, which results in an intact host gene protein and the HEG protein. Finally, free-standing homing endonucleases are independent ORFs, thus they do not undergo any splicing reactions for their expression. Figure adapted from (22).


Intron Encoded
Intein Encoded
Free-standing
HEG
ORF
HEG
ORF
ORF1
HEG
ORF2
Transcription
Transcription
Transcription
Exon 1
Intron
Exon 2
Splicing
HEG
ORF
N-extein
Intein
C-extein
HEG
Translation
Translation
Translation
Intein
N-Extein
C-Extein
HEG
ORF
Splicing
HEG
ORF
HEG

families (9,13,14). Typically, homing endonucleases have one of three general structures; homodimers and pseudosymmetric monomers, or an elongated structure consisting of distinct DNA binding and catalytic domain (9). Both of the aforementioned structures cause the homing endonuclease to have an extended target site of 14-36bp. However, the number of base-specific contacts made by homing endonucleases to their substrate are many fewer than are made by restriction enzymes, which have shorter target sites and more base-specific contacts to their substrates (9,15). The majority of these sequence-specific contacts are hydrogen bonds and nonpolar van der Waals interactions through the major groove of DNA (16). For example, the restriction enzyme EcoRI has a 6bp target site with 28 base-specific contacts, whereas the homing endonuclease I-HmuI has a 24bp target site with 14 base-specific contacts (15,17). Furthermore, restriction enzymes, unlike homing endonucleases, tend to be intolerant of substitutions in their target sites (15). The length of the homing endonuclease target site and the number of base-specific contacts highlights an important feature of homing endonucleases that is discussed further in section 1.7(i).

1.4 HNH homing endonucleases

The focus of my thesis is the HNH homing endonuclease family, which are named after the conserved histidine (H) and asparagine (N) amino acids found in their active sites (9). Homing endonucleases containing the HNH motif are predominantly found in phage, however, they have also been found in the chloroplasts of green algae. The key feature of the HNH endonucleases family is the type of DNA strand breaks they catalyze. Unlike the other families that catalyze DSBs, the HNH family has been shown to catalyze strand-specific DNA nicks (9,18-20). The homing pathway initiated by a DNA nick is poorly understood, when compared to the abundance of information regarding the homing pathway initiated by a DSB (see section 1.5) (21,22). The HNH family of homing endonucleases are members of a larger group of enzymes containing the same motif (9). These enzymes include tranposases, restriction endonucleases, and the closely related bacterial colicins. Typically, the HNH motif is found within a protein containing multiple domains. For example, the *Escherichia coli* colicin E9 is a microbial toxin that kills bacteria through the introduction of non-specific DNA nicks in their genomes (23,24). The colicin E9 protein contains three domains, an N-terminal translocation domain, a central surface receptor binding domain, and a HNH motif-containing C-terminal catalytic domain. Homing endonucleases containing the HNH motif have a two-domain structure, an N-terminal catalytic domain and a C-terminal DNA-binding domain (9).

The active site structure of HNH endonucleases is a conserved ββα-metal nuclease fold, which is responsible for the catalysis of strand-specific DNA breaks (9,17) (**Figure 2**). For nicking to occur, the nuclease fold requires a divalent metal ion to be co-coordinated by the HNH motif. The ββα-metal nuclease fold interacts non-specifically with the DNA phosphate backbone to nick the DNA substrate. To compensate for the lack of sequence specific contacts made by the ββα-metal nuclease fold, the catalytic domain of the homing endonuclease tends to contain additional secondary structure to provide more sequence specificity. This additional sequence specificity within the catalytic domain is presumed to prevent non-specific cleavage of the host genome.

Figure 2. The structure of the HNH homing endonuclease I-HmuI. The crystal structure is of *I-HmuI* with its DNA substrate is illustrated above from the PDB id: 1U3E (17). The ββα-metal nuclease fold is shown in red, the HNH motif residues are shown in blue, the metal ion is shown as a grey sphere, and the remainder of the catalytic domain is shown in cyan. Finally, the DNA binding domain is shown in green. Notably the whole enzyme has a very elongated structure that binds 25bp of DNA.

Notable HNH homing endonucleases include *I-TevIII* found in T-even phage, *I-HmuI* and *I-HmuII* from closely related *Bacillus subtilis* bacteriophages SPO1 and SP83, *I-CreII* from the chloroplast of *Chlamydomonas reinhardtii*, and *F-TflI*, *F-TflII* and *F-TflIV* from phage T5 (18,19,25-27). The divalent metal ion used for optimal cleavage varies for different homing endonucleases. For example, *I-HmuI* and *I-HmuII* nick optimally with Mg^{2+} or Mn^{2+}, and *I-TevIII* nicks optimally with Mg^{2+} or Ni^{2+} (17,25).

HNH homing endonucleases are generally thought to catalyze strand-specific DNA breaks in their target sites. However, there are few exceptions that have been shown to catalyze DSBs in its target sites (18,25,26). One example of a homing endonuclease that catalyzes DSBs is *I-TevIII*, which dimerizes to nick both strands of its target site (25). Dimerization is one mechanism that HNH enzymes utilize to introduce DSBs that lead to the initiation of the homing pathway. Additionally, the homing endonucleases *I-CreII* and *I-CmoeI* from the chloroplasts of *C. reinhardtii* and *C. moewusii* have been shown to catalyze DSBs in their target sites (26,28,29). However, the mechanism these HNH homing endonucleases use to catalyze DSBs is not understood.

1.5 Homing endonuclease gene mobility pathways

The key requirement for HEG mobility is the presence of the HEG containing (donor) and HEG lacking (recipient) genomes in the same cell or organelle (9,12,30). These genomes can be present in the same environment such as occurs when bacteriophage co-infect a bacterial cell, or upon the fusion of haploid yeast. An expressed homing endonuclease catalyzes strand-breaks in the recipient genome, which

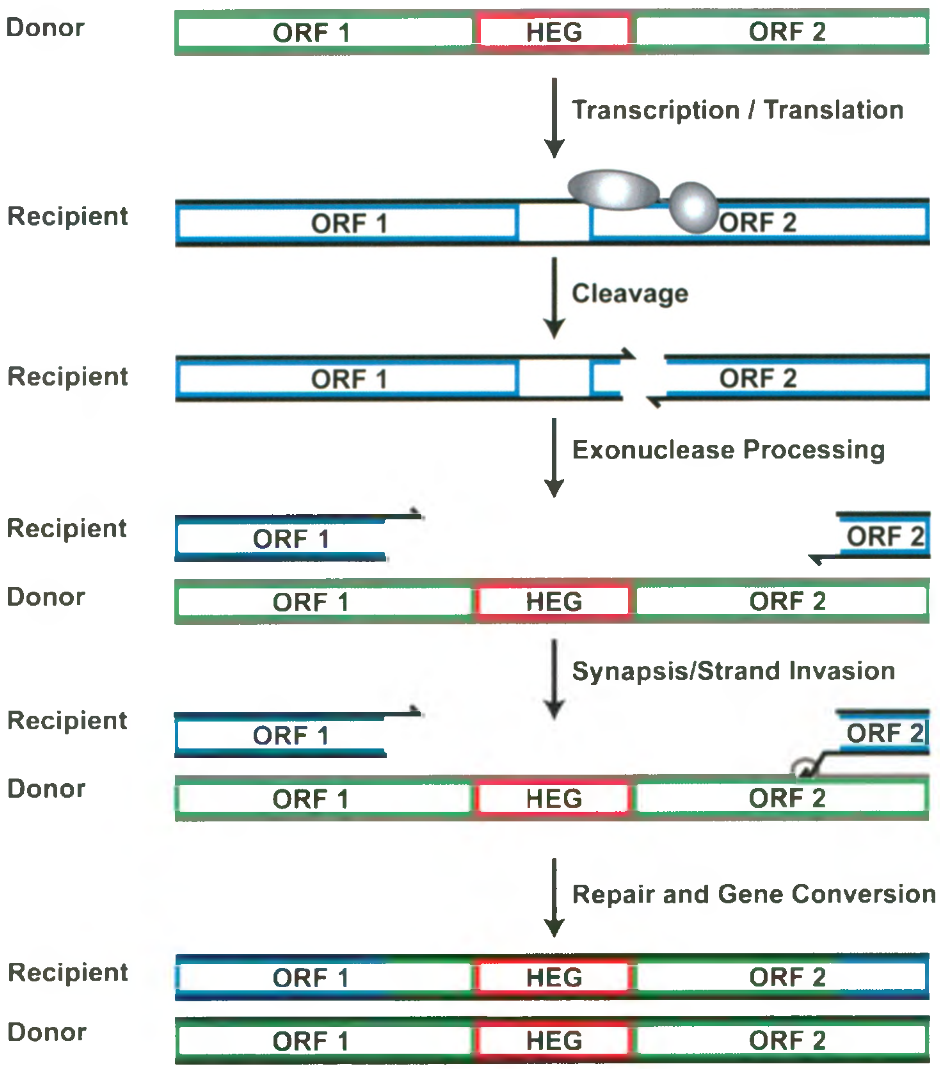
U.W.O. LIBRARY

are subsequently repaired by the host recombination-dependent DNA repair pathway. If the donor genome is used as a template for repair the HEG will be copied into the recipient genome as a byproduct of the DNA repair pathway (**Figure 3**). Currently, the rigorously characterized homing pathway begins with the introduction of a DSB in the recipient genome at the homing endonuclease target site.

After the introduction of a DSB by the homing endonuclease in the recipient genome, DNA exonucleases degrade the DNA in both directions from the DSB (21,31-33). This exonucleolytic degradation continues in both directions until a suitable region of homology is located between the donor and recipient genomes. In the *E. coli* phage T4, the exonucleolytic degradation initiated by a homing endonuclease has been shown to extend as far as 2kb from the location of the DSB (31,32,34). The final result of this degradation are 3' overhangs on each side of the degraded DNA, which are subsequently used for strand invasion of the donor genome (21). Finally, the donor genome is used as template to repair the degraded DNA with the host DSB repair machinery. The DSB repair causes duplication of the HEG into the recipient genome. As a consequence of the exonucleolytic degradation the repair pathway also causes gene conversion of genetic markers in the recipient genome flanking the HEG insertion site (21,31-36).

The homing pathway initiated by strand-specific breaks in the recipient genome has not been rigorously characterized. Previous work has shown that various nicking enzymes including *I-HmuI* and *I-HmuII,* are capable of homing and the initiation of gene conversion events in recipient phage genomes (18-20,27,37). The occurrence of gene conversion events suggest that homing events initiated by a DNA nick follow a similar pathway to the DSB initiated homing pathway (18-21,27,31-33,37-39). However, the

Figure 3. The homing pathway initiated by a DSB in the recipient genome. The homing pathway begins with transcription and translation of the HEG, followed by cleavage of the recipient allele by the homing endonuclease. Next, the cleaved region flanking the DSB is degraded by exonucleases until a suitable region of homology is reached with the donor genome. The degradation leaves 3' ends available for strand invasion of the donor genome, and the host DSB repair pathway is invoked. The DSB repair pathway causes duplication of the HEG and gene conversion of flanking genetic markers in the recipient genome. Figure adapted from (21).


Donor
ORF 1
HEG
ORF 2
Transcription / Translation
Recipient
ORF 1
ORF 2
Cleavage
Recipient
ORF 1
ORF 2
Exonuclease Processing
Recipient
ORF 1
ORF 2
Donor
ORF 1
HEG
ORF 2
Synapsis/Strand Invasion
Recipient
ORF 1
ORF 2
Donor
ORF 1
HEG
ORF 2
Repair and Gene Conversion
Recipient
ORF 1
HEG
ORF 2
Donor
ORF 1
HEG
ORF 2

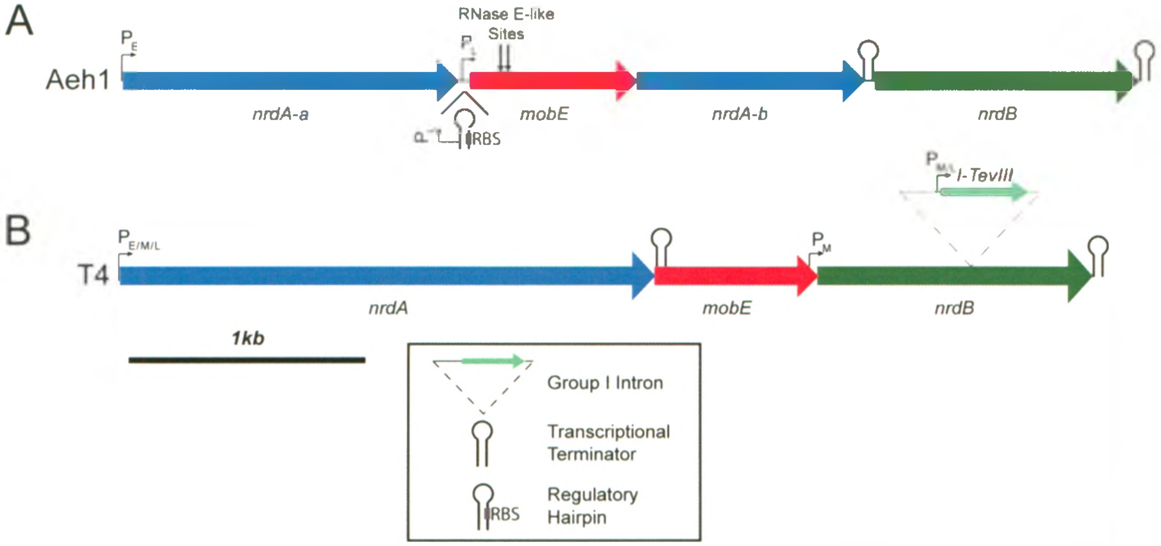
mechanism that causes a nick to initiate exonucleolytic degradation from the nicking site is not understood.

1.6 The regulation of homing endonucleases

To date, the regulation of homing endonuclease gene transcription and translation has been poorly studied. Gibb & Edgell presented the first rigorous characterization of the regulation governing HEG expression (40). They characterized the transcription and translation of *mobE* a putative free-standing HEG from the *Aeromonas hydrophila* phage. Strikingly, the *mobE* insertion within Aeh1 has split the *nrdA* gene into two parts *nrdA-a* and *nrdA-b* (41). This split *nrdA* is different than the *mobE* insertion in other phage, which is normally between the *nrdA* and *nrdB* genes (**Figure 4)** (40,41). Furthermore, the NrdA-a and NrdA-b polypeptides have been shown to interact and form a functional ribonucleotide reductase protein complex with NrdB (41).

Transcription during Aeh1 phage infection is controlled temporally by early and late promoters (42). Transcripts containing *mobE* were detected as early as five minutes post Aeh1 infection, these transcripts were derived from promoters upstream of *nrdA-a* (40). The presence of *mobE*-containing transcripts at early time points suggests that translation of the transcripts would also occur at early time points. Surprisingly, two separate mechanisms were identified that limited *mobE* expression to late time points. The first is the presence of a stem-loop structure in *mobE* transcripts derived from promoters upstream of *nrdA-a* (40,43). The stem-loop structure sequesters the *mobE* RBS inhibiting translation. However, when transcripts are initiated from the *mobE* late promoter, the stem loop structure would be disrupted and the RBS would no longer be

Figure 4. Schematic of the aerobic ribonucleotide reductase operon in the phages Aeh1 and T4. (A) The schematic of Aeh1 and **(B)** the schematic of T4. The promoters are indicated with a (P) and the class is indicated by a (E) for early, (M) for middle, and (L) for late. Transcriptional terminators are shown as stem loop structures and RNaseE-like sites are indicated with arrows. The regulatory hairpin that that sequesters the Aeh1 *mobE* RBS is also shown. The key differences between the two operons are the regulatory elements, and that Aeh1 *mobE* has split *nrdA* into two parts. Adapted from (40).


A
Aeh1
P_E
RNase E-like Sites
P_L
nrdA-a
RBS
mobE
nrdA-b
nrdB
B
T4
P_E/M/L
P_M/L
I-TevIII
P_M
nrdA
mobE
nrdB
1kb
Group I Intron
Transcriptional Terminator
RBS
Regulatory Hairpin

U.W.O. LIBRARY

sequestered. These transcripts lacking the stem-loop structure would allow for efficient MobE translation at late time points. Additionally, the translation of three intron-encoded endonucleases, I-TevI, I-TevII, and I-TevIII, present in phage T4 are limited to late time points after infection (44). This translational regulation is accomplished through a stem-loop structure upstream of the endonuclease ORFs, which sequesters their RBS in a mechanism that is similar to Aeh1 MobE (40,43,44). Furthermore, an RNase E like processing site in the 5' end of Aeh1 *mobE* was also identified (40). Cleavage of the *mobE* transcript by RNase E would prevent complete translation of the MobE protein.

Additional evidence for the down-regulation of HEG expression during phage T4 infection has also been observed (45). The intron containing *I-TevI* has a late promoter upstream of the *I-TevI* ORF and no other detectable regulatory elements. However, the Belfort lab observed that the I-TevI target site was similar to the late promoter containing operator site of *I-TevI (45).* Furthermore, I-TevI was shown to bind, but not cleave the operator site with gel-mobility shift assays, crystallography studies, and cleavage assays. This data suggests that I-TevI may function as a transcriptional autorepressor through the binding of its own operator site.

1.7 Evolutionary considerations regarding free-standing homing endonucleases

In this section and the remainder of my thesis I will focus on free-standing homing endonucleases and how they have evolved to minimize their impact on their host genomes (18,22,34,35,46-49). As mentioned previously, free-standing homing endonucleases are typically found between conserved genes in host genomes (50-52). There are a few common features a free-standing HEGs share with intron and intein encoded HEGs, such as the minimization of non-specific genome damage, and the previously mentioned down-regulation of their expression (9,30,40,43-45,51). There are some additional considerations with the insertion of a free-standing HEG compared to an intron or intein encoded HEG. The free-standing homing endonuclease typically replaces an intergenic region between two conserved genes that may contain essential regulatory elements for the proper expression of the operon harboring the HEG insertion (34,35). Without mechanisms to prevent the aforementioned effects the accumulation of HEG insertions may have an impact on host viability, and selective pressures for the host to lose the HEG insertion.

(i) Specificity versus Toxicity

As previously mentioned, homing endonucleases have extended target sites with relatively few base specific contacts when compared to restriction enzymes (9,30). The extended recognition site reduces the number of target site occurrences in the host and recipient genomes. Typically, homing endonucleases are thought to efficiently cut the recipient genome once, and most homing endonuclease aren't expected to efficiently cut the host genome (9,53). The rare occurrence of the homing endonuclease target site will reduce spurious cleavage of the host genome. Without this target site specificity, the homing endonuclease could induce non-specific DNA breaks in the host's genome, which would lead to a reduction of host viability. However, if the homing endonuclease is too specific the dissemination of the HEG to related genomes would be reduced. To increase their host range, homing endonucleases have longer target sites with fewer base-specific contacts. This reduction in specificity increases the number of genomes the homing endonuclease can introduce strand-specific breaks in. The requirement for the homing endonuclease to be both sequence specific and sequence tolerant highlights the two contrasting forces that drive homing endonuclease evolution.

(ii) Host resistance to homing endonucleases

Previous studies have shown that genomes harboring a HEG insertion are resistant to cleavage by the encoded homing endonuclease (9,30,34,35,38,54). Without this resistance the homing endonuclease would cleave the host genome reducing the host viability. Two different methods have been evolved by the HEG to prevent cleavage of the host genome. In the simplest case, the homing endonuclease cleavage is prevented through the disruption of the target site by the insertion of the intron or intein encoded HEG (9,30,38,54). The disruption of the target site can reduce cleavage or prevent binding of the target site by the homing endonuclease. However, in the case of a free-standing HEG, the target site is usually located within one of the conserved genes flanking the HEG insertion site (see section 1.5) (22).

Since the target site of the free-standing homing endonuclease and the HEG insertion site do not overlap, the insertion of the HEG won't cause target site disruption (22,34,35). Instead of disrupting the target site through HEG insertion, the host bearing the HEG insertion has accumulated nucleotide substitutions in the target site. These target site substitutions reduce the cleavage efficiency of the homing endonuclease. The cleavage resistant target site replaces the target site in the recipient genome during the gene conversion events associated with the homing pathway. The advantage of this mechanism is that the target site can be in close proximity to the insertion site, or hundreds of base pairs away. Previous studies have shown that the phage T4 homing endonuclease SegG cleaves the recipient *segG*$^-$ phage T2 in the flanking gene *gp32* 332bp downstream of the *segG* insertion site (34). Furthermore, phage T4 has accumulated nucleotide substitutions in the SegG target site relative to phage T2 that reduces SegG cleavage.

(iii) Target site location

The homing endonuclease target site tends to coincide with the DNA sequence of conserved amino acid residues in the gene targeted for cleavage (17,50). The first evidence for the target site overlapping conserved and functionally critical amino acid residues in the target gene was presented with *I-TevI* (50). I-TevI cleaves the thymidylate synthase (TS) gene within nucleotides encoding the conserved active site residues required for TS function. Following this observation with I-TevI, other homing endonuclease target sites were shown to overlap with nucleotides encoding conserved amino acids (9,34,35,51). For example, the target site for I-HmuI located in the DNA polymerase gene encodes the highly conserved amino acid residues **PN**A**QQ**F**P** (17). A second interesting observation was made regarding the nucleotide substitutions that were tolerated within homing endonuclease target sites. These sequence tolerant positions tend to coincide with non-conserved amino acids or the wobble position of conserved amino acids in the target gene (50,51). The advantage of the target site being located in conserved portions of an organism's genome is that the number of nucleotide substitutions tolerated in the target site will be reduced. This reduction in nucleotide substitutions in the target site increases the number of genomes the HEG can invade.

U.W.O. LIBRARY

(iv) Impact on the transcription of flanking genes

The insertion of a free-standing HEG typically replaces an intergenic region in the recipient genome (34,35,40). This intergenic region may contain regulatory elements essential for the proper expression of the operon targeted by the homing endonuclease for insertion. The regulatory elements replaced can include transcriptional terminators, ribosomal binding sites, and promoters. Without these regulatory elements, the flanking genes downstream of the HEG insertion may have altered expression affecting overall host fitness. To compensate for the missing lost regulatory elements, some HEG insertions have been shown to contain regulatory elements in their ORFs that are necessary for the proper expression of downstream genes. A notable example of a HEG containing regulatory elements within its ORF is *segG* from phage T4 (34). The insertion of *segG* in a related T-even phage T2 replaces two promoters that were present in the intergenic region between gp59 and gp32. However, *segG* contains two promoters in its ORF to replace the two from the intergenic region. Without these promoters proper expression of the essential phage protein gp32 would not occur, this would probably be a detriment to the overall fitness of the host.

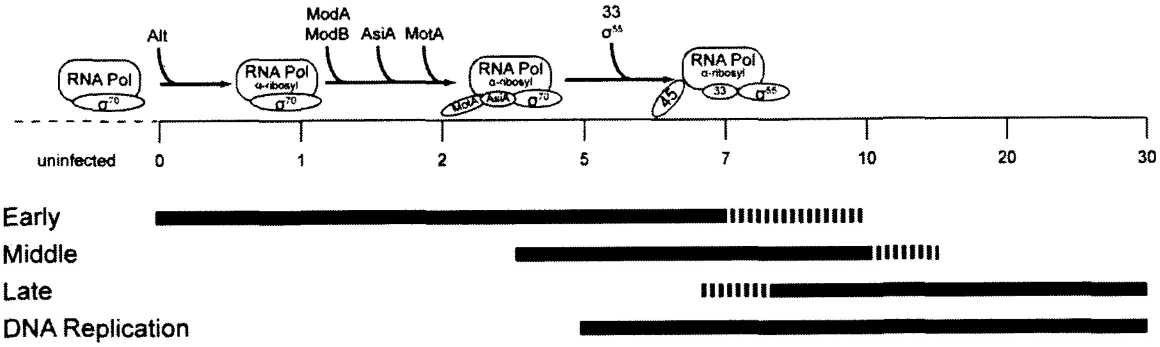
1.8 The *Escherichia coli* phage T4

The *Escherichia coli* phage T4 has a 168,903bp genome with 278 protein-coding genes and 10 structural RNAs (55,56). An interesting feature of the T4 genome is the incorporation of 5-hydroxymethylcytosine instead of cytosine (56,57). The 5-hydroxymethylcytosine bases are further modified by glucosylation of the cytosine in the α or β configuration. These DNA modifications inhibit the degradation of T4 genomic

DNA by *E. coli* nucleases. Efficient T4 infection at 37°C takes approximately 25 minutes from phage adsorption to lysis in logarithmically growing *E. coli.* From these infections 100-200 progeny phage are released into the environment from each productively infected *E. coli* cell.

The life cycle of the phage T4 begins with recognition and binding to the *E. coli* K-12 membrane receptor OmpC, or the lippolysaccharide of *E. coli* B (55,57,58). After receptor recognition, the phage protein gp5 creates a pore in the cell wall and the phage genome is injected into the cell. Once inside the cell the phage takes control of the host cellular machinery for phage production. The transcription of phage genes is temporally regulated by three classes of promoters; the so-called early, middle and late promoters (55-57) (**Figure 5**). Upon genome injection, early phage promoters that are similar to *E. coli* promoters compete with *E. coli* RNAP for binding. To skew the preference of RNAP for phage promoters, phage proteins catalyze different modifications to RNAP. The Alt protein is injected into the host with the phage genome, and Alt catalyzes ADP ribosylation of one of the two α-subunits of RNAP (59). After phage early transcription begins, the Alc protein inhibits the transcription of *E. coli* genes (57,60,61). The Alc protein binds to the *E. coli* genome acting as a site-specific termination factor preventing the transcription of *E. coli* genes. The 5-hydroxymethylcytosine bases in the T4 genome prevent Alc binding to the phage genome. Transcription of the middle mode promoters requires the translation of the early phage proteins AsiA and MotA (55-57).

Figure 5. The temporal regulation of transcription in phage T4. Phage T4 has three classes of promoters, the so-called early, middle and late promoters. The expression of these promoters is driven by *E. coli* RNAP, which is modified by phage-encoded proteins. T4 early promoters are recognized by *E. coli* RNAP, however, Alt ribosylates one of the α-subunits of RNAP to increase recognition of T4 early promoters. Next, the phage-encoded proteins AsiA and MotA bind to RNAP, which results in a change of RNAP specificity to T4 middle promoters. Finally, the host sigma factor is replaced by a phage-encoded sigma factor in conjunction with gp33 and gp45 to cause transcription of late promoters. The transcription of late phage promoters requires active replication of the phage genome. Figure adapted from (55,56).


ModA
ModB
AsiA
MotA
Alt
33
σ55
RNA Pol
σ70
RNA Pol
α-ribosyl
σ70
RNA Pol
α-ribosyl
MotA
AsiA
σ70
RNA Pol
α-ribosyl
45
33
σ55
uninfected
0
1
2
5
7
10
20
30
Early
Middle
Late
DNA Replication

Middle transcription is initiated two minutes after phage infection by the interaction of MotA and AsiA with RNAP (55-57). AsiA binds to the *E. coli* RNAP σ^{70} factor, which inhibits RNAP binding to T4 early promoters. MotA interacts with σ^{70} and acts as a class II activator by binding -30 sequence of the T4 middle promoter. This binding of MotA to the -30 sequence directs RNAP to T4 middle promoters. Additionally, ModA and ModB may also contribute to the change in specificity from early to middle promoters by the ADP ribosylation of the other α-subunits of RNAP (55). Late promoters control the transcription of the T4 proteins required for phage assembly, DNA packaging, and late phase DNA replication. Efficient late gene transcription requires DNA replication and additional modifications to RNAP. Late transcripts can be detected as early as seven minutes post phage infection (55-57). The interesting feature of late promoters is the absence of a -35 region and the presence of a conserved -10 sequence. During late transcription, middle promoters are out-competed by late promoters for RNAP binding. The phage-encoded sigma factor σ^{55}, gp33, and DNA replication is required for the initiation of transcription from late promoters.

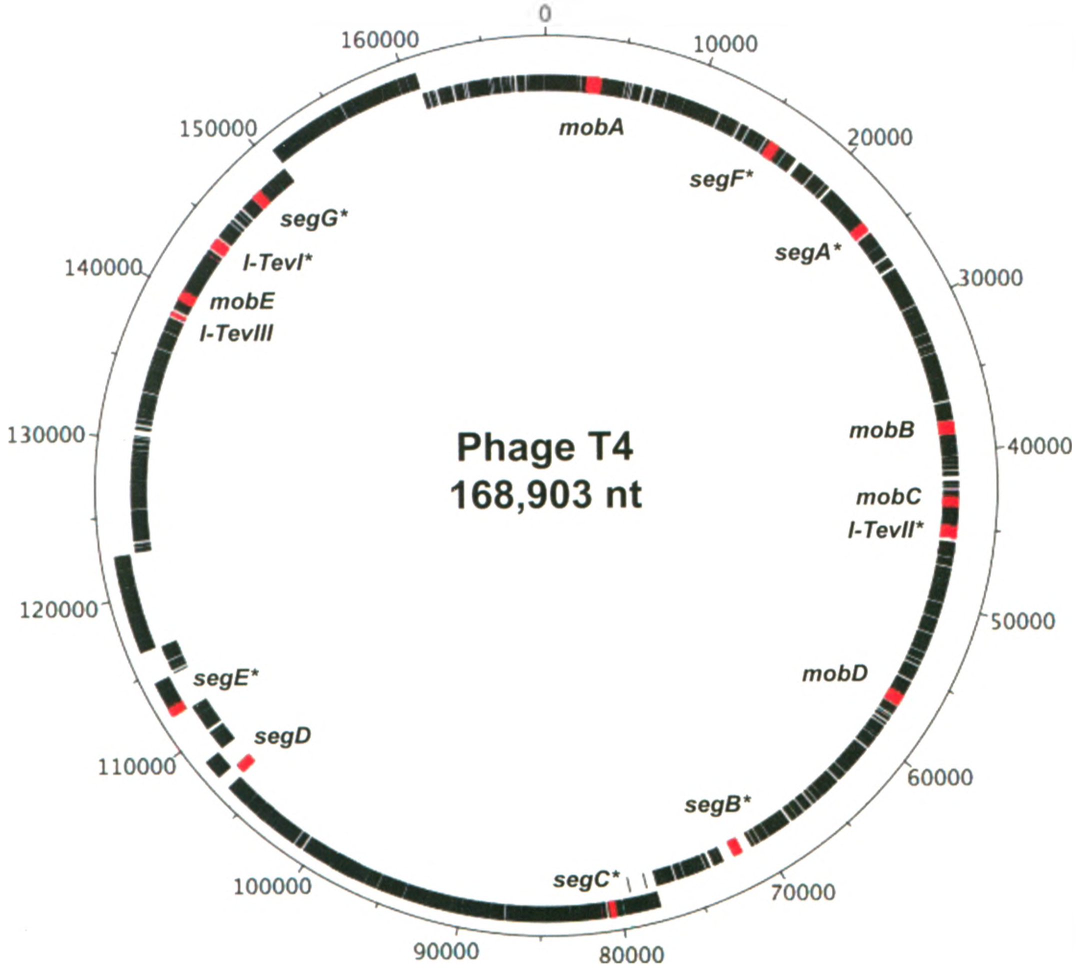
U.W.O. LIBRARY

1.9 Phage T4 as a model system for homing endonucleases

T-even phage genomes can have numerous HEG insertions, as evidenced by the phage T4 with 15 HEG insertions (**Figure 6**), constituting ~11% of the phage's genomic content (55). Twelve of these HEGs are free-standing and the remaining three are intron-encoded. The naming convention for the HEGs is based on their type and their homology to other HEGs (62,63). The three intron encoded HEGs are *I-TevI, I-TevII* and *I-TevIII,* for intron encoded T-even genes. Both *I-TevI* and *I-TevII* are members of the GIY-YIG

Figure 6. HEG insertions in the phage T4 genome.

The location of each of the 15 HEG insertions within the T4 genome is indicated in red on the circular map of the T4 genome. The HEGs that have been shown to be functional are marked with a (*), the outermost ring represents the (+) strand of the T4 genome, and the inner ring represents the (–) strand. Image generated using (81).


Phage T4
168,903 nt
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
100000
110000
120000
130000
140000
150000
160000
mobA
segF*
segA*
mobB
mobC
I-TevII*
mobD
segB*
segC*
segD
segE*
I-TevIII
mobE
I-TevI*
segG*

family, and *I-TevIII* is a defunct member of the HNH family due to an inactivating deletion in its catalytic domain (64,65). The free-standing HEG genes are the *seg* genes named for similarity to endonuclease encoded by group I introns and the *mob* genes for similarity to mobile endonucleases (62). There are seven *seg* genes, *segA* to *segG*, all members of the GIY-YIG family, and there are five *mob* genes, *mobA to mobE*, all members of the HNH family (34-36,47,48,55,66). Most of the detailed information known about the homing pathway initiated by a DSB was elucidated through studies with *I-TevI* (12,65). However, *I-TevII*, *segA – C*, and *segE – G* have also been shown to be mobile and encode functional homing endonucleases (34-36,47,48,67). To date, none of the *mob* genes have been shown to encode functional homing endonucleases.

T4 has been used as a model system for studying HEGs for multiple reasons including the ease of genetic manipulation, the short T4 life cycle, and a large library of T4 conditional lethal mutants (55-57,68-70). The conditional lethal T4 strains have amber, ochre, or opal mutations in essential phage genes, which prevent the proper translation of the mutated gene's mRNA. The mutations can be suppressed in specific *E. coli* strains expressing a tRNA that can decode the stop codon introduced in the mutation (71,72). For instance, *E. coli supF* encodes a tyrosine-suppressing tRNA. The mutated protein is expected to be restored to wild-type in the suppressor host which would restore phage viability. By utilizing the library of conditional lethal T4 mutants the enzymes required by the homing pathway could be elucidated (21).

1.10 Marker exclusion in the progeny from mixed T2 x T4 infections

An interesting effect was observed in mixed infections with T2 and conditionally lethal T4 phage stains (73). Phage T2 is a related T-even phage with high nucleotide identity to T4 but T2 lacks HEG insertions (73,74). Due to this high nucleotide identity, free-recombination is expected to occur between the T2 and T4 genomes during mixed infections. Mendellian inheritance predicts the progeny from these mixed infections would inherit a T4 specific marker at a rate of ~50%. The conditionally lethal mutants were used as genetic markers to study the inheritance of T2 and T4 specific markers during mixed infections. Progeny phage inheriting the lethal mutation would not survive in the non-suppressing host, providing a method to quantify inheritance of the genetic marker. However, when Russell and Huskey performed the experiments in 1974, the inheritance rate was skewed towards T4 markers. The T4 markers were present at rates of 70-100% in the progeny phage from these T2 x T4 mixed infections. This phenomenon was termed marker exclusion, and the reason for the observed effect still isn't fully understood. Interestingly, the regions with ~100% T4 marker inheritance were later shown to coincide with the location of HEG insertions (75). The high levels of marker exclusion in regions with HEG insertions suggest that HEGs participate in the marker exclusion phenomenon.

1.11 The putative homing endonuclease gene *mobE*

The remainder of this work will focus on the putative homing endonuclease gene *mobE* found in T-even phage. Specifically, my work will concentrate on the characterization of the putative free-standing homing endonuclease gene *mobE* from phage T4. The *mobE* gene is found as an ORF inserted between the conserved aerobic ribonucleotide reductase large (*nrdA*) and small subunit (*nrdB*) genes **(Figure 4B)**. The ribonucleotide reductase enzymes are an essential part of the DNA biosynthesis pathway, which is required for efficient phage infection (76). Without a functional NrdA/NrdB complex nucleotide biosynthesis during T4 infection is reduced to a level of ~25% relative to wild-type infections. This reduction in nucleotide biosynthesis reduced the number of phage released from each infected cell (77). The *mobE* insertion overlaps the coding region of *nrdA* and *nrdB,* which may decrease the potential for complete *mobE* excision.

The *mob* genes all contain a conserved N-terminal HNH domain, whereas the C-terminal domain, which is assumed to be the DNA binding domain, is specific to each protein (66). To date none of the *mob* genes in T4 have been shown to encode functional homing endonucleases. Previous experiments have shown that T4 specific markers in *nrdB* were inherited with ~100% efficiency in the progeny from T2 x T4 mixed infections (49). Additionally, the inheritance of the defunct homing endonuclease *I-TevIII*, which is located 675nts downstream of the *mobE* insertion has also been shown to be inherited in ~100% of the progeny from T2 x T4 mixed infections (78). However, whether or not the observed inheritance pattern is due to *mobE* activity or local marker exclusion effects has not been demonstrated.

Computational promoter predictions on the T4 genome have failed to detect any promoters upstream of the *mobE* ORF (55). These predictions suggest that *mobE* may be a promoterless ORF, but no experimental evidence has been presented to confirm this.

Moreover, no RBS was identified upstream of the *mobE* ORF suggesting that MobE translation depends on the translation of the upstream NrdA protein. Additionally, studies into the transcription of the *nrd* operon have suggested that a stem-loop structure in the 5' end of *mobE* may function as a rho-independent transcriptional terminator (79,80). Finally, the *mobE* ORF has been shown to contain two middle promoters and a RBS upstream of the *nrdB* ORF, these elements may be necessary for the proper expression of the *nrdB* gene.

The first part of my hypothesis is that *mobE* encodes a functional HNH homing endonuclease that catalyzes strand-specific breaks in DNA. These strand-specific breaks should initiate the recombination mediated homing pathway leading to the mobility of *mobE.* The second part of my hypothesis is that *mobE* is a promoterless ORF, and that *mobE* expression is down-regulated during phage T4 infection. The objective of my studies is to demonstrate that *mobE* is a mobile genetic element that has successfully integrated into the phage T4 *nrd* operon. Finally, I expect that the *mobE* nucleotide sequence has evolved to maintain the transcriptional program of the *nrd* operon.

Materials and Methods

2.1 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 sup^F were used for propagation of phage (34). *E. coli* strains were grown in LB medium supplemented with the appropriate antibiotics (100 μg/ml ampicillin or 50 μg/ml kanamycin).

2.2 Mutagenesis of phage T-even phage

Mutagenesis was performed by plasmid-to-phage recombination. Plasmids containing the desired mutations were generated by SOEing PCR reactions (81). 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 2. 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 an *E. coli* sup^0 strain, grown to an OD_{600} of 0.4 and infected with T4 at an multiplicity of infection of 4. The cells were incubated for 10 mins at 37°C then diluted 10^{-4} fold into pre-warmed LB medium and grown for an

Table I Bacterial strains, phage strains and plasmids

Bacteria	Genotype	Reference
E. coli DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Invitrogen
E. coli B40 sup^O	F^-, *argI140(am)*,str^R	(71)
E.coli B40 sup^F	F^-, *argI140(am)*,str^R, sup^F	(71)
Phage	Genotype	
T4D	wild-type	
T2L	wild-type	
T4K10	*am38, am51, denA, denB*	(69)
T4 Δ*mobE*	deletion of nucleoties 99-126	G. Wilson
T2 *mobE*R	T2 *nrdB* 1-141 replaced with T4 *nrdB* 1-141	G. Wilson
Plasmids		
pCR2.1	TA cloning vector	Invitrogen
pCRZap	pCR2.1 with a cloned in StuI site (DE-369/DE-370) using EcoRI	G. Wilson

Table II. Oligonucleotides used in this study.

Oligo	Purpose	5'-3' Sequence	Strand	Position Relative to *td* start codon	Phage
DE-30	*mobE* 3' end for mutant sequencing	CGGGATCCACTCATTTAGAATCCTTAAATTTAC	-	5397	T4
DE-193	RNA Adaptor 5'/3' RACE (5' phosphorylated)	GCUGAUGGCGAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA			N/A
DE-194	3' RACE adaptor specific	TTTCATCAAAGCCAGCAAACGC			N/A
DE-195	3' RACE adaptor specific	CAAACGCAGTGTTCATTCATCGCC			N/A
DE-196	5' RACE adaptor specific	GCTGATGGCGATGAATGAACACTG			N/A
DE-197	5' RACE adaptor specific	ACACTGCGTTTGCTGGCTTTGATG			N/A
DE-242	T2 *nrdB* southern probe	CTCATGGACGAGTCACTATC	+	3697	T2
DE-251	*mobE* specific probe	CTTCTGTTAAACGCAGACCTG	-	5078	T4
DE-256	*mobE*Δ33-62 segment 1	GAAAAATGCGAATATAACAGATTTTACCGGGCATCTAGGGATTATATGATG	-	4912	T4
DE-257	*mobE*Δ33-62 segment 2	CCGGTAAAATCTGTTATATTC	+	4883	T4
DE-259	T4 *mobE*Δ33-62 detection	TCCCTAGATGCCCGGTAAAATC	+	4785	T4
DE-345	T2 specific southern probe	AAATGGGCCTTGCGGCCCAC	-	4939	T2
DE-352	RNase protection *mobE*	GTAATACGACTCACTATAGGCCGTACTTCTGAACCTATAAAATGCTCTCTAG	-	4856	T4

	terminator + T7 promoter				
DE-353	RNase protection *mobE* terminator	CGTACTTCTGAACGTTCTG GTACTGATGATTAT	+	4630	T4
DE-363	*nrdA* northern probe	CTTATCTCCGTGATGGAAT GACAAC	+	2566	T4
DE-364	*nrdA* northern probe	TTGATGCAGTGCCATTCCA ATAGTC	-	2972	T4
DE-365	RT-PCR *mobE* / *mobE* northern probe	AACAGCTAGAGAGCATTT TATAGCG	+	4834	T4
DE-366	*nrdA* 3' RACE	TTCCAAAAGGAAAGGTTC CAATGTC	+	4531	T4
DE-367	RT-PCR *mobE* terminator / *mobE* northern probe	TCTTGTCGGAGAAATTGTA CCATAC	-	5239	T4
DE-368	5' RACE Outer *mobE* / *mobE* mutations segment 2 reverse	ATACTCTGCTCCTTGTTTA GTTCTC	-	5305	T4
DE-369	To clone StuI site into EcoRI digested pCR2.1	AATTCCGGCAGGCCTCGG			N/A
DE-370	To clone StuI site into EcoRI digested pCR2.1	AATTCCGAGGCCTGCCGG			N/A
DE-387	3' RACE *nrdA* / RT-PCR across *mobE* terminator	GAAACTCCTAAAGCCGAT GATTGTG	+	4656	T4

DE-388	*mobE* phage mutations/ pT2nrdB^R segment 1 forward	ATTATTCAGACACTCGTTGGTCTCG	+	4114	T4
DE-412	*I-TevI* specific probe (intron)	GAGGCCTGAGTATAAGGTG	+	555	T4
DE-413	*I-TevIII* specific probe (intron)	TGCGCCTTTAAACGGTTAG	+	6060	T4
DE-416	Sequencing progeny phage *nrdA* and T4 *mobE* phage mutations	TCGGTGTAGGTGTTACCAACTATGC	+	3958	T4
DE-421	*mobE* 5' end for progeny phage sequencing	TGAGCTCGGGAAATTAGGTC	-	4740	T4
DE-422	*mobE* 3' end for progeny phage sequencing	CAGCAGAGTATCCACCTTGC	+	5296	T4
DE-423	*I-TevIII* 5' end for progeny phage sequencing	TTCCATAAAGGACTTTCCAGC	-	6122	T4
DE-424	*I-TevIII* 3' end for progeny phage sequencing	GCAAAACAAGGTTCAACGAC	+	6496	T4
DE-465	T2 specific primer extension for *nrdB* 5' end	AAGCTGAAGATTGTTCATCC	+	3655	T2
DE-467	pT2nrdB^R segment 2	CAGGCAGAGGTAAAGAGAGC	-	4364	T2
DE-488	Progeny *nrdB* sequencing	CATCAACATCATTATCAATCTGTGC	-	7121 / 4881	T4/T2

DE-503	pT2nrdB^{R} segment 2	CGAGTCACTATCTATTAACTAAGTGGAAAATTTATGAGTACAGTTTTAATACAAATCCAG	+	3705	T2
DE-506	pT2nrdB^{R} segment 1	CACTTAGTTAATAGATAGTGACTCG	-	3729	T2
DE-521	T2*nrdB*^{R} phage detection	ACTCATTGAGCGGCAGATCAGTTTT	+	5505	T4
DE-522	Primer Extension +SacI site	CAGAGCCCTGAGGAAGCTTGTTAAATTGC	-	5592 / 3937	T4/T2

additional 3hr, treated with chloroform and incubated at room temperature overnight as previously described (34). 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 using a wide-ended pipette tip, which was used to infect *sup*^{O} cultures and the resultant phage were sequenced.

2.3 Construction of pT2nrdB^{R}

To create the pT2nrdB^{R} I used a SOEing PCR reaction described above, using two different phage strains as templates. For the first gene-specific segment I used T2 as template with the primers DE-388/DE-506. The second segment was generated using the primer pairs DE-503/DE-467, and one of the progeny phage from T2 x T4 mixed infections that had a recombination point near nucleotide 161 of the *nrdB* gene as template (**Figure 13**).

2.4 Isolation of phage DNA

Approximately 1 ml of a 1 × 10^{10} pfu/ml phage stock was precipitated as previously described (82), 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, expect that 1 ml of cells was collected prior to phage infection, and at 10, 20, and 30 mins post-infection. The culture was pelleted, flash frozen with ethanol and dry ice, and immediately 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%

sodium dodecyl sulfate, 10 mM ethylenediaminetetraacetic acid, 5 mM $CaCl_2$ and 0.3 mg/ml proteinase K), and incubated at 65°C for 1 hour followed by three phenol:chloroform extractions. The aqueous layer was treated with 1μl RNase A/T1 (Ambion) for 10 mins at room temperature, and DNA was precipitated with ethanol and resuspended in 100 μl 10mM TE pH 8.0.

2.5 Mixed phage infections and mobility assays

Mobility assays were performed using mixed phage infections. *E. coli sup^0* or *supF* strain was grown to an OD_{600} of 0.4 and infected with equal amounts of T2 and T4 at a multiplicity of infection (multiplicity of infection.) of 4. The infections were incubated at 37°C for 10 minutes, and subsequently diluted 1:10^4 into pre-warmed LB medium and grown at 37°C for 4 hr. 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 hershey-top agarose LB plates for 3-5 mins. The membrane was soaked in alkaline transfer buffer (1 M NaCl, 0.4 N NaOH) for 2 mins, and then neutralization buffer (1 M NaCl, 0.5 M Tris-HCl pH 7.2) for 5 mins. The membranes were air-dried, UV cross-linked using a stratalinker, and prehybridized in 0.5 M sodium phosphate pH 7.2, 1 mM ethylenediaminetetraacetic acid pH 8.0, 7% sodium dodecyl sulfate, 1% bovine serum albumin for 2 hr at 68°C. Fresh prehybridization buffer was then added with the end-labeled oligonucleotide (1x10^6 cpm). 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 ethylenediaminetetraacetic acid, 5% sodium dodecyl-sulfate, 0.5% bovine serum albumin) for 5 mins, and then with phosphate-SDS wash II buffer (40 mM sodium phosphate, 1 mM ethylenediaminetetraacetic acid, 1% sodium dodecyl sulfate) for 5 mins, and rinsed with 0.1 x saline-sodium citrate buffer. The membranes were dried briefly, and exposed to a phosphoimager screen (GE Healthcare).

2.6 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 (83) and a user-generated program created to display the nucleotide polymorphism graph shown in Figure 12.

2.7 RNA extraction and purification

Approximately 8.2×10^7 cells/ml of *E. coli sup^0* strain was infected with T4 at an multiplicity of infection of 8. Samples (6ml) 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 (purchased in 2008 catalog #AM2239). 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.

2.8 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 cooled on a mixture of ethanol and ice. Reaction mixtures were incubated at 42°C for 1 hour, and 75°C for 10 minutes with M-MuLV reverse transcriptase (New England Biolabs), with 0.5µl RNAsin (Promega), 1.25mM total dNTPs, and the manufacturers supplied buffer. The RT products were purified using PCR cleanup columns (QIAGEN). PCR amplification was performed using 5µl 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). The PCR cycle conditions were 94°C for 30s, 53.3°C for 30s, and 72°C for 1 min, with 30 cycles. The amplified products were resolved on 1% TAE-agarose gels.

2.9 5' RLM RACE

RNA-ligase mediated rapid amplification of 5' cDNA ends was performed as described with minor modifications (40). 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

2.10 3' RLM-RACE

3' RACE was adapted from Gibb and Edgell (40). 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.

2.11 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).

2.12 Northern blots

Northern hybridzations were performed as previous described with the following adaptations (40). 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 x 10^6 cpm of the nick-translated probe for hybridization. The membranes were washed in 100 ml of 0.1 x SSC, 0.2% sodium dodecyl sulfate for 10 mins at 46°C and three times in 100 ml of 0.5 x SSC, 0.2% sodium dodecyl sulfate for 10 mins at 68°C. The membranes were then dried and exposed on a phosphoimager screen.

2.13 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), 1.25mM 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 30s, annealing for 30s at temperatures specific for each primer, and 30s 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 multiplicity of infection. of 4. Samples (20 ml) samples were collected prior to infection, and at 5, 10, and 15 mins post-infection. Cells were pelleted at 5,000 x g for 5 minutes, 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 30s, 46.6°C for 30s, and 72°C for 30s for 51 cycles. The reactions were loaded on an 8% denaturing polyacrylamide gel.

2.14 Southern blots

Mixed infections were performed using the protocol described above, and 10 ml of cells was collected pre-infection, and 10 ml at 10, 20, and 30 minutes post-infection. The cells were mixed with 5ml of RNAlater and stored on ice. The genomic DNA was isolated using a Sigma GenElute bacterial genomic DNA kit following the manufacturers protocol. The GenElute columns were eluted in twice in 200 μl of elution buffer. The isolated genomic DNA was run on a 1% TAE-agarose gel to verify the quality and the concentration was quantified using a spectrophotometer. Overnight AseI (NEB) digestions were performed using 500 ng of genomic DNA at 37°C. The digestions were loaded on a 1% TAE-agarose gel, transferred to a positively charged nylon membrane (Millipore) using downward alkaline transfer, as outlined previously (84). The membranes were air-dried, UV cross-linked and stored at room temperature.

The probe used for southern hybridizations was a nick-translated PCR product representing T2 *nrdB*. The PCR reaction used primers DE-242 and DE-345 using Taq polymerase (NEB) with an annealing temperature of 47.6°C for 35 cycles. The PCR products were gel-purified (QIAGEN) and nick translated using the protocol described in the northern hybrization section. The probe was denatured at 95°C for 3 minutes and approximately $3x10^7$ CPM of the probe was used in each hybridization reaction. Hybridzation and washing was carried out using the same buffers and the hybridization temperature were 68°C for prehybridization and 42°C for hybridization. The membranes were washed once with phosphate-SDS wash I and twice with phosphate-SDS wash II, the membranes were rinsed with 0.1x SSC and exposed to a phosphoimager screen (GE).

2.15 Sequence comparison of T-even phage

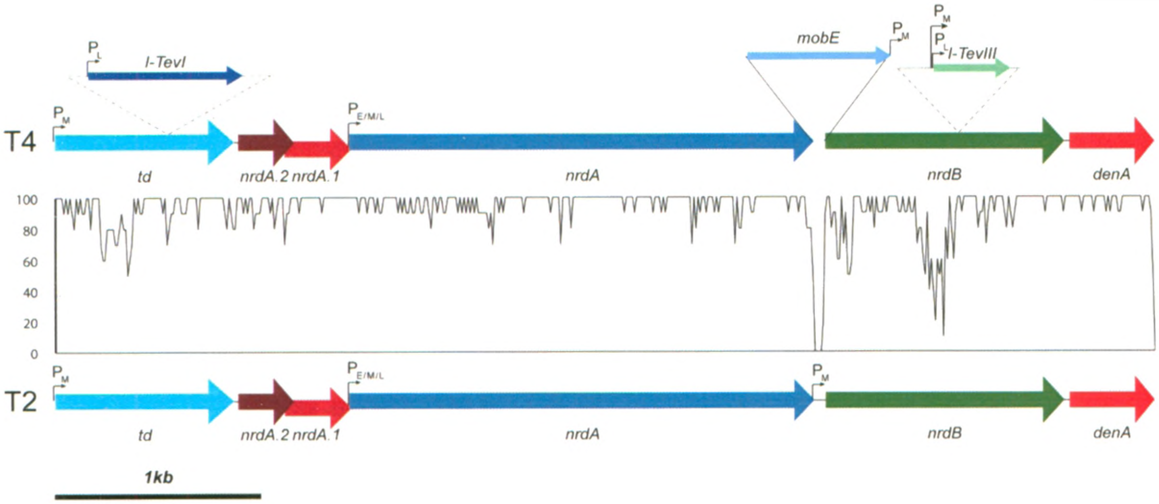
The *nrdB* nucleotide sequences of T-even were divided into two groups, phage containing *mobE* and phage without *mobE*. The region corresponding to the *mobE*-nicking site was aligned in both groups using ClustalW (83), and a sequence logo generated (85). The accession numbers or database sources for 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).

3.1 Computational analysis of the *nrd operon* in phages T4 and T2

A computational analysis of the *nrd* operon from the phages T2 and T4 was performed to identify potential regulatory elements and HEG insertions. T4 and T2 have similar gene *nrd* operon gene arrangements with high nucleotide identity (**Figure 7**). T4 has two HEG insertions, *mobE* between *nrdA* and *nrdB*, and *I-TevIII* an intron encoded HEG inserted within *nrdB*. Both *mobE* and *I-TevIII* encode putative HNH family homing endonucleases, however, *I-TevIII* contains an inactivating deletion that abolishes cleavage activity (64). The first adenine nucleotide of the *mobE* start codon is shared with the final adenine of the *nrdA* stop codon. The *mobE* ORF also overlaps *nrdB* by two nucleotides. T2 in contrast, has no HEG insertions in the *nrd* operon. Finally, the GIY-YIG homing endonuclease *I-TevI* is ~3.1kb upstream from the start codon of *mobE.*

Both T2 and T4 have early, middle and late promoters upstream of *nrdA* and middle promoters upstream of *nrdB*. The T2 *nrdB* middle promoters are contained within an intergenic region between *nrdA* and *nrdB*. In T4 the *mobE* ORF contains two middle promoters in the 3' end (79). Analysis of the sequence upstream of *mobE* yielded no discernable promoters or ribosomal binding sites. The lack of an upstream *mobE* promoter suggests that *mobE* expression is dependent on the promoters upstream of *nrdA* (80). The overlap of the *nrdA* and *mobE* ORFs and the lack of a ribosomal binding site upstream of *mobE* suggest that translation of *mobE* is coupled with *nrdA* translation. Translational coupling has been implicated in the translation of other T4 genes, for example *gp62* is coupled to the expression of *gp44* (86). The translation of the coupled gene *gp62* was shown to be four-fold less than the translation of *gp44.*

Figure 7. Overview of the genomic organization of the T4 and T2 *nrd* operon. The plot shown is a nucleotide identity comparison between T2 and T4 over a 10nt sliding window. Genes are indicated by arrows, dashed lines indicate self-splicing group I introns, and the solid lines indicate the T4 *mobE* insertion. Early (E), middle (M), and late (L) promoters are indicated by arrows with a (P) overtop.


P_L
I-TevI
mobE
P_M
P_M
P_L
I-TevIII
P_M
P_E/M/L
T4
td
nrdA.2 nrdA.1
nrdA
nrdB
denA
100
80
60
40
20
0
P_M
P_E/M/L
P_M
T2
td
nrdA.2 nrdA.1
nrdA
nrdB
denA
1kb

The translation of *mobE* may also be reduced due to the predicted translational coupling with *nrdA*.

Previous studies have suggested that a stem-loop structure within the *mobE* ORF may function as a Rho-independent transcriptional terminator (79,80). However, the study did not test if the terminator was functional. The stem-loop structure is located in the 5' end of the *mobE* ORF, 27nts upstream of the *mobE* start codon. The predicted structure of the stem-loop is consistent with the structure of a Rho-independent transcriptional terminator, including a 9nts stem, a 4nts loop, and a poly(U) tract after the stem. However, the poly(U) tract is 3nts long, which is indicative of a poor transcriptional terminator (87).

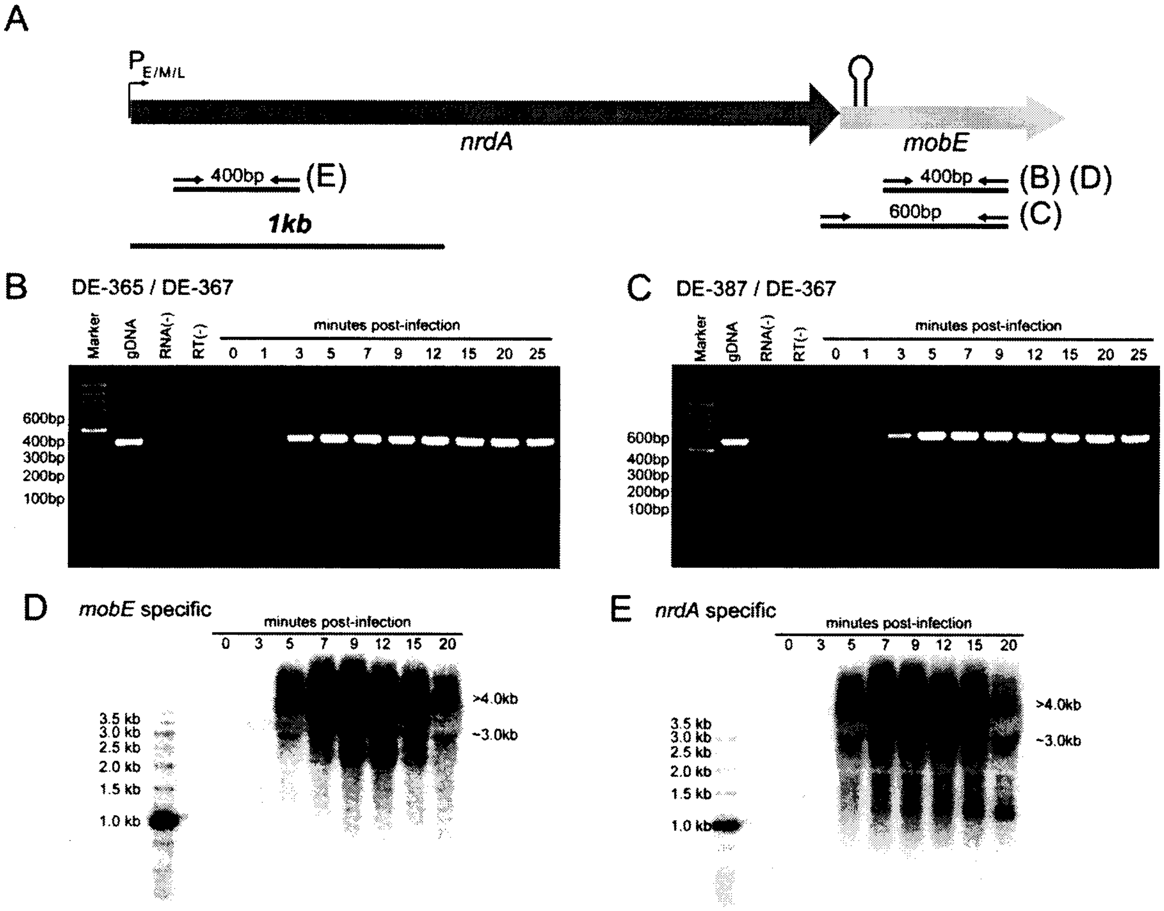
3.2 Profiling *mobE* expression during T4 infection.

Currently, there is no evidence that *mobE* is expressed during phage infection. To assay if *mobE* was expressed during T4 infection, RT-PCR was performed on RNA isolated prior to infection and at nine time points post infection (one to twenty-five minutes). The primer used for reverse transcription (DE-367) was designed complementary to sequences downstream of the predicted *mobE* transcriptional terminator. The location of DE-367 allows for the detection of transcripts that have read through the predicted transcriptional terminator in the 5' end of *mobE.* The first PCR reaction was performed using the primers DE-367/DE387, which are located downstream of the terminator (**Figure 8A**). Transcripts containing *mobE* were detected from three minutes post-infection to the last isolated time point of 25 minutes **(Figure 8B)**. The second PCR reaction used primers spanning *nrdA/mobE* (DE-365/DE367) allowing for the detection of transcripts containing both *nrdA* and *mobE.* Transcripts containing *nrdA/mobE* were detected at the same time points as the first PCR reaction **(Figure 8C)**. The presence of transcripts containing *nrdA* and *mobE* suggests that *nrdA* and *mobE* are present on a polycistronic mRNA.

To confirm that *nrdA* and *mobE* are present on a polycistronic mRNA, Northern blot hybridization was used with nick-translated *mobE* and *nrdA* specific probes **(Figure 8 D/E)**. One discrete hybridizing band was detected on both blots with an estimated size of 3kb, and a non-discrete set of bands with a size larger than 4kb were also detected. The smaller 3kb fragment may be a processed mRNA product containing both *nrdA* and *mobE*. The slower migrating bands represent previously described polycistronic mRNA messages that span as far downstream from *nrdA* as *frd* and as far upstream as *denA* (79,80).

To detect if there were any potential transcripts initiating upstream of the *mobE* ORF, I performed 5' RLM-RACE on RNA pooled from early (3,5,7 minutes) and late (9,12,15 minutes) time points in separate reactions. I was unable to detect any 5' ends upstream of *mobE* using 5' RACE. The 5' ends of transcripts initiating from promoters upstream of *nrdA* weren't detected, which may be due to the lack of processivity of M-MuLV reverse transcriptase. The same adaptor ligated RNA pool was used to map the 5' end of *nrdA* transcripts, where two clones were sequenced that mapped to the *nrdA* middle promoter **(Data not shown)**. The detection of *nrdA* 5' ends suggested that the lack of *mobE* 5' ends was not a failure of the experimental protocol. Collectively, these experiments provide evidence that *mobE* lacks an upstream promoter, and that *mobE* is dependent on the upstream transcription for expression.

Figure 8. ***mobE* and *nrdA* are co-transcribed.** (A) Schematic of the phage T4 *nrdA* and *mobE* genes. Promoters and putative RNA stem-loop structures are indicated. The positions of primers used for RT-PCR reactions, or for the generation of Northern blot probes are indicated by arrows, and their expected lengths are indicated. (B/C) RT-PCR reactions with primers specific to *mobE* (B) or to the *nrdA/mobE* junction (C). The RT-PCR reactions were loaded on 1% TAE-Agarose gels. gDNA, PCR with genomic DNA; RNA(-), RT-PCR without RNA; RT(-), RT-PCR without reverse transcriptase; 0 minutes, uninfected; 1-25 minutes, post-T4 infection. (D/E) Northern blot analyses using probes specific to *mobE* (D) or to *nrdA* (E), the positions of the probes used for the Northern blots are indicated in the schematic.

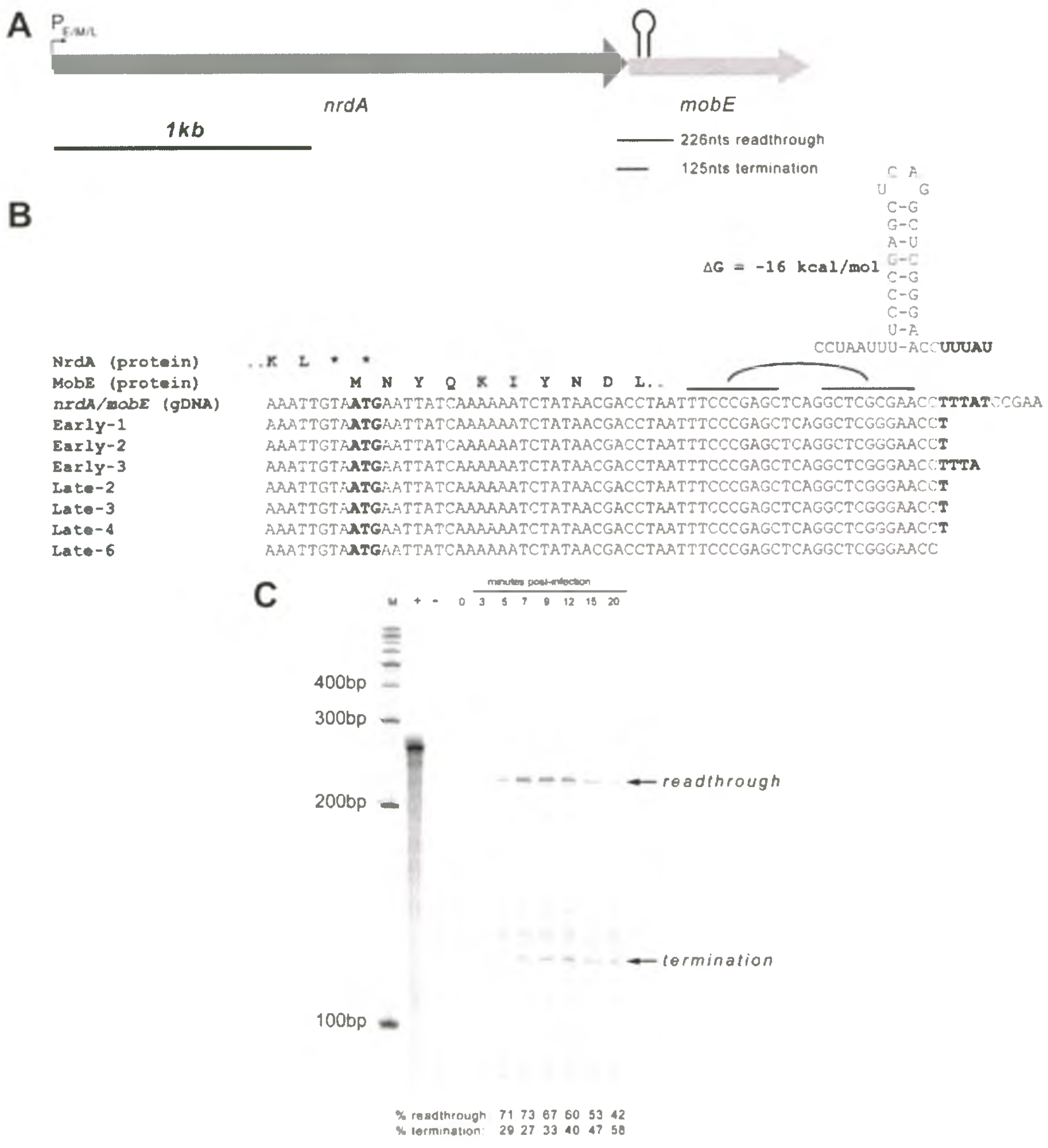

A
P E/M/L
nrdA
mobE
400bp (E)
1kb
400bp (B) (D)
600bp (C)
B DE-365 / DE-367
Marker
gDNA
RNA(-)
RT(-)
minutes post-infection
0 1 3 5 7 9 12 15 20 25
600bp
400bp
300bp
200bp
100bp
C DE-387 / DE-367
Marker
gDNA
RNA(-)
RT(-)
minutes post-infection
0 1 3 5 7 9 12 15 20 25
600bp
400bp
300bp
200bp
100bp
D mobE specific
minutes post-infection
0 3 5 7 9 12 15 20
3.5 kb
3.0 kb
2.5 kb
2.0 kb
1.5 kb
1.0 kb
>4.0kb
~3.0kb
E nrdA specific
minutes post-infection
0 3 5 7 9 12 15 20
3.5 kb
3.0 kb
2.5 kb
2.0 kb
1.5 kb
1.0 kb
>4.0kb
~3.0kb

3.3 The expression of *mobE* is regulated by a transcriptional terminator

To assay whether the *mobE* transcriptional terminator was functional I utilized 3' RLM-RACE to map the 3' ends of transcripts in the *nrdA/mobE* region. Two RNA pools were created by mixing early and late time points (3, 5, 7 minutes post infection or 9, 12, 15 minutes post infection). Multiple clones (five from early and six from late pools) were sequenced, and seven out of the eleven sequences mapped to the poly(U) tract downstream from the predicted transcriptional terminator (**Figure 9B**). The remaining sequences mapped to sites downstream of the transcriptional terminator. These sites were probably the result of read-through transcripts that were degraded during infection. The 3' RACE results show that there are 3' ends consistent with a functional terminator in the 5' end of *mobE.*

To estimate the efficiency of the predicted terminator in the 5' end of *mobE* I used RNase protection assays (RPA). Determining efficiency of termination will confirm that the terminator is indeed functional. Furthermore, read-through transcription appears to be the only way *mobE* can be expressed during phage infection. A 252 nt RNA probe was transcribed that contained 100 bp of sequence flanking both sides of the predicted terminator (**Figure 9A**). The probe also contained 13 nt of non-complementary sequence at each end as a control for RNAse activity. The non-complementary sequence allows us to distinguish full length probes from the read-through and termination products. The RPA assay was performed on RNA isolated pre-T4 infection, and at 9 time point's post-T4 infection. The efficiency of termination was measured at ~30% until late time points, where the efficiency increased to 50% (**Figure 9C**). The increase in termination at late

Figure 9. Transcriptional termination in the 5' end of *mobE.* (A) Schematic of the *nrdA* and *mobE* genes with the predicted Rho-independent terminator indicated as a stem-loop structure. The position of the probe and the predicted sizes of the RNase digestion products are indicated. (B) Sequences of the 3'-RACE clones from early or late RNA samples aligned with T4 genomic DNA sequence. Nucleotides highlighted in bold are part of the poly(U) tract following the terminator structure, or the *mobE* start codon (ATG). The terminator structure as predicted by Mfold is shown above the sequences (88). (C) RNase protection assay to determine the termination efficiency of termination. Shown is a representative experiment with the read-through and termination products indicated. The lanes are labeled as follows (M), 100-bp ladder; (+), probe only; (-), probe digested with RNase; 0 minutes, uninfected cells; 3-20, minutes post T4 infection.


A
P E/M/L
nrdA
mobE
1kb
226nts readthrough
125nts termination
B
C A
U G
C-G
G-C
A-U
G-C
C-G
C-G
C-G
U-A
CCUAAUUU-ACCUUUAU
ΔG = -16 kcal/mol
C
minutes post-infection
M + - 0 3 5 7 9 12 15 20
400bp
300bp
200bp
100bp
readthrough
termination
% readthrough: 71 73 67 60 53 42
% termination: 29 27 33 40 47 58

NrdA (protein)	..K L * *
MobE (protein)	M N Y Q K I Y N D L..
***nrdA/mobE* (gDNA)**	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGCGAACC**TTTAT**CCGAA
Early-1	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC**T**
Early-2	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC**T**
Early-3	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC**TTTA**
Late-2	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC**T**
Late-3	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC**T**
Late-4	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC**T**
Late-6	AAATTGTA**ATG**AATTATCAAAAAATCTATAACGACCTAATTTCCCGAGCTCAGGCTCGGGAACC

time points may be due to the accumulation of transcripts during infection. The RPA experiments demonstrate that the *mobE* terminator is functional but inefficient, which is consistent with the short poly(U) tract downstream of the steam-loop structure.

Finally, the presence of the terminator structure in other *mobE* containing T-even phage was verified using ClustalW and Mfold (83,88). Interestingly, the phages RB15 and LZ7 have accumulated nucleotide substitutions in the stem-loop structure relative to other *mobE*-containing phage (**Figure 10**). These polymorphisms shortened the stem and increased the loop length of the terminator structure, but the overall structure of the terminator was conserved. This data suggests that transcriptional terminator in *mobE* is an important regulatory mechanism for the *nrd* operon.

3.4 *mobE* encodes a functional and mobile endonuclease

To accumulate evidence that *mobE* encodes a functional homing endonuclease, T2 x T4 mixed infections were performed (**Figure 11A**). As previously mentioned, T2 has no HEG insertions and numerous HEGs from T4 have been shown to mobilize into T2 during mixed infections (34,35,48,49,73,74). The progeny phage from the T2 x T4 infections were screened for the presence of *mobE* using Southern blots of plaque lifts, and *mobE* specific probes, as described previously (**Figure 11B**) (34). The results from each phage cross were expressed as the ratio of the progeny phage containing *mobE* relative to the T4 input of each mixed infection (**Table III and Table IV**). If *mobE* is mobile the ratio will be much higher than the expected ratio of ~1.0 for Mendelian inheritance. In the T2 x T4 mixed infections, the *mobE* inheritance ratio was 2.1 +/- 0.1,

Figure 10. Conservation of the *mobE* transcriptional terminator in T-even phage. Shown above are nucleotide sequence alignments and the amino acid sequences of the 5' region of the *mobE* gene from various T-even phage. Nucleotide differences between T4 *mobE* and the other phage are highlighted with black boxes. The predicted terminator sequence is indicated by a rectangle and the terminator structures as predicted by Mfold are shown above (88). The terminating nucleotides mapped to the T4 *mobE* sequence by 3'-RACE are indicated with asterisks (*).

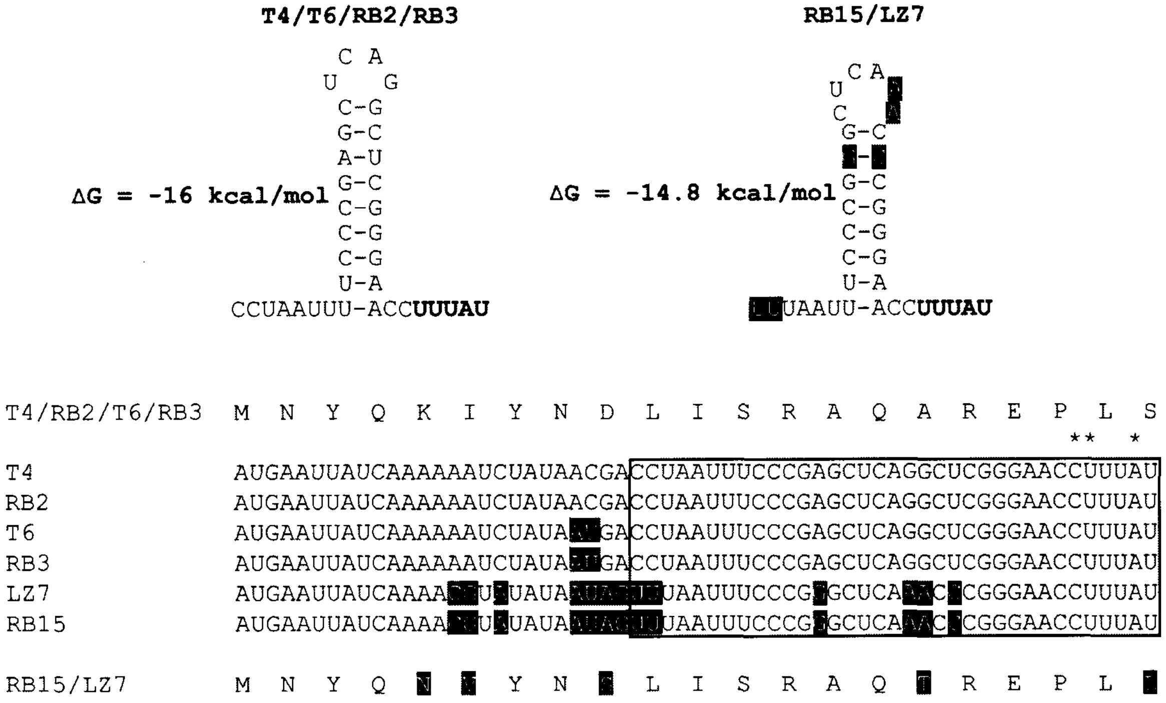

T4/T6/RB2/RB3
RB15/LZ7
ΔG = -16 kcal/mol
ΔG = -14.8 kcal/mol
CCUAAUUU-ACCUUUAU
UAAUU-ACCUUUAU
T4/RB2/T6/RB3
M N Y Q K I Y N D L I S R A Q A R E P L S
T4
RB2
T6
RB3
LZ7
RB15
AUGAAUUAUCAAAAAAUCUAUAACGACCUAAUUUCCCGAGCUCAGGCUCGGGAACCUUUAU
RB15/LZ7
M N Y Q Y N L I S R A Q T R E P L

Figure 11. T2 x T4 mixed infections. (A) Outline of the T2 x T4 mixed infection protocol. Equal amounts of T2 and T4, or T4Δ*mobE33-62*, phage were mixed, and subsequently used to infect *E. coli* sup^{O} cells at a multiplicity of infection of 4 for each phage for 10 minutes. The infected cells were diluted 1 x 10^4 fold into fresh LB media. The resultant phage were grid plated, transferred to a positively charged nylon membrane, and probed using plaque lifts. (B) Representative plaque lifts from T2 x T4 and T2 x T4ΔmobE mixed infections. The probes used for hybridization are indicated to the left of the blots, and the probe used to calculate T4 input were *mobE* specific. The T4 phage inputs are shown at the top followed by the progeny output for *mobE* and *I-TevIII* below.

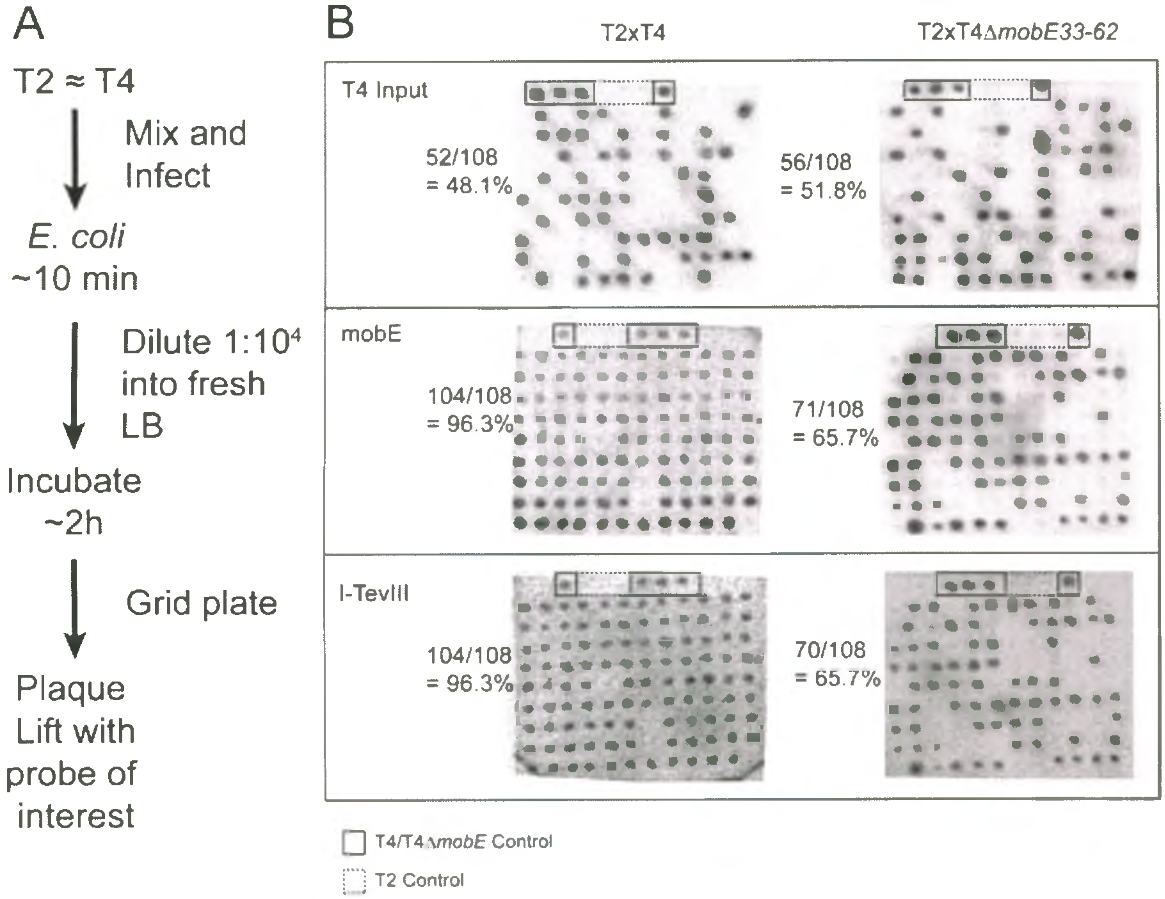

A
T2 ≈ T4
Mix and Infect
E. coli ~10 min
Dilute 1:10^4 into fresh LB
Incubate ~2h
Grid plate
Plaque Lift with probe of interest
B
T2xT4
T2xT4ΔmobE33-62
T4 Input
52/108 = 48.1%
56/108 = 51.8%
mobE
104/108 = 96.3%
71/108 = 65.7%
I-TevIII
104/108 = 96.3%
70/108 = 65.7%
T4/T4ΔmobE Control
T2 Control

Table III. Phage-to-phage mixed infections

Phages	Cross	%T4 Input	N	%mobE	%I-TevIII	%I-TevI
T2xT4	1	47.2	211	86.7	84.8	91.4
	2	48.1	216	97.7	97.7	96.3
	3	38.9	216	89.4	91.2	95.3
T2xT4*Δ*$mobE33$-*62*	1	49.1	216	71.7	N.D.	N.D.
	2	51.9	216	64.8	64.8	100
	3	44.4	216	81.9	82.4	99.5
T2*nrdB*^{R} x T4	1	56.4	216	90.7	N.D	N.D
	2	58.3	216	89.3	N.D	N.D

N.D. No Data

Table IV. Normalized phage-to-phage mixed infections

Phages	Cross	*mobE*/input[a]	*I-TevIII*/input[a]	*I-TevI*/input[a]
T2xT4	1	1.8	1.8	1.9
	2	2.0	2.0	2.0
	3	2.3	2.3	2.4
		2.1+/-0.1[b]	2.1+/-0.2[b]	2.1+/0.2[b]
T2xT4*ΔmobE33-62*	1	1.5	N.D.	N.D.
	2	1.2	1.2	1.9
	3	1.8	1.9	2.2
		1.6+/-0.2[b]	1.6[c]	2.1[c]
T2*nrdB*^{R} x T4	1	1.6	N.D	N.D
	2	1.5	N.D	N.D
		1.6[c]		

[a] The ratio of each endonuclease gene inherited in the progeny relative to the expected inheritance of each endonuclease gene. See Table III for raw data.

[b] The mean and standard for each set

[c] The mean for each set

which has the characteristics of non-Mendelian inheritance (34-36,48). Furthermore, the inheritance ratio for two intron-encoded HEGs , *I-TevI* and *I-TevIII,* were also high with ratios of 2.1 +/- 0.2 and 2.1 +/- 0.2 , respectively. Notably, the inheritance ratio of *mobE, I-TevIII,* and *I-TevI,* were not significantly different from each other. The high inheritance ratios of *mobE, I-TevIII,* and *I-TevI* suggest that all three of these genes are mobile.

To demonstrate that the high inheritance ratio of *mobE* required the predicted *mobE* HNH motif, I created an in-frame *mobE* internal deletion mutant. The internal deletion consisted of an in-frame deletion of 29 amino acids (positions 33-62), which encompass part of the conserved HNH motif. The resultant phage, T4*ΔmobE33-62* was used in mixed infections with T2, and the inheritance ratio of *mobE* in the progeny was reduced to 1.6 +/- 0.2, when compared to the wild-type mixed infections numbers **(Table III and Table IV)**. The observed ratio of 1.6 +/- 0.2 in the progeny phage rather than expected ratio ~1.0 based on the T2 and T4 sequence similarity is consistent with the exclusion of T2 markers in the progeny from T2 x T4 mixed infections (73). Furthermore, the observed ratio of 1.6 +/- 0.2 in the T4*ΔmobE33-62* mixed infections was significantly different from the wild-type infections ($p < 0.05$ using a one-tailed T-test). This data shows that a partial deletion within the HNH motif reduced *mobE* mobility. Finally, the inheritance ratio 2.1 for *I-TevI* from the T2 x T4*ΔmobE33-62* infections was high, which is similar to the ratio of 2.1 +/- 0.2 from T2 x T4 mixed infections, which is expected because *I-TevI* encodes a functional homing endonuclease. This result is important because it shows that the reduction in *mobE* inheritance in the T2 x T4*ΔmobE33-62* infections was not due to another effect. Collectively, this data demonstrates that *mobE* encodes a functional homing endonuclease that causes the high inheritance of *mobE* in mixed T2xT4 infections.

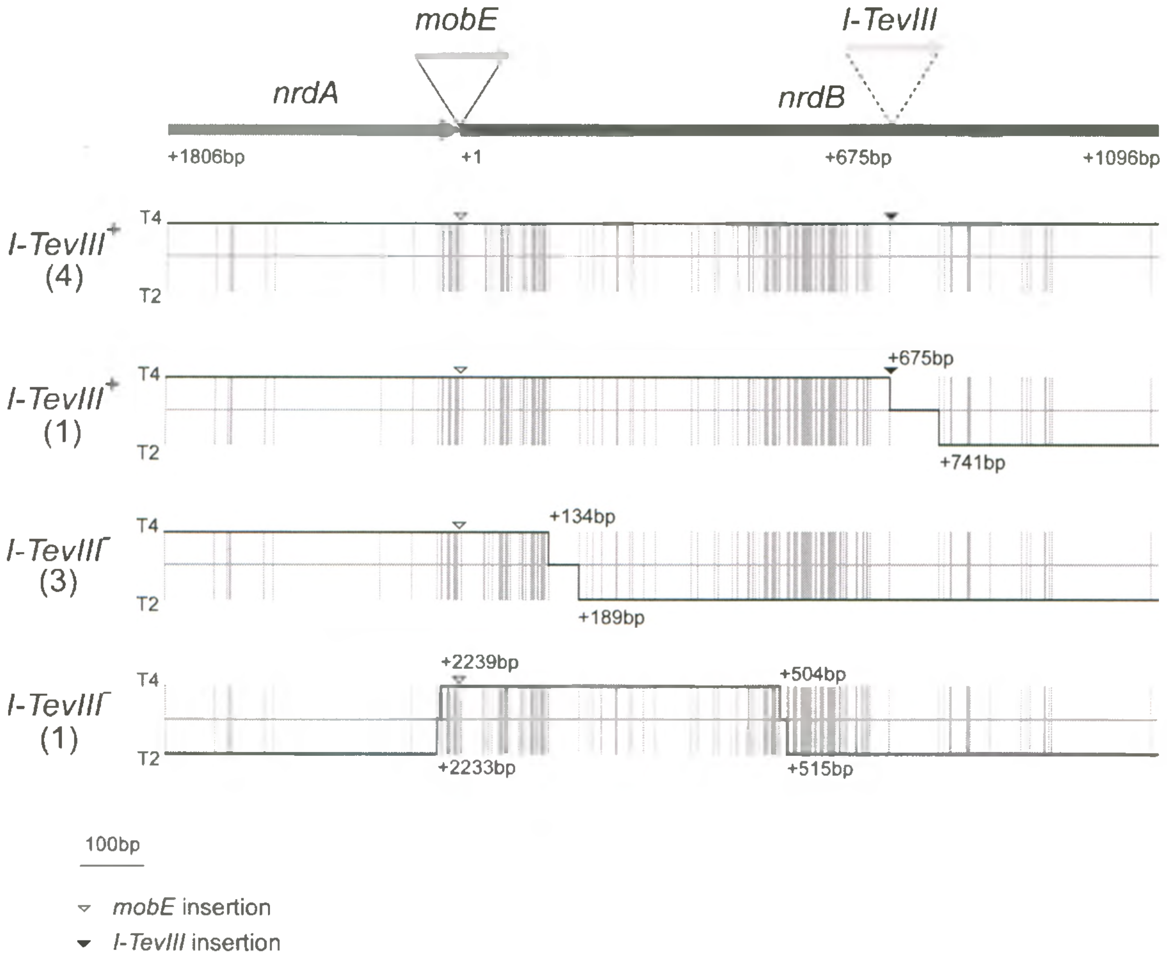
3.5 Inheritance of *I-TevIII* is dependent on *mobE* function

Previous, experimental evidence has suggested that *mobE* may be responsible for the high inheritance of *I-TevIII*, a defunct homing endonuclease gene encoded within an intron embedded in *nrdB* (78). The distance between *I-TevIII* and *mobE* insertions is only 675 bp, and previous studies with another homing endonuclease SegG have shown that the repair of DSBs introduced by SegG can result in 2 kb of gene conversion from the SegG cleavage site (31,32,34). This observation suggests that repair of strand-breaks introduced by MobE could result in inheritance of *I-TevIII.* To test this, progeny phage from mixed infection experiments with T2 x T4 and T2 x T4Δ*mobE33-62* were screened for *I-TevIII* inheritance (**Table III and Table IV**). In wild-type mixed infections the *I-TevIII* inheritance ratio was 2.1 +/- 0.2, which is not significantly different than the *mobE* ratio of 2.1 +/- 0.1. However, in mixed infections with T4Δ*mobE33-62* and T2, the *I-TevIII* inheritance ratio was reduced to 1.6, which is similar to the *mobE* ratio of 1.6 +/- 0.2. Furthermore, the *I-TevI* inheritance ratio of 2.1 from the T2 x T4Δ*mobE33-62* mixed infections was similar to the 2.1 +/- 0.2 inheritance ratio of *I-TevI* from the wild-type mixed infections. Since the *I-TevI* ratio remained high in the wild-type and T2 x T4Δ*mobE33-62* mixed infections, and the ratios for *mobE* and *I-TevIII* were reduced, I suggest that the inheritance of *mobE* and *I-TevIII* is linked. This evidence suggests that an intact *mobE*-gene is required for the high *I-TevIII* inheritance in the progeny phage from T2 x T4 mixed infections.

3.6 Sequencing of T2xT4 progeny localizes the MobE target site to T2 *nrdB*

While screening the progeny of the T2xT4 mixed infections for *I-TevIII* and *mobE* inheritance, I observed that most of these phage inherited *mobE* and *I-TevIII* (*mobE*$^{+}$*I-TevIII*$^{+}$). However, I also noticed that a few of these phage inherited *mobE* without *I-TevIII* (*mobE*$^{+}$*I-TevIII*$^{-}$). The presence of *mobE*$^{+}$*I-TevIII*$^{-}$ phage from these mixed infections suggested that a MobE cleavage site, and a recombination point was occurring in the 650 nucleotides between *mobE* the *I-TevIII* insertion sites. To estimate the location of these recombination points I sequenced five *mobE*$^{+}$*I-TevIII*$^{+}$ phage and four *mobE*$^{+}$*I-TevIII*$^{-}$ phage (**Figure 12**). The recombination point was localized based on the presence of unique polymorphisms between T2 and T4 in the 3' end of *nrdA* and the 5' end of *nrdB*. Three of the *mobE*$^{+}$*I-TevIII*$^{-}$ progeny had identical sequences with a recombination point between 134 and 189 nucleotides of *nrdB*. The other *mobE*$^{+}$*I-TevIII*$^{-}$ phage had two recombination points, one in the 3' end of *nrdA* between 2,233 and 2,239 nucleotides, and another in the 5' end of *nrdB*, near nucleotides 504 and 515. Four of the *mobE*$^{+}$*I-TevIII*$^{+}$ progeny contained T4 polymorphisms throughout the 3' region of *nrdA* and the 5' end of *nrdB*. Additionally, one of the *mobE*$^{+}$*I-TevIII*$^{+}$ phage sequenced, contained T4 polymorphisms in *nrdB* until the end of the *I-TevIII* insertion. Gene conversion events initiated by homing endonucleases extend in both directions from the strand-breaks (21,31,33). Therefore, I expect that the MobE recognition site is between the two recombination points closest to the *mobE* insertion site, which localizes the MobE target site to the last 26 nucleotides of T2 *nrdA*, or the first 134 nucleotides of T2 *nrdB*. However, there are more nucleotide polymorphisms in the first 134 nucleotides T2 *nrdB* than there are in the last 26 nucleotides of T2 *nrdA*. Previous studies have shown

Figure 12. Sequence analysis of progeny phage. Genomic DNA was isolated from the progeny phage of T2 x T4 mixed infections, and the 3' end of *nrdA* and the 5' end of *nrdB* was sequenced. All of the progeny phage inherited *mobE* (*mobE*$^+$), and inherited *I-TevIII* (*I-TevIII*$^+$) or didn't inherit *I-TevIII* (*I-TevIII*$^-$). The grey lines represent the positions of nucleotide polymorphisms between T2 and T4, and the black lines indicate which polymorphisms were contained in the sequenced progeny phage. The positions of the recombination points are indicated relative to the start codon of the gene they were contained in. The number of progeny phage sequenced for each map is indicated in brackets.


mobE
I-TevIII
nrdA
nrdB
+1806bp
+1
+675bp
+1096bp
I-TevIII+
(4)
T4
T2
I-TevIII+
(1)
T4
T2
+675bp
+741bp
I-TevIII-
(3)
T4
T2
+134bp
+189bp
I-TevIII-
(1)
T4
T2
+2239bp
+2233bp
+504bp
+515bp
100bp
mobE insertion
I-TevIII insertion

that homing endonucleases target sites in the host accumulate nucleotide substitutions relative to the recipient as a defense against homing endonuclease cleavage. From this, I hypothesize that the MobE target site is located in the first 134 nucleotides of T2 *nrdB* due to the number of polymorphisms relative to T4.

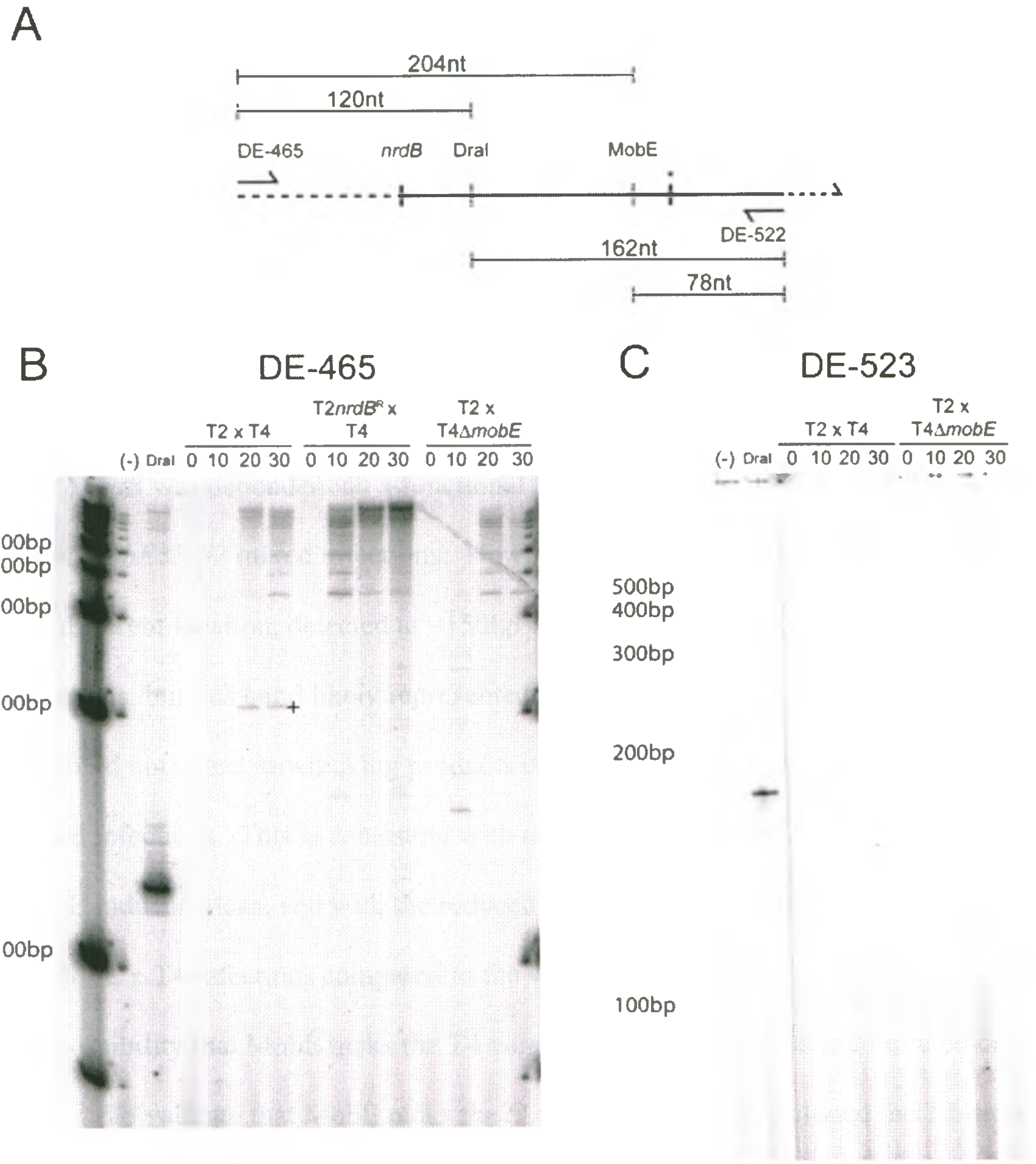
3.7 Creation of a MobE resistant T2 phage

The recombination point mapping in T2 x T4 mixed infections localized the MobE recognition site to the first 134 nucleotides of T2 *nrdB*. Previous studies with SegG demonstrated that the target site in T4 contained nucleotide polymorphisms relative to T2, which reduced the cleavage efficiency of SegG on the T4 sequence. Following from the idea of cleavage resistance, I decided to generate a T2 phage containing phage with T4 polymorphisms in the first 134 nucleotides of *nrdB*. The expectation is that the T2*nrdB*^{R} phage will be resistant to *mobE* mobility. The inheritance ratio of *mobE* from mixed infections with T2*nrdB*^{R} x T4 was 1.6 which is similar to the *mobE* inheritance ratio of 1.6 +/- 0.2 from the T2 x T4Δ*mobE33-62* mixed infections (**Table III and Table IV**). This suggests that T2*nrdB*^{R} is resistant to MobE catalyzed strand-breaks, and that the MobE recognition site is within the first 134 nucleotides of T2 *nrdB*.

3.8 Primer Extension Mapping of MobE-induced nicks in T2 *nrdB*

To confirm that the MobE recognition site is in the first 134 nucleotides of T2 *nrdB*, I mapped MobE-dependent breaks *in vivo* during T2 x T4 mixed infections (**Figure 13A**). Total DNA was isolated at various time points post mixed infection, and used in primer extension reactions. I used a primer specific to the 5' end of the intergenic region

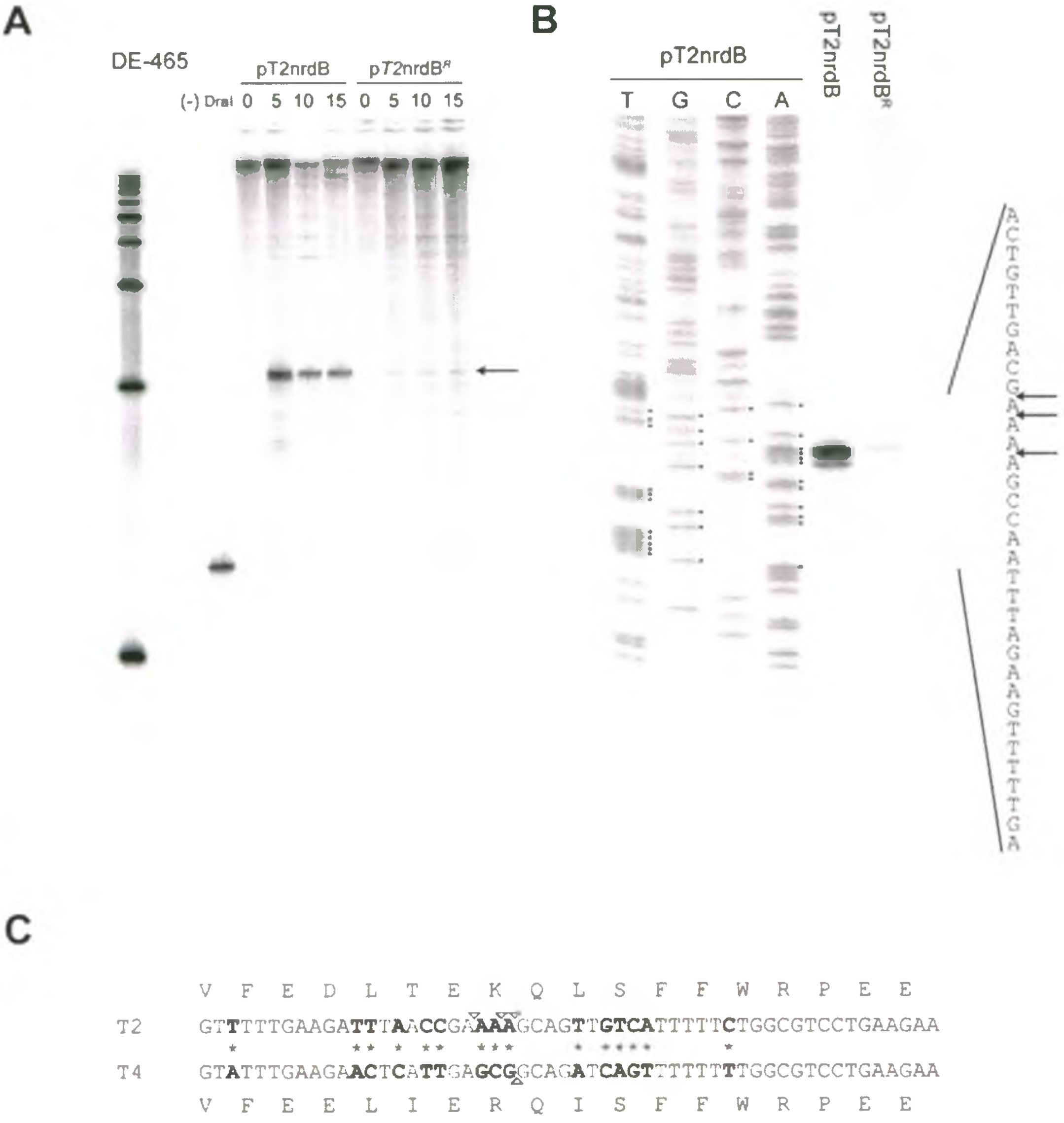
Figure 13. Mapping of MobE-dependent nicks *in vivo* using primer extension. (A) Schematic of the 5' end of T2 *nrdB* showing the positions of the primers used in the primer extension reactions, the DraI cleavage site, and the MobE nicking sites. The predicted lengths of the primer extension products are also indicated. The 134 nucleotide *nrdB* region replaced with T4 polymorphisms in T2*nrdB*^{R} encompasses the beginning of *nrdB* to vertical bar marked with an asterisk (*). (B) Representative gel of the primer extension reactions on the non-coding strand of genomic DNA isolated at various time points (minutes) from mixed infections with T2 x T4, T2*nrdB*^{R} x T4, and T2 x T4Δ*mobE33-62*. The sizes of a 100-bp DNA ladder are labeled on the left side of the gel. The (+) indicates the MobE nicking site present in wild-type but not the mutant phage infections. (-), no DNA added to the reactions; DraI, primer extension reactions on DraI-digested T2 genomic DNA. (C) Primer extension reactions on the coding strand, as labeled in panel (B). For this representative gel, the DraI primer extension reaction was diluted 3-fold prior to electrophoresis, and the T2*nrdB*^{R} x T4 mixed infections were not included.


A
204nt
120nt
DE-465
nrdB
DraI
MobE
DE-522
162nt
78nt
B
DE-465
T2 x T4
T2nrdB^R x T4
T2 x T4ΔmobE
(-) DraI 0 10 20 30 0 10 20 30 0 10 20 30
500bp
400bp
300bp
200bp
100bp
C
DE-523
T2 x T4
T2 x T4ΔmobE
(-) DraI 0 10 20 30 0 10 20 30
500bp
400bp
300bp
200bp
100bp

of T2 *nrdB* (DE-465) to detect nicks in the non-coding strand of *nrdB*, and a primer within the 5' end of *nrdB* (DE-523) to detect nicks on the coding strand. T2 genomic DNA digested with DraI, which cleaves in the 5' end of T2 *nrdB* was used as a positive control for the primer extension reactions. In DNA isolated from the T2 x T4 mixed infections, a nick was mapped on the non-coding strand of T2 *nrdB*, whereas no nicks were mapped to the coding strand of *nrdB* (**Figure 13B/C**). The nick was detected 10 minutes post T4 x T2 mixed infection, which coincides with the detection of *mobE* transcripts during phage infection (**Figure 8B**). The lack of any nicking products on the coding strand was not due to failure of the primer extension reaction, because the positive DraI control generated products of the expected size. Moreover, the strand-specific nick in T2 *nrdB* was dependent on a functional MobE, since no nicks were detected in the T2 x T4Δ*mobE33-62* mixed infections. However, there was a nick in the non-coding strand at a different location, detected at ~150bp 10 minutes post T2 x T4Δ*mobE* mixed infections, but this band likely represented a spurious primer extension product. Finally, we could not detect any nicking products on genomic DNA isolated from T2*nrdB*^{R} x T4 mixed infections. This is consistent with the expectation that T2*nrdB*^{R} will be resistant to MobE induced nicks, and with the reduced ratio of *mobE* inheritance (1.6) in the T2MobE x T4 infections compared to the wild-type infections (2.1 +/- 0.1). Nonetheless, the possibility that MobE nicks the T4 version of its target site cannot be ruled out.

To validate that MobE nicks the 5' region of T2 *nrdB*, I cloned the T2 nrdA/nrdB intergenic region and the first 661 nucleotides of the T2 nrdB gene to create pT2nrdB. Also, to try and resolve if MobE nicks T2*nrdB*^{R}, the plasmid pT2nrdB^R was used in the experiments. The high copy number of pT2nrdB and pT2nrdB^R (100-200 copies per cell)

Figure 14. Mapping of MobE-dependent nicks on plasmid substrates. Primer extension reactions on pT2nrdB and pT2nrdB^R plasmid substrates isolated at various time points (minutes) post T4K10 infection. The primer used in this experiment is the same primer used to map *in vivo* nicks on T2 *nrdB* as shown in Figure 13A. The MobE nicks are indicated by a left facing arrow. (B) High resolution mapping of the MobE nicks in pT2nrdB and pT2nrdB^R primer extension reactions. The reactions were run alongside a cycle sequencing ladder generated with the same primer and pT2nrdB. The pT2nrdB coding sequence is indicated to the right of the gel and the nick sites are indicated by left facing arrows. (C) Alignment of the *nrdB* nucleotide and amino acid sequences from phage T2 and T4 in the MobE nicking-site region. Nucleotide differences are indicated in bold and asterisks (*), and the nicking sites in T2 and the single nick site in T4 are indicated with triangles. Note that the sequences shown represent the coding strand and that the nicks were mapped to the non-coding strand.

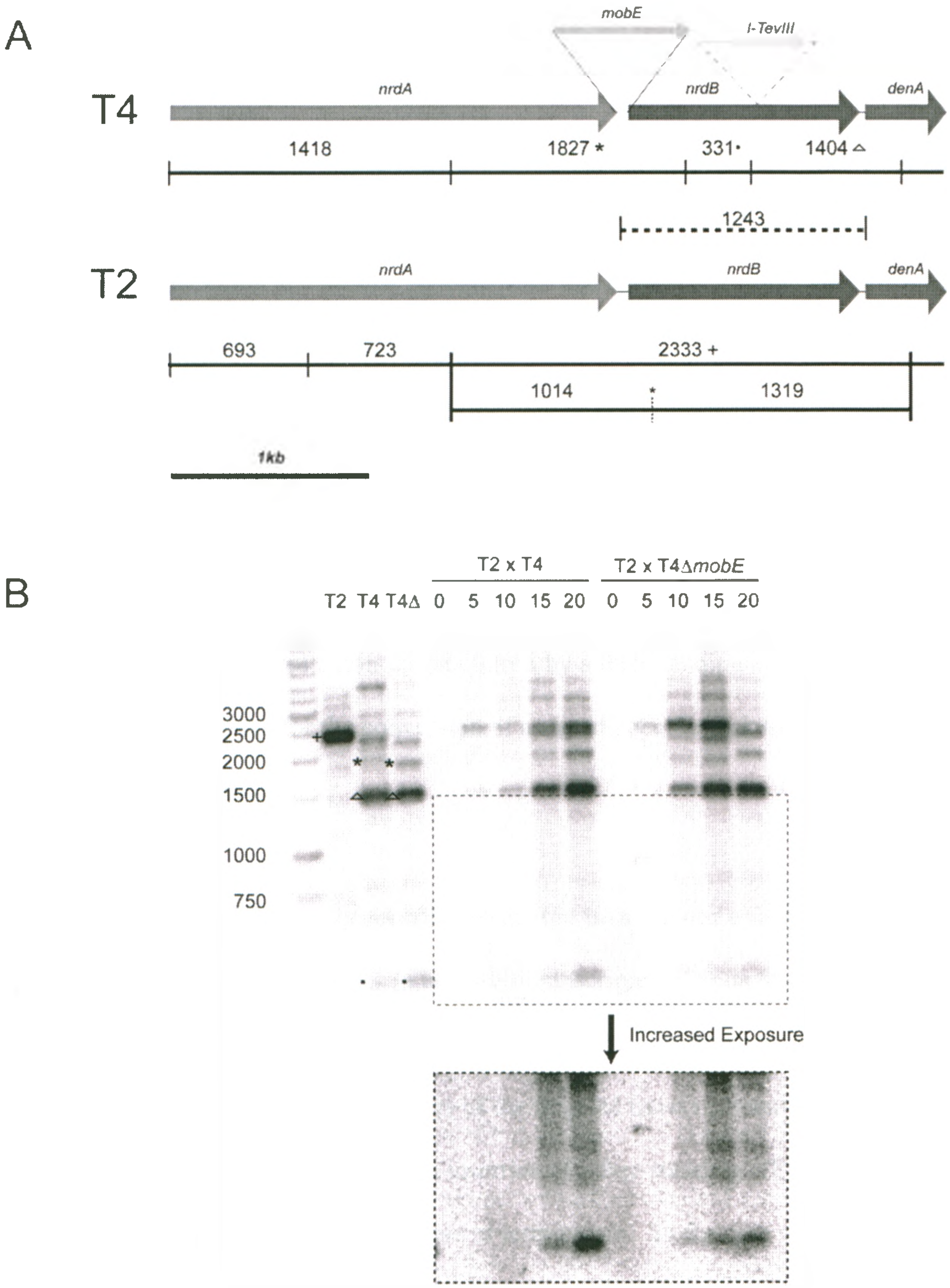

A
DE-465
pT2nrdB
pT2nrdB^R
(-) DraI 0 5 10 15 0 5 10 15
B
pT2nrdB
T G C A
pT2nrdB
pT2nrdB^R
ACTGTTGACGAAAAGCCAATTTAGAAGTTTTGA
C
V F E D L T E K Q L S F F W R P E E
T2 GTTTTTGAAGATTTAACCGAAAAGCAGTTGTCATTTTTCTGGCGTCCTGAAGAA
T4 GTATTTGAAGAACTCATTGAGCGGCAGATCAGTTTTTTTTGGCGTCCTGAAGAA
V F E E L I E R Q I S F F W R P E E

may help detect rare nicking events on the plasmids by MobE. Both pT2nrdB and $pT2nrdB^R$ were transformed into *E. coli* sup^O and infected with T4K10, a T4 phage that cannot degrade host DNA in sup^O. Plasmid DNA was isolated at various time points post T4K10 infection and used in primer extension reactions. Nicks on the non-coding strand were detected in both pT2nrdB and $pT2nrdB^R$ at all time points post-infection (**Figure 14A**). Three prominent nicking sites on pT2nrdB and one on $pT2nrdB^R$ were mapped between nucleotides 122-126 of *nrdB* by electrophoresis of the primer extension reactions alongside a sequencing ladder generated from pT2nrdB with DE-465 (**Figure 14B/C**). Finally, the nick site in pT2nrdB and T2 genomic DNA from T2 x T4 mixed infections was mapped to identical positions using gel electrophoresis (Data not shown). Additionally, it should be noted that the multiple nicking sites could have been due to the nucleotide addition activity of Taq DNA polymerase during the primer extension reactions (89).

3.9 Southern Blot mapping of double strand breaks *in vivo*

To elucidate whether the strand-specific nick generated by MobE is converted into a DSB Southern blots were performed on AseI digested genomic DNA isolated from mixed T2 x T4 and T2 x T4Δ*mobE33-62* infections. For hybridization, a 1243bp probe representing T2 *nrdB* gene was used, which also has high nucleotide identity to the T4 *nrdB* gene. Due to the high identity to T4 *nrdB* the probe is expected to hybridize to T4 DNA as well as T2. A schematic indicating the positions of the AseI sites and their respective sizes in T2 and T4 are indicated in figure 15A. Additionally, it should be noted that the 1827 bp T4 fragment will be 1741bp in T4Δ*mobE33-62* infections. I expect the T2 2333 bp AseI fragment to be cut into a ~1014bp and ~1319bp fragment if there is a MobE induced DSB. After hybridization, I could not detect fragments corresponding to a MobE induced DSB (**Figure 15B**). However, the probe I created may not have been sensitive enough to detect DSBs in these mixed infections.

Figure 15. AseI Southern blot of phage mixed infection genomic DNA. (A) Schematic of the *nrd* operon from phage T4 and T2, the location of AseI restriction fragments (vertical bar) and their respective sizes are indicated on the lines below the arrows. The T2 2333 bp fragment is further divided with the MobE nicking site (dashed line with a *) indicating the expected size of DSBs. The size and location of the T2 *nrdB* probe is indicated between the schematics with a dashed line. Finally, the fragments hybridized by the probe are indicated with symbols on the schematic and the blot. (B) Representative Southern blot of the AseI digested genomic DNA isolated from T2 x T4, and T2 x T4Δ*mobE33-62*. The lanes labeled T2, T4 and T4Δ represent lanes with AseI digested phage DNA. The sizes of the 1kb-DNA ladder products are indicated to the left of the blot. The region where DSBs were expected to occur is shown underneath the blot with increased exposure.


A
mobE
I-TevIII
T4
nrdA
nrdB
denA
1418
1827 *
331•
1404 △
1243
T2
nrdA
nrdB
denA
693
723
2333 +
1014
1319
1kb
B
T2 x T4
T2 x T4Δ*mobE*
T2 T4 T4Δ 0 5 10 15 20 0 5 10 15 20
3000
2500
2000
1500
1000
750
Increased Exposure

Discussion

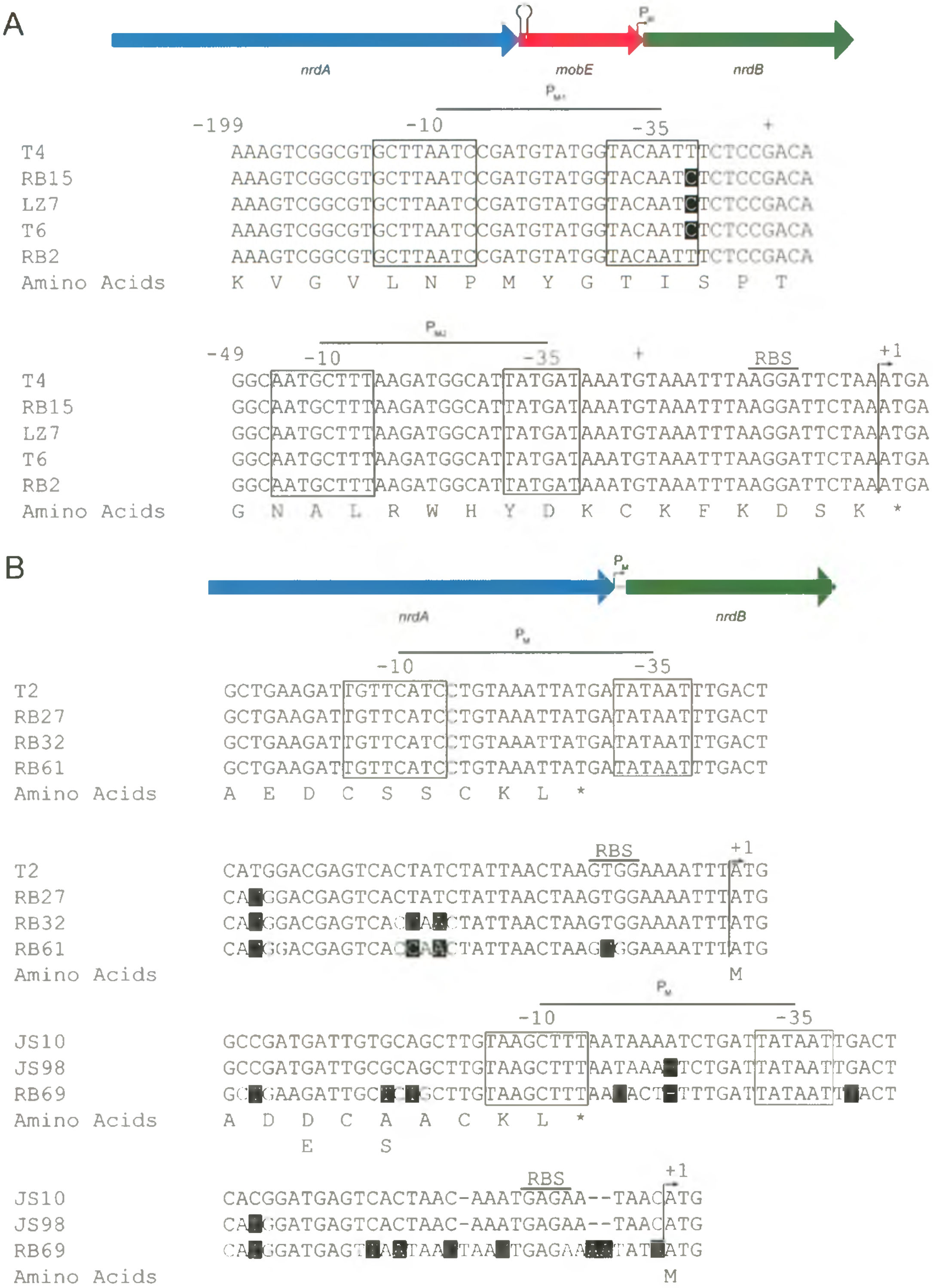
The overall goal of my study was to investigate the regulation and mobility of the *mobE* gene present in phage T4. I showed that *mobE* has integrated successfully into the transcriptional program of the *nrd* operon. Additionally, I showed that *mobE* expression is regulated by an intrinsic transcriptional terminator during phage infection. In this work, I showed that *mobE* is a mobile endonuclease that catalyzes a strand-specific break in T2 *nrdB*. However, further work is required to conclusively demonstrate that MobE is an endonuclease that catalyses a strand-specific nick or DSB. Furthermore, I showed that the mobility of *mobE* and the non-mobile *nrd* intron that contains *I-TevIII* depended on an intact *mobE* ORF. My results will be discussed in the context of the current knowledge regarding the homing pathway and HNH endonucleases. I also present a model for the *trans* homing of a non-mobile intron by a free-standing homing endonuclease, and a model for the conversion of a strand-specific nick to a DSB by a collapsed DNA replication fork.

4.1 The *mobE* insertion has successfully integrated into the *nrd* operon

The insertion of a HEG tends to be within transcriptional units that contain conserved genes that are critical for host function (50-52). The insertion of a free-standing HEG could perturb the expression of the targeted transcriptional unit through a number of mechanisms. For instance, the insertion of a free-standing HEG could replace an intergenic region that contains promoters or terminators. In the case of the *nrd* operon in *mobE*-containing T-even phage, the insertion of *mobE* is between the functionally critical *nrdA* and *nrdB* genes (55). The loss of NrdA and NrdB function has been shown to reduce the DNA replication levels during phage infection, which ultimately leads to a reduction in the number of progeny phage released from each infected cell (76,77). Current evidence suggests that the *mobE* nucleotide sequence has evolved to ensure that the transcriptional program of the *nrd* operon is not perturbed by the *mobE* insertion. These adaptations include the presence of middle promoters in the 3' end of the *mobE* ORF to drive *nrdB* expression (**Figure 16A**) (79). In comparison, *mobE*-lacking phage have an equivalent middle promoter within the 3' end of *nrdA*, or within an integenic region, which is located upstream of the *nrdB* ORF (**Figure 16B**). Furthermore, the presence of middle promoters in the *mobE* ORF may select against the deletion of *mobE*, as a deletion of the middle promoters may reduce the expression of *nrdB*. In addition to the middle promoters the *mobE* ORF overlaps *nrdA* and *nrdB*, which may also select against the complete excision of the *mobE* ORF. The deletion of the *mobE* ORF would disrupt the *nrdA* and *nrdB* ORFs by the loss of the last nucleotide of the *nrdA* stop codon and the first two nucleotides of the *nrdB* start codon. Collectively, this data suggests that the *mobE* ORF has integrated successfully within the *nrd* operon and that the deletion of *mobE* would disrupt the expression of the NrdA/NrdB complex.

The middle promoters within the *mobE* ORF may also compensate for the Rho-independent transcriptional terminator in the 5' end of the *mobE* ORF. Without the *mobE* middle promoters, the transcriptional terminator would reduce the *nrdB* transcript levels during phage infection, which could affect formation of the NrdA/NrdB complex. Interestingly, in phage lacking *mobE*, no transcriptional terminators were detected in the *nrdA-nrdB* intergenic region. In contrast, all of the known *mobE*-containing T-even phage possess a terminator in the 5' end of the *mobE* ORF (**Figure 10**). The

Figure 16. Conservation of the promoters upstream of *nrdB* inT-even phage. The nucleotide sequences upstream of *nrdB* from *mobE*$^+$ (A) and *mobE*$^-$ (B) T-even phage were aligned using clustalW . The middle promoter sequences were identified using the T4 middle promoter consensus sequences and previous experiments that identified the locations of the T4 *nrdB* promoters (55,79,80). A schematic showing the *nrd* operon for the *mobE*$^+$ and *mobE*$^-$ phage are shown above, and the amino acid sequence of MobE (A) or NrdA (B) is shown underneath the nucleotide alignment. The start codon of NrdB and the predicted RBS binding sites are also labeled. The -10 and -35 regions of the identified middle promoters are surrounded by boxes and the predicted nucleotide transcription would initiate from is indicated with a (+). Nucleotide polymorphisms in each alignment relative to the first sequence shown are highlighted (T4, T2, JS10).


A
nrdA
mobE
nrdB
P_ME
P_MJ
-199
-10
-35
+
T4 AAAGTCGGCGTGCTTAATCCGATGTATGGTACAATTTCTCCGACA
RB15 AAAGTCGGCGTGCTTAATCCGATGTATGGTACAATCTCTCCGACA
LZ7 AAAGTCGGCGTGCTTAATCCGATGTATGGTACAATCTCTCCGACA
T6 AAAGTCGGCGTGCTTAATCCGATGTATGGTACAATCTCTCCGACA
RB2 AAAGTCGGCGTGCTTAATCCGATGTATGGTACAATTTCTCCGACA
Amino Acids K V G V L N P M Y G T I S P T
-49
RBS
+1
T4 GGCAATGCTTTAAGATGGCATTATGATAAATGTAAATTTAAGGATTCTAAATGA
RB15 GGCAATGCTTTAAGATGGCATTATGATAAATGTAAATTTAAGGATTCTAAATGA
LZ7 GGCAATGCTTTAAGATGGCATTATGATAAATGTAAATTTAAGGATTCTAAATGA
T6 GGCAATGCTTTAAGATGGCATTATGATAAATGTAAATTTAAGGATTCTAAATGA
RB2 GGCAATGCTTTAAGATGGCATTATGATAAATGTAAATTTAAGGATTCTAAATGA
Amino Acids G N A L R W H Y D K C K F K D S K *
B
P_M
nrdA
nrdB
T2 GCTGAAGATTGTTCATCCTGTAAATTATGATATAATTTGACT
RB27 GCTGAAGATTGTTCATCCTGTAAATTATGATATAATTTGACT
RB32 GCTGAAGATTGTTCATCCTGTAAATTATGATATAATTTGACT
RB61 GCTGAAGATTGTTCATCCTGTAAATTATGATATAATTTGACT
Amino Acids A E D C S S C K L *
T2 CATGGACGAGTCACTATCTATTAACTAAGTGGAAAATTTATG
RB27
RB32
RB61
Amino Acids M
JS10 GCCGATGATTGTGCAGCTTGTAAGCTTTAATAAAATCTGATTATAATTGACT
JS98
RB69
Amino Acids A D D C A A C K L *
E S
JS10 CACGGATGAGTCACTAAC-AAATGAGAA--TAACATG
JS98
RB69
Amino Acids M

conservation of the *mobE* terminator suggests that the terminator is essential for the down-regulation of *mobE* expression.

In addition to the post-transcriptional down-regulation of *mobE* expression, there may also be down-regulation of MobE translation. The lack of a detectable *mobE* ribosomal binding site and the overlap of the *mobE* ORF with the *nrdA* gene suggests that *nrdA* and *mobE* are translationally coupled. Previous studies have shown that translational coupling of genes in T4 reduces the translation of the downstream gene (86). The stoichiometry of the NrdA to MobE translation is not known, nonetheless, I can speculate that the combination of the transcriptional terminator and translational coupling would reduce MobE expression. Collectively, the down-regulation of MobE expression suggests that over-expression of MobE may cause non-specific DNA damage to the host's genome.

4.2 An intact *mobE* gene is required for *mobE* mobility

Previous characterizations of the HEGs present in the *nrd* operon of T-even phage have provided genetic evidence that *mobE* is a functional HEG. However, efforts to detect MobE endonuclease activity using *in vitro* synthesized protein and PCR-generated substrates have not been successful (34,49)(G.W.W and D.R.E, unpublished data). One possible explanation for the lack of MobE cleavage with *in vitro* translated protein is the lack of DNA modifications that are normally present in T-even phage (49). In this work I have taken an *in vivo* approach to map a strand-specific nick in the T2 *nrdB* genes from T2 x T4 mixed infections (**Figure 13**). Furthermore, I could not detect any nicks in T2 *nrdB* when an in-frame deletion of the predicted HNH motif within *mobE* was used in the mapping assays. The requirement for an intact *mobE* ORF for nicking to occur in T2 *nrdB* coincides with the inheritance of *mobE* in the progeny phage from T2 x T4, and T2 x T4Δ*mobE* mixed infections (**Table III, Table IV**). Collectively, previous sequence analysis of MobE amino acid sequence, combined with the requirement for an intact MobE protein, suggests that MobE is a HNH endonuclease that promotes mobility by the catalysis of a strand-specific nick in T2 *nrdB* (49). However, the possibility that MobE catalyses a DSB break cannot be ruled out, since other HNH endonucleases such as *I-TevIII* from the T-even phage RB3 and *F-TflV* from phage T5 have been shown to catalyze DSBs (18,25).

4.3 MobE promotes the inheritance of *I-TevIII*

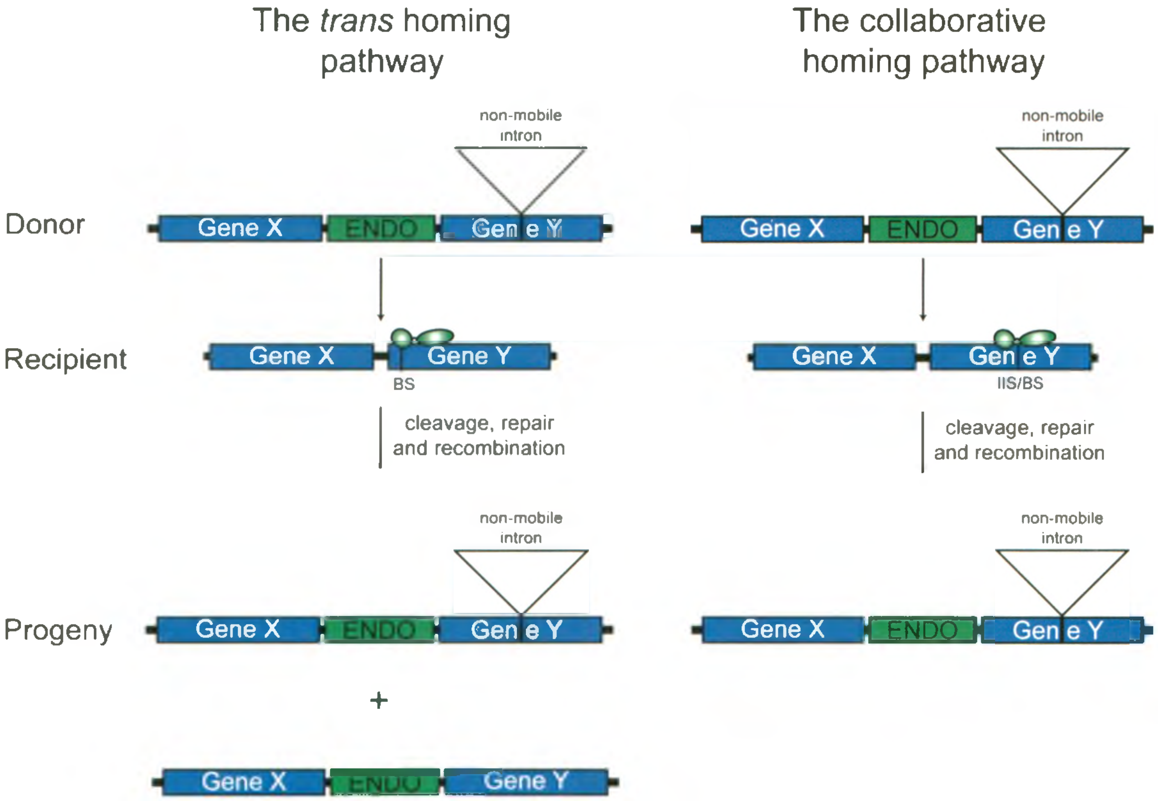
In addition to promoting the mobility of *mobE* in T2 x T4 mixed infections, an intact *mobE* gene is also required for the enhanced inheritance of the non-mobile *nrdB* intron and *I-TevIII* (**Table III, Table IV**). The dependence of *I-TevIII* mobility on an intact-*mobE* gene resolved a number of contradictory observations regarding the high frequency of *I-TevIII* inheritance, despite the *I-TevIII* encoded endonuclease being non-functional (64,67,78). Previous studies demonstrated that *I-TevIII* was non-functional using cloned fragments of the T4 *nrdB* intron in plasmid-to-phage homing assays (67). The lack of *I-TevIII* function was also supported by the observation that *I-TevIII* possessed an inactivating deletion compared to the functional I-TevIII protein found in the phage RB3 (64). Furthermore, contradictory evidence that *I-TevIII* and the *nrdB* intron are mobile in phage-to-phage crosses was demonstrated in data from this work and previous studies, despite the non-mobile nature of *I-TevIII* (78). The lack of *I-TevIII* inheritance observed in the plasmid-to-phage assays can be explained by the fact that the cloned T4 fragments used in the assays were missing the first 54 residues from the *mobE* ORF that contain the predicted HNH motif. However, in the phage-to-phage crosses the *mobE* gene is intact, and able to initiate the gene conversion events necessary for the high inheritance of the *nrdB* intron and *I-TevIII*.

Recently, Shub and colleagues presented a situation where a non-mobile intron and an adjacent free-standing HEG are interdependent on each other for homing (46,52). The free-standing homing endonuclease cleaves recipient HEG$^-$ genomes at the intron insertion site, which leads to the mobility of the intron and the HEG. The insertion of the non-mobile intron in the homing endonuclease recognition site prevents further cleavage of the host genome by the homing endonuclease. This collaboration between a non-mobile intron and a free-standing HEG has been termed collaborative homing, which is very similar to the mobilization of the *nrdB* intron by the adjacently encoded MobE endonuclease observed in this work. There are notable differences between the two pathways, one such difference is the position of the endonuclease cleavage site relative to the intron insertion site. The position of the cleavage site in the collaborative homing pathway encompasses the intron insertion site, whereas the cleavage site for MobE is within the *nrdB* coding region upstream of the *nrdB* intron insertion site. However, in both cases, cleavage or nicking by the endonuclease results in gene conversion that leads to the mobilization of the free-standing HEG and the non-mobile intron. Thus, these non-mobile introns can persist in the population without an intrinsic mechanism to promote their own mobility. This type of mobility is analogous to the to the mobilization of non-autonomous elements that lack the machinery required for their mobility, instead they are mobilized in *trans* by the machinery of a functional element (90-92). Furthermore, the *trans* mobility of the non-autonomous element is independent of the distance between the autonomous and non-autonomous elements. This distance independence is in contrast to the collaborative and *trans* homing pathways, which require the intron insertion site to be in close proximity to the free-standing endonucleases cleavage site (46,52).

For the intron to be mobilized the distance between the strand-break and the intron insertion site must be within the limits of the exonucleolytic degradation initiated from the break, otherwise recombination with the donor-genome will occur before the intron insertion site is reached (21,31-33). In the context of phage T4, the limit of the exonucleolytic degradation has been shown to be kb from the endonuclease cleavage site (21,31-34). Furthermore, the rate of inheritance of the intron is assumed to be reduced as the distance between intron insertion site and the free-standing endonuclease's recognition site increases. In the case of the collaborative homing pathway the intron and free-standing HEG are almost guaranteed to be mobilized together. The co-inheritance of the intron and free-standing HEG would occur due to the short 212bp distance between the free-standing HEG and the non-mobile intron in the model system studied. However, in the *trans* homing pathway with *mobE* and the *nrdB* intron, the distance between the intron insertion site and the MobE recogntion site is 553bp. This increased distance between the cleavage site and the intron insertion site causes the infrequent loss of the *nrdB* intron in *mobE*-containing progeny phage **(Figure 17)**.

The occurrence of the gene arrangement consisting of a functional homing endonuclease in close proximity to a non-mobile intron is expected to occur by chance. However, the occurrence of such an arrangement in small, endonuclease- and intron- rich

Figure 17. *trans* homing versus collaborative homing. Drawn schematically is a comparison of the *trans* (left) and collaborative (right) homing pathways, which cause the mobilization of a non-mobile intron (46,52). Both pathways begin with the transcription and translation of a free-standing homing endonuclease. The homing endonucleases catalyze strand-break(s) in their target site located in the cognate HEG minus allele of the recipient genome. The key difference between the two pathways is the location of the break site, in the case of *trans* homing the break site is located between the free-standing HEG and the non-mobile intron, whereas in collaborative homing the break site overlaps the intron insertion site. The repair of the strand breaks causes the mobilization of the free-standing HEG and the non-mobile intron in the progeny, however, in the *trans* homing pathway the distance between the HEG and non-mobile intron could be large enough that the intron will be infrequently lost.


The *trans* homing pathway
The collaborative homing pathway
non-mobile intron
Donor
Gene X
ENDO
Gene Y
Recipient
BS
IIS/BS
cleavage, repair and recombination
Progeny
+

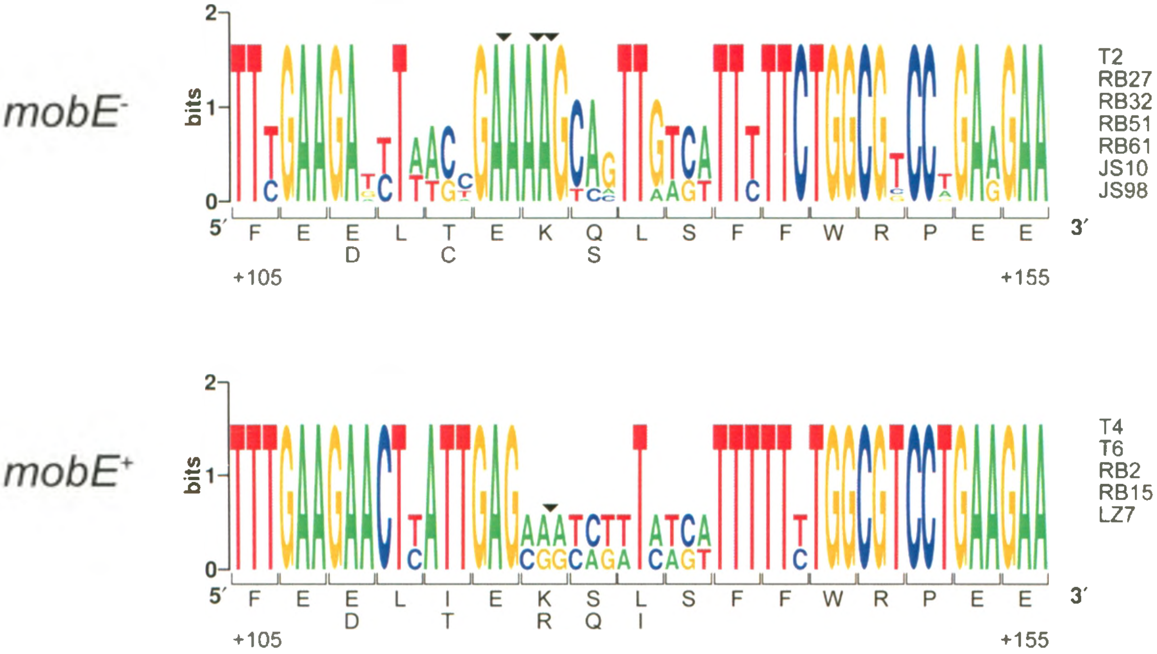
genomes may be more common, due to the finding that homing endonucleases and introns tend to be found in close proximity to conserved sequences of functionally critical host genes (50-52). These conserved sequences tend to occur rather infrequently. For example, phage T4 has 288 predicted genes and only 62 of these genes are considered essential under laboratory conditions (55). The low frequency of prospective HEG, and intron insertion sites, increases the probability that a free-standing endonuclease and an intron (or intein) will be within the limits of gene conversion events. Furthermore, the functional endonuclease could be encoded as a free-standing ORF or encoded within an intron or intein, and the *trans* homing pathway would still apply. In bacterial and phage genomes, there have been many occurrences of mobile and non-mobile introns (or inteins) in the same gene (37,39,93-95). These observations suggest that the *trans* mobility of a defunct or non-mobile intron (or intein), by an adjacent HEG may be common in bacterial and phage genomes.

4.4 The location of *mobE*-dependent nicks map to conserved regions of *nrdB*

I mapped the MobE-dependant nicking sites to a region of T2 *nrdB* that encodes conserved amino acids essential for the function of the NrdA/NrdB complex (**Figure 14 and 18**) (96,97). Specifically, the region contains a conserved tryptophan residue that is involved in the transfer of an electron radical from the iron/radical site of NrdB to the active site of NrdA. This radical transfer is essential for the reduction of ribonucleoside diphosphates to deoxynucleoside diphosphates. I identified three nicking sites in the T2 *nrdB* gene, one in glutamic acid 41 (GAA), and two in lysine 42 (AAG) (**Figure 14**). Moreover, the nucleotide sequences encoding the glutamic acid and the lysine are the same in all of the *mobE* lacking phage analyzed in this work (**Figure 18**). The equivalent codons and their nucleotide sequences in *mobE* containing phage are more varied; they are glutamic acid (GAG), followed by a lysine (AAG or AAA) or an arginine (CGG, CGA). Structural studies of the HNH homing endonucleases I-HmuI and I-TevIII have revealed a bipartite structure consisting of a separate N-terminal catalytic domain and a C-terminal DNA binding domain joined by a linker region (17,25). Current evidence suggests that MobE may also have a bipartite structure with an N-terminal DNA-binding domain and a C-terminal HNH containing catalytic domain (49). Based on these observations, I suggest that the MobE recognition site in *nrdB* is bipartite. The MobE DNA-binding domain would bind the nucleotides encoding the conserved amino acids FFWRPEE, and the N-terminal catalytic domain would bind the polymorphic region upstream of the conserved amino acids (**Figure 18**).

I suspected that the polymorphic region in the *mobE*-containing phage contained nucleotides that were suboptimal for MobE catalyzed strand breaks. This notion follows from similar observations with other free-standing homing endonucleases such as SegG (22,34-36,48). From these observations, I expected that in mixed infections with T4 and a T2 strain containing the T4 polymorphisms in the first 161bp of *nrdB* (T2*nrdB*^{R}) would be resistant to *mobE* mobility. As expected, the normalized inheritance ratio for *mobE* from T2*nrdB*^{R} x T4 mixed infections was similar to the ratios from T2 x T4Δ*mobE* mixed infections. Furthermore, no detectable nicks were mapped *in vivo* on genomic DNA isolated from T2*nrdB*^{R} mixed infections. However, I was able to map nicks in pT2nrdB^R, which corresponded to the location of the MobE recognition site in T2 *nrdB* (**Figure 13**

Figure 18. Logos representation of the *mobE*-dependent nicking sites in T-even phage. The logos of the nucleotides sequences flanking the *mobE*-dependent nicking sites in *mobE*$^+$ and *mobE*$^-$ phage. The amino acid sequence representing each codon and the nucleotide position in the *nrdB* gene is indicated below the logos. The *mobE*-dependent nick sites are shown by triangles. Note, the coding strand sequence is shown but the *mobE*-dependent nicks were mapped to the non-coding strand.

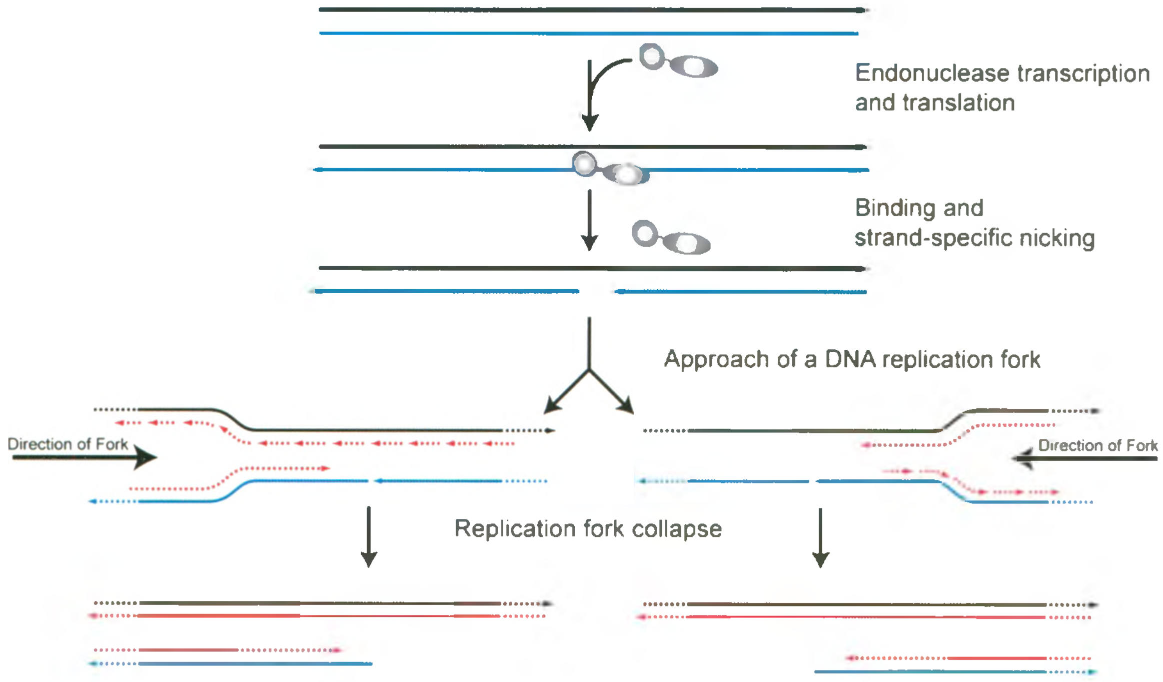

mobE-
bits
2
1
0
5′
F E E L T E K Q L S F F W R P E E
D C S
+105
+155
3′
T2
RB27
RB32
RB51
RB61
JS10
JS98
mobE+
bits
2
1
0
5′
F E E L I E K S L S F F W R P E E
D T R Q I
+105
+155
3′
T4
T6
RB2
RB15
LZ7

and 14). These observations suggest that MobE has reduced strand-break activity on the T4 version of the MobE recognition site compared to the T2 version, but I cannot conclude this until more experiments are carried out. The observation that MobE nicks the T4 version of its recognition site is not unique, other HNH homing endonucleases including *I-HmuI*, *I-HmuII*, *F-TflI* and *F-TflII*, have been shown to nick the host version of their recognition site less efficiently than the HEG$^-$ version (17-19,27). However, the significance of the modest nicking of the HEG-containing by the encoded endonuclease genome is not known.

4.5 Stand-specific nicks and the homing pathway

Evidence presented in my thesis suggests that MobE may also be a strand-specific DNA endonuclease, which nicks the non-coding strand of T2 *nrdB*. However, I was unable to detect DSBs specific to T2 *nrdB* in genomic DNA isolated from T2 x T4 mixed infections using Southern blot analysis. Previous studies have shown that endonuclease catalyzed strand-specific nicks are sufficient to cause gene conversion and HEG mobility (17,19,27). This leads to a central and unresolved question regarding the mechanism used by a DNA nick to initiate the homing pathway. The presence of gene conversion events associated with *mobE* mobility are consistent with the idea that the endonuclease catalyzed DNA nick may be converted to a DSB (19). One possible mechanism that would cause the conversion of a DNA nick to a DSB may be the collapse of a DNA replication fork as it approaches the nicked DNA (19,98). The replication fork collapse would cause a DSB in the daughter strand that uses the nicked DNA as template for replication (**Figure 19**). However, no experimental evidence has been presented to show

Figure 19. Model for the conversion of a strand-specific nick to a DSB. The model for the conversion of a homing endonuclease catalyzed strand-specific nick to a DSB by a collapsed DNA replication fork is illustrated. The template strands are shown in black and blue, the newly synthesized daughter strands are shown in red, and the direction of the approaching DNA replication fork is indicated. Additionally, the direction (5' to 3') of the DNA strands is indicated by an arrow. The endonuclease gene is transcribed and translated, and then the endonuclease nicks its target site in the HEG$^-$ genome. Next, a DNA replication fork approaches from either side of the strand-specific break, which leads to DNA replication fork collapse. The fork collapse leads to a DSB break in the daughter chromatid that uses the nicked template strand for replication.


Endonuclease transcription
and translation
Binding and
strand-specific nicking
Approach of a DNA replication fork
Direction of Fork
Direction of Fork
Replication fork collapse

that a collapsed DNA replication fork is part of the homing pathway initiated by a DNA nick.

An observation by Landthaler et al. showed that the nicking site of intron encoded HNH endonucleases is typically 5' to the HEG insertion site (37). The current reasoning for this observation is that the 3' OH of the nicked DNA strand is positioned to prime DNA replication towards the insertion site. The *mobE*-dependent strand-specific nick I mapped on the non-coding strand of the T2 *nrdB* gene is consistent with this observation. However, I also showed that *I-TevIII* located 3' to the nick site is inherited at a rate mirroring *mobE* in mixed infections, nonetheless, this observation does not rule out the possibility that the nick site location relative to the HEG insertion is significant.

4.6 Future directions and experiments

A next logical step regarding the regulation of *mobE* during phage infection would be to investigate *mobE* translation during phage infection. Current evidence suggests that *mobE* and *nrdA* are translationally coupled, but the stoichiometry of *mobE* translation compared to *nrdA* translation is not known. Since *mobE* and *nrdA* are present on polycistronic transcripts, western blots with antibodies specific to *nrdA* and *mobE* would provide evidence regarding the translation levels of *mobE* compared to the *nrdA*.

Furthermore, resolving the endonuclease activity of *mobE* is essential to demonstrate that *mobE* is functional endonucleases. The ideal situation would involve *in vitro* translation or expression and purification of MobE. Experimentation using MobE would determine if MobE catalyses single-strand nicks or DSBs.

The distance requirements for the *trans* homing pathway could be characterized by measuring the rate of inheritance of a non-mobile intron versus the distance of the intron from a homing endonuclease cleavage site. Measuring the rate of inheritance versus the distance could provide further insight into the gene conversion events associated with the homing pathway.

Finally, investigations into the homing pathway initiated by a strand-specific nick could provide insight into an unresolved question. Experiments could be performed to determine if the strand-specific nick is converted to a DSB through a mechanism independent of the homing endonuclease. Additionally, the importance of the nick site location versus the insertion site could be resolved.

References

1. Craig, N., Kidwell, M. and Lisch, D. (2002) In Craig, N., Craigie, R., Gellert, M. and Lambowitz, A. (eds.), *Mobile DNA II.* ASM Press, New York, pp. 3-11, 59-90.

2. McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K. *et al.* (2001) A physical map of the human genome. *Nature*, **409**, 934-941.

3. Miki, Y. (1998) Retrotransposal integration of mobile genetic elements in human diseases. *J Hum Genet*, **43**, 77-84.

4. Hurst, G.D. and Werren, J.H. (2001) The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet*, **2**, 597-606.

5. Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature*, **284**, 601-603.

6. Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604-607.

7. Hasler, J., Samuelsson, T. and Strub, K. (2007) Useful 'junk': Alu RNAs in the human transcriptome. *Cell Mol Life Sci*, **64**, 1793-1800.

8. Mahillon, J., Leonard, C. and Chandler, M. (1999) IS elements as constituents of bacterial genomes. *Res Microbiol*, **150**, 675-687.

9. Stoddard, B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys*, **38**, 49-95.

10. Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W. and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev*, **33**, 376-393.

11. Katayama, Y., Ito, T. and Hiramatsu, K. (2000) A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother*, **44**, 1549-1555.

12. Belfort, M., Derbyshire, V., Cousineau, B. and Lambowitz, A. (2002) In Craig, N., Craigie, R., Gellert, M. and Lambowitz, A. (eds.), *Mobile DNA II*. ASM Press, New York, pp. 761-783.

13. Dassa, B., London, N., Stoddard, B.L., Schueler-Furman, O. and Pietrokovski, S. (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res*, **37**, 2560-2573.

14. Zhao, L., Bonocora, R.P., Shub, D.A. and Stoddard, B.L. (2007) The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J*, **26**, 2432-2442.

15. Kim, Y.C., Grable, J.C., Love, R., Greene, P.J. and Rosenberg, J.M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307-1309.

16. Eastberg, J.H., McConnell Smith, A., Zhao, L., Ashworth, J., Shen, B.W. and Stoddard, B.L. (2007) Thermodynamics of DNA target site recognition by homing endonucleases. *Nucleic Acids Res*, **35**, 7209-7221.

17. 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.

18. Akulenko, N.V., Ivashina, T.V., Shaloiko, L.A., Shliapnikov, M.G. and Ksenzenko, V.N. (2004) [Novel site-specific endonucleases F-TflI, F-TflII and F-TflIV encoded by the bacteriophage T5]. *Mol Biol (Mosk)*, **38**, 632-641.

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. 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.

21. 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.

22. Edgell, D.R. (2005) In Belfort, M., Derbyshire, V., Stoddard, B. L. and Wood, D. W. (eds.), *Homing Endonucleases and Inteins*. Springer-Verlag, Heidelberg, pp. 147-160.

23. Pommer, A.J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kuhlmann, U.C., Cooper, A., Connolly, B.A., Hemmings, A.M., Moore, G.R. *et al.* (2001) Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J Mol Biol*, **314**, 735-749.

24. Cramer, W.A., Heymann, J.B., Schendel, S.L., Deriy, B.N., Cohen, F.S., Elkins, P.A. and Stauffacher, C.V. (1995) Structure-function of the channel-forming colicins. *Annu Rev Biophys Biomol Struct*, **24**, 611-641.

25. 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.

26. Kim, H.H., Corina, L.E., Suh, J.K. and Herrin, D.L. (2005) Expression, purification, and biochemical characterization of the intron-encoded endonuclease, I-CreII. *Protein Expr Purif*, **44**, 162-172.

27. 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.

28. Drouin, M., Lucas, P., Otis, C., Lemieux, C. and Turmel, M. (2000) Biochemical characterization of I-CmoeI reveals that this H-N-H homing endonuclease shares functional similarities with H-N-H colicins. *Nucleic Acids Res*, **28**, 4566-4572.

29. Corina, L.E., Qiu, W., Desai, A. and Herrin, D.L. (2009) Biochemical and mutagenic analysis of I-CreII reveals distinct but important roles for both the H-N-H and GIY-YIG motifas. *Nucleic Acids Res.*

30. Jurica, M.S. and Stoddard, B.L. (1999) Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci*, **55**, 1304-1326.

31. Mueller, J.E., Smith, D. and Belfort, M. (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Dev*, **10**, 2158-2166.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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 U S A*, **96**, 7005-7010.

40. 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.

41. Friedrich, N.C., Torrents, E., Gibb, E.A., Sahlin, M., Sjoberg, B.M. and Edgell, D.R. (2007) Insertion of a homing endonuclease creates a genes-in-pieces

ribonucleotide reductase that retains function. *Proc Natl Acad Sci U S A*, **104**, 6176-6181.

42. Nolan, J.M., Petrov, V., Bertrand, C., Krisch, H.M. and Karam, J.D. (2006) Genetic diversity among five T4-like bacteriophages. *Virol J*, **3**, 30.

43. Gibb, E.A. and Edgell, D.R. (2009) An RNA hairpin sequesters the ribosome binding site of the homing endonuclease mobE gene. *J Bacteriol*, **191**, 2409-2413.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. Bonocora, R.P. and Shub, D.A. (2009) A likely pathway for formation of mobile group I introns. *Curr Biol*, **19**, 223-228.

53. Belfort, M. and Roberts, R.J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res*, **25**, 3379-3388.

54. 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 U S A*, **98**, 7898-7903.

55. 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.

56. Matthews, C.K. (1994) In Karam, J. D., Drake, J. W., Kreuzer, K. N., Mosig, G., Hall, D. H., Eiserling, F. A., Black, L. W., Spicer, E. K., Kutter, E., Carlson, K. *et al.* (eds.), *Molecular Biology of Bacteriophage T4*. American Society for Microbiology, Washington, D.C., pp. 1-8.

57. Mosig, G. and Eiserling, F. (2006) In Calendar, R. (ed.), *The Bacteriophages* Oxford University Press, New York, pp. 225 - 267.

58. Yu, F. and Mizushima, S. (1982) Roles of lipopolysaccharide and outer membrane protein OmpC of Escherichia coli K-12 in the receptor function for bacteriophage T4. *J Bacteriol*, **151**, 718-722.

59. Goff, C.G. (1984) Coliphage-induced ADP-ribosylation of Escherichia coli RNA polymerase. *Methods Enzymol*, **106**, 418-429.

60. Drivdahl, R.H. and Kutter, E.M. (1990) Inhibition of transcription of cytosine-containing DNA in vitro by the alc gene product of bacteriophage T4. *J Bacteriol*, **172**, 2716-2727.

61. Kashlev, M., Nudler, E., Goldfarb, A., White, T. and Kutter, E. (1993) Bacteriophage T4 Alc protein: a transcription termination factor sensing local modification of DNA. *Cell*, **75**, 147-154.

62. 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 U S A*, **89**, 6658-6662.

63. 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.

64. 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.

65. Belfort, M. and Perlman, P.S. (1995) Mechanisms of intron mobility. *J Biol Chem*, **270**, 30237-30240.

66. 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.

67. Quirk, S.M., Bell-Pedersen, D. and Belfort, M. (1989) Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell*, **56**, 455-465.

68. Epstein, R.H., Bolle, A., Steinberg, C.M., Kellenberger, E., Tour, E.B.d.l. and Chevalley, R. (1964) Physiological Studies of Conditional Lethal Mutants of Bacteriophage T4D. *Cold Spring Harb Symp. Quant. Biol.*, **28**.

69. Selick, H.E., Kreuzer, K.N. and Alberts, B.M. (1988) The bacteriophage T4 insertion/substitution vector system. A method for introducing site-specific mutations into the virus chromosome. *J Biol Chem*, **263**, 11336-11347.

70. Kreuzer, H.W. and Kreuzer, K.N. (1994) Integration of plasmids into the bacteriophage T4 genome. *Genetics*, **138**, 983-992.

71. Goodman, H.M., Abelson, J., Landy, A., Brenner, S. and Smith, J.D. (1968) Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA. *Nature*, **217**, 1019-1024.

72. Eggertsson, G. and Soll, D. (1988) Transfer ribonucleic acid-mediated suppression of termination codons in Escherichia coli. *Microbiol Rev*, **52**, 354-374.

73. Russell, R.L. and Huskey, R.J. (1974) Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. *Genetics*, **78**, 989-1014.

74. Russell, R.L. (1974) Comparative genetics of the T-even bacteriophages. *Genetics*, **78**, 967-988.

75. Edgell, D.R. (2002) Selfish DNA: new abode for homing endonucleases. *Curr Biol*, **12**, R276-278.

76. Chiu, C.S., Cox, S.M. and Greenberg, G.R. (1980) Effect of bacteriophage T4 nrd mutants on deoxyribonucleotide synthesis in vivo. *J Biol Chem*, **255**, 2747-2751.

77. Yeh, Y.C. and Tessman, I. (1972) Control of pyrimidine biosynthesis by phage T4. II. In vitro complementation between ribonucleotide reductase mutants. *Virology*, **47**, 767-772.

78. Sandegren, L. and Sjoberg, 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.

79. 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.

80. 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.

81. Horton, R.M. (1997) In vitro recombination and mutagenesis of DNA. SOEing together tailor-made genes. *Methods Mol Biol*, **67**, 141-149.

82. Miller, E.S. and Carlson, K. (1994) In Karam, J. D. (ed.), *Molecular Biology of Bacteriophage T4*. ASM Press, Washington, D.C., pp. 455-456.

83. 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.

84. Sambrook, J. and Russell, D.W. (2001). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Vol. 1, pp. 6.39 - 36.46, 36.50 - 36.57.

85. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.

86. Torgov, M.Y., Janzen, D.M. and Reddy, M.K. (1998) Efficiency and frequency of translational coupling between the bacteriophage T4 clamp loader genes. *J Bacteriol*, **180**, 4339-4343.

87. Jeng, S.T., Lay, S.H. and Lai, H.M. (1997) Transcription termination by bacteriophage T3 and SP6 RNA polymerases at Rho-independent terminators. *Can J Microbiol*, **43**, 1147-1156.

88. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, **31**, 3406-3415.

89. Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res*, **16**, 9677-9686.

90. Dewannieux, M., Esnault, C. and Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*, **35**, 41-48.

91. 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.

92. 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 U S A*, **104**, 10962-10967.

93. 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 SPbeta. *Proc Natl Acad Sci U S A*, **95**, 1692-1697.

94. 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.

95. 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.

96. Nordlund, P. and Eklund, H. (1993) Structure and function of the Escherichia coli ribonucleotide reductase protein R2. *J Mol Biol*, **232**, 123-164.

97. 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.

98. Kuzminov, A. (2001) Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc Natl Acad Sci U S A*, **98**, 8241-8246.