To the Graduate Council:

I am submitting herewith a dissertation written by Scott A. Rice entitled "The Distribution and Inheritance of Retrons in the Proteobacteria and the Use of msDNA as a Diagnostic Probe." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Microbiology.

Bert Lampson

Bert C. Lampson, major Professor

We have read this dissertation and recommend its acceptance:

Bette C Mullin

Gary Stacey

David Brian

Accepted for the Council:

C W Minkel

Associate Vice Chancellor and
Dean of the Graduate School

The Distribution and Inheritance of Retrons in the Proteobacteria and the Use of msDNA as a Diagnostic Probe

A Dissertation
Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

Scott A. Rice
May 1996

Acknowledgments

First I would like to thank the members of my committee, David Brian, Beth Mullin, and Gary Stacey, for their guidance and suggestions. I would especially like to thank Bert Lampson for his constant patience and guidance throughout my studies at the University of Tennessee. I would also like to say thank you to my friends in the program for their support and our communal stress breaks, which have enriched my life as a graduate student.

I would also like to thank my parents, Gay Lynn and Delmar Hasselbacher and Bill Rice, for their encouragement during the course of my extensive academic career. Finally, I thank my wife, Diane McDougald, who has put up with many hardships to see me through this degree.

Abstract

This study presents information regarding the discovery of new retrons, their distribution in the proteobacteria, a mechanism of inheritance, and the first use of msDNA as a diagnostic probe to detect myxobacteria in soils. Approximately 10% of proteobacteria tested from the alpha and gamma groups produce msDNA. In contrast, >98% of myxobacteria strains (delta group) tested positive for msDNA production. DNA hybridization experiments revealed that myxobacteria retron elements have a clustered distribution, being found exclusively within particular myxobacteria subgroups. A retron element of the Mx162 type was cloned from *Melittangium lichenicola* and its DNA sequence compared with similar elements in *M. xanthus* and *Stigmatella aurantiaca*. Together, the degree of sequence diversity, the codon bias of the reverse transcriptase genes, and the clustered distribution of these retrons suggest a possible evolutionary scenario in which a common ancestor of the *Myxococcus* subgroup may have acquired this retroelement. It was found that some *Nannocystis exedens* strains contain mulitple, partial copies of the *msd* gene and suggests that msDNA is capable of generating short repeats which may have significant effects on the evolution of these bacteria.

Very few reports exist regarding the number of myxobacteria present in natural soils, or in contaminated soils. In this study, msDNAs were used as probes which indicated that DNA sequences specific to members of the *Myxococcus* and *Nannocystis* subgroups were present in both the contaminated and uncontaminated soils and represented 0.01% and 0.11% of the total detectable eubacteria in these soils respectively. Members of the third subgroup, *Chondromyces*, were not detected in any of the soil samples used in this study. Both the total number of bacteria detected and the number of myxobacteria decreased with increasing soil depth, and no myxobacteria were detected at a depth of 60 centimeters (cm).

Table of Contents

Chapter **Page**

I. Bacterial Reverse Transcriptase and msDNA 1
Introduction 1
msDNA has unique features 2
msDNA is produced by reverse transcriptase 5
Primary transcript 10
Retron RTs use a novel primer 11
Distribution of retron-encoded reverse transcriptase in bacteria 12
Bacterial reverse transcriptase 15
Are retrons mobile? 20
Do msDNA and RT have a function? 21

II. Diversity of Retron Elements in a Population of Rhizobia and Other Gram-Negative Bacteria 22
Introduction 22
Materials and Methods 23
Bacteria Strains 23
Detection of msDNA 23
Sequencing of msDNA 23
Hybridization experiments and generation of probes 27
Results and Discussion 27
Detection of retron elements among different bacterial groups 27
Prevalence of msDNAs among a population of rhizobia 29
Rhizobial retrons are unique 30
Proteus mirabilis 1174b msDNA is atypical 33
Why are retrons so rare? 35
Acknowledgments 39

III. Phylogenetic Comparison of Retron Elements Among the Myxobacteria: Evidence for Vertical Inheritance 40
Introduction 40
Materials and Methods 41
Bacterial strains and plasmids 41
Detection of msDNA 44
Detection of homologous retron elements 44
Cloning of a retron homologous to Mx162 44
DNA Sequencing 45
Silent substitutions and codon bias 45
Nucleotide sequence accession number 46
Results 46
Retron elements in the Myxobacteria 46
Homology to known retrons 48
Cloning and sequencing of a Melittangium retron 50
Retron junctions 61
Sequence divergence 63
Discussion 66
Acknowledgments 68

IV. Detection and Quantification of Myxobacteria in Natural and Contaminated Soils Using DNA Hybridization 69
Introduction 69
Materials and Methods 70
Bacterial strains and plasmids and primers 70
Probes 71
Soil samples 71
DNA preparation and hybridization 72
Determination of cell number 72
Mineralization assays 73
Results 74
Detection of myxobacteria in natural and contaminated soils 74
Quantification of bacterial cell numbers 81
Release of 14CO2 from labeled aromatic compounds 83
Chromosomal DNA from various myxobacteria hybridizes with known catabolic genes 83
Cloning of a nahA homologue 85
Discussion 87
Acknowledgments 90

V. Multiple Repeats of the Retron Encoded *msd* Gene in *Nannocystis exedens* 91
Introduction 91
Materials and Methods 92
Bacterial strains and plasmids 92
Detection of the chromosomally-encoded retron locus 92
Cloning of repeat sequences 92
DNA sequencing 93
Results 93
Nannocystis exedens strains contain two msDNAs 93
Some N. exedens strains contain multiple chromosomally encoded msd genes 95
Cloning and sequencing of reiterated retron sequences 98
Sequence analysis of msDNA repeat elements 101
Discussion 102
Summary 108
Acknowledgments 110
References 111
Vita 121

List of Tables

TABLE	PAGE
1.1 Characterized retrons	14
2.1 Rhizobial isolates tested for msDNA	24
2.2 Summary of hybridizations using rhizobial msDNAs as probes	32
3.1 Survey of the myxobacteria for retron elements	42
3.2 Primers used for sequencing of retron ML162	57
3.3 Rates of change at silent sites for RT and other bacterial genes	65
4.1 Quantification of bacteria in soil samples	82
4.2 $^{14}CO_2$ released from radiolabeled compounds in the presence of metabolically active cells	84
5.1 *Nannocystis exdens* Na e434 clones containing msDNA A repeat sequences	99

List of Figures

FIGURE PAGE

1.1 An ethidium bromide stained msDNA..... 3
1.2 Conserved secondary structures of msDNA..... 4
1.3 Structure of the unbranched form of msDNA..... 6
1.4 The organization of the chromosomal locus, known as a retron, coding for msDNA production..... 7
1.5 The mechanism for biosynthesis of msDNA by reverse transcriptase..... 8
1.6 Distribution and prevalence of retron elements among the proteobacteria..... 13
1.7 Amino acid alignment of bacterial RT..... 16
1.8 Domain organization of RT-encoding elements..... 18
2.1 Detection of ^{32}P-labeled msDNAs by the RT extension method..... 28
2.2 Southern hybridization with purified msDNA as a probe..... 31
2.3 Alkali treatment of ^{32}P end-labeled 1174b msDNA..... 34
2.4 DNA sequence of the msDNA from *Proteus mirabilis* 1174b..... 36
2.5 The proposed secondary structure of the msDNA stem-loop from *P. mirabilis* 1174b..... 37
2.6 DNA sequence alignment of msDNA from *P. mirabilis* 1174b with the plasmid pUC9..... 38
3.1 Detection of ^{32}P-labeled msDNA by the RT extension assay..... 47
3.2 Southern hybridization of nine members of the *Myxococcus* subgroup with a labeled Mx162 probe..... 49
3.3 Phylogenetic distribution of msDNA and Mx162 retrons in the myxobacteria..... 51
3.4 The sequence and secondary structure of msDNA from *Melittangium lichenicola* Fx e1..... 53
3.5 Sequencing strategy of the retron ML162..... 56
3.6 The complete nucleotide sequence of retron ML162 from *Melittangium lichenicola*..... 58
3.7 DNA sequence alignment of the 3' ends of ML162 and Mx162..... 62
3.8 The distal 3' region of retron ML162 from *Melittangium lichenicola*... 64

FIGURE PAGE
4.1 Detection of total eubacteria in soil by DNA hybridization..... 75
4.2 Detection of *Myxococcus* subgroup bacteria in soils..... 76
4.3 Detection of *Nannocystis* subgroup bacteria in soils..... 79
4.4 Detection of *Chondromyces* subgroup bacteria in soil..... 80
4.5 Southern hybridization of myxobacteria chromosomal DNA with *nah*A..... 86
5.1 Ethidium bromide stained msDNA from *Nannocystis exedens* Na e434..... 94
5.2 Sequence and secondary structure of msDNA from *Nannocystis exedens* Na e434..... 96
5.3 Southern hybridization of *Nannocystis* chromosomal DNA using msDNA A as a probe..... 97
5.4 Hybridization of *Nannocystis* strains with the cloned fragment, p1A.42, from strain Na e434..... 100
5.5 Alignment of *msd* repeat sequences with the DNA portion of msDNA A..... 103
5.6 Model for insertion of *msd* repeats..... 105

List of Abbreviations

aa	amino acid
avg	average
BIMEs	bacterial intrespersed mosaic elements
BLAST	basic local alignment tool
bp	base pairs
cm	centimeter
dN	any deoxyribonucleotide
ENV	envelope domain
ERICs	enterobacterial repetitive intergentic consensus sequences
gag	the genes encoding the internal structural proteins of the virus
GCG	genetics computer group
h	hour
HIV	human immunodeficiency virus
IN	integrase domain
IS	insertion sequence
kbp	kilobase pairs
LINE	long interspersed element
LTR	long terminal repeat
M-MLV-RT	moloney murine leukemia virus reverse transcriptase
mRNA	messenger RNA
msDNA	multicopy, single-stranded DNA
P	protease
pol	polymerase
ORF	open reading frame
RFLP	restriction length polymorphism
rG	guanine ribonucleotide
RT	reverse transcriptase
RNase A	ribonuclease A
RNase H	ribonuclease H
RRADs	repetitive retron associated elements
rRNA	ribosomal RNA
SINE	short interspersed element

List of Abbreviations

TCE	trichloroethylene
tRNA	transfer RNA
yr	year
Zn	zinc finger domain

Chapter I

Bacterial Reverse Transcriptase and msDNA

Introduction

The discovery of reverse transcriptase (RT), over 20 years ago (Baltimore, 1970) was important for two reasons. First, it suggested that genetic information could flow back and forth from DNA to RNA and from RNA to DNA, thus explaining the replication of some viruses. Second, the presence of RT suggested a mechanism for the shift from an RNA based life (Joyce, 1989, Waldrop, 1989), to one revolving around the DNA double helix. Since the discovery of RT in retroviruses, open reading frames encoding similar RT proteins have been described in association with a wide variety of viral and non-viral sources; LINE elements (Dombroski *et al.*, 1993), Ty elements (Boeke *et al.*, 1985), R2Bm Luan *et al.*, 1993, group II introns (Moran *et al.*, 1992), and telomerases (Blackburn, 1993). Despite the finding of many non-viral sources of RT, all reported RTs were found in eukaryotes, suggesting that eubacteria and the archeabacteria were devoid of RT activity. This suggested several possibilities. First, either RT evolved in eukaryotes after their divergence from eubacteria and archeabacteria, or second, that RT emerged prior to the divergence of the three groups and was subsequently lost by the eubacteria and archeabacteria. Then, in 1984 the discovery of an unusual extrachromosomal DNA, called multicopy, single-stranded DNA (msDNA) in the developmental bacterium, *Myxococcus xanthus* (Yee *et al.*, 1984), lead to the proposal that eubacteria had RT (Dhundale *et al.*, 1987). This theory was proven correct with the discovery of a chromosomal locus, termed a retron (Temin, 1989), which is responsible for the production of msDNA. The retron contains sequences coding for msDNA and an open reading frame (ORF). Subsequent homology searches with the ORF led to the discovery that the ORF codes for a reverse transcriptase similar to the polymerase of retroviruses. The discovery of retron-encoded RT supported the hypothesis that msDNA synthesis requires RT activity (Lampson *et al.*, 1989b, Lim and Maas, 1989), and the conservation of amino acid domains of both the eukaryotic and eubacterial RT suggest that the polymerases are derived from a common, ancient RT. This hypothesis could place RT at a critical stage in evolution from the

proposed RNA-based life. Therefore, if bacterial RTs are most reminiscent of the first RT, then the study of bacterial reverse transcriptase may be helpful in understanding the evolution from RNA to DNA-based information storage. After the initial discovery of msDNA in *M. xanthus*, retrons were found to be ubiquitous in the myxobacteria (Dhundale *et al.*, 1985, Lampson *et al.*, 1991a, Rice and Lampson, 1995) and in approximately 10% of all alpha and gamma proteobacteria tested (Herzer *et al.*, 1990, Kawaguchi *et al.*, 1992, Lim *et al.*, 1990, Rice *et al.*, 1993, Sun *et al.*, 1991). What follows is a review of the literature on bacterial RT associated with retrons along with new and relevant findings related to the distribution and evolution of RT in bacteria.

msDNA has unique features

msDNA, is a small, functionally obscure, extra-chromosomal nucleic acid, that is present in 500-700 copies per cell in *M. xanthus* (Furuchi *et al.*, 1987b). When total DNA prepared from bacterial cells is isolated on an acrylamide gel, msDNA can be directly observed as a fast migrating band, ranging in size from 65 to 163 bases (Figure 1.1), depending on the source bacterium. Among the different bacterial groups in which it is found, msDNA is highly diverse in its primary nucleotide sequence. However, all msDNAs share certain features. One feature of msDNA is a single-strand of DNA linked to a single-strand of RNA (Figure 1.2). Both the RNA and DNA components show considerable secondary structure consisting of extensive, stable stem-loops (Dhundale *et al.*, 1987, Furuchi *et al.*, 1987a). Second, a conserved, invariant guanine ribonucleotide (rG) near the 5' end of the RNA strand provides the link that covalently joins the two strands together (Figure 1.2). The 2'-OH of this rG residue is covalently linked through a phosphodiester bond to the 5'-P of the DNA strand. The third feature is the hybridization or overlap at the 3' ends of the RNA and DNA strands (Figure 1.2). This overlap is the result of the termination of reverse transcription of the msDNA and will be discussed further below. Nomenclature is based on the species of origin for the retron and the length of the DNA portion of msDNA. Thus the retron found in *Melittangium lichenicola*, with 162 bases in the DNA strand, is labeled ML162 (Rice and Lampson, 1995).

While these are highly conserved features, considered to be diagnostic of msDNA, a recent report presents the discovery of msDNA elements in *Escherichia coli* where msDNA lacks a 2'-5' phosphodiester linkage and is not associated with a single-stranded RNA (Lim, 1992). These msDNA elements, called "unbranched msDNA", have so far only been identified in *E. coli*. While the unbranched msDNAs have a slightly different

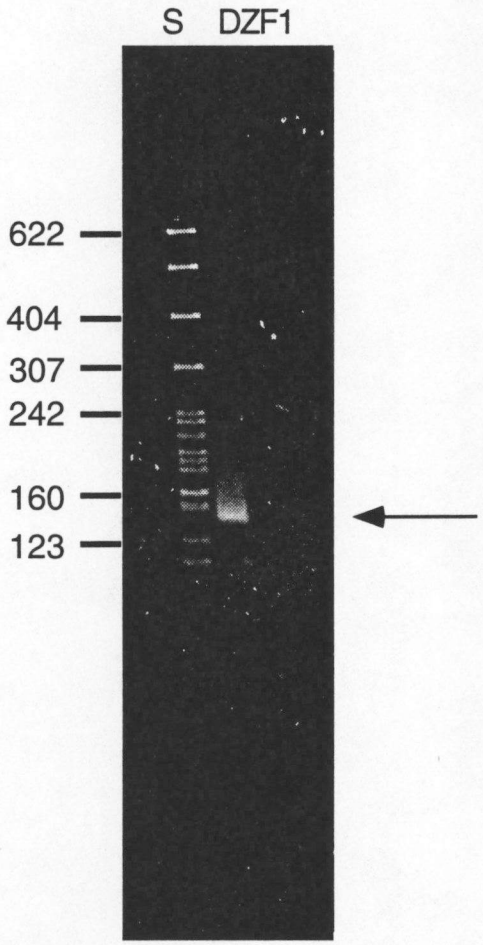


Figure 1.1. **Detection of msDNA in an ethidium bromide stained acrylamide gel.** DNA was extracted from the myxobacterium *Myxococcus xanthus* DZF1, treated with RNase A, electrophoresed on a 5% acrylamide gel, and stained with ethidium bromide. Molecular weight standards (S) are given in base pairs.

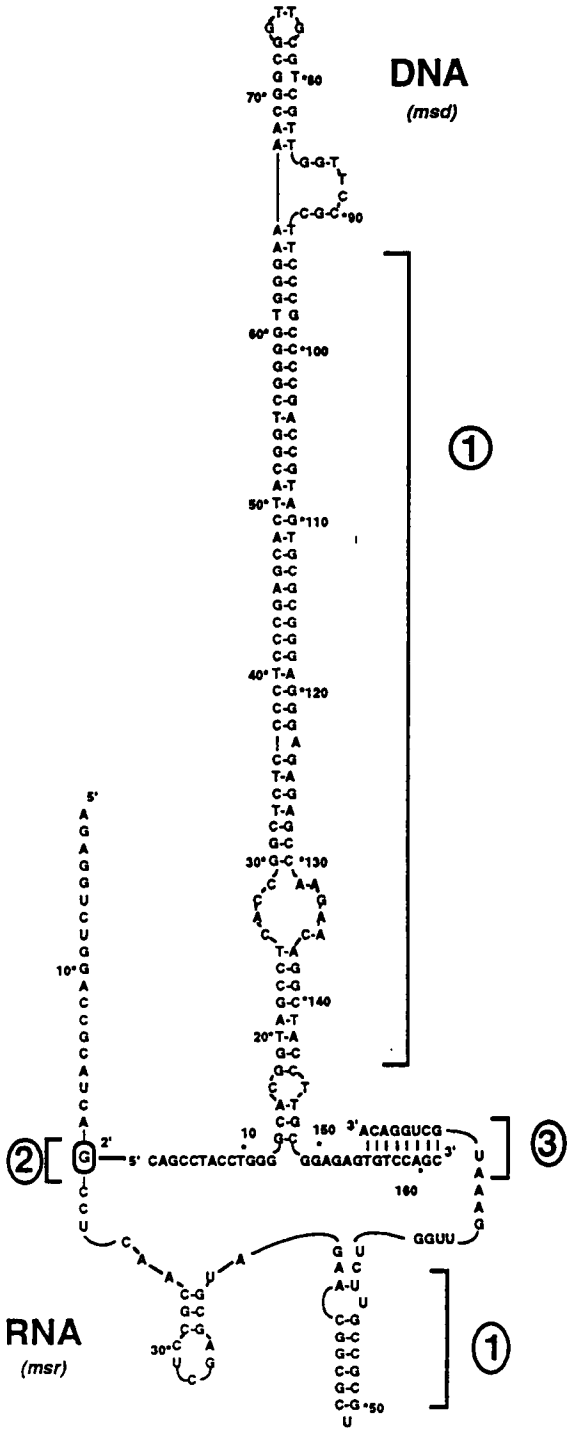


Figure 1.2. **Conserved secondary structures of msDNA.**
While highly variable in their primary nucleotide sequence, all msDNAs share the secondary structures numbered as shown: 1. The single strands of DNA and RNA form extensive, stable, stem and loops, 2. A 2'-5' phosphodiester bond joins the 5' end of the DNA strand to an internal G residue in the RNA strand, 3. The 3' ends of the DNA and RNA form a short hybrid. The structure of msDNA-ML162, from *Melittangium lichenicola* is shown.

final form, it appears that unbranched msDNA is first synthesized as a "typical" msDNA, and the DNA portion is subsequently processed by cleaving the DNA strand at the 5' end, near the branched G linkage to release the RNA strand (Figure 1.3). It should be noted that the cleavage involves breaking of a 3'-5' bond, not the 2'-5' linkage. While it is unclear how the processing occurs, it has been speculated that the cleavage is carried out by either an endonuclease activity of the RT, or that the msDNA itself is autocatalytic. The significance of the unbranched form is unknown at this time.

As stated above, msDNA and RT are chromosomally encoded genes linked in a single loci termed a "retron" (Temin, 1989). Figure 1.4 shows a schematic of the organization of retrons. In every case to date, the gene order of the retron is the same. The RNA and DNA strands of msDNA are encoded by the region denoted *msr* and *msd* respectively; note the overlap of *msr-msd* which corresponds to the overlapping 3' ends of the RNA and DNA strands in the msDNA. The RT encoding ORF is separated from *msr-msd* by a small spacer region and is at the 3' end of the retron locus. A pair of imperfect, inverted repeats, denoted as a1 and a2, flank the *msr-msd* region (Figs. 1.4 and 1.5A) and are required for proper folding of the transcript for msDNA synthesis (Dhundale *et al.*, 1987, Lampson *et al.*, 1989a). Aside from a1-a2, the retron locus seems to lack any other direct or inverted repeats, and in the case of the myxobacteria, lack well defined 5' and 3' ends (Rice and Lampson, 1995). In contrast, the ends of some *E. coli* retrons have been analyzed and are discussed below.

msDNA is produced by reverse transcriptase

Formation of msDNA *in vivo* is dependent on transcription of the entire retron locus, translation of the RT, and proper folding of the mRNA to be used as primer and template for the RT (Figure 1.5A). Closer examination reveals several distinct features of retron elements and msDNA formation that separate them from other RT elements. The long terminal repeat (LTR)-containing retrotransposons have RNA pol II promoters within the 5' LTR (Boone and Skalka, 1993) which initiate transcription downstream from the promoter, resulting in an mRNA lacking the 5' promoter site. Regeneration of the 5' LTR is achieved by reverse transcription using the 3' LTR as a template. In lacking LTRs and associated promoters, some of the non-LTR elements rely on insertion near a host promoter for control of transcription (Eickbush, 1992). Not all non-LTR retroposons rely on insertion near host promoters, and are able to initiate transcription upstream of an internal promoter (Eickbush, 1992). Retrons do not contain LTRs, but do appear to have their own

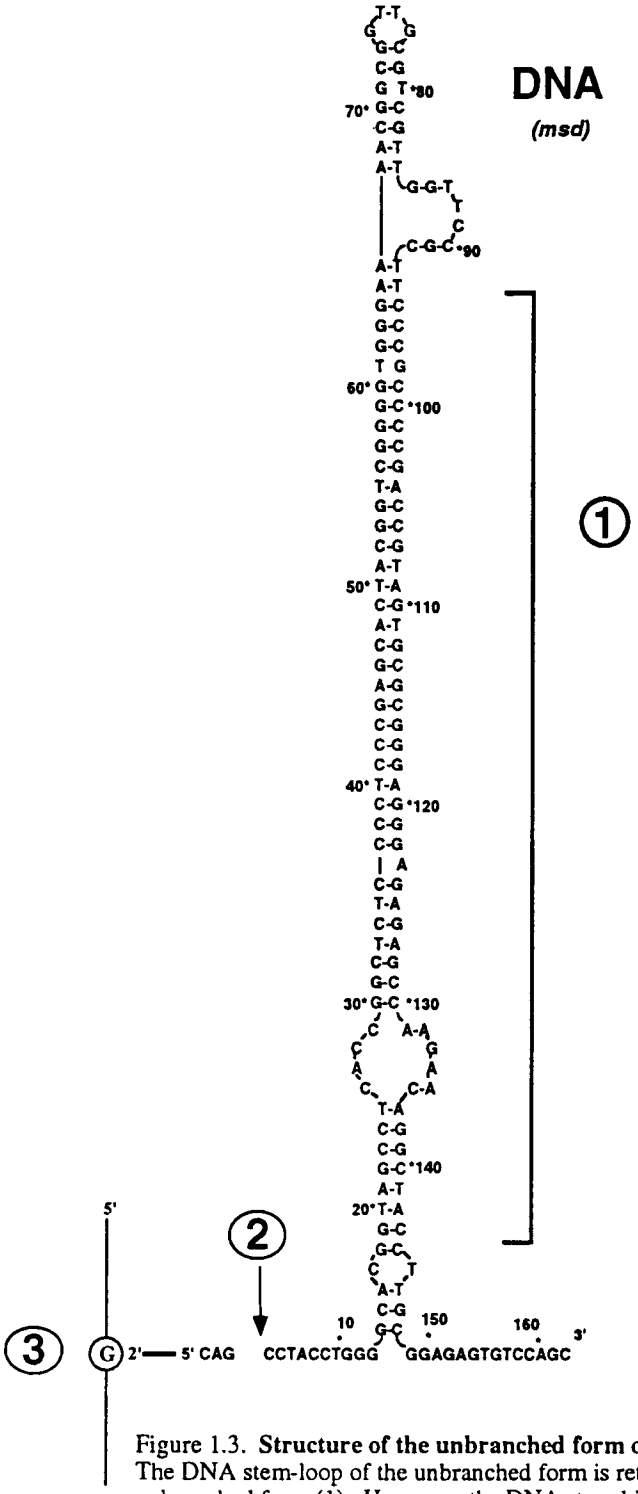


Figure 1.3. **Structure of the unbranched form of msDNA.** The DNA stem-loop of the unbranched form is retained in the unbranched form (1). However, the DNA strand is cleaved at a 3'-5' phosphodiester bond near the 5' end of the DNA strand (2). Thus, 2-3 deoxyribonucleotides remain linked to the RNA strand through the 2'-5' linkage (3), and are subsequently lost, along with the RNA strand, shown as a thin line.

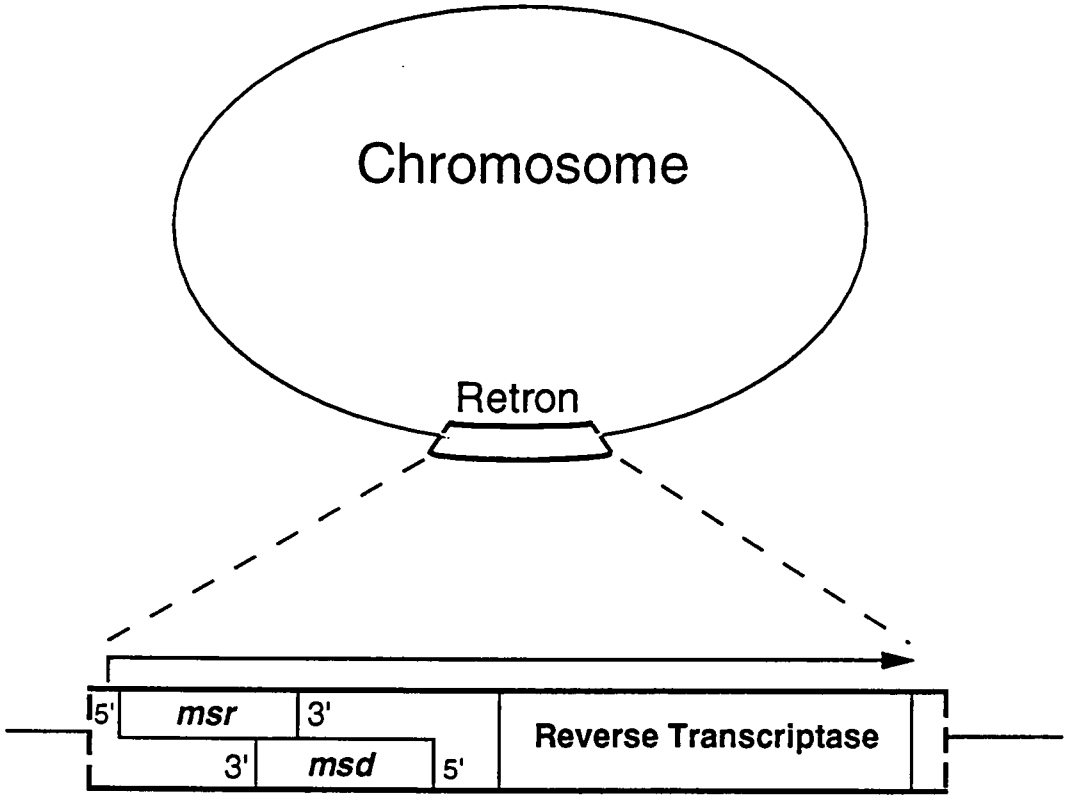
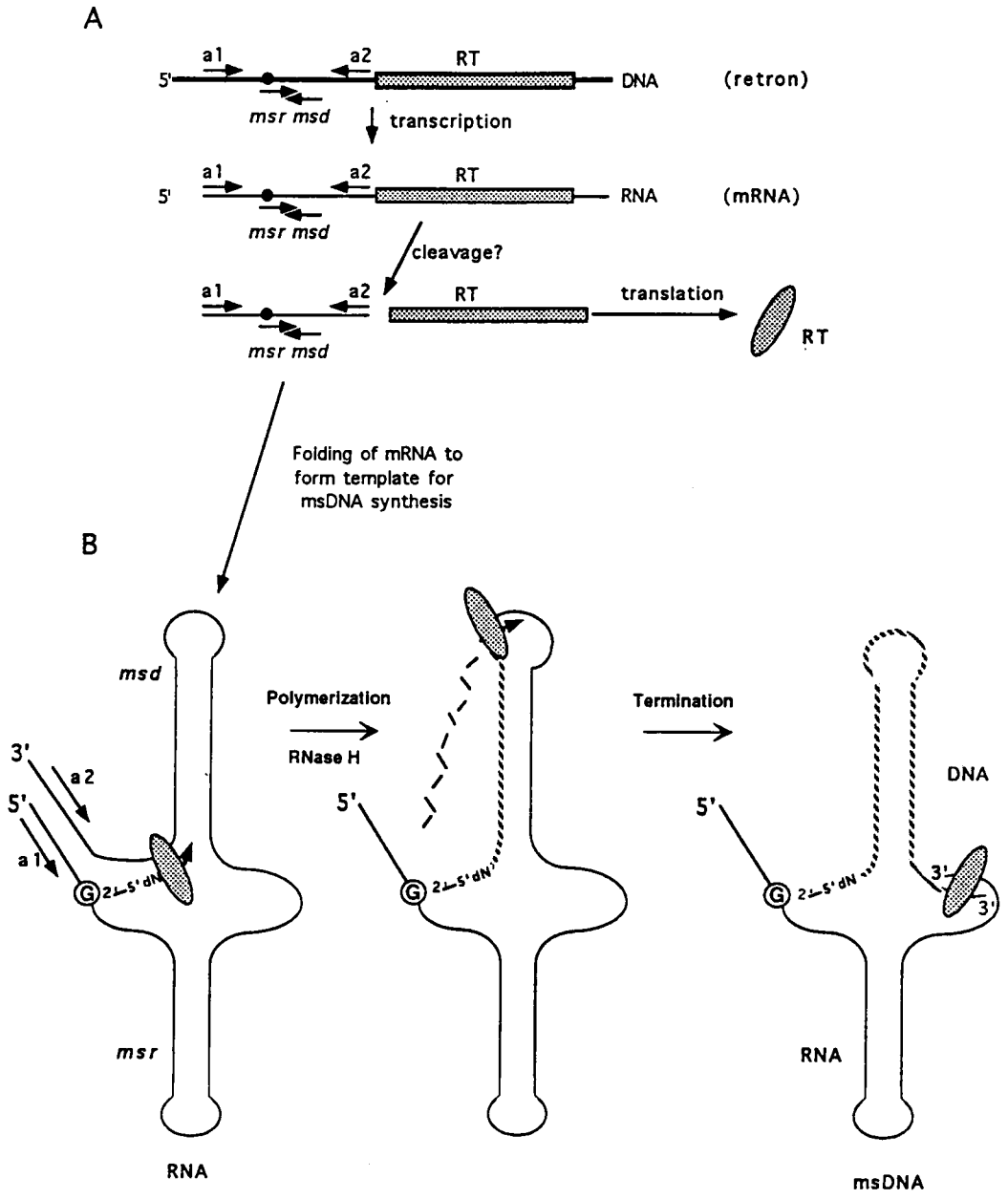


Figure 1.4. **The organization of the chromosomal locus, known as a retron, coding for msDNA production.** Direction of transcription is denoted by the arrow. A single transcript coding for both msDNA and RT is made from the chromosomal locus. *msr* and *msd* correspond respectively to the RNA and DNA strands of the msDNA. The RT-ORF is always found 3' of the *msr-msd* region. The 5' and 3' ends are marked with dashed lines to indicate the lack of specific information about the ends of most retrons.

Figure 1.5. **The mechanism for biosynthesis of msDNA by reverse transcriptase.** (A) Transcription of a retron coding for msDNA and RT. A single transcript is made that is initiated at or near a1, and includes both the *msr-msd* and the RT regions. After the initial mRNA is made, it is unclear whether or not the transcript is cleaved between the *msr-msd* and RT coding region. The RT mRNA is translated and the mRNA which serves as a template for msDNA synthesis folds on itself, using the inverted repeats a1 and a2; this also positions the branch G residue to prime cDNA synthesis. The inverted repeats a1 and a2 are denoted by arrows to indicate their orientation. Arrows are also used to indicate the 5' to 3' orientation of *msr* and *msd* in relation to the RNA and DNA strands of msDNA. The filled circle represents the position of the conserved guanine residue which forms the 2'-5' phosphodiester linkage in msDNA (B) RT recognizes the folded template RNA and initiates cDNA synthesis of the *msd* region using the 2'-OH of the branched rG for a primer. As cDNA synthesis proceeds, host encoded RNase H activity removes the RNA from the growing DNA strand. Finally, RT terminates cDNA synthesis, leaving a molecule composed of a single-strand of DNA, covalently linked through a 2'-5' phosphodiester bond to a single-strand of RNA. dN represents the first deoxyribonucleotide linked to the branched G, (circled G). The shaded oval is the RT protein; thin lines represent RNA; the hatched line is DNA; *msr* and *msd* code for the RNA and DNA strands of msDNA respectively.


A
a1
a2
RT
5'
DNA
(retron)
msr msd
transcription
RNA
(mRNA)
cleavage?
translation
RT
Folding of mRNA to
form template for
msDNA synthesis
B
msd
3'
a2
5'
a1
msr
RNA
Polymerization
RNase H
Termination
DNA
msDNA

internal promoter since plasmid clones of retrons are capable of msDNA production (Dhundale *et al.*, 1987, Eickbush, 1992, Lim and Maas, 1989). S1 mapping of primary transcripts from retron Mx162 indicated that transcription is initiated approximately 70 base pairs (bp) 5' of the *msr* region at the a1 site. A putative promoter has been identified for *E. coli* retron Ec107, which has similarity to other *E. coli* promoters (Herzer *et al.*, 1992). Promoters have been mapped for the myxobacteria retrons, but specific sequences have not been identified. Based on the position of the retron promoter, the primary transcript for msDNA production does not appear to contain the entire element. Thus, if retrons transpose through an RNA intermediate, it is unclear how full length elements, including the promoter, could be regenerated.

Primary transcript

A single mRNA transcript of the retron, beginning at or near the a1 site, is made which serves as both transcript for translation of RT and as template for msDNA synthesis (Figure 1.5A) (Dhundale *et al.*, 1987, Lampson *et al.*, 1989a, Lim and Maas, 1989). This is supported by S1 mapping and LacZ fusion constructions. Whether the RT-encoding region of the mRNA is cleaved prior to msDNA synthesis is unknown. It is known that the region of mRNA coding for msDNA folds into a complex secondary structure by using the a1 and a2 inverted repeats. Mutations in either repeat sequence that disrupt base pairing between the two, block or sharply reduce msDNA synthesis. However, if compensatory mutations are made that restore the base pairing, synthesis of msDNA is restored (Hsu *et al.*, 1989, Lampson *et al.*, 1989a). Base pairing by the a1 and a2 units is also vital for positioning the branched guanine residue such that RT can recognize and utilize it as a primer for msDNA synthesis (Figure 1.5B). The RT then presumably binds at the a1-a2 stem site, and using the 2'-OH of the conserved guanine residue as a primer, begins reverse-transcription of the *msd* region. As reverse transcription proceeds, the RNA is removed from the growing DNA strand by host encoded RNase H activity (Lampson *et al.*, 1989a, Shimamto *et al.*, 1995b). Reverse transcription then terminates at the end of the *msd* region, leaving the RNA-DNA complex linked through a 2'-5' phosphodiester bond. The reason for termination is not currently known, but the RT always terminates at the same site, suggesting that either additional proteins (Viswanathan *et al.*, 1989) or RNA secondary structure are involved in blocking further reverse-transcription. After reverse transcription is terminated, the RT may remain associated with the msDNA (Lampson *et al.*, 1991b), although the reason for this is not known.

Retron RTs use a novel primer

The primer, supplied by the correct positioning of the 2'-OH of a rG, is required for msDNA production (Figure 1.5B). Exchange of the rG for rC or rA blocks the synthesis of msDNA (Hsu *et al.*, 1989). This is the only reported occurrence of a 2'-OH priming cDNA synthesis, and contrasts sharply with retroviral and LTR retrotransposon priming mechanisms which rely on the use of a 3'-OH supplied by a host tRNA. Until the discovery of retrons, the 3'-OH priming mechanism for retroviral cDNA synthesis was unchallenged. Shimamoto *et al.* (1995a) demonstrated through *in vitro* studies that HIV reverse transcription is initiated by a 3'-OH and is incapable of using a 2'-OH, while the bacterial RT is capable of using the 2'-OH. Finally, *in vitro* synthesis experiments, using only exogenously added mRNA template and purified RT, demonstrated that the 2'-OH of the rG is the only potential primer, and subsequent treatment of the product with yeast debranching enzyme specifically cleaves the 2'-5' linkage (Shimada *et al.*, 1994). Yeast debranching enzyme is involved in debranching RNA lariats formed during mRNA splicing. Because the 2'-OH allows for self-priming it has been proposed that bacterial RTs are ancient RTs, and possibly represent a progenitor RT enzyme (Shimada *et al.*, 1994, Temin, 1989). However, in the Mauriceville plasmid of *Neurospora* mitochondria, reverse transcription also is self-primed. For the Mauriceville plasmid, self-priming is initiated not from a 2'-OH, but from a 3 '-OH provided by the 3' terminus which resembles tRNA in secondary structure (Wang and Seeger, 1992). This is similar to retroviral RTs which rely on the 3'-OH of a tRNA for priming of cDNA, synthesis. Thus, while eukaryotic RTs use a variety of priming mechanisms (self-priming, protein primer, tRNA primers), they all rely on a 3'-OH. The ability to use a 2'-OH primer appears to be unique to the bacterial RTs which produce msDNA.

Synthesis of msDNA has been shown through *in vitro* experiments to be absolutely dependent on the presence and activity of RT (Hsu *et al.*, 1992a, Lampson *et al.*, 1989b, Shimamoto *et al.*, 1995a). In the presence of a conventional primer, such as a DNA oligonucleotide, purified retron RT will reverse transcribe a non-specific template, such as 5S RNA (Lampson *et al.*, 1990). However, it appears that *in vivo*, retron RT is specific for its corresponding RNA template (e.g. RT-Ec73 will only use the Ec73 *msr-msd* template). Shimamoto *et al.* (1993) demonstrated that the specificity of a retron RT is dictated by the *msr* sequence in the template RNA (e.g. RT-Mx162 will only recognize *msr* from Mx162), while the *msd* sequence can be supplied from any retron. The specificity of RT for its template is likely due to secondary structure and specific sequences around the

branched G residue in the template RNA. Finally, it appears that as long as a stable stem is maintained in the DNA strand of msDNA, msDNA synthesis is not effected by the specific sequence of *msd*. Thus, alterations in the size of the DNA stem seem to have little effect on production of msDNA. Indeed, removal of almost all of the stem, or addition of extra sequences in the loop region, does not block synthesis of msDNA (Mirochnitchenko *et al.*, 1994, Miyata *et al.*, 1992, Shimada *et al.*, 1994). Therefore, it may be possible to use msDNA as a system for expression of heterologous DNA sequences, such as anti-sense sequences, in a variety of cell types.

Distribution of retron-encoded reverse transcriptase in bacteria

As noted above, RT was first found in bacteria as an open reading frame linked to the locus encoding msDNA (Lampson *et al.*, 1989b, Lim and Maas, 1989). Since the primary discovery, with the exception of *Cystobacter ferrugineus* Cb fe18, msDNA has been found in every myxobacterium tested, (Dhundale *et al.*, 1987, Lampson *et al.*, 1991a, Rice and Lampson, 1995) (Figure 1.6). Hybridization studies of the myxobacteria retrons indicated similar msDNA elements were shared by members of the same myxobacteria subgroup. For example, all myxobacteria in the *Myxococcus* subgroup have retrons that share significant nucleotide identity with retron Mx162. Thus, particular msDNAs seem to be clustered within particular subgroups.

In contrast to the myxobacteria retrons, msDNA is only found in approximately 10% of all *E. coli* strains tested (Figure 1.6) (Herzer *et al.*, 1990, Kawaguchi *et al.*, 1992, Lim and Maas, 1989, Sun *et al.*, 1989), and 6-10% of all *Klebsiella*, *Proteus*, *Salmonella*, and *Rhizobia* strains tested (Rice *et al.*, 1993). These newly discovered retrons did not cross-hybridize with any of the previously characterized retrons (Table 1.1), or with each other, and are presumed to be novel elements. While it appears that there is a lower incidence of retrons in the non-myxobacteria proteobacteria, the number of retron-encoding elements in *E. coli* and other strains may be grossly underestimated. The newly discovered unbranched forms of msDNA are not easily detected and thus may be overlooked. In addition, the current screening method will not detect retrons unless msDNA is produced, thus non-msDNA-producing retrons would also be overlooked. Retrons have not been reported in archeabacteria or eukaryotes. However, using a clone of retron Ec67, under control of host specific promoters, it was demonstrated that both *Saccharomyces cerevisiae* (Miyata *et al.*, 1992) and mammalian cells (Mirochnitchenko *et al.*, 1994) are capable of supporting msDNA synthesis. Interestingly, RT has been found in an archeabacterium,

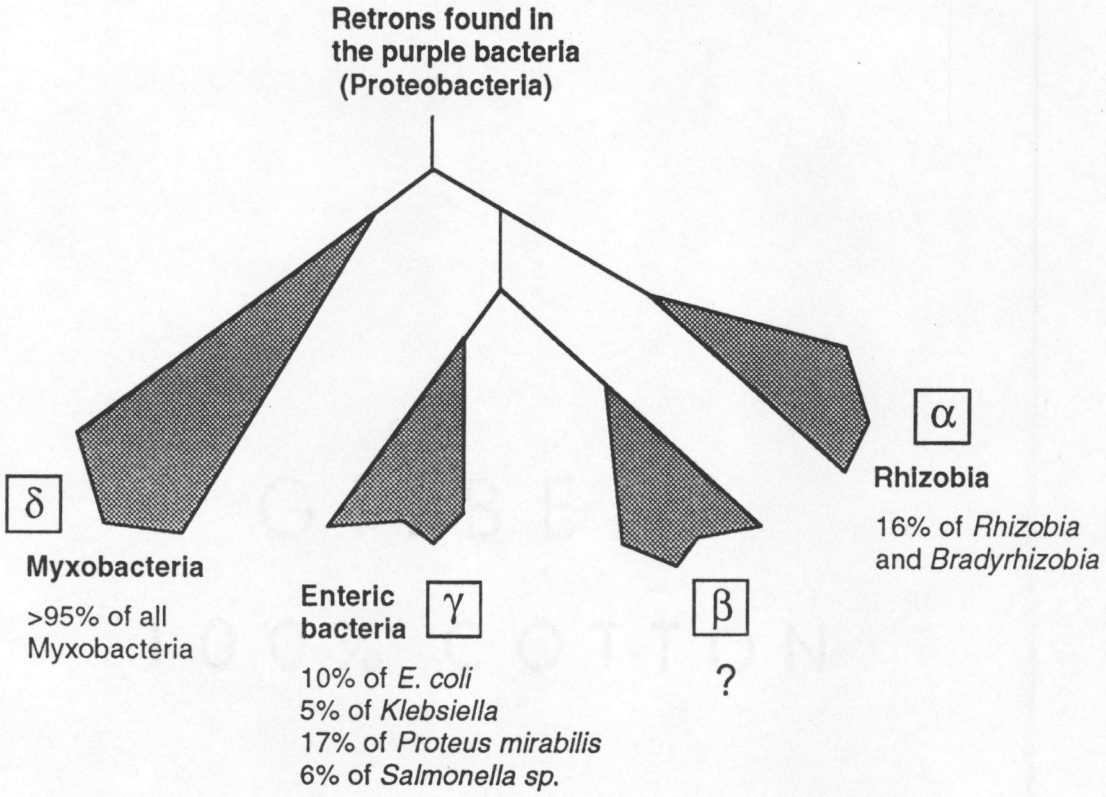


Figure 1.6. **Distribution and prevalence of retron elements among the proteobacteria**. An abbreviated phylogenetic tree of the purple bacteria, based on 16S rRNA analysis by Woese (1987), is presented. The tree is divided into four major branches, α, β, γ, and δ, with the specific bacterial groups screened for retron elements shown. The prevalence for each group is reported as the percentage of strains from each group tested that produce msDNA. At present, no bacterial strains from the β branch have been screened for retrons.

Table 1.1. Characterized retrons

Retron	Size of msDNA strand (nucleotides)	Source	Insertion site	RT ORF (amino acids)	Other features	Reference
Mx65	65	*Myxococcus xanthus*	chromosome	427	135 aa[c] amino terminus	Dhundale *et al.*,1988, Inouye *et al.*, 1990
Mx162	162	*Myxococcus xanthus*	chromosome	485	170 aa amino terminus	Dhundale *et al.*, 1987, Inouye *et al.*, 1989
Sa163	163	*Stigmatella aurantiaca*	chromosome	480	170 aa amino terminus	Furuichi *et al.*, 1987a, 1987b, Hsu *et al.*, 1992b
ML162	162	*Melittangium lichenicola*	chromosome	482	170 aa amino terminus	Rice and Lampson, 1995
Ec67	67	*E. coli*	prophage (186-like)	586	potential RNase H domain	Lampson *et al.*, 1989b
Ec73	73	*E. coli*	prophage (P4-like)	320	extra ORF	Sun *et al.*, 1988
Ec86	86	*E. coli*	prophage (P2-like)	320	extra ORF	Lim and Maas, 1989
Ec107	107	*E. coli*	chromosome	319	no	Herzer *et al.*, 1992
Ec83	83[a]	*E. coli*	ND[b]	313	no	Lim, 1992
Ec78	78[a]	*E. coli*	ND	ND	ND	Lim, 1992

[a]debranched msDNA
[b]ND, not determined
[c]aa, amino acids

Calothrix PCC7061 (Cal.x1), a cyanobacterium, and in an *Azotobacter* strain, but those RTs were found to be associated with group II introns (Ferat and Michel, 1993). The relationship between the intron-encoded RTs and retron RTs is discussed below.

Bacterial reverse transcriptase

While the different reverse transcriptases of bacterial retrons show little overall identity at either the nucleotide or amino acid level, analysis of the amino acid sequences reveals the presence of the seven conserved structural domains indicative of reverse transcriptases (Figure 1.7). Most notable of these is domain five, containing the YXDD box, which for the bacterial reverse transcriptases is YADD. This region has been implicated in the catalytic site for RT (Wakefield *et al.*, 1992). Interestingly, information presented by Xiong and Eickbush (1990) indicated that domain five is YADD for the majority of the non-LTR retrotransposons of eukaryotes. While retron RTs share features with the non-LTR elements, an amino acid alignment that includes the newly reported bacterial group II intron RT demonstrates that retron RTs have a conserved group of amino acids in domain 7 that distinguish them from other RT-encoding elements (Fig. 1.7 domain 7 is VTG for retron RT). Phylogenetic analysis (Xiong and Eickbush, 1990) of the RTs indicates that the group II introns are among the closest relatives of the retron RTs. However, even the bacterial group II intron elements seem to be significantly different from retrons in that the intron RTs have a zinc finger domain at the C terminus of the protein that is not found in the retron RTs (Figure 1.8).

In addition to the reverse transcriptase domain, eukaryotic RTs have other domains that are linked to RT. The most common of these is an RNase H domain, C terminal to the RT domain (Figure 1.8) (McClure, 1991). RNase H functions by removing RNA from the RNA-DNA hybrid after reverse transcription. Of the retron-encoded RTs, only one retron, Ec67, has a C terminal region which could be an RNase H, although RNase H activity has not been demonstrated (Lampson *et al.*, 1989a). Thus, for all retrons except possibly Ec67, RNase H activity is provided by the host-encoded enzyme, *rnh*A, and is required for production of the "mature" form of msDNA consisting of single-stranded DNA (Shimamoto *et al.*, 1995b). The myxobacteria RTs have an additional 135-165 amino acids at the N-terminus that are not observed in *E. coli* retron RTs. However, the function of this region is not known. No other definable motifs have been observed and thus retron RTs represent the simplest forms of RT known.

Figure 1.7. **Amino acid alignment of bacterial RT.** The boxed regions, numbered I-VII are the seven conserved reverse transcriptase domains. Amino acids that are conserved are in bold, with the corresponding amino acid at the top of the column. Capital and lower case letters, respectively, represent amino acids either found in all the RTs shown or that are specific for bacterial RTs. References and accession number for sequences used are as follows: Mx65 (Hsu *et al.*, 1989, J03763), Mx162 (Inouye *et al.*, 1989, M24392), ML162 (Rice and Lampson, 1995, L36722), Sa163 (Furuichi *et al.*, 1987b, M86352), Ec67 (Lampson *et al.*, 1989b, M55249), Ec73 (Inouye *et al.*, 1991, M64113), Ec86 (Lim, 1991, X60206, X60207), Ec107 (Herzer *et al.*, 1992, X62583), Ec161 (Lim, 1992, Z12832), Cal is the RT sequence from the group II bacterial intron (Ferat and Michel, 1993, X71404). A retroviral RT, (HIV-1) is included for comparison (Wain-Hobson *et al.*, 1985, M38432).

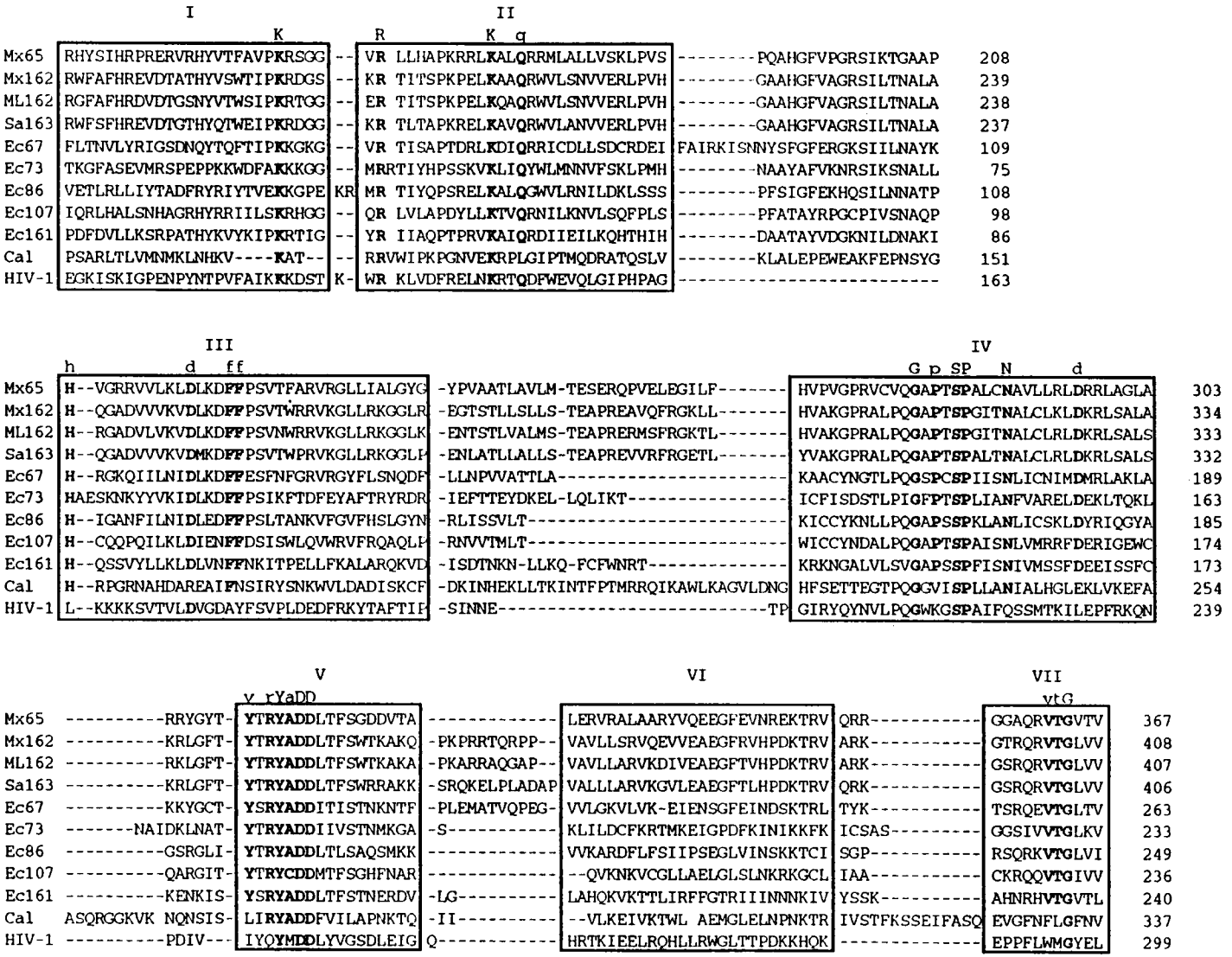
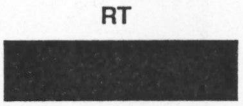

I
II
K
R
K q
Mx65 RHYSIHRPRERVRHYVTFAVPKRSGG -- VR LLHAPKRRLKALQRRMLALLVSKLPVS -------- PQAHGFVPGRSIKTGAAP 208
Mx162 RWFAFHREVDTATHYVSWTIPKRDGS -- KR TITSPKPELKAAQRWVLSNVVERLPVH -------- GAAHGFVAGRSILTNALA 239
ML162 RGFAFHRDVDTGSNYVTWSIPKRTGG -- ER TITSPKPELKQAQRWVLSNVVERLPVH -------- GAAHGFVAGRSILTNALA 238
Sa163 RWFSFHREVDTGTHYQTWEIPKRDGG -- KR TLTAPKRELKAVQRWVLANVVERLPVH -------- GAAHGFVAGRSILTNALA 237
Ec67 FLTNVLYRIGSDNQYTQFTIPKKGKG -- VR TISAPTDRLKDIQRRICDLLSDCRDEI FAIRKISNNYSFGFERGKSIILNAYK 109
Ec73 TKGFASEVMRSPEPPKKWDFAKKKGG -- MRRTIYHPSSKVKLIQYWLMNNVFSKLPMH -------- NAAYAFVKNRSIKSNALL 75
Ec86 VETLRLLIYTADFRYRIYTVEKKGPE KR MR TIYQPSRELKALQGWVLRNILDKLSSS -------- PFSIGFEKHQSILNNATP 108
Ec107 IQRLHALSNHAGRHYRRIILSKRHGG -- QR LVLAPDYLLKTVQRNILKNVLSQFPLS -------- PFATAYRPGCPIVSNAQP 98
Ec161 PDFDVLLKSRPATHYKVYKIPKRTIG -- YR IIAQPTPRVKAIQRDIIEILKQHTHIH -------- DAATAYVDGKNILDNAKI 86
Cal PSARLTLVMNMKLNHKV----KAT-- -- RRVWIPKPGNVEKRPLGIPTMQDRATQSLV -------- KLALEPEWEAKFEPNSYG 151
HIV-1 EGKISKIGPENPYNTPVFAIKKKDST K- WR KLVDFRELNKRTQDFWEVQLGIPHPAG ---------- 163
III
IV
h d ff
G p SP N d
Mx65 H--VGRRVVLKLDLKDFFPSVTFARVRGLLIALGYG -YPVAATLAVLM-TESERQPVELEGILF------- HVPVGPRVCVQGAPTSPALCNAVLLRLDRRLAGLA 303
Mx162 H--QGADVVVKVDLKDFFPSVTWRRVKGLLRKGGLR -EGTSTLLSLLS-TEAPREAVQFRGKLL------- HVAKGPRALPQGAPTSPGITNALCLKLDKRLSALA 334
ML162 H--RGADVLVKVDLKDFFPSVNWRRVKGLLRKGGLK -ENTSTLVALMS-TEAPRERMSFRGKTL------- HVAKGPRALPQGAPTSPGITNALCLRLDKRLSALS 333
Sa163 H--QGADVVVKVDMKDFFPSVTWPRVKGLLRKGGLP -ENLATLLALLS-TEAPREVVRFRGETL------- YVAKGPRALPQGAPTSPALTNALCLRLDKRLSALS 332
Ec67 H--RGKQIILNIDLKDFFESFNFGRVRGYFLSNQDF -LLNPVVATTLA---------- KAACYNGTLPQGSPCSPIISNLICNIMDMRLAKLA 189
Ec73 HAESKNKYYVKIDLKDFFPSIKFTDFEYAFTRYRDR -IEFTTEYDKEL-LQLIKT---------- ICFISDSTLPIGFPTSPLIANFVARELDEKLTQKL 163
Ec86 H--IGANFILNIDLEDFFPSLTANKVFGVFHSLGYN -RLISSVLT---------- KICCYKNLLPQGAPSSPKLANLICSKLDYRIQGYA 185
Ec107 H--CQQPQILKLDIENFFDSISWLQVWRVFRQAQLH -RNVVTMLT---------- WICCYNDALPQGAPTSPAISNLVMRRFDERIGEWC 174
Ec161 H--QSSVYLLKLDLVNFFNKITPELLFKALARQKVD -ISDTNKN-LLKQ-FCFWNRT---------- KRKNGALVLSVGAPSSPFISNIVMSSFDEEISSFC 173
Cal H--RPGRNAHDAREAIFNSIRYSNKWVLDADISKCF -DKINHEKLLTKINTFPTMRRQIKAWLKAGVLDNG HFSETTEGTPQGGVISPLLANIALHGLEKLVKEFA 254
HIV-1 L--KKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIP -SINNE---------- TP GIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQN 239
V
VI
VII
v rYaDD
vtG
Mx65 ----------RRYGYT- YTRYADDLTFSGDDVTA ---------- LERVRALAARYVQEEGFEVNREKTRV QRR---------- GGAQRVTGVTV 367
Mx162 ----------KRLGFT- YTRYADDLTFSWTKAKQ -PKPRRTQRPP-- VAVLLSRVQEVVEAEGFRVHPDKTRV ARK---------- GTRQRVTGLVV 408
ML162 ----------RKLGFT- YTRYADDLTFSWTKAKA -PKARRAQGAP-- VAVLLARVKDIVEAEGFTVHPDKTRV ARK---------- GSRQRVTGLVV 407
Sa163 ----------KRLGFT- YTRYADDLTFSWRRAKK -SRQKELPLADAP VALLLARVKGVLEAEGFTLHPDKTRV QRK---------- GSRQRVTGLVV 406
Ec67 ----------KKYGCT- YSRYADDITISTNKNTF -PLEMATVQPEG- VVLGKVLVK-EIENSGFEINDSKTRL TYK---------- TSRQEVTGLTV 263
Ec73 -------NAIDKLNAT- YTRYADDIIVSTNMKGA -S---------- KLILDCFKRTMKEIGPDFKINIKKFK ICSAS--------- GGSIVVTGLKV 233
Ec86 ----------GSRGLI- YTRYADDLTLSAQSMKK ---------- VVKARDFLFSIIPSEGLVINSKKTCI SGP---------- RSQRKVTGLVI 249
Ec107 ----------QARGIT- YTRYCDDMTFSGHFNAR ---------- --QVKNKVCGLLAELGLSLNKRKGCL IAA---------- CKRQQVTGIVV 236
Ec161 ----------KENKIS- YSRYADDLTFSTNERDV -LG---------- LAHQKVKTTLIRFFGTRIIINNNKIV YSSK---------- AHNRHVTGVTL 240
Cal ASQRGGKVK NQNSIS- LIRYADDFVILAPNKTQ -II---------- --VLKEIVKTWL AEMGLELNPNKTR IVSTFKSSEIFASQ EVGFNFLGFNV 337
HIV-1 ----------PDIV--- IYQYMDDLYVGSDLEIG Q---------- HRTKIEELRQHLLRWGLTTPDKKHQK ---------- EPPFLWMGYEL 299

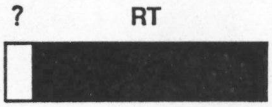
Figure 1.8. **Domain organization of RT-encoding elements.** (A) Prokaryotic RTs. The protein region for Mx162, marked with a "?", indicates the 130-170 amino acid domain found in the amino terminus of the three myxobacteria RTs, Mx162, ML162, and Sa163. The exact identity and function of that domain is unknown. Note the absence of protease, integrase and LTRs. (B) Eukaryotic RT encoding elements. RT, reverse transcriptase (dark box); RH, RNase H (hatched box); Zn, zinc finger motif (horizontal lined box); G, gag region (stippled box); P, protease domain (diagonal lined box); IN, integrase domain (horizontal lined box); ENV, envelope domain (open box); LTR, long terminal repeats (vertical lined box).

A.

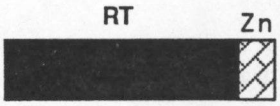
Retron Ec107
(bacteria)



Retron Mx162
(bacteria)

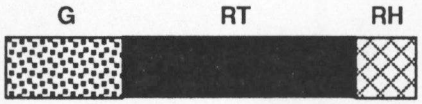


Group II intron
(bacteria)

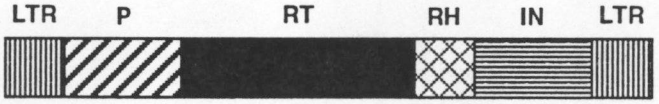


B.

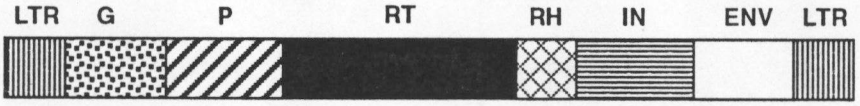
LINE of L1
(Human)



Retrotransposon
gypsy (insect)



Retrovirus
(mammals)



Did the first RT elements originally contain most of the domains, including *gag*, *env*, and *int*, recognizable in retroviruses and the LTR retrotransposons, or was the primordial RT similar to retron RT, lacking these extra domains (Figure 1.8)? Xiong and Eickbush (1990) have proposed that the progenitor RT was most likely similar in domain organization to the non-LTR retrotransposons, consisting of a *gag*, RT, RNase H, and *int* domains. This scenario implies that retron RT is descended from a non-LTR RT, having lost both the *gag* and *int* domains.

Are retrons mobile?

Flavel (1995) suggests that all RT-encoding elements are not present for the benefit of the host, but rather that RT-encoding elements are primarily selfish entities that exist only to propagate and spread. Indeed, nearly every class of retroelement has been shown to be mobile, from the retroviruses, to Ty elements (Boeke, 1985), to the elements of *B. mori* (Luan *et al.*, 1993), to LINE elements (Dombroski *et al.*, 1993). Currently, it is unknown if retron elements are mobile since transposition of a retron has not been directly demonstrated. In fact, for the myxobacteria, based on the distribution of retrons, codon usage, and the synonymous substitution rate among the RT genes, it appears that retron Mx162 is an ancient gene in the myxobacteria genome and is most likely inherited vertically instead of through horizontal transfer (Rice and Lampson, 1995).

For *E. coli,* as well as other bacteria, retrons appear to be recent additions to the genome. In the alpha and gamma groups of the proteobacteria, retrons are found in approximately 10% of the strains tested (Rice *et al.*, 1993) (Figure 1.6). Sequence analysis of the *E. coli* retrons indicates the RT gene has a G+C content and codon usage that differs from "typical" *E. coli* genes and implies recent acquisition of the retron. This information would suggest that retrons are not ancestral *E. coli* genes. Three of the *E. coli* retrons that have been cloned and sequenced, Ec67 (Lampson *et al.*, 1989b), Ec73 (Dhundale *et al.*, 1988b), and Ec86 (Lim and Maas, 1989) (Table 1.1), have all been determined to be associated with phage sequences. For example, retrons Ec67 and Ec86 have both inserted independently into the same site of a 186-like prophage (Lim, 1991). In the presence of P2 helper phage, retron Ec73 has been shown to be mobilized with its associated phage, Rϕ73 (Inouye *et al.*, 1991, Sun *et al.*, 1991). Thus, one mode of transmission of retrons seems to be through the lytic replication cycle of a host prophage, although how the retrons became associated with the phage has not been resolved.

In contrast, *E. coli* retron Ec107 (Herzer *et al.*, 1992) (Table 1.1), is not associated with phage sequences. Analysis of the sequences flanking retron Ec107 indicates that the element has inserted in between the *pyr*E and *ttk* genes of *E. coli*, and replaced a 34 bp region separating the two genes. In *E. coli* strains lacking Ec107, the 34 bp region is present. This 34 bp region is present in some strains that are closely related, having common parent strains, while Ec107 can be found in distantly related strains (Herzer *et al.*, 1990, Kawaguchi *et al.*, 1992). This data indicated that Ec107 was recently introduced into those *E. coli* strains rather than having been lost by strains lacking Ec107. Indirectly, this suggests that Ec107 may be mobile and such mobility is apparently not dependent on phage sequences, although it must be stressed that mobility of Ec107 has not been demonstrated experimentally. Additionally, sequences defining the 5' and 3' ends of Ec107 show neither direct or indirect repeats, nor target site duplications, and thus do not indicate a conventional method of transposition.

Do msDNA and RT have a function?

Deletion of a retron from the chromosome has no observable phenotype on either vegetative cells, or on the development of myxobacteria (Inouye *et al.*, 1990) and thus the function of the element remains unknown. Is msDNA a mechanism for generating genetic diversity? Maas *et al.* (1994) have recently presented data that suggests msDNA may play a role in mutagenesis. Specifically, the bulge regions in the *msd* stem recruit proteins involved in mismatch repair and in doing so may increase the mutation rate of the host cell, or perhaps msDNA, in conjunction with the RT serves as a primer for reverse transcription of other mRNAs. As noted above, RT seems to remain associated with the msDNA at the 3'-OH and thus the msDNA-RT complex may serve as a primer, although cDNA synthesis by the RT-msDNA complex has not been observed. Alternatively, the typical, branched msDNA may in fact be a defective species, a theory which is supported by the discovery of unbranched msDNA. Could the unbranched form be active in some hitherto unknown capacity? Interestingly, yeast debranching enzyme is capable of cleaving the 2'-5' linkage *in vitro* (Dhundale *et al.*, 1987, Furuchi *et al.*, 1987a). Chapman and Boeke (1991) demonstrated debranching activity was required for efficient retrotransposition of Ty1. Does msDNA function in some way to regulate the (postulated) mobility of retron elements? Since retrons have been incorporated into the genome of several phages, does msDNA or RT affect the replication of those viruses? Many intriguing questions about retrons and the affects of retrons on their hosts remain to be answered.

Chapter II

Diversity of Retron Elements in a Population of Rhizobia and Other Gram-Negative Bacteria

Introduction

A new class of retroelements which encode a reverse transcriptase (RT) structurally similar to the polymerase found in retroviruses has recently been discovered in bacterial cells (Inouye and Inouye, 1992, Lampson *et al.*, 1989a, Lim and Maas, 1989). These genetic elements, termed retrons, are found on the chromosome of several different strains of *Escherichia coli* and the soil bacterium *Myxococcus xanthus*, and represent the first known example of a RT-bearing element in a prokaryotic organism (Inouye and Inouye, 1992).

All known retrons produce an unusual RNA-DNA satellite molecule called multicopy single-stranded DNA (msDNA). In *M. xanthus*, msDNA-Mx162 consists of a 162-base single-stranded DNA joined to a 77 base single-stranded RNA (Dhundale *et al.*, 1987). However, the DNA strand is, surprisingly, linked to the RNA strand at a specific internal guanosine residue such that a unique 2'-5' phosphodiester bond (branching out from the 2' position of this G) joins the RNA to the 5' end of the DNA strand (see Lampson *et al.* 1991b for a review of the structure of msDNA). Although msDNAs are produced in abundant quantity (up to 500 copies per cell), their function remains obscure. However, it is now well established that the retron-encoded RT is responsible for its synthesis (Lampson *et al.*, 1989a, Lim and Maas, 1989).

Presently retron elements have been reported to exist in only two bacterial groups: *E. coli* and a few members of the myxobacteria such as *M. xanthus* and *Stigmatella aurantiaca* (Dhundale *et al.*, 1985, Herzer *et al.*, 1990, Hsu *et al.*, 1992a, Lim and Maas 1990, Sun *et al.*, 1989). Here we report the presence of msDNA-producing retron elements in a number of diverse bacterial groups including the rhizobia and demonstrate the highly diverse nature of these elements among a natural population of rhizobial strains.

Materials and Methods

Bacteria Strains

The rhizobia screened in this study are listed in Table 2.1 and were kindly donated by Peter van Berkum (USDA, Beltsville, MD). The *Proteus mirabilis*, *Klebsiella pneumonia*, *Enterococcus* strains, were obtained from David Bemis (University of Tennessee, College of Veterinary Medicine, Knoxville, TN). The *Salmonella* strains, donated by Noel Smith and Robert Selander, were from the SARB reference collection (Beltran *et al.*, 1991, Selander and Smith, 1990) (Pennsylvania State University). The myxobacterium *Nannocystis exedens* Na e1 was obtained from Lawrence Shimkets (University of Georgia, Athens). *Escherichia coli* B was also used. *E. coli*, *P. mirabilis*, *K. pneumonia*, and *Salmonella* strains were grown at 37°C in modified Luria-Bertani medium (Maniatis *et al.*, 1989). *N. exedens* Na e1 was grown on MD1 medium (Reichenbach and Dworkin, 1992) at 30°C.

Detection of msDNA

Total RNA was purified by the guanidinium thiocyanate method (Chromcyzynski and Sacchi, 1987) from freshly grown cells or frozen cell pellets. Purified RNA (this fraction also contains msDNA) was stored in ethanol at -80°C. msDNA was detected by the RT extension method (Lampson *et al.*, 1991a, Lampson *et al.*, 1990, Lim and Mass, 1989). Labeled products were divided in half, with RNaseA treatment of one set of samples, and separated on 4% Acrylamide-8M Urea gels. Gels were dried and exposed to x-ray film at -80°C.

Sequencing of msDNA

The sequence of the DNA portion of msDNA was determined directly. Briefly, DNA was extracted from growing cells by the alkaline lysis method (Birnboim and Doly, 1979), RNase A-treated, and electrophoresed on a 5% acrylamide gel. The ethidium bromide-stained msDNA band was subsequently cut out of the gel, purified, and 5' end labeled using [γ^{32}P] ATP (New England Nuclear) and T4 Polynucleotide Kinase (New England Biolabs, Beverly, MA). The ^{32}P-labeled product was subsequently isolated by electrophoresis on a 5%-8M Urea acrylamide gel. The purified, ^{32}P-labeled, msDNA was then sequenced by the chemical cleavage method of Maxam and Gilbert (1980).

Table 2.1. Rhizobial isolates[a] tested for msDNA.

Name (legume host)	USDA strain no.	geographic source/date	produce[b] msDNA
Rhizobium sp. (Acacia)	3002	Brazil/1959	+
Rhizobium sp. (Acacia)	3003	Africa/1950	
Rhizobium sp. (Acacia)	3325	Morocco/1974	
Rhizobium sp. (Acacia)	3838	1976	+
Bradyrhizobium sp. (Aeschynomene)	3516	Florida/1972	+
Bradyrhizobium sp. (Aeschynomene)	4362		
Bradyrhizobium sp. (Albizia)	3004	Maryland/1952	+
Bradyrhizobium sp. (Apios)	3240	Maryland/1939	
Bradyrhizobium sp. (Arachis)	3339	Thailand/1979	
Bradyrhizobium sp. (Arachis)	3341	Hawaii/1978	
Rhizobium sp. (Astragalus)	3854	Alaska/1962	
Rhizobium sp. (Cajanus)	3472		
Bradyrhizobium sp. (Canavalia)	3317	Brazil/1794	
Rhizobium sp. (Cicer)	3378		
Rhizobium sp. (Cicer)	3379	Mexico/1963	
Bradyrhizobium sp. (Coronilla)	3165	Virginia/1935	
Bradyrhizobium sp. (Coronilla)	3167	1961	
Bradyrhizobium sp. (Crotalria)	3384	Brazil/1967	
Bradyrhizobium sp. (Desmodium)	3225	Ecuador/1948	
Bradyrhizobium sp. (Erythrina)	3241		
Bradyrhizobium sp. (Erythrina)	3242	Maryland/1939	+
Rhizobium fredii	191	China/1979	
Rhizobium leguminosarum	2370	Illinois/1933	
Rhizobium leguminosarum	2429	Hawaii/1978	
Rhizobium leguminosarum	2435	Holland/1955	

Table 2.1. (continued)

Name (legume host)	USDA strain no.	geographic source/date	produce msDNA
Rhizobium leguminosarum	2489		
Rhizobium sp. (Lens)	2426		
Rhizobium sp. (Lens)	3404	Columbia/1979	
Rhizobium loti	3084	Maryland/1946	
Rhizobium loti	3468	New Zealand/1961	+
Rhizobium loti	3469		
Rhizobium loti	3471		
Rhizobium loti	3503		
Rhizobium loti	3669	California/1968	
Bradyrhizobium sp. (Lotus)	3074	Minnesota/1954	
Bradyrhizobium sp. (Lotus)	3470	California/1916	
Rhizobium sp. (Lupinas)	3040	Florida/1940	
Bradyrhizobium sp. (Lupinas)	3045	Florida/1946	
Bradyrhizobium sp. (Macrotyloma)	3451	Zimbabwe/1960	
Rhizobium sp. (*Medicago lupulina L.*)	1097	North Dakota/1948	
Rhizobium meliloti	1011	Maryland/1933	
Rhizobium meliloti	1021a	North Dakota/1948	
Rhizobium phaseoli	2667	Washington/1948	
Rhizobium phaseoli	2669		
Rhizobium phaseoli	2674	Brazil	
Rhizobium phaseoli	2676	Columbia/1972	
Rhizobium phaseoli	3256	Illinois/1941	
Rhizobium sp. (Robinia)	3436		
Bradyrhizobium sp. (Stylosanthes)	3441	Brazil	
Bradyrhizobium sp. (Stylosanthes)	3477	Columbia/1976	
Rhizobium trifolii	2046	Virginia/1934	
Rhizobium trifolii	2048	Illinois/1934	+
Rhizobium trifolii	2063	Florida/1939	
Rhizobium trifolii	2065	Alabama/1952	+
Rhizobium trifolii	2116	S. Carolina/1944	
Rhizobium trifolii	2134	1974	
Rhizobium trifolii	2145		
Rhizobium trifolii	2156	California/1920	
Rhizobium sp. (Trigonella)	1177	Florida/1939	
Rhizobium tropici	2744	Brazil	

Table 2.1. (continued)

Name (legume host)	USDA strain no.	geographic source/date	produce msDNA
Bradyrhizobium sp. (Vigna)	3447	Thailand/1979	+
Bradyrhizobium sp. (Vigna)	3456	Wisconsin/1966	

[a] all strains are from the USDA-Beltsville Rhizobium Culture Collection, provided by Peter van Berkum.

[b] as defined by detection of radiolabeled msDNA using the RT extension method.

Hybridization experiments and generation of probes

Chromosomal DNA from growing cells was extracted as described (Maniatis *et al.*, 1989). Purified DNA was digested with the restriction endonuclease *Pst* I. Digested DNA was subsequently electrophoresed in 0.7% agarose and transferred to nitrocellulose membranes. Nitrocellulose filters (BA-S NC Schleicher and Schuell, Keene, NH) were probed with ^{32}P-labeled msDNA under conditions of high stringency (50% Formamide-5X Denhardt's solution-5XSSPE-0.3% SDS) at 42°C. msDNA was gel-isolated from 5% acrylamide gels, and labeled with [α^{32}P] dCTP (New England Nuclear) by the random hexamer procedure (Feinberg and Volgstein, 1983).

Results and Discussion

Detection of retron elements among different bacterial groups

Retron elements were discovered by detecting the presence of msDNA using the RT extension method (Lampson *et al.*, 1989a, Lampson *et al.*, 1990). Using this method, the DNA portion of msDNA is specifically radiolabeled (^{32}P) from a total RNA preparation extracted from each bacterial strain. Twenty or more isolates of *Proteus mirabilis*, *Klebsiella pneumoniae*, *Salmonella* species, rhizobial species and enterococcal species were screened by this method. Small molecular weight bands (Figure 2.1) indicate the presence of small labeled DNAs after polyacrylamide gel electrophoresis and autoradiography of the labeling reactions. In addition, half of each labeling reaction was also treated with RNase A, causing a shift to a faster migrating band indicating that the labeled DNA is also associated with RNA, a hallmark of the msDNA molecule. msDNAs were readily detected in *P. mirabilis* strain 1174b (Figure 2.1, lanes 9, 10), *K. pneumoniae* strain 912b (lanes 3, 4), *Salmonella* sp. strain SARB-3 (not shown), *Nannocystis exedens* strain Na e1 (lanes 7, 8) and *Rhizobium trifolii* strain USDA 2065 (lanes 5, 6). Four out of 23 *P. mirabilis* isolates screened produced msDNA, while only one strain out of 21 *K. pneumoniae* isolates and one out of 34 *Salmonella* isolates screened produced msDNA. However, msDNA was not detected in any of the 30 enterococcal strains screened by this method.

The bacterial genera shown in Figure 2.1 to contain msDNA producing retron elements are representatives of three of the four major subdivisions of the purple bacteria (or *Proteobacteria*) (Woese, 1987). They are *Proteus*, *Klebsiella* and *Salmonella* of the

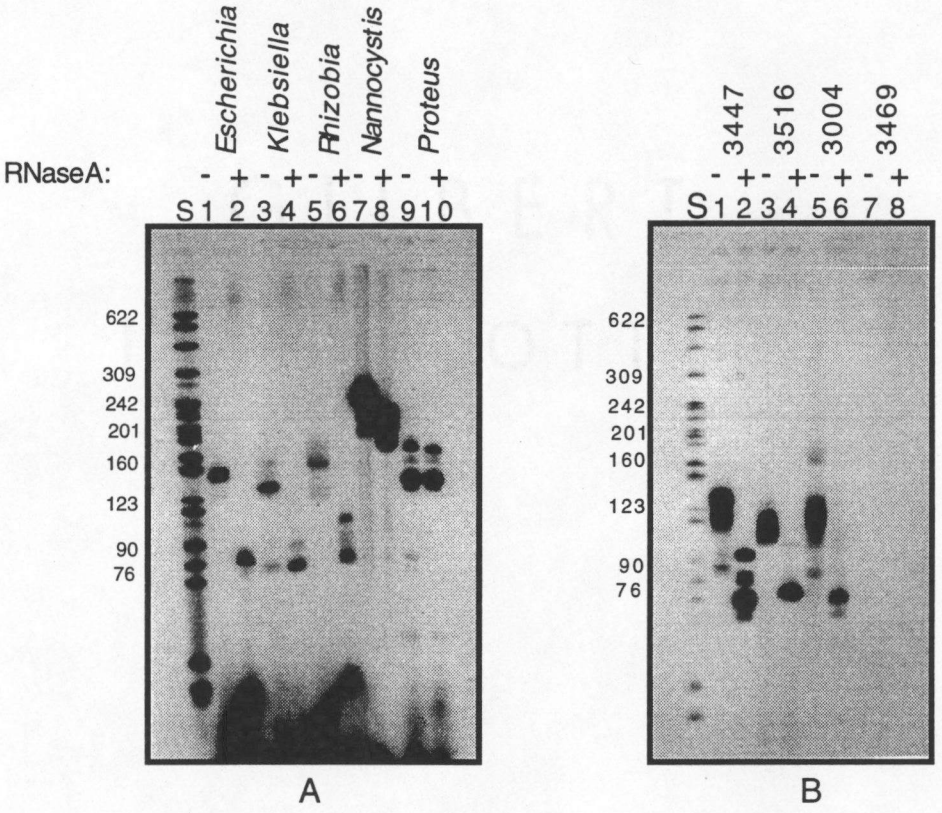


Figure 2.1. **Detection of ^{32}P-labeled msDNAs by the RT extension method**. A. msDNAs from total RNA prepared from each bacterial strain were specifically labeled with ^{32}P by the RT extension method (Lampson *et al.*, 1991, Lampson *et al.*, 1990). An aliquot from each labeling reaction mix was treated with RNase A (lanes 2, 4, 6, 8, and 10) prior to electrophoresis to detect a shift to a faster-migrating species, indicating that each labeled DNA is also associated with RNA. B. Similarly labeled msDNAs from three additional rhizobial strains. These strains include *Bradyrhizobium* spp. 3447 (lanes 1 and 2), 3516 (lanes 3 and 4), and 3004 (lanes 5 and 6). *R. loti* 3469 appears to lack an msDNA-producing element (lanes 7 and 8). Sizes are shown in base pairs.

gamma subdivision; *Rhizobium* and *Bradyrhizobium* from the alpha subdivision; and *Nannocystis* (a myxobacterium) from the delta subdivision.

All previously characterized retron elements have been found in the bacterial chromosome. Chromosomal DNA, therefore, was isolated from each of the msDNA-producing strains described above, and subjected to a Southern (DNA) hybridization experiment with all previously cloned retron genes. Except for *N. exedens*, no cross hybridization was detected with the four *E. coli* retrons (Ec67, Ec86, Ec73, and Ec107) or the two myxobacterial retrons (Mx65 and Mx162) (data not shown). A strong hybridization signal was detected between an 11-kilobase pair (kbp) *PstI* chromosomal restriction fragment from *N. exedens* and DNA sequences from the retron Mx162 (not shown). The Mx162 retron was originally discovered and cloned from another myxobacterium, *Myxococcus xanthus* (Dhundale *et al.*, 1987, Inouye *et al.*, 1989). Thus, retron elements appear to be widely prevalent, at least among the purple bacteria, occurring in three of the four major subdivisions. In addition, each group appears to have its own unique set of retron elements with little similarity among these retrons.

Prevalence of msDNAs among a population of rhizobia

The discovery of msDNA in *R. trifolii* (Figure 2.1) extends, for the first time, the distribution of retron elements to a new phylogenetic subdivision of the purple bacteria, namely, the alpha subdivision. We were interested, therefore, in examining the prevalence of these elements among a large population of rhizobia and to determine if any rhizobial retron elements are similar or related to the retrons found in *E. coli* or the myxobacteria. The detection of related retrons in the rhizobia might identify a horizontal transfer event and provide an experimental tool for making evolutionary comparisons between homologous RT genes present in distantly related phyla.

A collection of 63 rhizobia isolates, (Table 2.1) was screened for the presence of msDNA by the RT extension method. This collection (obtained from P. van Berkum, USDA, Beltville, MD) represents isolates obtained at different times, from different legume hosts, and from different geographic locations. Of 63 isolates, msDNAs were detected in only 8 (13%) strains (Table 2.1). However, all 8 positive isolates gave strong, clearly labeled bands with a typical shift to a faster migrating band after treatment with RNase A, indicating the presence of RNA and DNA in the labeled molecule (Figure 2.1B). The eight retron-encoding rhizobia strains include both fast (*Rhizobium*) and slow (*Bradyrhizobium*) growing rhizobia. Total DNA from each of the eight strains clearly cross-hybridizes with a

nod Y, A, B (1.6-kbp *Eco*RI fragment) gene probe derived from *B. japonicum* (Banfalvi *et al.*, 1988), confirming that these strains are members of the *Rhizobiaceae*.

Rhizobial retrons are unique

Southern blots were prepared of *Pst*I-digested chromosomal DNAs derived from each of the 8 rhizobia strains which produce msDNAs. The blots were screened with hybridization probes derived from retron sequences, including the RT gene, from previously cloned elements of *E. coli* and the myxobacteria. However, no cross hybridization signal was detected at either high or low stringency conditions with these probes. Likewise, chromosomal DNA dot blots containing DNA from all 63 strains of the rhizobia collection were also screened with the same hybridization probes. Here again no cross hybridization signal was detected (not shown).

Rhizobial msDNAs were also used as hybridization probes to identify related retron elements among the rhizobia. The DNA portion of the msDNA molecule was purified from total DNA extracts prepared from seven msDNA-producing rhizobia strains. Each purified msDNA was labeled with ^{32}P and hybridized against *Pst*I-digested chromosomes of the other retron-encoding rhizobia strains. For example, when the ^{32}P-labeled msDNA, isolated from strain USDA 3242, was used as a hybridization probe against a Southern blot of digested chromosomal DNAs (Figure 2.2A and 2.2B) a single *Pst*I restriction fragment was detected from the corresponding host cell chromosome (strain 3242) (Figure 2.2B, lane 3242). In addition, a single restriction fragment from another strain, USDA 3004, also cross hybridized with this particular msDNA, indicating that these two strains probably contain the same or very similar retron element (Figure 2.2B, lane 3004). These two strains, 3242 and 3004, appear to be very closely related having similar restriction fragment length polymorphism (RFLP) profiles and similar geographic source. Six other rhizobia msDNAs were similarly isolated and used for hybridization probes. From these hybridizations another pair of strains, 2065 and 2048 (Table 2.1) also appear to share a related or closely similar retron element (Table 2.2). The other remaining msDNAs hybridized to single *PstI* restriction fragments from their respective host cell chromosomal DNA only (Table 2.2). Thus out of eight msDNA-producing rhizobia strains, there appears to be six unique retron elements which show little nucleotide sequence similarity to previously characterized retrons from *E. coli* and the myxobacteria.

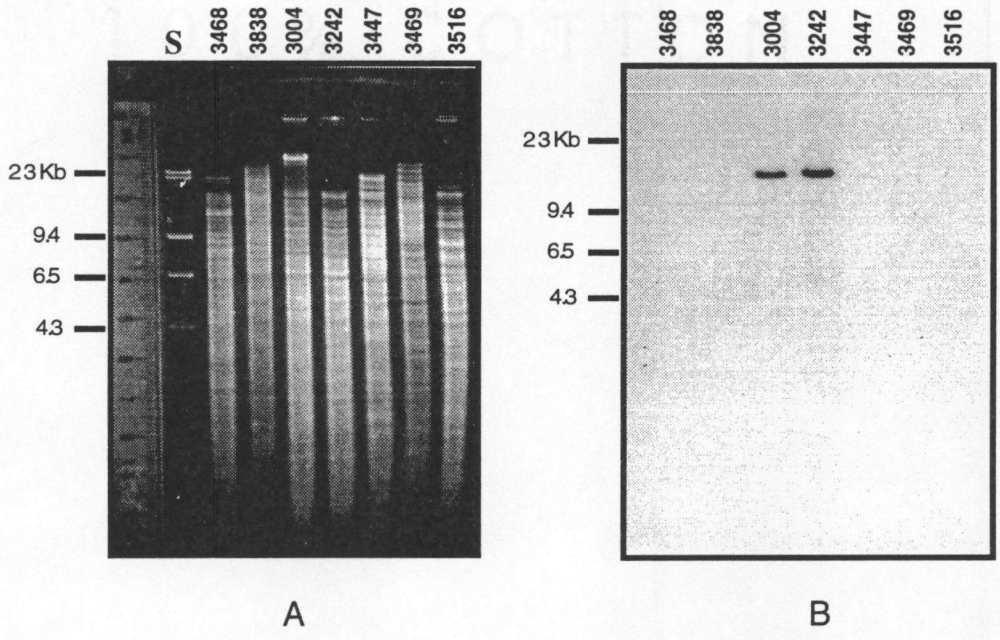


Figure 2.2. **Southern hybridizations using purified msDNA as a probe**. A. Total chromosomal DNA from 7 rhizobial strains was digested with *Pst* I, separated on a 0.7% agarose gel, and transfered to a nitrocellulose filter. B. The DNA from the gel shown in panel A was transfered to a nitrocellulose membrane and the filter was subsequently hybridized with ^{32}P labeled msDNA from *Bradyrhizobium* sp. 3242. Lane numbers correspond to the USDA strain numbers listed in Table 2.1. The lane designated S contains lambda DNA digested with *Hind* III as size standards. Sizes are given in kilobase pairs.

Table 2.2. Summary of hybridizations using rhizobial msDNAs as probes

msDNA probe[a]	Hybridization[b]	Chromosome fragment size[c] (kbp)
3447	3447	1.1 kbp
3838	3838	18.5 kbp
3242	3242, 3004	14 kbp
2065	2065, 2048	4.3 kbp
3516	3516	8.1 kbp
3004	3004, 3242	14 kbp
3468	3468	2.6 kbp
2048	ND[d]	ND

[a] msDNAs are identified here by the bacterial host strain (USDA) number from which the molecule was originally isolated.

[b] chromosomal DNAs which hybridized with the indicated msDNA probe are identified by their respective bacterial host strain (USDA) number.

[c] size in kilobase pairs of the *PstI*, chromosomal restriction fragment which hybridizes with the indicated msDNA probe.

[d] ND, not determined.

Proteus mirabilis 1174b msDNA is atypical

One of the characteristics of msDNA, as detected by the RT method discussed above, is that RNase A-treatment of the RT-extended msDNA, yields a product that migrates slightly faster than the sample not treated with RNase A (e.g. Figure 2.1, lanes 2 and 3). This is due to the presence of a 5' RNA arm associated with the untreated sample, which is RNase A sensitive (see Figure 2.2). However, RNase A treatment of the *P. mirabilis* 1174b msDNA did not produce a faster migrating species (Figure 2.1, lanes 9 and 10). There are several possible explanations of this result. Either the labeled product from 1174b is not a *bona fide* msDNA, or the msDNA does not have a long enough 5' RNA arm to yield a noticeably faster migrating species after RNase A treatment, or there is not a 5' RNA arm associated with the 1174b msDNA. RT labeling of msDNA is dependent on the reverse transcriptase using the RNA strand as a template. Pretreatment of the 1174b sample with RNase A prior to the RT extension reaction blocks production of ^{32}P labeled bands (data not shown), and suggests that ability to ^{32}P label the msDNA from 1174b is dependent on an RNA template.

To determine the presence of the 5' RNA arm, msDNA was collected, purified, and 5' end labeled with T4 Polynucleotide Kinase and [$\gamma\,^{32}P$] ATP. The purified labeled product was treated with 0.5 N NaOH at 90°C. If the branched guanine containing the 2'-5' linkage with msDNA is present, the ^{32}P label will be at a 5' triribonucleotide which is alkali sensitive. However, if this 2'-5' linkage has been lost, then the 5' labeled nucleotide will be a deoxyribonucleotide and will be alkali resistant. Figure 2.3 shows the effect of alkali treatment of 5' end labeled msDNA. The msDNA Ec67 has been shown to contain the 2'-5' linkage and residual 5' RNA arm after RT labeling. Alkali treatment at 90°C for increasing duration removes the ^{32}P label (Figure 2.3, lanes 1-4), while a similarly 5' end labeled deoyxribonucleotide (67mer) (Figure 2.3, lanes 10-13, was alkali resistant. The labeled 1174b msDNA was alkali resistant (Figure 2.3, lanes 5-9), indicating the msDNA from *P. mirabilis* 1174b was labeled at a 5' deoxyribonucleotide. This suggested that the msDNA from 1174b did not retain a 2'-5' phosphodiester linkage characteristic of most msDNAs. Recently, "unbranched" forms of msDNA in *E. coli* have been reported which similarly lack the 2'-5' phosphodiester bond (Lim, 1992), which possibly occurs by processing of the msDNA. These forms of msDNA have 5' deoxyribonucleotides, but lack an associated RNA component and therefore cannot be ^{32}P labeled by the RT extension method.

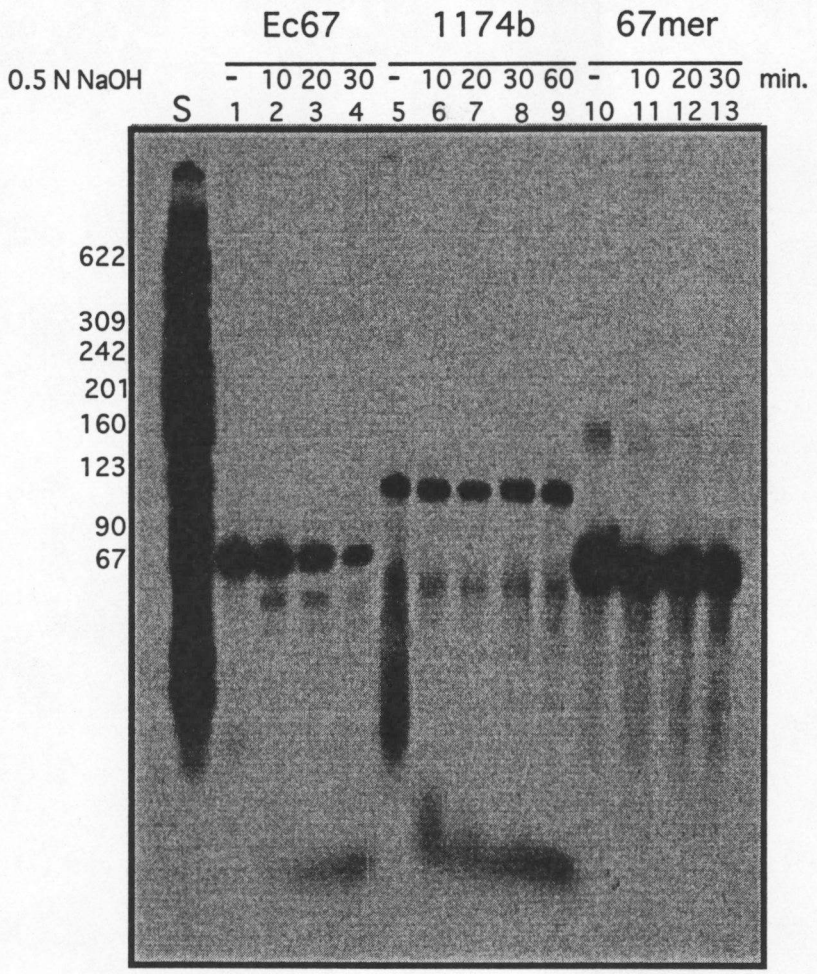


Figure 2.3. **Alkali treatment of ^{32}P end-labeled 1174b msDNA.** msDNA from *E. coli* CL-1 (Ec67), 1174b, and a 67 base oligonucleotide (67mer), were end labeled with g-^{32}P to look for release of label in the presence of 0.5 N NaOH at 90°C. Untreated samples (lanes 1, 5, and 10) show a single prominant labeled band. Labeled DNAs were exposed to 0.5 N NaOH for 10 (lanes 2, 6, and 11), 20 (lanes 3, 7, and 12), 30 (lanes 4, 8, and 13), or 60 (lane 9) minutes. Plasmid pBR322 digested with *Msp* I (S), was used as a moleular weight marker, with sizes listed in base pairs.

The msDNA from 1174b was end labeled as described and sequenced directly by the chemical cleavage method (Maxam and Gilbert, 1980) (Figure 2.4). The msDNA consisted of 103 nucleotides, and can base pair to form a stable central stem of 40 bases (Figure 2.5). Therefore, based on the data presented here, and the report of "unbranched" forms of msDNA, the msDNA found in *P. mirabilis* 1174b was concluded to be an msDNA having a unique unbranched structure.

Attempts were made to clone the chromosomal retron locus from *P. mirabilis* 1174b. Chromosomal DNA from *P. mirabilis* 1174b was digested with *Pst* I, and the resulting DNA fragments were size fractionated by electrophoresis in 0.7% agarose, blotted to nitrocellulose, and hybridized with a random hexamer labeled msDNA probe (data not shown). The chromosomal fraction that hybridized to the msDNA probe was cloned into the *Pst* I site of pUC9 (Yanisch-Perron *et al.*, 1985). To identify retron encoding clones, colony hybridizations were performed using the msDNA probe. Unexpectedly, almost every colony probed positive, yet analysis of the plasmids indicated that none had an insert that would hybridize with the 1174b probe. A homology search using the BLAST algorithm (Altschul *et al.*, 1993) indicated that nucleotides 22-50 of the 1174b msDNA showed significant homology to the β-lactamase gene in pUC9 (Figure 2.6), and thus could account for the false-positive hybridization signals observed. No further attempts were made to clone the retron from 1174b.

Why are retrons so rare?

The infrequent occurrence and heterogeneous nature of these retrons, observed among this presumably diverse collection of rhizobia, is reminiscent of the retrons found in populations of *E. coli*. A survey conducted on an *E. coli* reference collection of 72 strains (the ECOR collection) similarly showed about 13% of these strains contained a retron (Herzer *et al.*, 1990). And among these positive strains, there appears to be four distinct types of retron elements.

What is the explanation for the existence of these highly diverse genetic elements in only a few selected strains of the *Rhizobiaceae*? Like *E. coli* populations, selective lineages of rhizobia may have independently acquired or "picked up" a retron element from some unknown source. This transfer of retrons into a few genomes of the rhizobia probably occurred relatively recently with subsequent horizontal exchange of some retrons possibly occurring between strains of rhizobia as well. Indeed, studies of *E. coli* retrons provide some evidence for this idea. For example, the amino acid codon usage for RT genes found

```
              *10          *20          *30          *40          *50
5' GAATGCGCTT  TAGACTTTTA  GACGGCTCCT  GAATGGCACA  AACCTGCGCA

   GTGAAATTCA  GGATCTGGCT  AAAAGTGCCA  AAGCGCAACC  CTAAATGTCN  NNA  3'
              *60          *70          *80          *90          *99
```

Figure 2.4. **DNA sequence of the msDNA from *Proteus mirabilis* 1174b.**
The DNA component of purified msDNA from *P. mirabilis* 1174b was ^{32}P -end labeled and directly sequenced by the method of Maxam and Gilbert (1980)

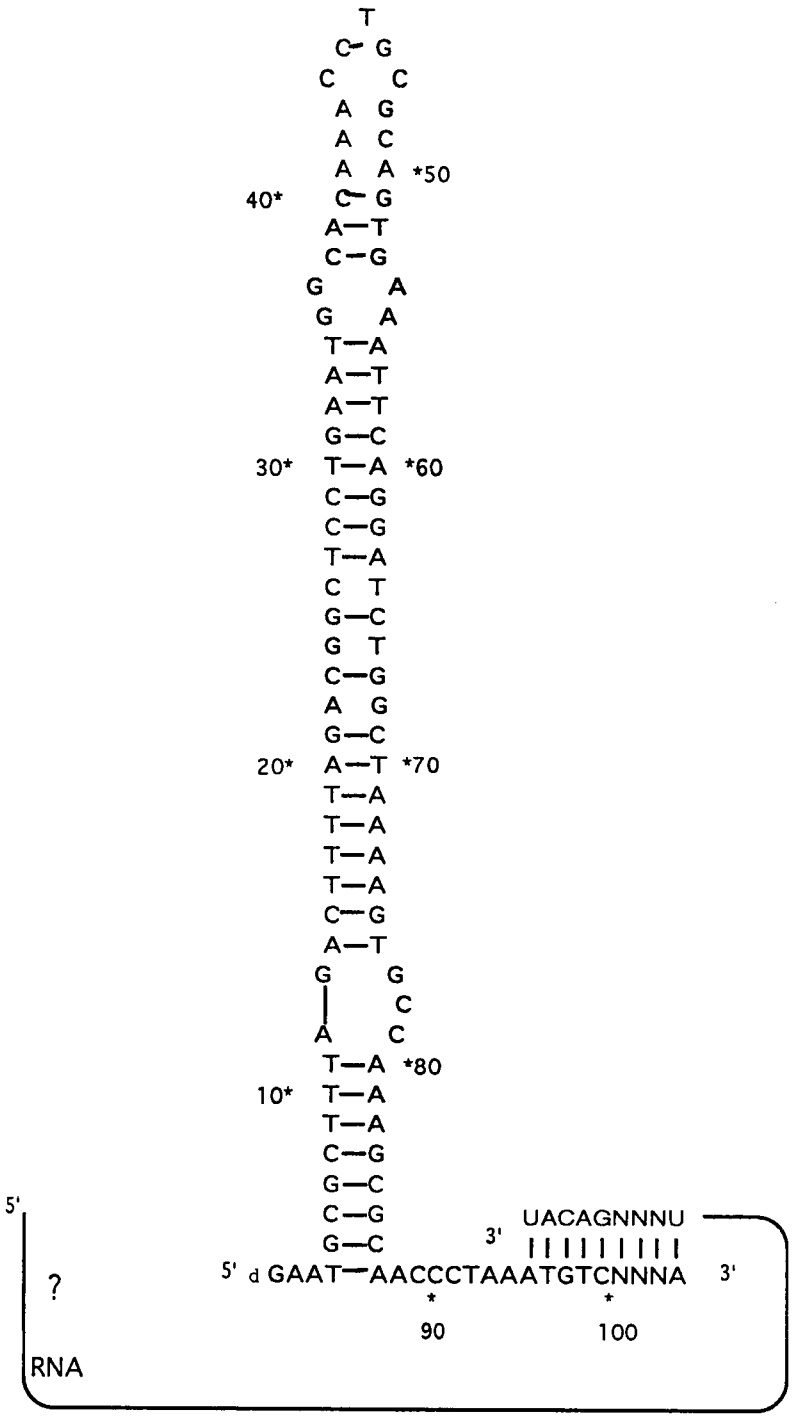


Figure 2.5. **The proposed secondary structure of the DNA stem-loop from *P. mirabilis* 1174b.** Note that the 2'-5' phosphodiester linkage is left open to indicate this msDNA may be unbranched.

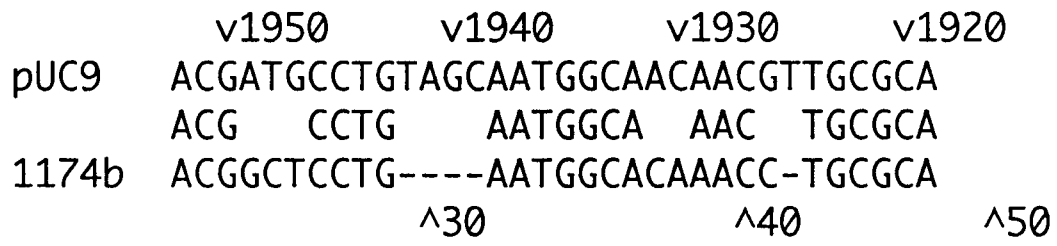


Figure 2.6. **DNA sequence alignment of msDNA from *P. mirabilis* 1174b with the plasmid pUC9.** Using the BLAST protocol (Altschul *et al.*, 1993), significant nucleotide identity was found to exist between msDNA from *P. mirabilis* 1174b (1174b) and the β-lactamase gene in the plasmid pUC9.

in *E. coli* retrons is clearly different from most protein encoding genes of the host chromosome indicating that retrons are foreign to *E. coli* (Inouye *et al.*, 1989). Likewise, the lack of nucleotide sequence diversity among several individual (homologous) retrons of the Ec107 type, which are found in several different phylogenetic strains of the ECOR collection, indicates a recent exchange or dissemination of this retron among several *E. coli* strains (Kawaguchi *et al.*, 1992).

Features of retron elements described here among the rhizobia and previously for *E. coli* are suggestive of a mobile element. However, this remains to be demonstrated directly in the laboratory. If retrons are mobile, perhaps by a retrotransposition mechanism, this appears to be a rare event among natural populations of rhizobia and *E. coli* due to their infrequent occurrence. Contrast this with a survey of the ECOR collection for insertion sequences (IS), which found that almost every strain contains multiple copies of one or more IS elements (Sawyer *et al.*, 1987).

As revealed for the rhizobia in this paper and previously for *E. coli*, retron elements appear to be a recent guest of the bacterial genome. But where did these retrons come from and what is the evolutionary relationship among the different types of elements? Further study of the prevalence and kinship of retron elements among bacterial groups like the rhizobia and the myxobacteria may provide some clue, as well as clarify questions about the origin of RT and retroelements in general.

Acknowledgments

The authors wish to thank Peter van Berkum (USDA, Beltsville, MD) for the rhizobia collection, Noel Smith and Robert Selander (Penn. State University) for the *Salmonella* collection, David Bemis (U. of Tennessee, Knoxville) for the other enteric bacterial strains, and Lawrence Shimkets (U. of Georgia, Athens) for *N. exedens* strain Na e1.

Chapter III

Phylogenetic Comparison of Retron Elements Among the Myxobacteria: Evidence for Vertical Inheritance

Introduction

In 1984, Yee *et al.*, (1984) discovered a small, high copy number, satellite DNA associated with the gram-negative soil bacterium *Myxococcus xanthus*. This DNA consisted of a single strand of DNA covalently linked by a 2'-5' phosphodiester bond to an internal guanine residue of a single-stranded RNA (Dhundale *et al.*, 1987, Furuchi *et al.*, 1987a). These molecules, termed multicopy, single-stranded DNA (msDNA), are encoded by a chromosomal locus called a retron (Temin, 1989). The original retron, found in *M. xanthus*, is designated Mx162 (Inouye *et al.*, 1989, Yee *et al.*, 1984), and a second unique retron element, designated Mx65, coresides with Mx162 on the chromosome of *M. xanthus* (Dhundale *et al.*, 1988b). In addition to coding for the RNA and DNA regions of msDNA, the chromosomal locus also contains an open reading frame (ORF) that shows significant similarity to retroviral reverse transcriptases (RT) (Inouye *et al.*, 1989). These genes associated with msDNA were the first prokaryotic RTs to be discovered (Inouye *et al.*, 1989, Lampson *et al.*, 1989b, Lim and Maas, 1989). Other types of RT also appear to exist in prokaryotes such as the newly discovered intron encoded RTs (Ferat and Michel, 1993). Synthesis of msDNA is postulated to occur by reverse transcription and is dependent on the production of a functional RT (Inouye *et al.*, 1989, Lampson *et al.* 1989a, Lim and Maas, 1989). Deletion of the retrons in *M. xanthus* does not affect growth or development of the bacterium (Dhundale *et al.*, 1988a, Inouye *et al.*, 1989). Additionally, since the elements are not ubiquitous among bacterial strains, it is apparent that the elements are not required for survival, and despite the widespread nature of retrons, no function has been ascribed to them.

In addition to *M. xanthus* and the myxobacterium *Stigmatella aurantiaca* (Furuichi *et al.*, 1987), retrons have been reported in 9-15% of all *Escherichia coli* strains tested (Herzer *et al.*, 1990, Lampson *et al.*, 1989b, Lim *et al.*, 1990, Sun *et al.*, 1991). Recently, we reported the presence of retron elements in several distantly related bacterial groups suggesting that msDNA and RT genes are widely prevalent in the prokaryotes (Rice *et al.*,

1993). Like *E. coli*, only a small percentage (5-15%) of strains of *Klebsiella pneumoniae*, *Proteus mirabilis*, *Salmonella sp.*, *Rhizobium sp.*, and *Bradyrhizobium sp.* tested contain a retron (Rice *et al.*, 1993).

A number of characteristics distinguish the retron elements of the myxobacteria from retrons found in the rest of the proteobacteria. First, unlike other bacterial groups where retrons occur rarely, the elements appear to be nearly ubiquitous among natural *M. xanthus* strains (Lampson *et al.*, 1991a). Second, DNA sequence analysis of retrons from *M. xanthus* and *S. aurantiaca* indicates the codon bias of the RT gene is similar to the codon bias observed for other known myxobacterial genes (Inouye *et al.*, 1989). These features aroused speculation that bacterial RTs of the myxobacteria are "ancient elements" and that they evolved prior to the emergence of RT and retroelements in eukaryotes (Inouye and Inouye, 1992). These intriguing evolutionary questions prompted a closer examination of the retroelements of the myxobacteria.

This report presents the findings of an extensive survey of the retron elements found among the myxobacteria. All but one of the 28 strains examined contained an msDNA producing element. In addition, several different myxobacteria groups were found to contain a retron element similar to Mx162; a retron previously characterized from *M. xanthus*. One of the retron elements, from the myxobacterium *Melittangium lichenicola*, was cloned and its DNA sequence compared with similar elements in *M. xanthus* and *S. aurantiaca*. Together, the degree of sequence diversity, the codon bias of the RT genes, and the distribution of these retrons suggests a possible evolutionary scenario in which a common ancestor of the *Myxococcus* subgroup probably acquired this retroelement.

Materials and Methods

Bacterial strains and plasmids

The myxobacteria strains used for this study and their sources are listed in Table 3.1. *E. coli* JM109 and *E. coli* DH5α were used for cloning in conjunction with plasmids pBR322 (Bolivar *et al.*, 1977) and pUC19 (38 Yannish-Perron *et al.*, 1985). *M. xanthus* DZF1 was grown at 30°C in CYE broth (Reichenbach and Dworkin, 1992). *Chondromyces* and *Sorangium* strains were grown on VY/2 plates (0.5% bakers' yeast; 0.1% $CaCl_2 \cdot 2H_2O$; 1.5% agar) at 30°C (Reichenbach and Dworkin, 1992). All other myxobacteria strains were grown at 30°C on MD1 plates (0.3% Difco Casitone; 0.07% $CaCl_2 \cdot 2H_2O$; 0.2% $MgSO_4 \cdot 7H_2O$; trace elements; 1.5% agar) (Reichenbach and

Table 3.1. Survey of the myxobacteria for retron elements

Organism screened	msDNA production	hybridizations Mx162	Mx65	Na e434	reference	location/strain source
Subgroup						
Myxococcus						
Angiococcus disciformis						
ATCC 33172	+	+	-	-	this study	
Archangium gephyra						
ATCC 25201	+	+	-	-	this study	
Corallococcus						
(*Myxococcus*) *coralloides*						
DK817	+[a]	+[b]	ND	ND	Dhundale *et al.*, 1988a	Yosemite, CA/ Dale Kaiser
ATCC 25202	+	+	-	-	this study	
Cystobacter ferrugineus						
Cb fe17	+	+	ND	ND	this study	/Hans Reichenbach
Cb fe18	-	-	-	-	this study	/Hans Reichenbach
Cystobacter fuscus						
Cb f17	+	+	ND	ND	this study	/Hans Reichenbach
Cystobacter violaceus						
Cb vi4	+	+	ND	ND	this study	/Hans Reichenbach
Melittangium lichenicola						
ATCC 25946	+	+	-	-	this study	/Hans Reichenbach
Myxococcus fulvus						
ATCC 25199	+	+	-	-	this study	
Myxococcus macrosporus						
ATCC 29619	+	+	+	-	this study	
Myxococcus stipitatus						
ATCC 29611	+	+	-	-	this study	
Myxococcus virescens						
ATCC 25203	+	+	+	-	this study	
Myxococcus xanthus						
DZF1	+	+	+	-	Lampson *et al.* 1991a	David Zusman
19 other strains	+	+	+/-	-	Lampson *et al.* 1991a	Hans Reichenbach and Dale Kaiser
Stigmatella aurantiaca						
DW4	+	+	-	-	Lampson *et al.* 1991a	Minnieapolis, MN/ David White

Table 3.1. (continued)

Organism screened	msDNA production	hybridizations Mx162	Mx65	Na e434	reference	location/strain source
Subgroup						
Chondromyces						
Chondromyces apiculatus						
Cm a15	+	-	-	-	this study	
Cm a16	+	-	-	-	this study	
*Chondromyces pedicalatus*Cm p5	+	-	-	-	this study	Montes Claros, Brazil/ Hans Reichenbach
Sorangium cellulosum						
So ce5	+	-	-	-	this study	Cefalù, Sicily / Hans Reichenbach
So ce7	+	-	-	-	this study	Delphi, Greece/ Hans Reichenbach
So ce11	+	-	-	-	this study	Yucatan, Mexico/ Hans Reichenbach
Subgroup						
Nannocystis						
Nannocystis exedens						
Na e1	+	-	-	+	this study	Hans Reichenbach
Na e3	+	-	-	+	this study	Nebraska, USA/ Hans Reichenbach
Na e24	+	-	-	+	this study	Teneriffe, Spain/ Hans Reichenbach
Na e30	+	-	-	+	this study	Calpe, Spain/ Hans Reichenbach
Na e39	+	-	-	+	this study	Cefalù, Sicily / Hans Reichenbach
Na e434	+	-	-	+	this study	Jerusalem, Israel/ Hans Reichenbach
Na e465	+	-	-	+	this study	Hans Reichenbach
Na e585	+	-	-	+	this study	Pissouri, Cyprus/ Hans Reichenbach
Na e619	+	-	-	+	this study	Mallorca, Spain/ Hans Reichenbach
Na e642	+	-	-	+	this study	Bornholm, Denmark/ Hans Reichenbach

[a] screened by ethidium bromide staining
[b] probed using *msd* region only for probe

Dworkin, 1992). *E. coli* was grown at 37°C in modified Luria-Bertani medium (Maniatis *et al.*, 1989). When necessary, tetracycline (12 μg/ml) or ampicillin (50 μg/ml) was added for selection of plasmids.

Detection of msDNA

Total RNA was purified by the guanidinium thiocyanate method (Chomcyzynski and Sacchi, 1987) from freshly grown cells or frozen cell pellets. Purified RNA (this fraction also contains msDNA) was stored in ethanol at -80°C. msDNA was detected by the RT extension method (Lampson *et al.*, 1991a, Lampson *et al.*, 1990, Rice *et al.*, 1993). Labeled products were divided in half, with RNase A treatment of one set of samples, and separated on 4% Acrylamide-8M Urea gels. Gels were dried and exposed to x-ray film at -80°C.

Detection of homologous retron elements

Chromosomal DNA was isolated from each myxobacteria strain and digested with *Pst*I restriction endonuclease as described (Rice *et al.*, 1993). Chromosomal DNA digests were separated on 0.7% agarose gels and transferred to nitrocellulose filters. Southern hybridizations were carried out at high stringency (50% formamide-5X SSPE-5X Denhardt's-0.3% SDS) at 42°C. The Mx162 probe was prepared as a 1.3 kilobase pair (kbp) *Xho*I-*Rsa*I subfragment from plasmid pmsSB7 (Rice *et al.*, 1993). This fragment contains part of the *msd* gene, which codes for the DNA portion of msDNA, and about two-thirds of the N-terminal region of the RT-ORF. The Mx65 probe was prepared as a 1.2-kbp *Eco*RI-*Bam*HI fragment from plasmid pBPV-1 (Dhundale *et al.*, 1988b, Lampson *et al.*, 1991a). Gel purified restriction fragments were labeled for probes with [α-^{32}P] dCTP using the random hexamer labeling technique (Feinberg and Volgstein, 1983), and purified using Sephadex G-50 columns. msDNA from *Nannocystis exedens* strain Na e434 was prepared for use as probe as follows. msDNA was specifically labeled with ^{32}P by the RT-extension method, RNase A treated, then isolated by electrophoresis on a 4% acrylamide-8M urea gel. The radiolabeled msDNA was cut out of the gel and recovered by elution.

Cloning of a retron homologous to Mx162

Melittangium lichenicola chromosomal DNA was digested with *Pst*I and separated by electrophoresis on a 0.7% agarose gel. *Pst*I fragments in the range of 9-kbp to 11-kbp

were recovered, purified by electroelution, and ligated into the dephosphorylated *Pst*I site of pBR322. Tetracycline resistant, ampicillin sensitive colonies were screened by colony hybridization (Maniatis *et al.*, 1989) using a 0.75-kbp *Eco*RI-*Hind*III fragment of retron Mx162 as a probe. This probe hybridized to plasmid pNaeF4.15, containing a 10.8-kbp fragment. From this positive clone, a 3.8-kbp *Bam*HI fragment was sub-cloned into a dephosphorylated *Bam*HI site in pUC19 (designated pNaeBB4.0.A4). Overlapping fragments of pNaeBB4.0.A4 were subcloned into pUC19 for sequencing.

DNA Sequencing

DNA sequencing of pNaeBB4.0.A4 subclones was performed by the dideoxy chain termination method of Sanger *et al.* (1977) using ΔTaq Version 2.0 sequencing kit from USB (United States Biochemical, Cleveland, OH). Overlapping clones or synthetic primers were used to ensure that sequencing runs extended through all cloning sites. The DNA sequence was determined for both strands. Double-stranded templates were prepared by alkali denaturation as described in the sequencing kit (USB). The nucleotide sequence of the DNA portion of the msDNA was determined directly by the method of Maxam and Gilbert (1980). msDNA was isolated from freshly grown cells, treated with RNase A, gel purified on a 5% acrylamide gel and labeled at the 5' end with [γ-^{32}P] ATP using T4 Polynucleotide Kinase (New England Biolabs, Beverly, MA). Electrophoresis of sequencing reactions was performed using 6% (wt/vol) acrylamide-8M urea gels. Sequences were analyzed using DNA inspector (version IIe), the Genbank sequence database (version 7.3), and the Genetics Computer Group (GCG) of the University of Wisconsin (Devereux *et al.*, 1984).

Silent substitutions and codon bias

The degree of nucleotide sequence variation between two homologous genes is the result of the accumulation of nucleotide substitutions over a period of time since the two genes diverged from a common ancestor. For most genes, changes tend to occur at a uniform frequency or in a "clock-like" fashion (Ochman and Wilson, 1987, Wilson *et al.*, 1977). One measure of gene divergence is the number of changes per unit time or the rate of change for that gene. To calculate this rate, two types of mutations are considered within protein coding regions, synonymous and non-synonymous substitutions. A non-synonymous substitution is one that changes a codon and subsequently the encoded amino acid. Synonymous substitutions, or silent substitutions, change a particular codon, but do

not change the encoded amino acid; this is the result of degeneracy of the genetic code. While it has been proposed that mutations in bacteria occur at a regular, clock-like rate, mutations do not accumulate at the same rate for the two types of sites (Li *et al.*, 1985). Non-synonymous substitutions may have detrimental effects, and thus be eliminated from the observed population. However, since synonymous substitutions do not alter the amino acid sequence, these substitutions are presumed to be selectively neutral, and accumulate in the population over time at relatively constant rates. The percentage of synonymous substitutions is determined by dividing the number of potential synonymous sites by the number of changes at those sites. Calculation of percent synonymous base substitutions was based on the method of Li *et al.* (1985). The rate is determined by dividing percent substitution by the evolutionary distance (time) between the two species, based on 16S rRNA.

Nucleotide sequence accession number

The sequence for retron ML162 was deposited into the Genbank database under accession number L36722.

Results

Retron elements in the Myxobacteria

Based on 16S rRNA sequence comparison, the myxobacteria, that is the taxonomic order *Myxococcales*, fall into three major subgroups: *Myxococcus*, *Chondromyces*, and *Nannocystis* (Shimkets and Woese, 1992). In this study, 28 strains of myxobacteria, with representatives from each of the three major subgroups, were screened for the presence of retron elements by looking for the production of msDNA. The RT extension method was used to detect msDNA. Briefly, this technique uses Moloney Murine Leukemia Virus reverse transcriptase (M-MLV-RT) and [α-^{32}P] dCTP to specifically radiolabel msDNA by extending the DNA strand, using the RNA strand as a template (Lampson *et al.*, 1991a, Lampson *et al.*, 1990). Extension continues to the branch-guanine residue where reverse transcription stops, leaving an RNA arm at the 5' end. Subsequent treatment of the extended and labeled product with RNase A removes the 5' RNA arm, producing a faster migrating species compared to the untreated sample. Figure 3.1 presents typical results from an RT extension assay. Strains So ce5, So ce7, So ce11, and Cm a16 (Table 3.1) all produced distinctly labeled bands in the size range of 145 to 245 nucleotide bases (Figure

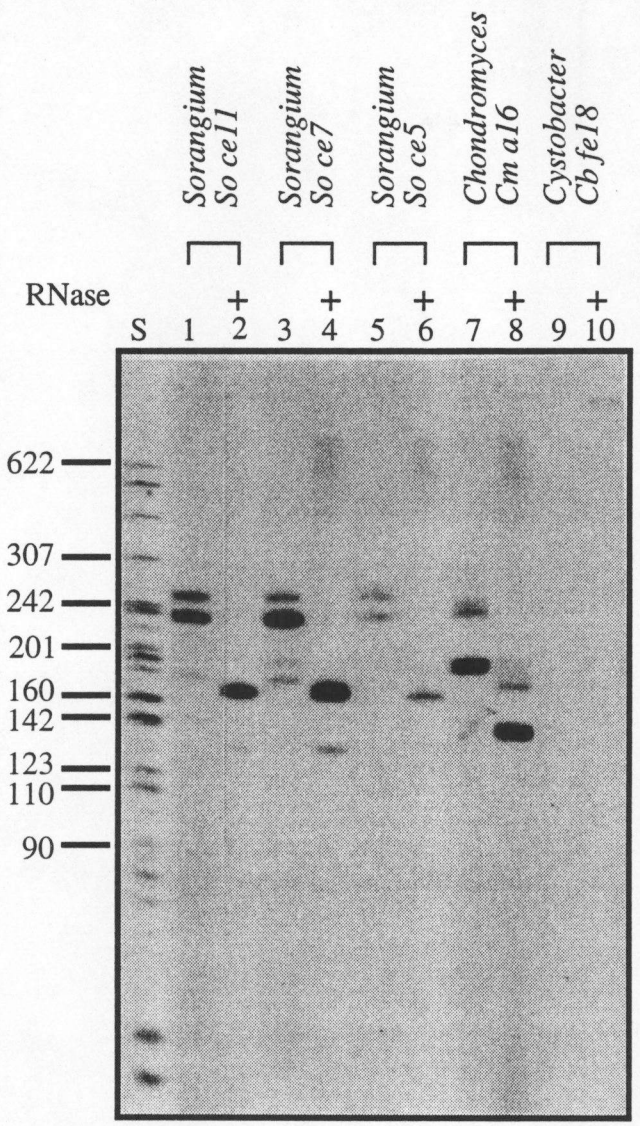


Figure 3.1. **Detection of ^{32}P-labeled msDNA by the RT extension assay.** The molecular weight standard (S) is radiolabeled pBR322 digested with *Msp*I, and sizes are given in base pairs. Strains are listed by genus and strain designation. Half of each labeling reaction was treated with RNase A, designated with a (+), prior to electrophoresis. Treated and untreated samples were eletrophoresed in adjacent lanes. Note that RNase A-treated msDNAs (radiolabeled bands) migrate faster (lanes 2, 4, 6, and 8) than the untreated samples (lanes 1, 3, 5, and 7). Strains So ce11, So ce7, So ce5, and Cm a16 are positive for the presence of msDNA, while Cb fe18 produced no labeled band and thus is considered to be negative for msDNA production.

3.1, lanes 1-8). The presence of msDNA indicates the presence of the corresponding chromosomally encoded gene, the retron. The shift in mobility when the extended products were treated with RNase A (Figure 3.1, lanes 2, 4, 6, and, 8) is characteristic of msDNA, and the size difference reflects the length of the 5' RNA arm. The double bands present in some strains represent either the presence of more than one msDNA, or heterogeneity of extension by the RT. No labeled bands were present for strain Cb fe18 (Figure 3.1, lanes 9 and 10), indicating that retron elements are absent from this *Cystobacter* strain. We screened several species of myxobacteria which have not been previously studied for the presence of retron elements, including: 3 *Chondromyces*, 4 *Cystobacter*, 3 *Sorangium*, 1 *Melittangium lichenicola*, 1 *Myxococcus virescens*, 1 *Myxococcus fulvus*, 1 *Myxococcus macrosporus*, 1 *Myxococcus stipitatus*, 1 *Angiococcus disciformis*, 1 *Archangium gephyra*, 1 *Corallococcus* (*Myxococcus*) *coralloides*, and 10 *Nannocystis exedens* strains. The positive and negative strains as determined by the RT extension method are summarized in Table 3.1. All of the strains tested positive for the production of msDNA except for *Cystobacter ferrugineus* Cb fe18. It appears that strains Cm p5, Cm a15, and Cm a16, contain two different msDNAs while So ce5, So ce7, and So ce11 contain only a single msDNA (data not shown). In addition, we also screened 10 *N. exedens* strains using the RT extension assay. All ten strains tested positive (data not shown), with the msDNAs ranging in size from 120 to 180 nucleotides. Based on the different sizes of msDNAs in the *Nannocystis* strains, it is apparent that some strains may also produce more than one species of msDNA. With the exception of one *Cystobacter* strain, retrons are ubiquitous among the myxobacteria reported here (Table 3.1).

Homology to known retrons

Identification of retron elements similar to the two elements, Mx65 and Mx162, previously described from *M. xanthus*, was accomplished by DNA hybridization. Radiolabeled fragments derived from plasmid clones of these retrons (Dhundale *et al.*, 1985, Dhundale *et al.*, 1988b, Dhundale *et al.*, 1987) were hybridized to *Pst*I digested chromosomal DNA from each myxobacteria strain. In addition, msDNA isolated from *N. exedens* strain Na e434 was radiolabeled and also used as a hybridization probe.

All of the strains in the *Myxococcus* subgroup cross hybridized with the Mx162 probe, indicating that these species share retrons that are similar in primary nucleotide sequence to Mx162, (Figure 3.2, lanes 1-8). In contrast, the Mx65 based probe cross hybridized with only three species of the genus *Myxococcus* (data not shown) (Table 3.1);

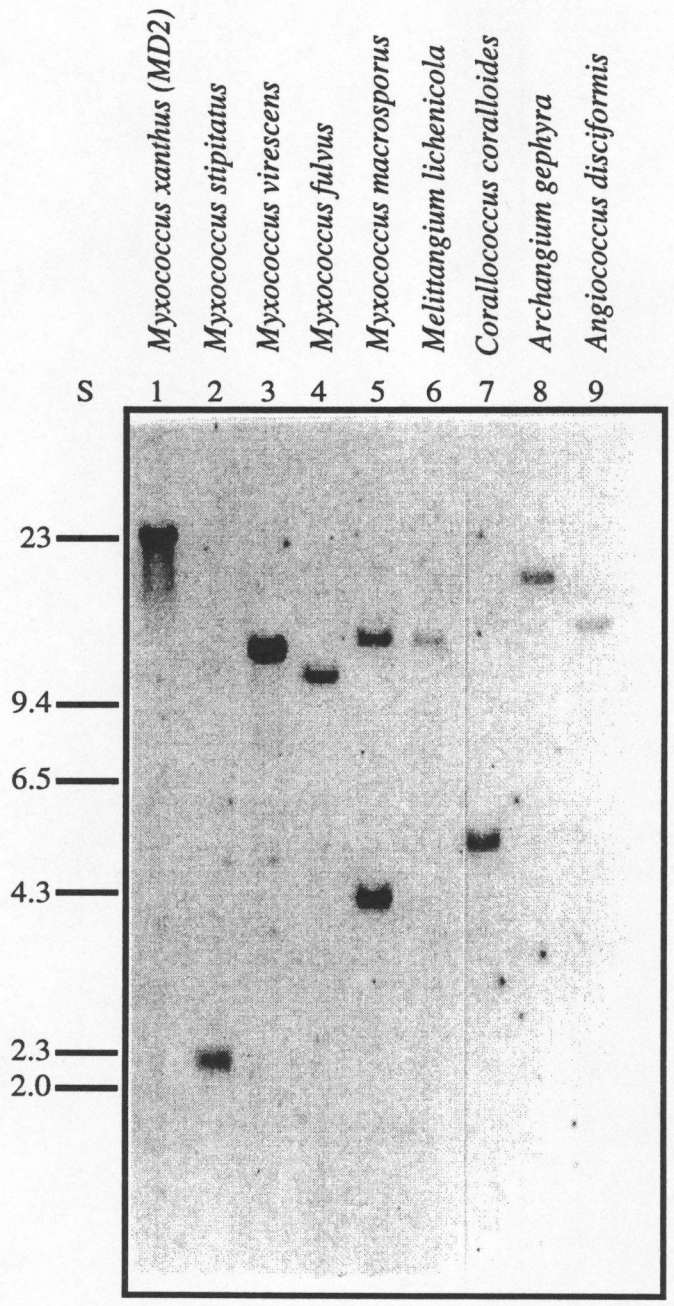


Figure 3.2. **Southern hybridization of nine members of the *Myxococcus* subgroup with labeled Mx162 probe.** Chromosomes were digested with *Pst*I, except for *C. coralloides*, which was digested with *Bam*HI. Digested chromosomes were electrophoresed on a 0.7% agarose gel, transfered to nitrocellulose, and hybridized with a ^{32}P-labeled internal fragment of the RTgene from Mx162. Molecular weight standards (S) are given in kilobase pairs. All strains tested reveal a single band which hybridized with the Mx162 probe, except *M. macrosporus*,which shows two distinct bands. Weak, but clearly visible signals appear for *M. lichenicola*, *A. gephyra*, and *A. disciformis* (lanes 6, 8, and 9).

M. xanthus, *M. virescens*, and *M. macrosporus*. Additionally, chromosomal DNA from strains in the subgroups *Chondromyces* and *Nannocystis* did not hybridize with either the Mx162 or Mx65 probes, indicating that these retrons are restricted to the *Myxococcus* subgroup (Table 3.1). Figure 3.3 presents a phylogenetic tree of the myxobacteria showing the clustered distribution of retron elements similar to Mx162 among the seven genera that make up the *Myxococcus* subgroup. Purified msDNA, from *Nannocystis exedens* Na e434, was also used as a hybridization probe. While the *Myxococcus*, *Chondromyces*, and *Sorangium* strains did not hybridize to this probe, all ten strains of *Nannocystis* surveyed hybridized to the Na e434 msDNA probe (Table 3.1). This suggests that members of the *Nannocystis* subgroup likely contain the same, or a very similar, retron.

An msDNA, hybridization probe was also prepared from a member of the *Chondromyces* subgroup (*Sorangium* strain So ce11). However, due to difficulty in preparing this probe, only a small number of myxobacteria strains were screened with this probe (not shown). Nevertheless, chromosomal DNA from strains So ce7, So ce11 and Cm p5 (*Chondromyces*) hybridized strongly to the msDNA probe from So ce11. DNA from *Nannocystis* Na e1, and *M. xanthus* DZF1 gave a weak hybridization signal, and DNA from *Nannocystis* Na e434 and *Melittangium lichenicola* gave no hybridization signal (not shown).

Cloning and sequencing of a *Melittangium* retron

Because the chromosome digest from *Melittangium lichenicola* gave a particularly weak hybridization signal with the Mx162 probe (Figure 3.2, lane 6), we were interested in cloning and sequencing the retron from *M. lichenicola* to compare sequence diversity with similar retrons found in two other genera within the same myxobacteria subgroup. The objective was to be able to make some determination about the evolutionary history of the retron; is it an ancient element, or was it recently introduced into these three species.

Maxam and Gilbert (1980) chemical sequencing of the DNA portion of the msDNA from *M. lichenicola* revealed the DNA strand to have 162 bases and has been designated ML162. ML162 msDNA showed significant nucleotide identity to Mx162 in the DNA region, with only 17 base differences, 10 of which are found in the central stem region which forms a stable hair-pin structure due to complementary secondary folding of the DNA strand (Figure 3.4). The majority of these base differences are complementary mutations on opposite sides of the stem that help to maintain the stem-loop structure. In

Figure 3.3. **Phylogenetic distribution of msDNA and Mx162 homologues in the myxobacteria.** The phylogenetic tree of the myxobacteria, as determined by 16S RNA sequencing, is abbreviated from Shimkets and Woese (1992). Strains are listed as positive (+) or negative (-) for msDNA production and for Southern hybridization with a ^{32}P-labeled internal fragment of the RT gene from Mx162. ND indicates those strains were not tested. It should be noted that *Desulfovibrio*, *Desulfosarcinia*, and *Bdellovibrio* are not myxobacteria and serve as out groups.

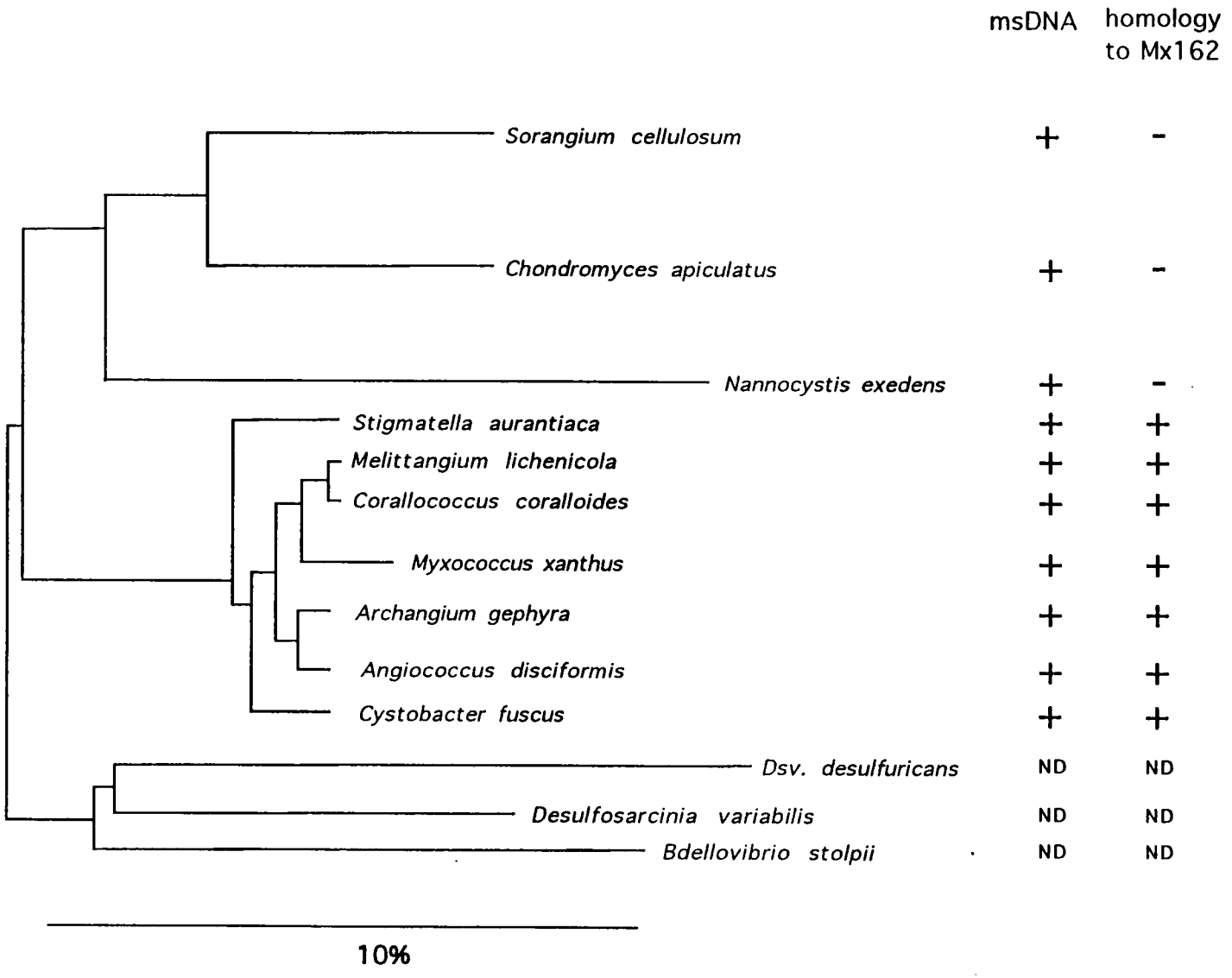
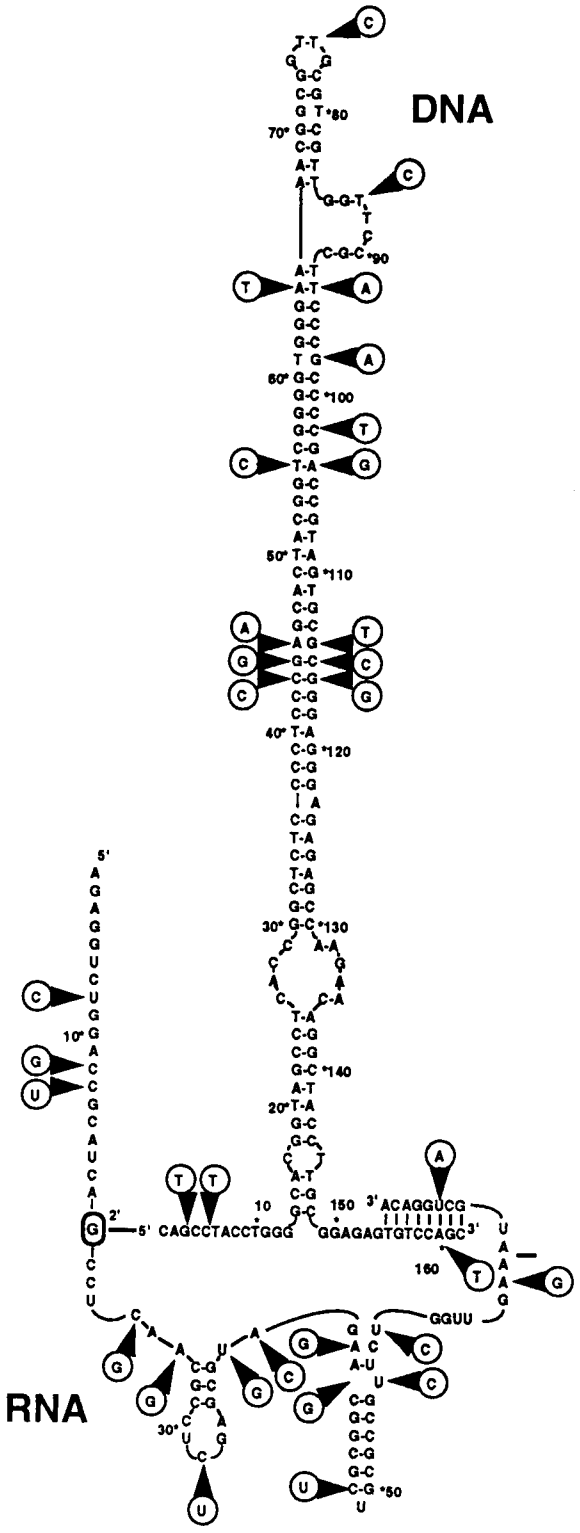

msDNA
homology to Mx162
Sorangium cellulosum
+
-
Chondromyces apiculatus
+
-
Nannocystis exedens
+
-
Stigmatella aurantiaca
+
+
Melittangium lichenicola
+
+
Corallococcus coralloides
+
+
Myxococcus xanthus
+
+
Archangium gephyra
+
+
Angiococcus disciformis
+
+
Cystobacter fuscus
+
+
Dsv. desulfuricans
ND
ND
Desulfosarcinia variabilis
ND
ND
Bdellovibrio stolpii
ND
ND
10%

Figure 3.4. **The sequence and secondary structure of msDNA from *Melittangium lichenicola*.** The sequence of the DNA portion of *M. lichenicola* msDNA was determined directly by the chemical cleavage method (Maxam and Gilbert, 1980), while the sequence of the RNA strand is based on the *msr* sequence as determined by the chain termination method (Sanger *et al.*, 1977). The predicted secondary structure is shown. Differences in the msDNA sequence from ML162 as compared to the msDNA sequence of Mx162 (from *M. xanthus*) are shown. Base substitutions are indicated by inward pointing arrows.


DNA
RNA
5'
3'
2'
CAGCCTACCTGGG
GGAGAGTGTCCAGC
ACAGGUCG
GGUU
10
150
160
30°
70°
50°
60°
40°
20°
*80
*90
*100
*110
*120
*130
*140
*50

addition to the nucleotide differences in the DNA strand, based on the *msr* sequence, 15 of the 76 predicted nucleotides that make up the RNA strand are different.

Based on hybridization of Mx162 DNA to *M. lichenicola* (Fig. 3.2, lane 6), *Pst*I chromosomal fragments of *M. lichenicola* in the range of 9 to 11-kbp were cloned into the *Pst*I site of pBR322. The recombinant plasmids were transformed into *E. coli* and screened by colony hybridization. One positive clone was picked (pNaeF4.15) and confirmation of the retron clone by Southern hybridization revealed a 10.8-kbp *Pst*I fragment that hybridized with the Mx162 probe. This primary clone was subsequently digested with *Bam*HI and a 3.8-kbp fragment was subcloned into the *Bam*HI site of pUC9 (pNae BB4.0.A4). This clone was then further subcloned for sequencing (Figure 3.5). In addition, synthetic oligonucleotides, listed in Table 3.2, were used to sequence regions that could not be subcloned.

Using the Maxam and Gilbert-generated sequence of the msDNA (Figure 3.4), we were able to locate the *msd* gene from the chromosomal clone, and subsequently identify the predicted *msr* gene encoding the RNA strand of the msDNA. Two 17 base pair (bp) inverted repeats were found, one (a2) is located just upstream of the predicted branched guanine residue which forms the 2'-5' RNA-DNA linkage, and the other (a1) is just downstream of the *msd* gene (Figure 3.6). These inverted repeats have been demonstrated to be required for the proper folding of the primary transcript to initiate the self-priming reaction that begins msDNA synthesis (Hsu *et al.*, 1989, Lampson *et al.*, 1989a, Shimmamoto *et al.*, 1993). Analysis of the sequence in the vicinity of the *msr* and *msd* genes revealed five ORFs. The fourth ORF, beginning at position +1897, showed significant nucleotide identity (>70%) to the reverse transcriptases found in the myxobacteria retrons Mx162 and Sa163 (Hsu *et al.*, 1992, Inouye *et al.*, 1989). The predicted protein from ORF 4 shows the seven conserved amino acid sequence domains, diagnostic for RT including the highly conserved YADD sequence of domain 5 (Xiong and Eickbush, 1990). Indeed, this ORF shows 79% nucleotide identity to the RT gene in retron Mx162. The RT-ORF did not show significant nucleotide identity to the RT from Mx65. This is expected since Mx65 shows only 35% nucleotide identity to Mx162 (Dhundale *et al.*, 1988a, Inouye *et al.*, 1990). The other four ORFs, ranging in size from 152 to 441 amino acids showed no significant homology to any previously reported genes when searched through the Genbank database. Comparison of the DNA sequence from a2 to the 3' end of the RT-ORF with the Mx162 sequence revealed 77% nucleotide identity overall between the two retrons, with identity being highest in the *msr-msd* regions (83%-

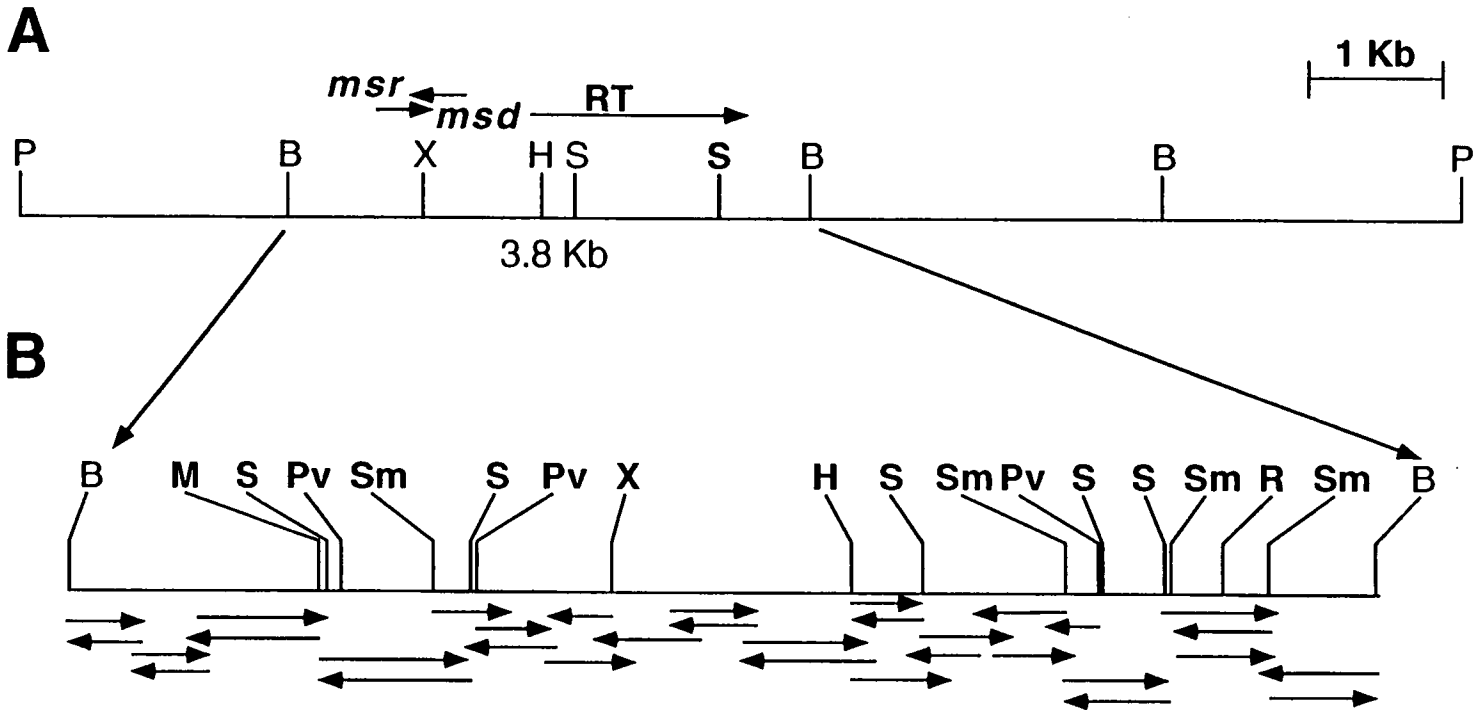


Figure 3.5. **Sequencing strategy of the retron ML162.** A. The primary genomic clone (pNaeF4.15) of retron ML162 from *Melittangium lichenicola* Fx e1was obtained by inserting a 10.8 kilobase pair *Pst*I chromosomal fragment from *M. lichenicola* into the *Pst*I site of pBR322. Arrows indicate the orientation of the genes that makeup the retron element: *msr* encodes the RNA portion of msDNA, *msd* codes for the DNA portion of msDNA, and RT encodes the reverse transcriptase.
B. A *Bam* H1 secondary clone of ML162 (pNae.BB4.0.A4) from which the retron was mapped and sequenced. The arrows indicate the sequencing strategy for ML162 using either subclones or synthetic primers. P, *Pst*I; B, *Bam*HI; X, *Xho*I; H, *Hinc*II S, *Sty*I; M, *Msp*I; Pv, *Pvu*II; Sm, *Sma*I; R, *Rsa*I.

Table 3.2. Primers used for sequencing of retron ML162

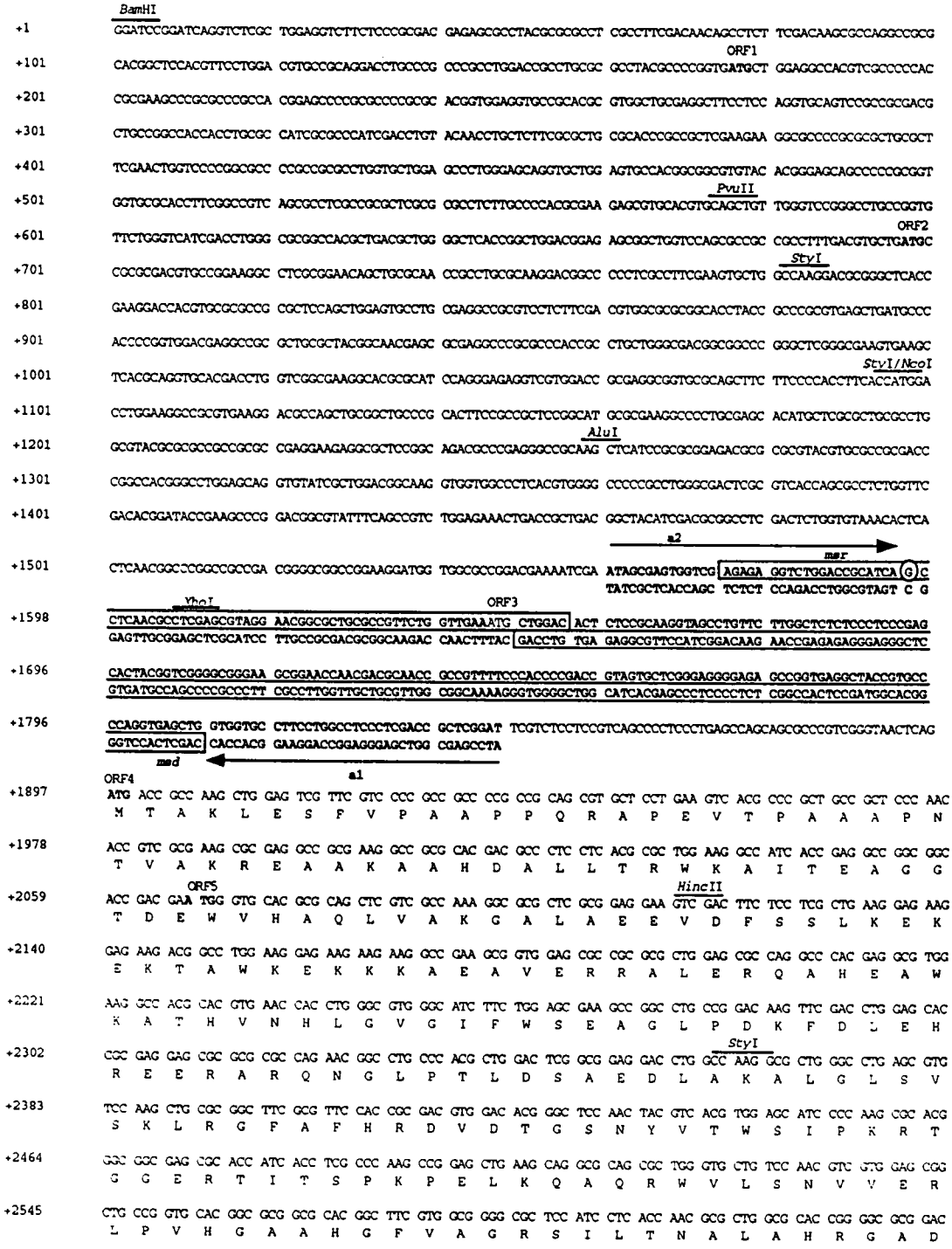
Primer name	Size[a]	Sequence (5'-3')	Position[b]	Strand[c]
Hincdown	18	CTCCTCCGTTCGCCCCTCC	1841-1858	coding
Hincup	18	GGCCTTCTTCTTCTCCTT	2149-2166	complementary
Xhodown	17	ACATCGACGCGGCCTCG	1459-1475	coding
Xhoup	17	GCGGTCGAGGGAGGCCA	1814-1830	complementary
Pcrstyup	17	ACGCCTGCTGGAGTTG	3356-3371	complementary

[a] number of bases in the synthetic oligonucleotide

[b] indicates the nucleotide position of the primer as reported in figure 3.6

[c] indicates the strand on which the primer runs 5'-3'

Figure 3.6. **The complete nucleotide sequence of retron ML162 from *Melittangium lichenicola*.** Potential ORFs are labeled and the ATG start codons are in bold. The msDNA coding region for ML162 starts at nucleotide +1662, includes the RNA (*msr*) and DNA (*msd*) coding region for msDNA (boxed), the branched guanine residue at position +1597 (circled), and the a2 and a1 inverted repeats at +1662 and +1841 (arrows). The RT-ORF (ORF 4) begins at +1896 and ends at position +3342.


BamHI
+1
ORF1
+101
+201
+301
+401
PvuII
+501
ORF2
+601
StyI
+701
+801
+901
StyI/NcoI
+1001
+1101
AluI
+1201
+1301
+1401
a2
mer
+1501
YhcI
ORF3
+1598
+1696
+1796
mer
a1
ORF4
+1897
ORF5
+2059
HincII
+2140
+2221
StyI
+2302
+2383
+2464
+2545

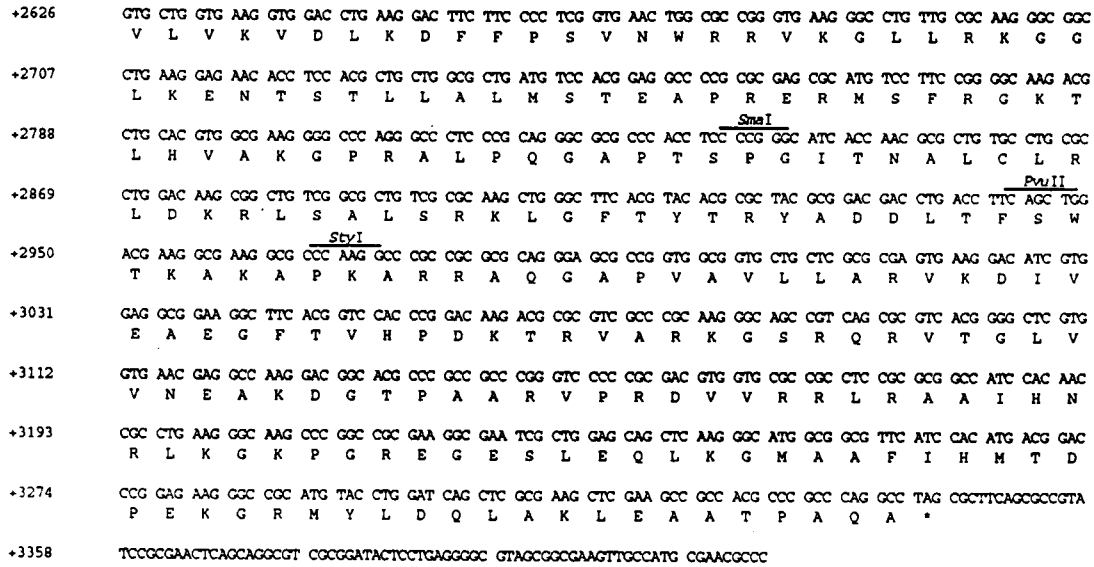


Figure 3.6. (continued)

90%) and lowest in the spacer region which is located between *msd* and the RT-ORF (39.7%).

The G+C content of the myxobacteria is 67-72% (Reichenbach and Dworkin, 1992). This high G+C content is reflected in the preferential usage of codons that are G-C rich. The RT-ORF of the retron found in *M. lichenicola* shows a heavy bias towards codons that use G and C over A and T. For example, there are six possible codons for the amino acid leucine (CTA, CTC, CTG, CTT, TTA, and TTG), yet of the 46 leucine residues in the RT-ORF of ML162, 12 are CTC, 33 are CTG, and 1 is TTG. This bias towards G-C rich codons is pervasive throughout the RT-ORF and is consistent with an organism with a high G+C genome. Such a GC codon bias has also been noted previously for the RT-ORF of retron Mx162 (Inouye *et al.*, 1989).

Retron junctions

At present, the precise ends of retrons in the myxobacteria are unknown. Therefore the ends of a retron element can only be defined as unique DNA sequences which are associated with the genes encoding msDNA and RT. If two similar or isogenic chromosomes exist, in which one of the chromosomes lacks the retron element, then comparison of the corresponding region between the two chromosomes should reveal the junctions of the retron element. However, for *M. xanthus*, there are no known strains in which the Mx162 retron element is absent. Alternatively, a different strategy can be used to identify the ends of this genetic element by comparison with the ML162 retron from *M. lichenicola*. Since the two retrons from *M. lichenicola* and *M. xanthus* are very similar in DNA sequence, and if the chromosomal region in which these elements have inserted is significantly different in DNA sequence between *Melittangium* and *Myxococcus*, then it could be possible to approximate where the 5' and 3' ends of this element lie. Figure 3.7 presents an alignment of flanking DNA sequence from *M. xanthus* and *M. lichenicola* around the 3' end of the RT-ORF of their respective retrons. In this 3' region, sequence identity rapidly degenerates from >70% to almost zero and extends for several hundred bp of sequence downstream from the RT-ORF. This 3' junction appears to be very near the stop codon of the RT-ORF (Figure 3.7). Interestingly, at the 5' end of the retron significant nucleotide identity is maintained over 300 bp upstream of the *msr* gene. This indicates the retron could be inserted in a gene common to both *M. xanthus* and *M. lichenicola* or probably in some homologous location on each chromosome. Alternatively, the retron could have been transmitted vertically from a common ancestor of *M. xanthus*

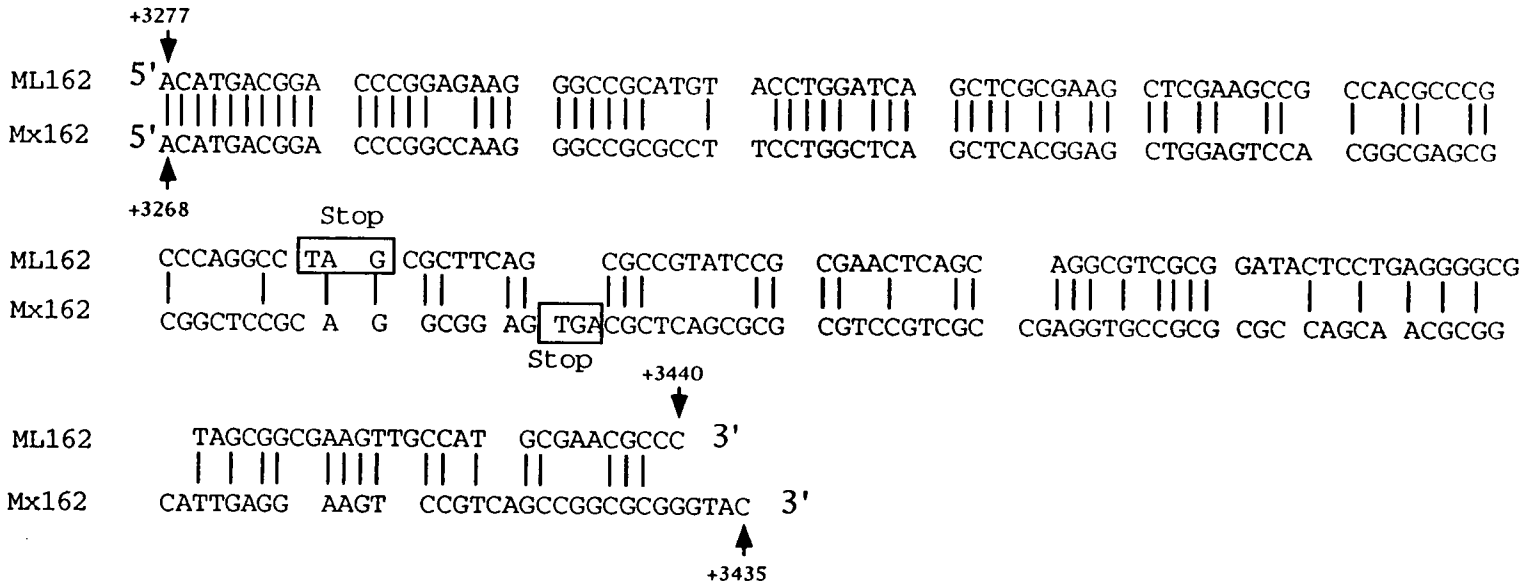


Figure 3.7. **Alignment of the 3' ends of ML162 and Mx162.** The alignment was created using "bestfit" program of the GCG software (Devereux *et al.*, 1984). Stop codons for each coding sequence are boxed. The sequences shown begin at positions 3277 and 3268 for ML162 and Mx162 respectively. Likewise, they end at positions 3440 and 3435. A close match between the two sequences appears to end around the stop codon for their respective RT-ORFs.

and *M. lichenicola* in which case the junction sequences would not be distinguishable from host sequences if both diverged at similar rates. A similar situation is observed with the DNA sequence flanking the retron Sa163 from *Stigmatella* (Hsu *et al.*, 1992b), where significant nucleotide identity is maintained upstream of the *msr* gene but not downstream of the RT gene. Sequence data for *M. lichenicola* not submitted to genbank is reported here and includes 352 bases 3' (Figure 3.8) of the published sequence (Figure 3.6). The 3' sequence was compared against several sequence databases using the BLAST algorithm (Altschul *et al.*, 1988) and no significant nucleotide identity to reported nucleotide sequences, or sequence similarity to reported peptide sequences was found (data not shown).

Sequence divergence

By calculating the rate of nucleotide change at synonymous sites for the RT gene encoded by the Mx162 type retrons, the divergence of this gene can be compared to other bacterial genes. Such comparisons may provide a hint about the origin and age of this RT gene found in the myxobacteria genome.

The average rate of silent substitution found for genes common to *Salmonella* and *E. coli* is 0.335 substitutions per million years (Sharp, 1991) (Table 3.3). The fastest and slowest rates of substitution between *E. coli* and *Salmonella* are 0.63 substitutions per million years and 0.08 substitutions per million years respectively. The myxobacteria *Melittangium* and *Myxococcus* are estimated to have diverged from a common ancestor about 90 million years ago. This is based on comparison of their 16S rRNA sequences (Shimkets and Woese, 1992) and a universal substitution rate calculated for bacterial genomes (Ochman and Wilson, 1987). The percent divergence of the RT genes from retrons Mx162 and ML162 were compared using the method of Li *et al.* (1985) (Table 3.3). ML162 and Mx162 show 46% divergence at silent sites, which translates into a rate of 0.26 silent substitutions per million years. This is close to the average rate of change (0.335) for genes reported in *Salmonella* and *E. coli*, which presumably diverged about 140 million years ago (Sharp, 1991). In fact, it can be argued that the degree of sequence variation observed among the Mx162 RT genes is more significant than what is reflected in the calculated rate of silent substitutions. This is because of the high G+C content of the myxobacteria genome which results in strong codon bias in protein coding regions. This in turn may reduce the rate at which substitutions at silent sites accumulate by limiting the number of alternative codons that can be used. Thus the amount of sequence variation

+3427	GGGATGCGCC	ACGTTTGCGC	ATACCCCGCG	ACGTGTGTGA
+3467	ACAGCTCATC	CAGGGCAACG	GCTGCCCGGA	CTCCTCCGTC
+3507	CCCGCGCACG	ATGAACAGGG	GCCGGGGCTT	GATGTGGTCC
+3547	ACCGCGTCGA	TGGTCCGCAC	CTCGTCCAGG	GCAATCCCCC
+3587	GCCGCCAGAA	GGGAATCAAA	GCCCCCGTCT	GCNTCACCAC
+3627	GCCGAAGCGC	CCGGAATGCA	TACGCCGCCG	CCAGCCACAG
+3667	CGTGTTGAAC	GGAGACAGCA	GCACCACCGC	CGCACCTGCG
+3707	GGTCCTTCGC	CGCCACCTCC	GCCACCGCCG	CGGACCCGAT
+3747	GGAAAACCCC	AGCGCCCCCA	CGCGAGCCGG	ATC

Figure 3.8. **The distal 3' region of retron ML162 from *M. lichenicola*.** This region, which starts at position 3427, is the 3' half of the *Sma*I site at which the reported sequence, (Figure 3.6), ends. No significant nucleotide identity or protein homology was noted when this sequence was searched against the genbank database.

Table 3.3. Rates of change at silent sites for RT and other bacterial genes

Organisms compared	Time of divergence (million yr ago)	Gene	% Divergence at silent sites (corrected)	Rates of substitution at silent sites/million yr
Melittangium sp.-*Myxococcus* sp.	90	RT	46	0.26
E. coli-*Salmonella* sp.	140	*trp*A	177	0.63
	140	*pts*H	23	0.08
	140		94 (avg)	0.335 (avg)

observed between the RT genes of *Myxococcus* and *Melittangium*, although relatively small, is probably evolutionarily significant; having accumulated changes at the same rate as most other genes since the divergence of *Myxococcus* and *Melittangium* from a common ancestor.

Discussion

It is evident from findings presented here and previously (Dhundale *et al.*, 1985, Lampson *et al.*, 1991a, Rice *et al.*, 1993) that the family of retrons found in the myxobacteria are unique from retrons found in other bacteria and that the origin and acquisition of these retrons in the myxobacteria may be different from those of the other proteobacteria. The first indication is based on the distribution of retrons in the myxobacteria as compared to other proteobacteria, such as *E. coli*. Retrons are ubiquitous, or nearly so, in the myxobacteria, being absent in only one *Cystobacter* strain out of 28 myxobacteria strains tested in this study. In contrast, only about 10% of all *E. coli*, *Klebsiella*, *Salmonella*, *Proteus*, *Rhizobium*, and *Bradyrhizobium* strains tested have been demonstrated to contain retrons (Herzer *et al.*, 1990, Rice *et al.*, 1993). The distribution of a particular retron type within the myxobacteria family appears to cluster within a phylogenetic subgroup (Figure 3.3). For example, elements very similar to retron Mx162 (originally discovered in *M. xanthus*) were found in 5 different species of *Myxococcus*, 3 species of *Cystobacter* as well as in *Melittangium lichenicola*, *Corallococcus coralloides*, *Archangium gephyra*, *Angiococcus disciformis*, and previously from *Stigmatella aurantiaca* (Table 3.1) (Furuichi *et al.*, 1987a). Based on 16S rRNA analysis (Shimkets and Woese, 1992), all of these bacteria form a phylogenetically related cluster or subgroup within the myxobacteria family (Figure 3.3). Likewise, all ten strains of the *Nannocystis* subgroup surveyed appear to share the same novel retron element (Table 3.1).

This clustered distribution of the Mx162 retron within the *Myxococcus* subgroup implies a vertical rather than a horizontal transmission of this particular element. In other words, an evolutionary scenario can be envisioned in which the genome of a common ancestor acquired a Mx162 type retron element with subsequent inheritance of this element by all the different genera during the divergence and evolution of this subgroup.

Other evidence also tends to support this scenario. For example, a homologue of retron Mx162 found in *Melittangium lichenicola*, retron ML162, was cloned and its DNA sequence was compared with two previously characterized examples of this retron type

(Mx162 and Sa163). Overall nucleotide identity between retrons Mx162 and ML162 is 77%. In addition, the silent site substitution rate for the RT-ORF from these two retrons is calculated to be 0.26 substitutions per million years (Table 3.3). These figures are consistent with a gene that has diverged at least as long as the genera *Myxococcus* and *Melittangium* shared a common ancestor about 90 million years ago. Therefore, since it appears that retrons were acquired by members of the *Myxococcus* subgroup prior to their divergence, it is likely that *C. ferrugineus* Cb fe18 has lost either the ability to synthesize msDNA or the entire retron element.

In contrast, when the retron elements found in other proteobacteria such as *E. coli* are compared with the findings presented here for the myxobacteria, it appears unlikely that the Mx162 retron is a recent acquisition by the myxobacteria genome. For example, in *E. coli* the retron element Ec107 is widely distributed among ECOR strains; a reference collection of natural isolates (Herzer *et al.*, 1992, Herzer *et al.*, 1990, Kawaguchi *et al.*, 1992). However, its occurrence is sporadic and this element is found in distantly related phylogenetic branches of the ECOR collection. Also, DNA sequence determination of several individual examples of the Ec107 element revealed little or, in some cases, no nucleotide sequence diversity among these retrons (Kawaguchi *et al.*, 1992). In addition, codon usage within the RT-ORF of most *E. coli* retrons is atypical for native *E. coli* genes (Inouye *et al.*, 1989). All of these features appear to mark the Ec107 retron as a recent addition to the genome of certain lineages of *E. coli* with the likely horizontal spread to other strains.

The features described here for the Mx162 class of retron found in the myxobacteria do not resemble those of the *E. coli* retrons and appear to be inconsistent with a recent inheritance in the genome. Indeed, an analysis of the codon usage for the RT-ORF from the retron of *Melittangium lichenicola* revealed a heavy bias towards codons that predominantly use G and C and is typical of the codons used for other known myxobacteria genes (Inouye *et al.*, 1989, Shimkets, 1993). Although recent acquisition from a GC rich organism can not be discounted, this characteristic does suggests that the retron encoded RT gene has resided in the myxobacteria genome for millions of years, at least since the speciation of the *Myxococcus* subgroup. However, such inheritance may not be true for other myxobacteria retrons. Retron Mx65 was found in only two species other than *Myxococcus xanthus*, and may have been introduced into the *Myxococcus* genome at a later time.

Evidence continues to indicate that, like the eukaryotes, RT-encoding genetic elements of diverse types are probably widespread and a nascent characteristic of the prokaryotic world. This idea is supported by the prevalence of diverse retron elements, among a wide variety of bacteria, as documented here and in a previous paper (Rice *et al.*, 1993). Additional evidence comes from the recent discovery of RT-encoding group II introns in several groups of the eubacteria (Ferat and Michel, 1993). Many questions about the retron elements need to be explained, including their origin in the bacterial genome, potential mobility, and their relationship with other classes of retroelements. Based on the prevalence of RT genes, has the process of reverse transcription played a role in the generation of genetic variation or otherwise influenced the evolution of the bacterial genome? Indeed, the chromosomes of the myxobacteria appear to have a particular predilection for retron elements and may be a good place to look for clues to answer these questions. Because retrons can be deleted from the chromosome of *M. xanthus* without affecting their growth or development in the laboratory, no function has been linked to these genetic elements. It can not be discounted, however, that these RT-encoding elements have a function which has provided a selective advantage for the myxobacteria. This could be an alternative explanation for why retron elements have been retained by the myxobacterial genome for millions of years.

Acknowledgments

We would like to thank Larry Shimkets and Hans Reichenbach for providing many of the myxobacteria strains used for this study. We would also like to thank Tom Whittam for critical reading of this manuscript and help in generating the synonymous substitution rates.

CHAPTER IV

Detection and Quantification of Myxobacteria in Natural and Contaminated Soils Using DNA Hybridization

Introduction

The myxobacteria are aerobic, gliding, Gram-negative eubacteria, endemic to most aerated soils throughout the world (Reichenbach and Dworkin, 1992). While the multicellular behavior and developmental life cycle of these bacteria are well studied (Dworkin, 1991), little is known about the ecology and distribution of the myxobacteria in the environment. It has been previously noted (Reichenbach and Dworkin, 1992) that studies which focus on soil microbe communities rarely mention the isolation of myxobacteria, thus leading to a lack of information or under reporting of data concerning myxobacteria in soil communities. Even studies that concentrate on myxobacteria may underestimate the number of myxobacteria present since those studies typically rely on growth and isolation of myxobacteria cells. This approach is extremely time intensive, taking from several weeks to years to isolate and identify myxobacteria species, and may underestimate myxobacteria numbers due to the long doubling time of myxobacteria, which puts them at a competitive disadvantage with other microorganisms. An approach that utilizes molecular techniques such as gene probing may overcome these limitations and facilitate the study of the ecology, enumeration, and distribution of the myxobacteria. Gene probing has been utilized to detect and enumerate bacteria from very diverse environments and should also be applicable to the study of myxobacteria (Sayler *et al.*, 1991, Stahl *et al.*, 1988).

DNA sequences based on msDNA, an unusual satellite DNA, were used as probes for this study to specifically detect and quantify myxobacteria in soil environments. In 1984, Yee *et al.* (1984) reported a high copy number, small, satellite DNA associated with the myxobacterium *Myxococcus xanthus*. This DNA consisted of a single strand of DNA covalently linked via a 2'-5' phosphodiester bond to an internal guanine residue of a single-stranded RNA (Dhundale *et al.*, 1987, Furuichi *et al.*, 1987b). These molecules, termed multicopy, single-stranded DNA (msDNA), are encoded by a chromosomal locus called a retron (Lampson, 1993). In addition to producing msDNA, retrons also code for reverse

transcriptase (RT). While much is known about retrons and msDNA regarding biosynthesis and distribution in bacteria, no function has been demonstrated for retrons or msDNA. It was demonstrated previously that retrons are present in the myxobacteria in a clustered distribution, with each of the three subgroups containing its own "type" retron (Rice and Lampson, 1995). In addition, almost all natural isolates examined (>99%) from each subgroup contained their respective "type" retron. Thus each of the three "type" retrons represent a characteristic and specific set of DNA sequences for each myxobacteria subgroup and may be useful as a diagnostic probe for detecting myxobacteria in environmental samples.

In addition to producing a unique extra chromosomal element, myxobacteria also exhibit an interesting multicellular behavior during vegetative growth where cells clump together and feed in large groups, often described as the "wolf pack effect" (Dworkin, 1973). In conjunction with multicellular feeding, the myxobacteria can excrete extracellular enzymes capable of breaking down complex macromolecules such as cellulose, chitin, starch, and protein and are also capable of preying on other bacteria (Reichenbach and Dworkin, 1992). In light of their prevalence in the soil, and the detection of myxobacteria in contaminated soils, as shown in this study, the potential of the myxobacteria to attack and degrade aromatic and polyaromatic hydrocarbons was also examined. Additionally, DNA hybridization experiments were performed using genes which have been previously determined to be involved in the biotransformation of naphthalene and catechol to determine if myxobacteria have similar genes.

Materials and Methods

Bacterial strains and plasmids and primers

The myxobacteria strains used include: *Sorangium cellulosum* strains, So ce5, So ce7, and So ce11, *Chondromyces apiculatus* strains, Cm a5, Cm a15, Cm a16, *Cystobacter fuscus* Cb fe18, *Melittangium lichenicola*, *Nannocystis exedens* strains, Na e1, Na e3, Na e24, Na e30, Na e39, Na e434, Na e585, Na e619, Na e642, and *Myxococcus xanthus* DZF1. The soil bacterium, *Flexibacter elegans* Fx e1, was also included. Myxobacteria strains were kindly donated by Hans Reichenbach and Larry Shimkets. All myxobacteria species and Fx e1 were grown on MD1 medium (Reichenbach and Dworkin, 1992) at 30°C with shaking. *E. coli* DH5α was used as the host strain for cloning in conjunction with plasmid pUC19. Transformants were selected on LB plates

supplemented with 50 μg/ml ampicillin. *E. coli* DH5α chromosomal DNA was also used as a negative control for Southern blots. Plasmids used included: pDTG601 (Zylstra *et al.*, 1989) and pGSH2836 (Suzuki *et al.*, 1991). Primers specific for *nah*A, based on the published sequence (Simon *et al.*, 1993), were: nahA3 5' CCCTAGCGCGTAACTACCCC 3' and nahA4 5' GGTCCAGACCTCGGTGGTC 3'.

Probes

The 16S eubacteria probe was a 15 base pair oligonucleotide (rDNA) (5' ACGGGCGGTGTGTRC 3'), which is based on a region near position 1400 in *E. coli* 16S rRNA that is complimentary to nearly every eubacterial 16S rRNA tested so far (Stahl *et al.*, 1988). This synthetic oligonucleotide was 5' end-labeled using T4 Polynucleotide Kinase and [γ-^{32}P] ATP (New England Nuclear, Boston, MA). The *Myxococcus* subgroup-specific probe was prepared as a 0.3 kilobase pair *XhoI-Bam*HI fragment from the plasmid clone, pmsAlu (Inouye *et al.*, 1989), of retron Mx162 first isolated from the myxobacterium *Myxococcus xanthus* DZF1. This probe contains the *msd* region, the coding region for the DNA portion of msDNA encoded by retron Mx162. The *Nannocystis* Na e434, *Sorangium* So ce11, and *Flexibacter* Fx e1, probes were prepared as follows: msDNA was extracted from total DNA and gel purified on a 5% acrylamide gel. Purified msDNA and DNA restriction endonuclease fragments were labeled with [α-^{32}P] dCTP (New England Nuclear, Boston, MA) by the random hexamer labeling method (Feinberg and Volgstein, 1983).

Soil samples

Soil samples were collected from four sites at the Savannah River Nuclear facility. Two samples were from the rhizosphere of a legume, Lespedeza, in soil with either a high concentration (LH) of contaminating trichloroethylene (TCE) or low concentration (LN) of TCE. Two samples were similarly collected from the rhizosphere of pine seedlings in areas of either high (PH) or low (PN) TCE concentration. At all four sites, soil samples were collected at 5, 30, and 60 centimeter (cm) depths (Ringleberg and White, 1992). A fifth soil sample was collected from an uncontaminated farm soil at the University of Tennessee (FS).

DNA preparation and hybridization

Prior to DNA extraction, soil samples LH, LN, PH, and PN were extracted for total lipids as described (Kehrmeyer *et al.*, 1995) and the analysis and quantification of the extracted lipids was reported elsewhere (Ringleberg and White, 1992). Following lipid extraction, DNA was extracted from the same soil samples by the method of Ogram *et al.* (1987) and the analysis of those DNA samples is reported here. This represents a simultaneous lipid-DNA extraction technique. Chromosomal DNA was isolated from myxobacteria strains as described previously (Lampson *et al.*, 1991). For Southern blots, DNA digested with restriction endonucleases *Pst*I or *Sal* I, was separated on 0.7% agarose, and subsequently transferred to nitrocellulose membrane (BA-S NC Schleicher and Schuell, Keene, NH) using a vacuum blotter apparatus (Milliblot-V-System, Millipore, Bedford, MA). All other filters were prepared by transfer to BioTrace HP nylon membranes (Gelman Sciences, Ann Arbor, MI) by vacuum blotting using a slot blot manifold (BioRad, Melville, NY). Hybridization for nitrocellulose membranes was carried out at low stringency [40% formamide, 5X SSPE (1X-0.01 M NaH_2PO_4-0.001 M EDTA-0.004 M NaOH-0.18 M NaCl), 5X Denhardt's solution (1X-0.2 g ficoll 400-0.2 g PVP-0.2 g BSA per liter), 0.3% $NaDodSO_4$] at 37°C. The hybridization solution for nylon membranes was 1 mM EDTA-0.5 M $NaHPO_4$, pH 7.2-7% $NaDodSO_4$ (Church and Gilbert, 1984). Hybridization was carried out at high stringency (65°C) for msDNA probes and at low stringency (37°C) for the 16S rDNAprobe. Hybridization was visualized by autoradiography at -70°C.

Determination of cell number

For the mineralization assays, cell density of the myxobacteria was determined indirectly by protein concentration. Protein was extracted from cell pellets of *N. exedens* and *S. cellulosum* by sonication, and were centrifuged to remove cellular debris. Protein concentration was determined spectrophotometrically by the method of Bradford *et al.* (1976) using a BioRad protein assay kit (BioRad, Melville, NY). Protein concentration was used to estimate initial cell numbers of myxobacteria added in the mineralization studies. The number of bacteria in soil samples was determined indirectly by DNA hybridization, as described (Applegate *et al.*, 1995a, Applegate *et al.*, 1995b). The amount of target DNA (the specific DNA sequences to which the probe hybridizes) is determined by the following formula (Applegate *et al.*, 1995a, Applegate *et al.*, 1995b): target DNA μg/μl=[kilobase of target / (kilobase of vector + kilobase of target)]X[total

DNA μg/μl]. Based on the intensity of the hybridization signal of the probe to standards of known concentration (target DNA), a standard curve is generated which can be used to determine the concentration of target DNA in the experimental sample (nanograms hybridized). Once the concentration of target DNA in the unknown sample is calculated, the total number of bacterial cells in the unknown samples can be determined using the following formulas (Applegate *et al.*, 1995a, Applegate *et al.*, 1995b): (a) nanograms of one target=[(number of base pairs in target)X(660 grams/mole base pair)X(1 mole base pair / 6.023 $X10^{23}$ molecules)X(10^9 nanograms / g)], (b) number of cells=(nanograms hybridized / nanograms of one target)X(1 / copy number of target)X(1 / extraction efficiency)X(1 / sample fraction loaded). Extraction efficiency is measured by the inclusion of an internal standard to the sample prior to initiation of extraction. The internal standard should be sequences which are not normally found in the sample, such as lambda DNA. Since msDNA is present in multiple copies, target for the msDNA probe was estimated at 500 copies per cell (Yee *et al.*, 1984). Densitometry of slot blot soil hybridizations was performed using a model GS-670 imaging densitometer (BioRad, Melville, NY) and the densitometry software Molecular Analyst version 2.0 (BioRad, Melville, NY).

Mineralization assays

[UL-^{14}C] toluene (9.7 mCi/mmol), [1-^{14}C] naphthalene (8.0 mCi/mmol), [UL-^{14}C] phenol (8.0 mCi/mmol), and [UL-^{14}C] catechol (17.7 mCi/mmol) were purchased from Sigma Chemical Co., St. Louis, Mo. Myxobacteria strains were grown to log phase in MD1 medium and 1 $X10^8$ cells were added to 4 ml of MD1 medium in sterile, 25 ml mineralization vials (Pierce Chemical Co., Rockford, IL). Mineralization vials contained an 8 ml glass vial to serve as a CO_2 trap containing 1 ml of 0.5 N NaOH. Radiolabeled compound was added at approximately 2 $X10^6$ cpm per vial. Vials were sealed with Teflon-silica septa and incubated at 30°C with shaking (200 rpm) for 7 days. Reactions were stopped by injecting 1 ml of 4 N H_2SO_4 to the medium through the septa and allowing cells to continue shaking for 1-2 hours longer. 0.5 ml of the CO_2 trap was then removed, added to 1 ml H_2O, and 10 ml Liquiscint, allowed to sit in the dark for 24 h, and counted. 0.5 ml of 4 N H_2SO_4 was added to the CO_2 trap and allowed to sit for 1 h open, at which time, the remaining contents of the trap were collected and counted as described above. Results are presented as the percentage of labeled compound released as $^{14}CO_2$.

Results

Detection of myxobacteria in natural and contaminated soils

Soil samples collected from four sites (LH, LN, PH, PN), at three different depths (5, 30, and 60 centimeters), from the Savannah River Nuclear facility were extracted for total DNA by the method of Ogram *et al.* (1987). Three samples from each site and depth were collected. One sample of uncontaminated farm soil was also extracted for total DNA. The purified DNA was subsequently transferred to nylon membranes for DNA hybridization. A probe derived from 16S rDNA was used to detect the presence of eubacteria in the various soil samples studied. Hybridization of this 16S rDNA oligonucleotide to one of the slot blots containing various soil DNAs is shown in Figure 4.1. While the 16S probe hybridized strongly to all of the various soil DNAs at both the 5 cm depth (Figure 4.1) and the 30 cm depth (data not shown), the hybridization signal was much weaker for the 60 cm depth (data not shown). Thus, bacteria appear to be present, especially in the shallower depths, in all of the contaminated and uncontaminated soil samples studied. The 16S rDNA probe also hybridized to the control DNA of the myxobacterium *Sorangium cellulosum* So ce11 (Figure 4.1), which was expected since the So ce11 control was total chromosomal DNA from this bacterium. This probe did not hybridize to the *Nannocystis exedens* control DNA as anticipated since this control DNA is a plasmid clone containing just the msDNA gene, *msd.*

The specific detection of members of the *Myxococcus* subgroup was accomplished using a DNA probe containing just the *msd* gene from the retron element Mx162. The *msd* gene encodes the DNA portion of msDNA and was originally cloned from the myxobacterium *Myxococcus xanthus* (Yee *et al.*, 1984). However, a recent study (Rice and Lampson, 1995) indicates that all members of a large group within the myxobacteria, called the *Myxococcus* subgroup, have this same chromosomally encoded *msd* sequence. Hybridization of the various soil DNAs with this *Myxococcus* probe indicated that DNA sequences specific to members of the *Myxococcus* subgroup were present in all four of the contaminated soil sites at both the 5 cm (Figure 4.2A) and 30 cm (data not shown) depths. No hybridization was observed at the 60 cm depth (Figure 4.2B). It should be noted that this *Myxococcus* probe did not hybridize to either the *E. coli* or the lambda control DNAs; only to the Mx162 control DNA (Figure 4.2A and B).

In a similar manner, another probe was made using purified msDNA isolated from the myxobacterium *Nannocystis exedens*, strain Na e434. Again, it appears that most all

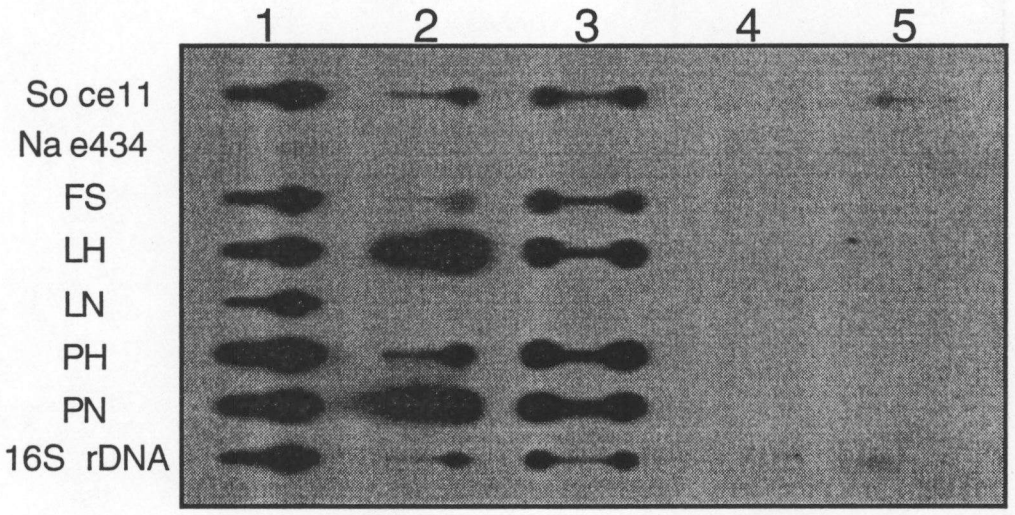
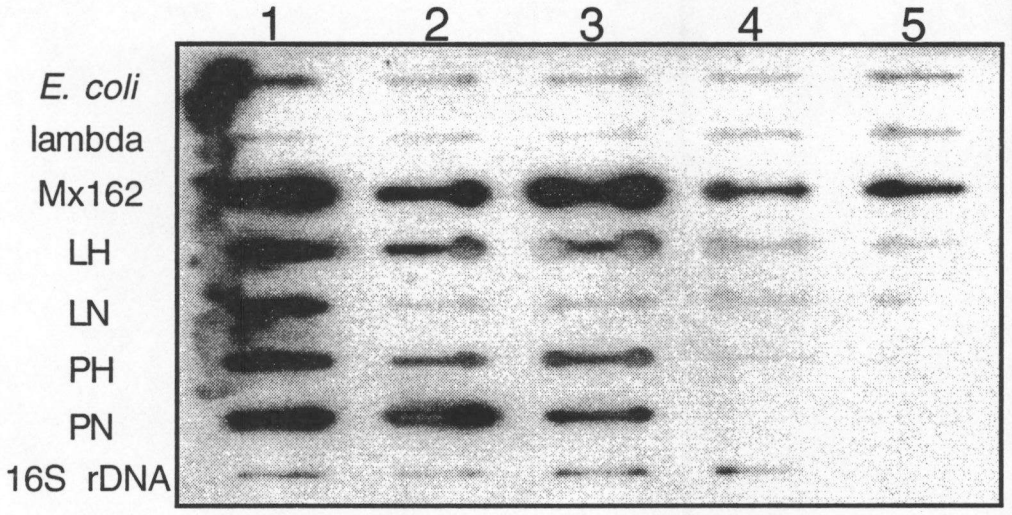
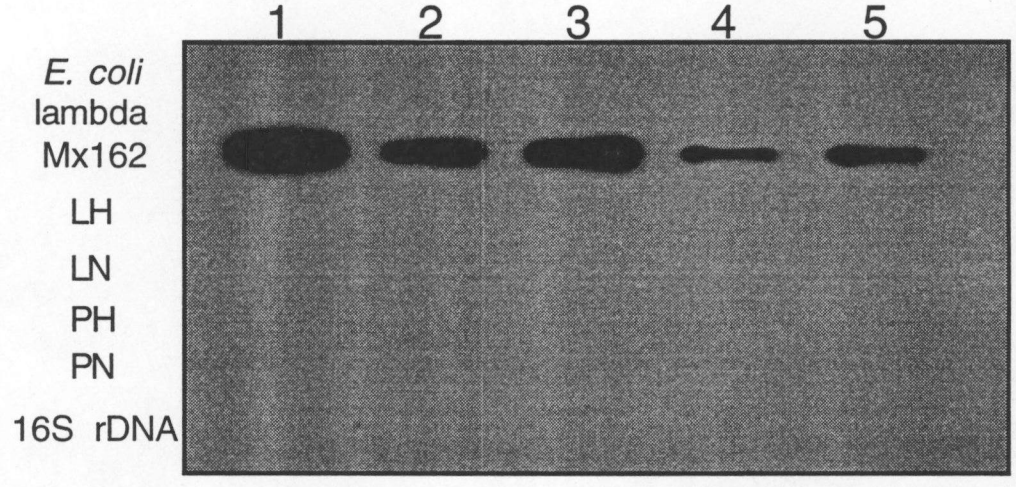


Figure 4.1. **Detection of total eubacteria in soil by DNA hybridization.** DNA samples were transferred to a nylon membrane filter using a slot blot manifold as described (see materials and methods) in the following concentrations: the three control DNAs, So ce11, Na e434, and the 16S rDNA were loaded at concentrations equivalent to 10, 1, 3, 0.1, and 0.01 ng of target sequence to columns 1-5 respectively. The control DNA for So ce11 was total chromosomal DNA from the myxobacterium *Sorangium cellulosum* So ce11. The control DNA for Na e434 was a plasmid clone of the retron element from *Nannocystis exedens* Na e434 and the control DNA for the 16S rDNA was a plasmid clone of the 16S rDNA gene from *E. coli*. Total DNA extracted from 1, 0.1, 0.5, 0.01, and 0.001 grams of farm soil (row FS) was loaded in columns 1-5 respectively. Total DNA extracted from 5 grams of soil from each of the four Savannah River Nuclear facility soil samples (rows LH, LN, PH, PN) were loaded, in triplicate, to columns 1-3. There is no DNA loaded in columns 4 and 5 for rows LH, LN, PH, and PN. The slot blot was then hybridized with a ^{32}P labeled 16S rDNA oligonucleotide probe. Hybridization signals indicate the presence of bacterial DNA in the various soils and 16S rDNA sequences in the two control DNAs, So ce11 and 16S rDNA.

Figure 4.2. **Detection of *Myxococcus* subgroup bacteria in soils.** A 0.3-kilobase pair *Xho*I-*Bam*HI fragment from the plasmid clone, pmsAlu (Inouye *et al.*, 1989), of the retron Mx162 was used as a *Myxococcus* specific hybridization probe against total DNA from soil samples from 5 cm (A) and 60 cm (B) depths. Total DNA extracted from 10^8, 10^4, 10^6, 10^5, and 10^7 *E. coli* cells/ml was added in columns 1-5 respectively. Lambda DNA, Mx162 clone pmsAlu, and a plasmid clone of the 16S rDNA, were loaded at concentrations equivalent to 10, 1, 3, 0.1, and 0.01 ng of target sequence DNA in columns 1-5 respectively. Standards were loaded in this order intentionally to separate standards of high concentration so hybridization signals do not mask each other. Rows LH, LN, PH, and PN were as in Figure 4.1. Hybridization signals indicate the presence of *Myxococcus* DNA in all four of the Savannah River Nuclear facility samples at the 5 cm (A) but not at the 60 cm depth (B).



A



B

Nannocystis strains contain msDNA sequences specific to the *Nannocystis* subgroup in their genome (Rice and Lampson, 1995). Hybridization of soil extracted DNA with this *Nannocystis* probe gave similar results as with the *Myxococcus* probe. That is, hybridization occurred against all of the contaminated soil samples at the shallow depths (5 and 30 cm) and against DNA from the uncontaminated farm soil (Figure 4.3). Again, the *Nannocystis* probe did not cross hybridize with the So ce11 chromosomal DNA or the 16S rDNA controls.

The last set of hybridization experiments were done using two probes; one based on purified msDNA from the myxobacterium *Sorangium cellulosum* So ce11, and the other based on msDNA isolated from *Flexibacter elegans* Fx e1. *F. elegans* is not a myxobacterium but is a Gram-negative, gliding bacterium isolated from habitats similar to those occupied by the myxobacteria. In contrast to the earlier probes, neither the *Sorangium* probe (Figure 4.4) nor the *Flexibacter* probe (data not shown) hybridized with any of the soil derived DNAs. Both probes did, however, hybridize to their respective positive control DNAs (Figure 4.4).

The probes used in these hybridization experiments appear to be specific for the myxobacteria with no cross-hybridization with negative control DNA and strongly suggests that hybridization is due to the presence of myxobacterial DNA sequences in the soil samples tested. Therefore, these data indicate that two of the three myxobacteria subgroups, *Myxococcus* and *Nannocystis*, are present in the soils tested and that either no *Chondromyces* or *F. elegans* species are present or these bacteria are present below the detection limit of the assay, which for these experiments was 10^3 cells per gram soil. In addition, hybridization of the soils with the 16S rDNA, *M. xanthus*, and *N. exedens* probes, indicated that the concentration of detectable bacterial DNA decreased with increasing soil depth (Figure 4.1).

In addition to the msDNA-derived probe, other myxobacteria genes were also used in both Southern and slot blot hybridizations. These genes included: *veg*A (Komano *et al.*, 1987), *csg*A (Li *et al.*, 1992), *dsg* (Cheng *et al.*, 1986), a sigma factor (Inouye, 1990), and protein S (Downard *et al.*, 1985). Southern hybridization against myxobacteria from all three subgroups, indicated that none of the aforementioned gene probes hybridized with a wide enough range of strains to be inclusive of all the myxobacteria, or even a single subgroup (data not shown). Slot blot hybridization with these probes indicated that they detect myxobacteria in soils, but in light of the Southern hybridization data, these probes likely underestimate the number of myxobacteria present.

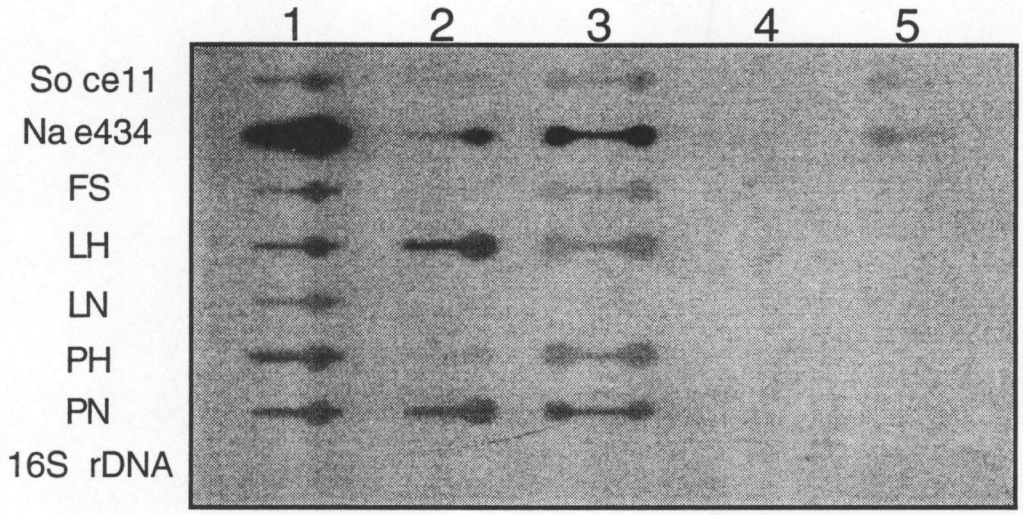


Figure 4.3. **Detection of *Nannocystis* subgroup bacteria in soils.** The same slot blot as in Fig. 4.1 was stripped and hybridized with a ^{32}P labeled msDNA from *N. exedens* 434. Hybridization signals indicate the presence of *Nannocystis* DNA in both the uncontaminated farm soil (row FS) and the TCE contaminated soils (rows LH, LN, PH, and PN).

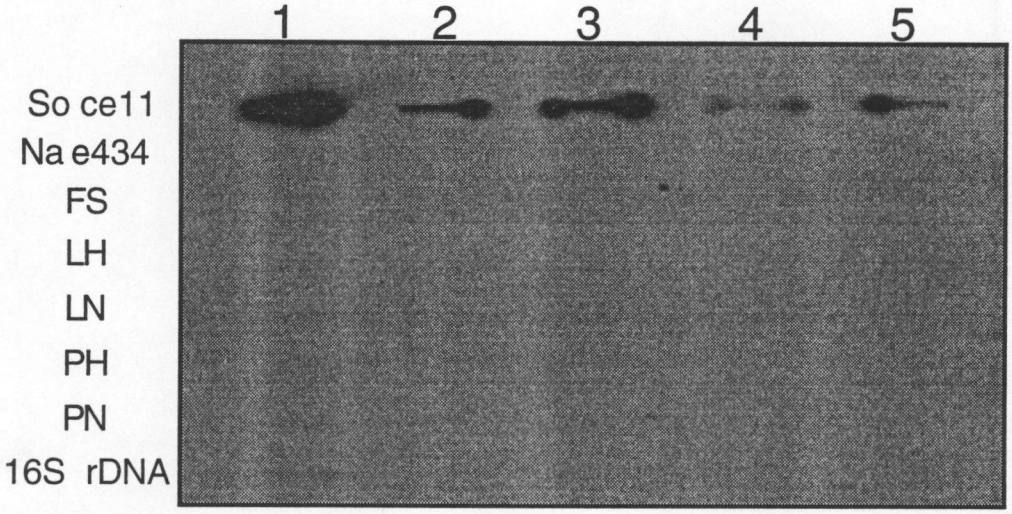


Figure 4.4. **Detection of Chondromyces subgroup bacteria in soil.** The same slot blot filter as in Fig. 4.1 was stripped and hybridized with the ^{32}P labeled msDNA probe from *Sorangium cellulosum* So ce11. Aside from the positive control, no hybridization signal and therefore presumably no *Chondromyces* specific DNA, was detected in any of the soil samples.

Quantification of bacterial cell numbers

An estimation of the number of bacterial cells present in the soil samples tested can be made from the DNA hybridization experiments described above. Using a known concentration of target DNA to calibrate hybridization signal intensity, it is possible to calculate the total number of cells per gram of soil based on the intensity of the hybridization signal of the probe to an environmental DNA sample (see methods) (Applegate *et al.*, 1995a, Applegate *et al.*, 1995b). Using this procedure, the total number of eubacteria present in the farm soil was estimated at 2.5×10^{11} cells per gram of soil based on the hybridization with the 16S rDNA probe (Table 4.1). Hybridization of the 16S rDNA probe to the TCE-contaminated environmental samples indicated that the total eubacterial population at 60 cm varies from 9×10^{8} to 3.5×10^{9} cells per of gram soil in three of the four sites tested, and was undetectable in the fourth site (Table 4.1). At depths of 30 cm and 5 cm, the total eubacteria population ranged from 4×10^{8} to 2×10^{10} cells per gram of soil respectively. Using the *Myxococcus* specific probe, members of this subgroup were estimated to vary from 2×10^{4} cells per gram of soil at 30 cm to 3×10^{4} cells per gram of soil at the 5 cm depth in the contaminated soils (Table 4.1). In comparison to the total number of eubacteria present, members of the *Myxococcus* subgroup made up no more than 0.003% of the total detectable cells. The *Myxococcus* subgroup-specific probe also hybridized strongly with uncontaminated farm soil DNA (data not shown), but accurate determination of cell number was not possible.

Hybridization using the *N. exedens* msDNA probe estimated numbers of *Nannocystis* cells (Table 4.1) in one of the contaminated soil samples at both the 60 and 30 cm depths, at 8.8×10^{5} and 2.2×10^{5} cells per gram of soil respectively. At a depth of 5 cm, *Nannocystis* cells were estimated at 1×10^{6} to 3×10^{6} cells per gram of soil in each of the four contaminated sites tested, representing less than 1% of the total eubacteria present in those soils. *N. exedens* accounted for less than 0.1% (9.8×10^{7} cells per gram of soil) of the total eubacteria present (2.5×10^{11}) in the farm soil sample. The number of *N. exedens* cells either decreased or were absent at depths deeper than 5 cm, except for the 60 cm LH sample (Table 4.1), which indicated a cell number of 8.8×10^{5} cells per gram of soil. The cell density of the Nannocystis subgroup was highest in the soils contaminated with TCE.

Table 4.1. Quantification of bacteria in soil samples

Soil	Depth	No. of bacteria per gram soil			
		Total Eubacteria	*Myxococcus* subgroup	*Nannocystis* subgroup	*Chondromyces* subgroup
LH	5 cm	2.6×10^{10}[a]	1.4×10^{4} (<0.001%)[b]	3.2×10^{6} (0.01%)	ND
	30 cm	3.7×10^{8}	9.2×10^{3} (0.003%)	2.2×10^{5} (0.06%)	ND
	60 cm	3.5×10^{9}	ND[c]	8.8×10^{5} (0.03%)	ND
LN	5 cm	1.0×10^{10}	1.5×10^{4} (<0.001%)	1.3×10^{6} (0.14%)	ND
	30 cm	4.0×10^{8}	7.3×10^{3} (0.002%)	ND	ND
	60 cm	9.2×10^{8}	ND	ND	ND
PH	5 cm	2.3×10^{10}	1.6×10^{4} (<0.001%)	1.0×10^{6} (0.0.9%)	ND
	30 cm	1.1×10^{9}	3.2×10^{3} (<0.001%)	ND	ND
	60 cm	1.1×10^{9}	ND	ND	ND
PN	5 cm	2.0×10^{10}	3.6×10^{4} (<0.001%)	3.2×10^{6} (0.29%)	ND
	30 cm	1.1×10^{9}	2.2×10^{4} (0.002%)	ND	ND
	60 cm	ND	ND	ND	ND
Farm Soil	5 cm	1.4×10^{11}	NA[d]	9.8×10^{7} (0.07%)	ND
	5 cm	2.2×10^{11}	NA	6.8×10^{6} (0.003%)	ND
	5 cm	3.8×10^{11}	NA	1.4×10^{7} (0.004%)	ND

[a] calculated number of cells per gram of soil
[b] percentage of myxobacteria present compared to the total number of eubacteria detected
[c] ND, below detection limit
[d] NA, not quantifiable

Release of $^{14}CO_2$ from labeled aromatic compounds

Since the soil hybridization data suggested that myxobacteria may be present in soils contaminated with TCE, the potential of myxobacteria to mineralize aromatic and polyaromatic hydrocarbons was investigated. Three myxobacteria species, representing the three myxobacteria subgroups were tested: *M. xanthus* DZF1, (*Myxococcus* subgroup), *N. exedens* Na e434, (*Nannocystis* subgroup), and *S. cellulosum* So ce11 (*Chondromyces* subgroup). In addition, one non myxobacteria strain, *F. elegans* Fx e1 was examined. The compounds chosen for the mineralization experiments were catechol, naphthalene, phenol, and toluene. Three separate experiments were done for each organism and compound, and each individual experiment was done in triplicate. Results, given in Table 4.2, are reported as the percentage of radiolabeled compound added that was subsequently released as $^{14}CO_2$. After several replicates of these experiments, three strains, *M. xanthus*, *F. elegans*, and *S. cellulosum*, appeared to consistently demonstrate low levels of activity in degrading catechol as indicated by the release of ^{14}C (as $^{14}CO_2$). None of the strains tested produced levels of $^{14}CO_2$ above the killed controls for phenol or naphthalene indicating these organisms do not appear to attack these compounds under these conditions. However, since the ^{14}C of naphthalene is at the 1 position, it may be that these organisms are attacking naphthalene, but are unable to completely mineralize the compound. Strain *S. cellulosum* So ce11 was the only strain to produce consistent detectable levels of $^{14}CO_2$ from toluene. It should be noted that the levels of $^{14}CO_2$ were low, and do not appear to be significant (see discussion). The *Nannocystis* strain Na e434 did not demonstrate significant mineralization for any of the compounds tested.

Chromosomal DNA from various myxobacteria hybridizes with known catabolic genes

The pathways of biotransformation have been characterized for several compounds most notably naphthalene degradation by *Pseudomonas sp.* (Sanseverino *et al.*, 1993, Yen *et al.*, 1985, Yen *et al.*, 1988). This process is mediated by two distinct operons contained on a large catabolic plasmid, NAH7. Since three of the myxobacteria strains tested appeared to attack catechol and since conservation of sequences among dioxygenase genes has been reported (Harayama *et al.*, 1992), the naphthalene dioxygenase (*nah*A) (Simon *et al.*, 1993) and catechol-2, 3-dioxygenase (*nah*H) (Ghosal *et al.*, 1987) genes from the NAH7 plasmid were used as probes to look for potentially similar genes in a variety of

Table 4.2. $^{14}CO_2$ released from radiolabeled compounds in the presence of metabolically active cells

Strain	% of $^{14}CO_2$ released from ^{14}C labeled compounds			
	catechol	naphthalene	phenol	toluene
M. xanthus DZF1	1.57%	0[a]	0	0
N. exedens Na e434	0.06%	0.51%	0.14%	0
S. cellulosum So ce11	0.78%	0	0	2.09%
F. elegans Fx e1	0.79%	0.15%	0.134%	0

[a] indicates no $^{14}CO_2$ above killed control levels

myxobacteria species. Chromosomal DNA, digested with restriction endonucleases *Pst*I or *Sal* I, was probed with these genes under conditions of low stringency.

Several myxobacteria strains showed weak hybridization signals with the *nah*A probe (Figure 4.5). Most notably, one species, *F. elegans* Fx e1 (Figure 4.5; lane 10), showed a single band that hybridized strongly with the *nah*A probe. Additionally, *S. cellulosum* strains So ce7 and So ce11 and strain *C. fuscus* Cb fe18 (Figure 4.5; lanes 2, 3, and 7 respectively) showed single bands that hybridized much more weakly with *nah*A. Strains *S. cellulosum* So ce5, *C. apiculatus* Cm a5, Cm a15, Cm a16, *M. lichenicola*, and *M. xanthus* DZF1 all showed multiple bands that hybridized to varying degrees with *nah*A. This indicates that all of these myxobacteria strains may have DNA sequences that share some degree of relatedness to *nah*A. None of the strains tested showed hybridization to the *nah*H probe (data not shown).

N. exedens Na e434, *M. xanthus* DZF1, *S. cellulosum* So ce11, and *F. elegans* Fx e1 were also hybridized with several other well characterized catabolic genes of *Pseudomonas sp.*. These probes included: *nah*G, encoding salicylate hydroxylase (You *et al.*, 1991), *xyl*A, a subunit of the xylene monooxygenase encoded on the TOL plasmid pWWO (Suzuki *et al.*, 1991), and *tod*C1, a gene from the toluene degradation pathway of *P. putida* F1 (Zylstra *et al.*, 1989). Probes based on *xly*A, *tod*C1, and *nah*G did not hybridize to any of the strains tested at low stringency.

Cloning of a *nah*A homologue

A single chromosomal DNA fragment of *F. elegans* Fx e1 showed a strong hybridization signal with *nah*A (Figure 4.5, lane 10). The Fx e1 gene was subsequently cloned to determine the degree of nucleotide similarity to *nah*A. *Pst*I chromosomal fractions in the range of 5.7 to 6-kbp were shotgun cloned into the *Pst*I site of pUC19. A single transformant, pFL6B.86, was obtained that hybridized to the *nah*A gene. While the gene cloned from *F. elegans* hybridized strongly to *nah*A, sequence analysis of the *F. elegans* clone revealed no significant nucleotide identity with the *nah*A gene. The observed hybridization could be due to a region of 30 base pairs at the 5' end of the cloned gene that shows 60% nucleotide identity with *nah*A (data not shown).

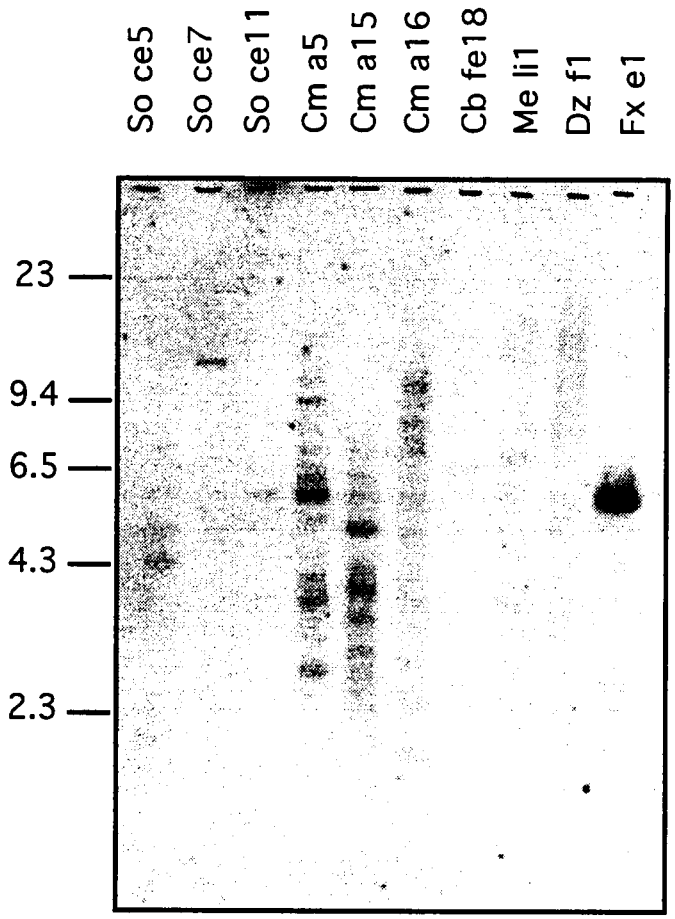


Figure 4.5. **Southern hybridization of myxobacteria chromosomal DNA with *nah*A.** Chromosomal DNAs, digested with the restriction endonuclease *Pst*I, from nine myxobacteria, and one non-myxobacterium, were hybridized with a PCR generated, ^{32}P-labeled probe based on *nah*A, the naphthalene dioxygenase gene from *Pseudomonas putida*. Molecular weight standards are given in kilobase pairs (kbp). The myxobacteria tested (lanes 1-10 respectively) were: *Sorangium cellulosum* So ce5, So ce7 and So ce11, *Chondromyces apiculatus* Cm a5, Cm a15, and Cm a16, *Cystobacter fuscus* Cb fe18, *Melittangium lichenicola* Me li1, *Myxococcus xanthus* DZF1, and *Flexibacter elegans* Fx e1.

Discussion

Using the DNA hybridization method described here, bacterial DNA sequences appear to be present in all the soil samples tested, in both uncontaminated soil and soil heavily contaminated with TCE. Probes designed to specifically detect myxobacteria DNA sequences also revealed myxobacteria to be present in all of the soil samples examined, including those with high levels of TCE. However, the myxobacteria appear to be restricted to the shallower soil depths (5 and 30 cm).

DNA hybridization experiments were also used to estimate the cell density of myxobacteria in uncontaminated soil at around 10^7 cells per gram of soil. This is much higher than previous estimates of 10^3 to 10^5 cells per gram of soil, which were based on culture techniques (McCurdy, 1969, Singh, 1947). Therefore, the DNA hybridization method appears to be more sensitive for detecting myxobacteria in environmental samples than traditional culture based methods. In comparison to the total bacterial population, members of the *Nannocystis* subgroup made up the largest percentage of myxobacteria present in the soils tested, from 0.01% for the LH sample to a high of 0.29% for the PN sample. The *Myxococcus* subgroup comprised less than 0.01%, while the *Chondromyces* subgroup and *F. elegans* Fx e1 were undetectable. Thus, in these soils, *N. exedens* strains appeared to be the dominant myxobacteria subgroup. The dearth of *Chondromyces* members might be explained by regional variation in the isolation of myxobacteria (Reichenbach and Dworkin, 1992). It should be noted that the myxobacteria form resistant spores under adverse conditions, thus it is not clear whether the myxobacterial DNA detected in these soils is derived from metabolically active, vegetative cells, or derived from their spores that have been deposited in the soils. Interestingly, the percentage of *Nannocystis* appeared to be much higher in the contaminated soils than in the uncontaminated farm soil. This could be due to better survival and persistence of these bacteria compared with the general bacterial population in the contaminated soils. Thus, high survival in the presence of aromatic hydrocarbons, plus the observation that *Nannocystis* increased in cell density during the *in vitro* mineralization experiments, suggests that some myxobacteria may be capable of metabolic activity in the presence of aromatic hydrocarbons. The lack of myxobacteria gene sequences at deeper depths could make them suitable control markers for the study of deep subsurface soils obtained from core samples. Thus the detection of myxobacteria sequences from a deep core sample may

suggest contamination of the sample by bacteria from shallower depths instead of being part of the community of deeper subsurface microorganisms.

As noted in the materials and methods, the contaminated soil samples LH, LN, PH, and PN were first extracted for total lipids and then subsequently extracted for total DNA. The quantity and types of lipids recovered from a soil sample can also be used to estimate the number of bacteria present and provide a general idea of the different types of microorganisms found. Thus the DNA analysis of these soil samples, reported here, complements the information gained from the lipid analysis reported earlier (Ringleberg and White, 1992) and represents the first direct field application of the simultaneous lipid-DNA extraction technique described by Kehrmeyer *et al.* (1995). The lipid analysis of these soil samples reported the total number of bacteria as 5.9×10^8 and 4.4×10^5 cells per gram of soil for the 5 and 60 cm depths respectively. These numbers contrast sharply with the cell numbers based on DNA analysis, which are reported here as 2×10^{10} to 1×10^9 respectively for the same samples (Table 4.1). In addition, Kehrmeyer *et al.* (1995) indicated that only 40-50% of DNA is recovered from soil after first being extracted for lipids, which would lead to underestimation of bacterial cell numbers based on that technique. While these investigators show a close correlation between cell numbers calculated from lipid and DNA analysis when bacteria are added exogenously to soil samples, it may be that in the natural environment, the extraction efficiency of the two classes of macromolecules is not equal. Therefore, this method of dual lipid/DNA extraction may need to be reevaluated using a variety of natural soil samples to validate that both techniques yield similar results in regard to quantification of cell numbers.

Of the four strains tested, two myxobacteria strains, *M. xanthus* DZF1 and *S. cellulosum* So ce11 and one non myxobacterium, *F. elegans* Fx e1 demonstrated consistent, low-levels of catechol mineralization (Table 4.2). The low level of catechol and toluene mineralization observed in these experiments may have been due to the presence of a dilute source of carbon and energy (casitone) in the mineralization medium in addition to the ^{14}C labeled aromatic substrate, thus the poor utilization of the aromatic compound could be due to preferential use of the casitone, or catabolite repression. It is also possible that no true mineralization occurred and that the $^{14}CO_2$ released was from trace amounts of ^{14}C labeled contaminants present in the radiolabeled aromatic compound, as the compounds are graded at 98% radiopurity (Sigma Chemical Co., St. Louis, MO). Rosseló-Mora *et al.* (1994) suggested that the ability of *Pseudomonas stutzeri* strains to mineralize naphthalene is enhanced in those strains isolated from environments "enriched"

for aromatic contaminants. Thus, it may be necessary to isolate myxobacteria from contaminated sites which may lead to the discovery of strains that are more efficient at degrading aromatic hydrocarbons than the strains tested here.

From Figure 4.5, it appears that the *nah*A probe hybridized strongly to a single restriction fragment of *F. elegans* Fx e1. The chromosomal DNA from *F. elegans* Fx e1 that hybridized to *nah*A was cloned and sequenced. Sequence comparison of the gene cloned from Fx e1 with *nah*A failed to reveal any nucleotide identity, therefore, it is possible that the clone is not a true *nah*A homologue. The region of *nah*A used as a probe was part of the iron ferredoxin center of the naphthalene dioxygenase enzyme. It is possible that the gene isolated from *F. elegans* is an iron sulfur protein, or other protein involved in redox reactions such as a member of the electron transport chain. It should be noted that a protein and nucleotide homology search of sequences in the Genbank database did not indicate any significant homology to the gene cloned from *F. elegans* with either DNA or proteins in that database. Therefore, studies that rely on gene probing for the detection of homologous genes in environmental samples may need to bolster hybridization results with additional data, such as DNA or protein sequences, or enzyme assays to confirm the detection of homologous gene sequences.

In conclusion, this paper presents the first use of DNA hybridization to detect and quantify myxobacteria in soils. There are several advantages of using msDNA for such an application. Unlike other characterized myxobacteria genes, msDNA sequences are specific for particular subgroups of myxobacteria, yet, are inclusive of greater than 98% of all myxobacteria strains. For example, every member of the *Myxococcus* subgroup tested, representing seven different genera, has the same Mx162 retron (Rice and Lampson, 1995). In addition, msDNA is made in large quantity by the bacterium and is easy to detect, isolate and label. Since cloning of the msDNA is not necessary, a hybridization probe can be made which lacks contaminating vector sequences that might otherwise lead to non-specific hybridization. This study also reports the initial assessment of the capacity of myxobacteria to mineralize aromatic hydrocarbons and their prevalence in contaminated soils. Since the myxobacteria make up a significant portion of the soil environment, further studies into the full ecological diversity and metabolic capabilities of this group of bacteria may become an important avenue of study in the search for better and more efficient bioremediative bacteria and genes.

Acknowledgments

The authors would like to thank Dr. Charles Stuart for the farm soil sample, Ray Stapleton for the 16S oligo and standards, and Sue Pfiffner for DNA extractions of the Savannah River Facility soil samples.

This project was supported by the WMREI/Center of excellence, University of Tennessee, grant number R01253654 (to B. L.) and by DOE-WSRS grant number R0125389/01993837 (to G. S.).

CHAPTER V

Multiple Repeats of the Retron Encoded *msd* Gene in *Nannocystis exedens*

Introduction

A recent study suggests that insertion of retrons Ec67 and Ec86, both of which are inserted into identical sites in a phage, resulted in the loss of host (phage) sequences (Lim, 1995). Thus, based on the distribution in *E. coli* and the similarity of the RT gene to retrotransposable elements, retrons could be a new class of selfish DNA. In contrast to the distribution of retrons observed in *E. coli*, a study of the myxobacteria indicated that retrons were inherited in a vertical rather than horizontal fashion (Rice and Lampson, 1995). Thus, based on the distribution in *E. coli*, retrons appear to be mobile elements, which, based on the studies of the myxobacteria, are stably inherited genes once inserted into a host genome.

A common feature of retrons, in addition to the *msr*, *msd*, and RT genes, is that they appear to be single-copy, chromosomally encoded elements. While some *Myxococcus xanthus* strains have been shown to harbor two different retrons, Mx65 and Mx162 (Dhundale *et al.*, 1988b, Inouye *et al.*, 1990), no bacterium has been demonstrated to have more than a single copy of the same retron locus. Nor has data been previously presented indicating the presence of partial or truncated copies of a retron (i.e. a chromosomal retron incapable of msDNA synthesis). Thus, if retrons are mobile elements, they may have specific insertion sites, much like the R2Bm element of *Bombyx mori* (Luan *et al.*, 1993) or group II introns (Moran *et al.*, 1992) This study presents new information suggesting that retrons in the myxobacteria may be mobile elements, generating multiple, partial copies of the *msd* gene in some *Nannocystis exedens* strains. Sequence analysis of the *msd* repeats is presented along with speculation about the origins and implications of the *msd* repeats regarding the evolution of the *Nannocystis* subgroup.

Materials and Methods

Bacterial strains and plasmids

The myxobacteria strains used were: *Myxococcus xanthus* DZF1, *Melittangium lichenicola*, *Nannocystis exedens* strains Na e3, Na e24, Na e30, Na e39, Na e434, Na e465, Na e585, Na e619, and Na e642. All *Nannocystis* strains were kindly donated by Hans Reichenbach. *Melittangium lichenicola* was provided by Larry Shimkets (University of Georgia, Athens, GA). *E. coli* strain DH5α was used as a host bacterium in conjunction with the plasmid pUC9 (Yannish-Perron *et al.*, 1985) for cloning experiments. *M. xanthus* DZF1 and *Melittangium lichenicola* were grown at 30°C in CYE broth (Reichenbach and Dworkin, 1992). *Nannocystis* strains were grown at 30°C on MD1 plates (0.3% Difco Casitone; 0.07% $CaCl_2 \cdot 2H_2O$; 0.2% $MgSO_4 \cdot 7H_2O$; trace elements; 1.5% agar) (Reichenbach and Dworkin, 1992). *E. coli* was grown at 37°C in modified Luria-Bertani medium (Maniatis *et al.*, 1989). When necessary, ampicillin (50 μg/ml) was added for selection of plasmids.

Detection of the chromosomally-encoded retron locus

Chromosomal DNA was isolated from each myxobacteria strain and digested with the restriction endonuclease *Pst*I as described (Rice *et al.*, 1993). Chromosomal DNA digests were separated on 0.7% agarose gels and transferred to nitrocellulose filters. Southern hybridizations were carried out at high stringency (50% formamide-5X SSPE-5X Denhardt's-0.3% SDS) at 42°C. msDNA from *Nannocystis exedens* strain Na e434 was prepared for use as probe as follows. DNA was extracted from *Nannocystis* cell pellets by the method of Birnboim and Doly (1979), treated with RNase A, and separated by electrophoresis on a 5% acrylamide gel. The isolated msDNA band was subsequently excised from the gel and purified by electroelution. Purified msDNA and restriction endonuclease fragments of retron clones were labeled for use as hybridization probes with [α-^{32}P] dCTP using the random hexamer labeling procedure (Feinberg and Volgstein, 1983).

Cloning of repeat sequences

Chromosomal DNA from *Nannocystis exedens* Na e434 was digested with the restriction endonuclease *Pst*I, electrophoresed on 0.7% agarose, and size fractionated. Fractions were electroeluted, electrophoresed on 0.7% agarose, and the resulting "ladder"

of chromosomal DNA was transferred to a nitrocellulose membrane. Fractions that hybridized with the msDNA probe were cloned into the dephosphorylated *Pst*I site of pUC9. Resulting transformants were screened by colony hybridization using the msDNA as a hybridization probe.

DNA sequencing

DNA sequencing of retron clones in pUC9 was performed by the dideoxy chain termination method of Sanger *et al.* (1977) using ΔTaq Version 2.0 sequencing kit from USB (United States Biochemical, Cleveland, OH). The DNA sequence was determined for both strands. Double-stranded templates were prepared by alkali denaturation as described in the sequencing kit (USB). The nucleotide sequence of the DNA portion of the msDNA was determined directly by the method of Maxam and Gilbert (1980). msDNA was isolated from freshly grown cells, treated with RNase A, gel purified on a 5% acrylamide gel, and labeled at the 5' end with [γ-^{32}P] ATP using T4 Polynucleotide Kinase (New England Biolabs, Beverly, MA). Electrophoresis of sequencing reactions was performed using 6% (wt/vol) acrylamide-8M urea gels. Sequences were analyzed using DNA Inspector (version IIe), the GenBank sequence database (version 7.3), and the Genetics Computer Group of the University of Wisconsin (Devereux *et al.*, 1984).

Results

Nannocystis exedens strains contain two msDNAs

It was reported previously (Rice and Lampson, 1995) that all nine strains of *Nannocystis exedens* contain msDNA, and thus presumably contain a corresponding retron responsible for msDNA production. All of the strains tested (e.g. *N. exedens* strain Na e434) appeared to harbor two msDNA species. When msDNA was isolated from strain Na e434 and electrophoresed on a 5% acrylamide gel, two distinct extrachromosomal, satellite bands were observed (Figure 5.1) and thus either represent two distinct msDNAs or two different forms of the same msDNA. Based on molecular weight markers, the two msDNAs, labeled "A" and "B", are approximately 180 bases and 140 bases respectively. It should be noted that when the msDNAs are electrophoresed on the acrylamide gel, they are present as single-stranded DNAs with complex secondary structures and therefore may not migrate true to their molecular weight. To clarify if msDNAs A and B are different versions of the same msDNA or if the two are unrelated elements, the DNA sequence of

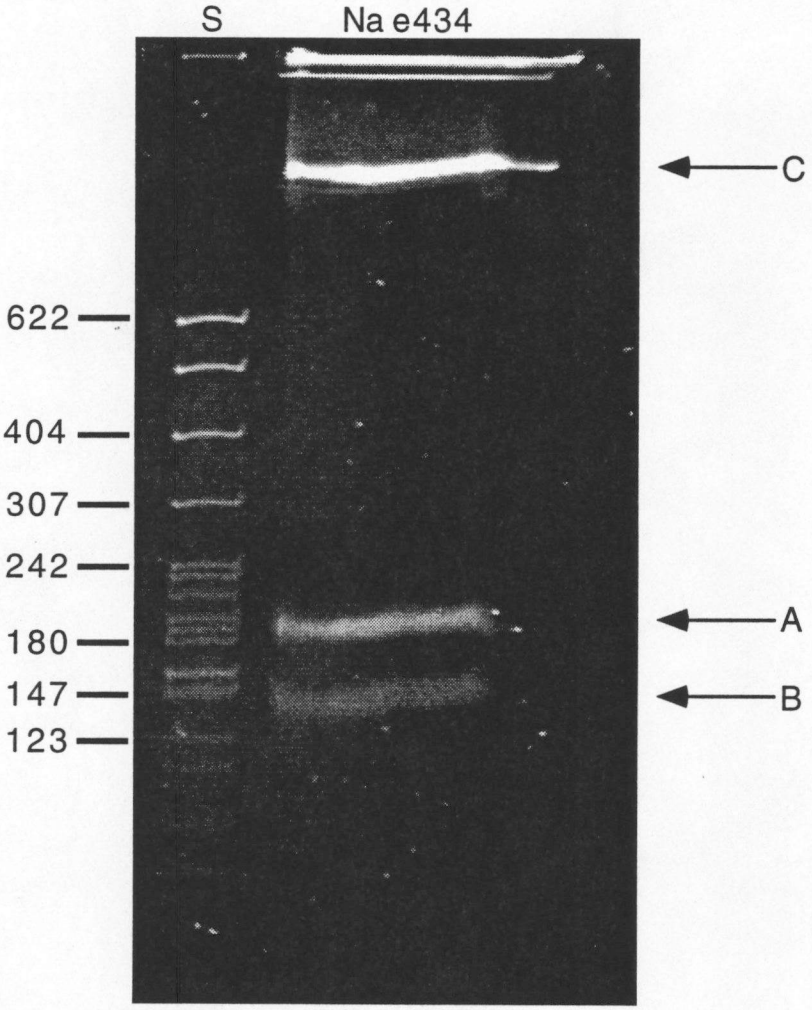


Figure 5.1. **Ethidium bromide stained msDNA from *Nannocystis exedens* Na e434.** DNA was extracted from Na e434, treated with RNase A, electrophoresed on a 5% acrylamide gel, and stained with ethidium bromide. The high molecular weight band (C) is chromosomal DNA, and (A) and (B) denote msDNAs A and B. Molecular weight standards (S) are given in base pairs.

both was determined. Gel purified msDNA types A and B were 5' end-labeled, and their sequences determined by the chemical cleavage method of Maxam and Gilbert (1980). This was done for two reasons. First, if the nucleotide sequence of msDNA is known, specific analysis of the differences between the two msDNAs could be made. Second, since the DNA portion of msDNA is encoded by the *msd* gene of the retron locus, knowing the msDNA sequence would aid in identifying the retron locus when the chromosomal gene is cloned and sequenced. Sequence analysis of msDNA A indicated that this msDNA consisted of 160 bases and had the potential to form a very stable stem-loop structure, typical of msDNAs (Figure 5.2). While sequencing of msDNA B was incomplete, enough sequence data was obtained for comparison with msDNA A, which indicated no nucleotide identity between the two, and therefore it was concluded that msDNA A and msDNA B are separate elements (data not shown). These data therefore indicate that *N. exedens* strain 434 contains two independent msDNA-encoding retrons.

Some *N. exedens* strains contain multiple chromosomally encoded *msd* genes

Southern hybridizations of chromosomal digests were performed, using either ^{32}P-labeled msDNA A or msDNA B as probes to identify the chromosomal fragment containing the retron locus. It should be noted that until this time, all retrons studied have been shown to be present as single copy elements and typically yield a single chromosomal restriction fragment that hybridizes with the msDNA probe (Rice and Lampson, 1995, Rice *et al.*, 1993). This was also found to be the case when msDNA A was hybridized against total chromosomal DNA from strains Na e3, Na e24, Na e39, and Na e 642 (Figure 5.3, lanes 2, 3, 5, and 10). A single restriction fragment hybridized with this probe, indicating the presence of a single copy of the chromosomally encoded retron. Strain Na e30 (Figure 5.3, lane 4) showed two bands that hybridized to the msDNA A probe, but the corresponding retron could still be present in a single copy, with a *PstI* site in the *msd* gene, which codes for the DNA portion of msDNA. The alternative is that there are two copies of the retron present. In contrast, strains Na e434, Na e465, Na e585, and Na e619 (Figure 5.2, lanes 6, 7, 8, and 9) showed multiple bands that hybridized to the msDNA A probe. This suggested that, unlike any other retron-encoding bacteria previously described, these strains contain multiple copies of the same retron. Additionally, there was variability in the hybridization intensity of the different bands. The difference in hybridization intensity could indicate that there was some nucleotide differences between the multiple copies of the retron. To confirm that the multiple banding was not artifactual,

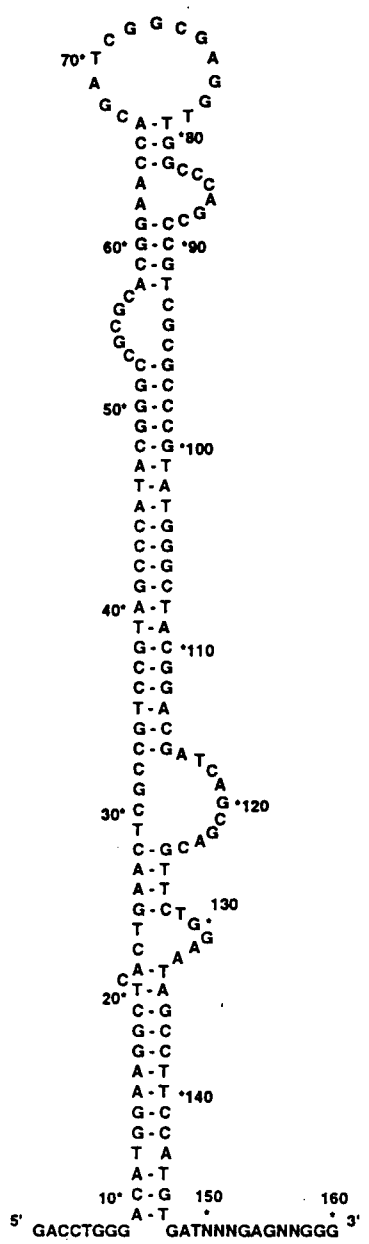


Figure 5.2. **Sequence and secondary structure of msDNA A from *Nannocystis exedens* Na e434.**. The DNA portion of msDNA A from *N. exedens* Na e434 was determined by the method of Maxam and Gilbert (1980). Bases designated "N" were ambiguous by this sequencing method. Note the presence of a stem structure, with bulge regions and a central 14 base loop.

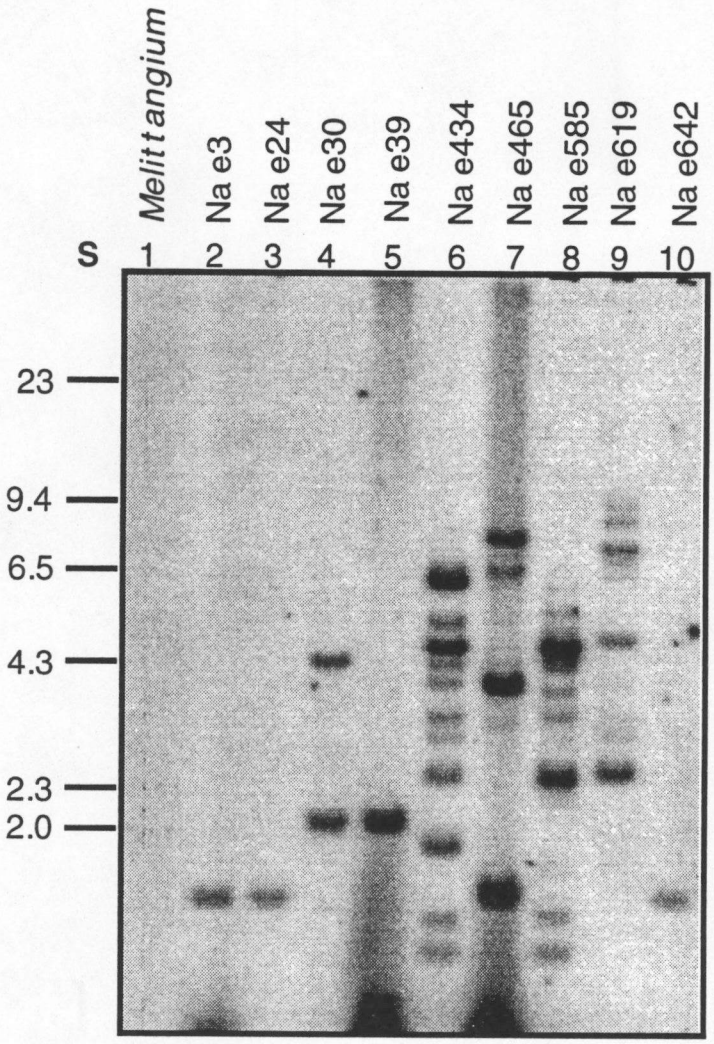


Figure 5.3. **Southern hybridization of *Nannocystis* chromosomal DNA using msDNA A as a probe.** Chromsomal DNA from *Melittangium lichenicola* and nine *Nannocystis* strains was digested with the restriction endonuclease *Pst*I, electrophoresed on 0.7% agarose, and transferred to a nitrocellulose membrane. As expected, the negative control *M. lichenicola*, did not hybridize with the msDNA A probe. Molecular weight standards (S) are given in kilobase pairs.

the same chromosomes were digested separately with two other restriction endonucleases, *Bam*HI and *Sal*I (data not shown). Again, strains Na e434, Na e465, Na e585, and Na e619, showed multiple bands when hybridized with msDNA A, indicating that these strains did indeed have multiple copies of the msDNA coding region.

Hybridization against the same *Pst*I chromosomal digests, shown in Figure 5.2, using msDNA B as a hybridization probe yielded similar results. Strain Na e3 showed a single restriction fragment that hybridized with the msDNA B probe, while strains Na e434 and Na e585 had multiple bands that hybridized with this probe (data not shown). It should be noted that while the results were similar, the pattern of bands hybridizing with msDNA A and msDNA B was different, reinforcing that these two msDNAs are the product of two independent chromosomally encoded retrons. The negative control, *M. lichenicola* (Figure 5.2, lane 1), did not hybridize with either msDNA probe as expected based on previous work (Rice and Lampson, 1995).

Cloning and sequencing of reiterated retron sequences

Initially, the presence of multiple bands was taken to indicate that multiple copies of the retron were inserted in different locations in the chromosome, and that these copies were due to replicative transposition of the element. Additionally, since there was variability in hybridization intensity of different bands, it was thought that the transposition events had occurred some time ago and that divergence of the copies would account for the differences in hybridization. Therefore, to test these hypotheses, six chromosomal fragments from *N. exedens* Na e434 that hybridized to the msDNA A probe were cloned into the *Pst*I site of plasmid pUC9 (Table 5.1), with the intention of sequencing the retrons to determine the degree of sequence divergence and to analyze the 5' and 3' ends to look for direct or inverted repeats which might be suggestive of a mechanism of insertion.

To confirm that each of these clones represented full length retron elements, a 2.5-kbp *HindIII-Bgl*II restriction endonuclease fragment from the clone designated p1A.42 was radiolabeled and used as a hybridization probe against total chromosomal DNA which had been digested separately with the restriction endonucleases *Pst*I, *Sal*I, and *Bam*HI. Surprisingly, only a single 5-kbp *Pst*I fragment in Na e434 hybridized with this probe (Figure 5.4, lane 4). This 5-kbp fragment is the same size as the original fragment from which the 1A.42 fragment was cloned (Figure 5.3, lane 6; Table 5.1). This probe also only hybridized to a single fragment in strain Na e585 (Figure 5.4, lanes, 8, 9, and 10). An extremely weak hybridizing band was detected in strains Na e3 (Figure 5.4, lanes 1, 2,

Table 5.1 *Nannocystis exedens* Na e434 clones containing msDNA A repeat sequences

Clone	Insert size	Repeat length[a]	Position[b]
7A.140	5.0-kbp	160	whole element
1A.42	5.0-kbp	22	5-26
1A.21	4.2-kbp	51	6-56
11B.23	3.0-kbp	31	1-32
12A.30	2.0-kbp	42	74-116
13A.152	1.7-kbp	41	25-65

[a] length is reported as the number of base pairs

[b] relative to the 5' end of the msDNA sequence as determined by the method of Maxam and Gilbert (1980)

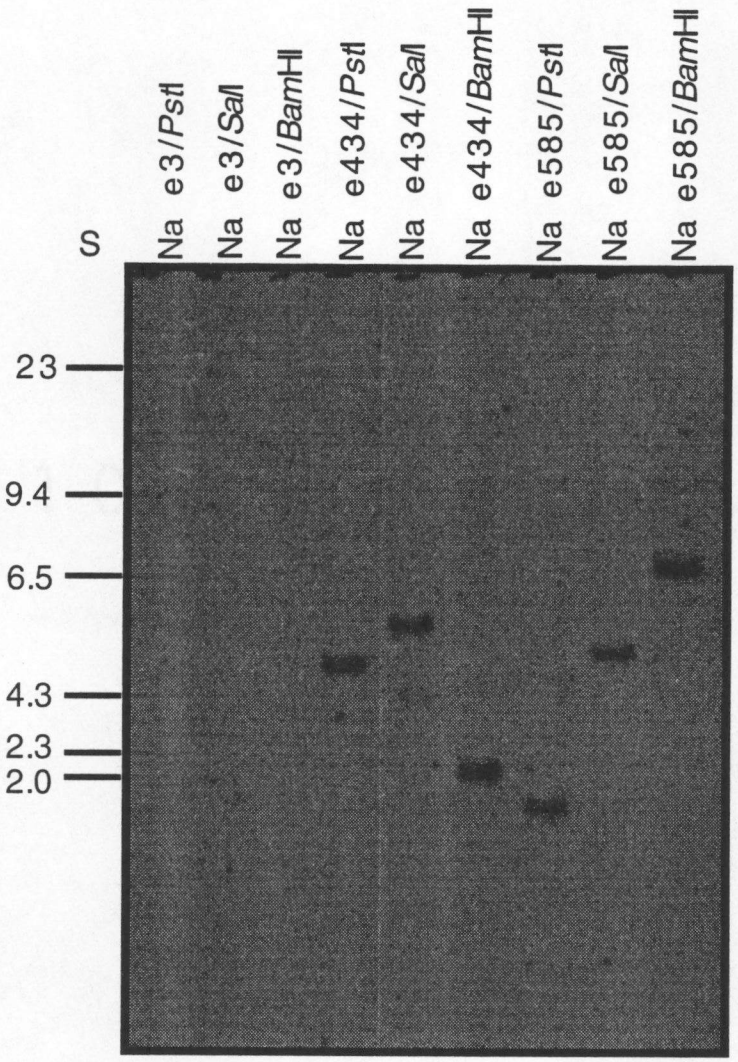


Figure 5.4. **Hybridization of *Nannocystis* strains with the cloned fragment, p1A.42, from strain Na e434**. Chromsosmes from three strains of *Nannocystis exedens*, Na e3, Na e434, and Na e585, were digested with restriction endonucleases *Pst*I, *Sal*I, and *Bam*HI, electrophoresed on 0.7% agarose, transferred to nitrocellulose, and hybridized with a 2.5 kilobase pair (kbp) *Bgl*II-*Hind*III fragment from p1A.42. Molecular weight standards (S) are given in kbp.

and 3) and a more strongly hybridizing band in strain Na e39 (data not shown). These results contrast with the data based on the msDNA probes where multiple bands were observed when hybridized against chromosomal DNA. Interestingly, the 4.2-kbp fragment from strain Na e39 that hybridized with the p1A.21 probe did not hybridize to either msDNA A or B.

Since the msDNA probe is relatively small (160 bases) compared to the p1A.42 probe (2.5-kbp), it was postulated that the reason only single bands were observed when using the 2.5-kbp *Bgl*II-*Hind*III fragment of p1A.42 as a hybridization probe was due to the extensive amount of flanking DNA sequence in the probe. This implied that hybridization was not due to the presence of multiple, full-length retrons, but rather to the presence of short repeats of the *msd* gene which codes for the DNA portion of msDNA. Thus, most of the clones would predominately include flanking sequences not involved in msDNA production. Therefore, the hybridization of the 2.5-kbp probe from p1A.42 to chromosomal DNA from Na e39 and the lack of hybridization using the msDNA probe can be explained if strains Na e39 and Na e434 have similar chromosomal regions that differ by the presence (Na e434) or absence (Na e39) of the msDNA repeat.

Similarly, when the clone designated p12A.30 was used as a hybridization probe against chromosomal DNA, a single restriction fragment in Na e434 was observed (data not shown), corresponding to the fragment from which it was cloned. This probe also hybridized to a single restriction fragment in strain Na e585, but showed no hybridization with any of the other strains (data not shown). Therefore, it was postulated that the multiple hybridizing bands observed in strains Na e434, Na e465, Na e585, and Na e615 were due to the presence of short repeats of the *msd* gene and was not due to the presence of multiple copies of full length retrons. It should be noted though, that at least one full length copy of the retron, including the *msr*, *msd*, and RT genes, must be present to produce msDNA (Dhundale *et al.*, 1988a, Inouye *et al.*, 1989).

Sequence analysis of msDNA repeat elements

To test the prediction of multiple, short repeats, and to determine which clone contained the full length retron, several of the clones from Na e434 were sequenced by the chain termination method (Sanger *et al.*, 1977). Six clones (Table 5.1) were sequenced and analyzed for sequence identity with the DNA portion of msDNA and for putative open reading frames (ORFs) that show amino acid homology to bacterial reverse transcriptases (RTs). Five of the clones were found to contain short regions, ranging from 22 to 51 bp,

that showed nearly 100 percent nucleotide identity to the DNA portion of msDNA A (Figure 5.5). Analysis of the 5' and 3' regions flanking the msDNA repeat sequences of clones p1A.21, p12A.30, p13A.152, p1A.42, and p11B.23 revealed no direct or inverted repeats, nor was there any conservation of flanking sequences when the different clones were compared. To identify the clone containing the full length retron, a synthetic sequencing primer, complementary to the 3' end of the msDNA A was made. Using this primer, one clone, p7A.140, yielded a sequencing ladder. In addition, the observed sequence contained the whole DNA region of msDNA (Figure 5.5). Therefore, it is speculated that the clone p7A.140 contains a full length retron element responsible for production of msDNA A in *N. exedens* strain Na e434. These experiments have not yet been performed for the msDNA B repeat sequences, but it is likely, based on the hybridization results, that similar results would be obtained.

Discussion

That strains of *Nannocystis exedens* contain two different msDNAs is not novel in itself, since it has been reported previously that some *Myxococcus xanthus* strains also contain two distinct msDNAs, Mx162 and Mx65 (Dhundale *et al.*, 1988b, Inouye *et al.*, 1990). What makes the discovery of msDNA in the *Nannocystis* strains novel and very exciting is that these strains appear to contain multiple, short repeats on the chromosome that are identical to part of the *msd* gene. However, only sequences corresponding to the *msd* gene were used as a hybridization probe. Thus, it may also be that the RNA coding region, *msr*, may similarly be reiterated in these strains. These studies are currently in progress.

The discovery of multiple copies of the *msd* gene in *Nannocystis* suggests that retrons may be transposable elements, generating multiple abbreviated copies much in the same manner as LINE (Di Nocera *et al.*, 1994) or SINE (Oshima *et al.*, 1993) elements, or they may be related to the interspersed DNA families which include ERICs (Hulton *et al.*, 1991) and BIMEs (Bachellier *et al.*, 1994). The possibility that retrons are mobile is supported by the analysis of distribution patterns and codon bias of the retrons in *E. coli*. However, *E. coli* retrons are present solely as single copy, full length elements; no msDNA repeats similar to those presented in this paper have been reported before in any other bacteria. The *msd* repeats do not appear to be the result of incomplete retrotransposition, as is observed in conjunction with LINE and SINE elements. Abortive retrotransposition of

```
msDNA A   AGACCTGGGA CATGGAAGGC TCACTGAACT CGCCGTCC GT AGCCCATACG GGCCGCGCACG GAACCACGAT CGGCGAGGTT
p7A.140   AGACCTGGGA CATGGAAGGC TCACTGAACT CGCCGTCC GT AGCCCATACG GGCCGCGCACG GAACCACGAT CGGCGAGGTT
p1A.42        CTGGGA CATGGAAGGC TCACTG
p11B.23    GACCTGGGA CATGGAAGGC TCACTGAACT CG
p1A.21         TGGGA CATGGAAGGC TCACTGAACT CGCgGcCC GT AGCCCATACG GGCCGC
p13A.125                          TGtcCT CGCCGTtC GT AGCCCATACG GGCCGCGCACG GAAC
p12A.30                                                                        CGAGGTT


msDNA A   GGCCCAGCCC GTCGCGCCCG TAT-GGGCTAC GGACGATCAG CGACGTTCTG GAATAGCCTTC CTTGTGATNN NGAGNNGGG
p7A140    GGCCCAGCCC GTCGCGgCCC Gta-GGGCTAC GGACGATCAG CGACGTTCTG GAATAGCCTTC CTTGTtATGG AtgGTGGGG
p12A.30   GGCCCAGCCC GTCGCGgCCC GTAtGGGCTAC GGACG
```

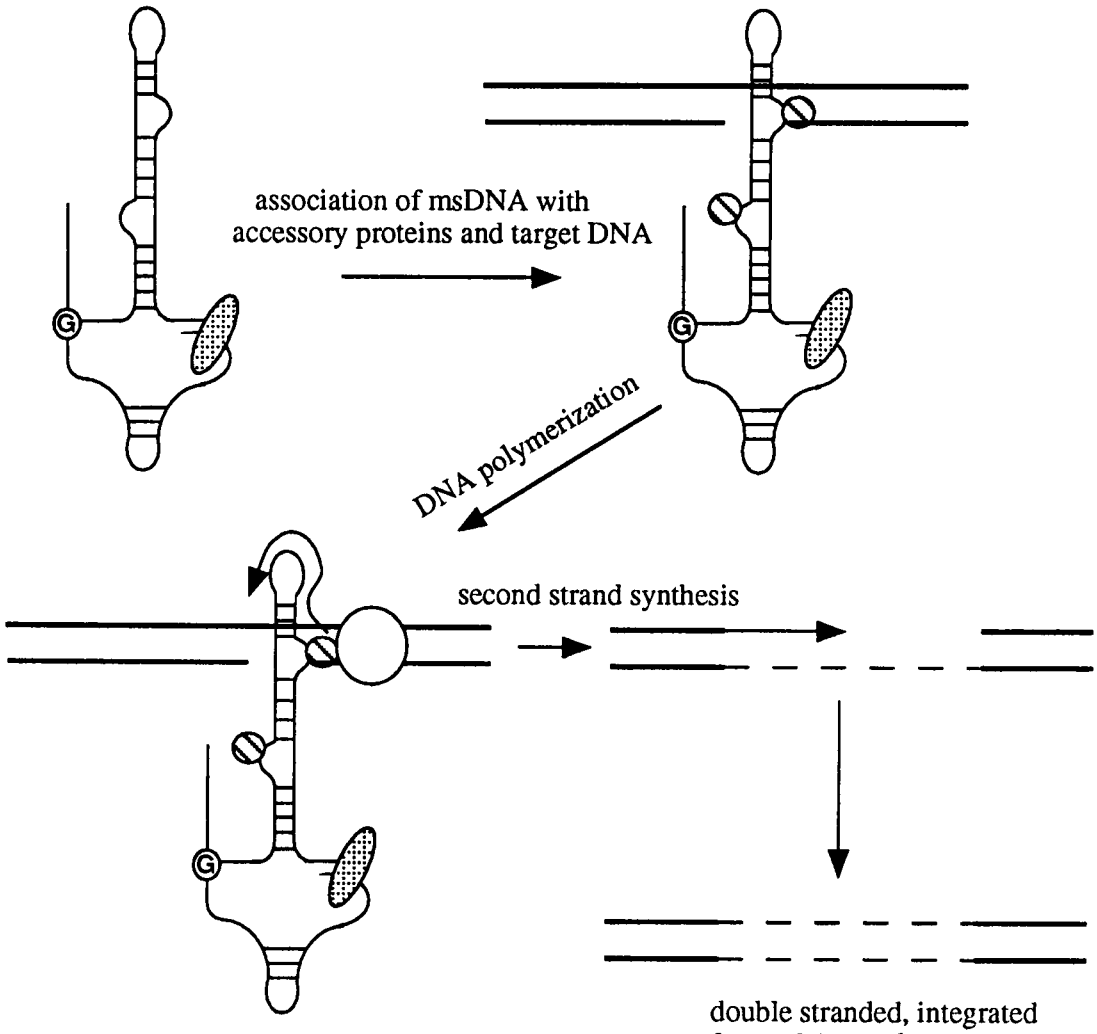
Figure 5.5. **Alignment of *msd* repeat sequences with the DNA portion of msDNA A.** The six clones from *N. exedens* strain Na e434 are compared with the sequence of the DNA region of msDNA. Bases in the repeats that differ from the msDNA sequence are shown in lower case letters.

LINEs and SINEs results in integrated repeats which have conserved 3' ends and variable 5' ends. Also, unlike the families of previously described repetitive DNAs, the *msd* repeats do not appear to consist of palindromic sequences, lack dyad symmetry, are not flanked by direct or inverted repeats, and show no conserved flanking sequences. Thus, the *msd* repeats may represent a new class of bacterial repetitive DNA, RRADs (Repetitive Retron Associated DNAs).

The lack of conserved flanking sequences also does little to suggest a mechanism for the generation of the *msd* repeats or identify a preferred insertion site. Does insertion of the *msd* repeat result in a loss of host sequences as is the case for the retrotransposable element R2Bm (Luan *et al.*, 1993), or do the repeats insert directly from the 3' end of nicks in the host chromosome? Study of similar strains in which one strain has an *msd* repeat, and the other strain does not, such as strains Na e39 and Na e434, may help elucidate a mechanism of insertion of the msDNA repeat elements. This type of study was performed for *E. coli* retrons, and lead to the discovery that insertion of retron Ec107 was associated with the deletion of a 34 bp intergenic region (Herzer *et al.*, 1992, Kawaguchi *et al.*, 1992). Similarly, insertion of retrons Ec67 and Ec86 resulted in the loss of 2.5-kbp of the associated phage sequences (Lim, 1995).

By drawing on previous work with other mobile elements and some additional data concerning retrons, a mechanism for generation of the *msd* repeats is proposed here (Figure 5.6). It has been demonstrated that the reverse transcriptase of R2Bm (Luan *et al.*, 1993) and group II introns (Zimmerly *et al.*, 1995) has DNA endonuclease activity which is responsible for single-stranded nicking of the target site for insertion of the element. Second strand cleavage seems to be similarly mediated by RT, but requires an RNA cofactor (template RNA?). Interestingly, it has also been suggested that either msDNA, the retron-encoded RT, or the msDNA-RT complex has endonuclease activity and is capable of cleaving a single-stranded 3'-5' DNA phosphodiester bond in msDNA (Lim, 1992). msDNA also appears to bind proteins involved in mismatch repair of DNA (Maas *et al.*, 1994). Likewise, it has been shown that bacterial repetitive elements are capable of binding various cellular proteins such as gyrase (Yang and Ames, 1988), integration host factor (Oppenheim *et al.*, 1993), and DNA polymerase I (Gilson *et al.*, 1990). There is also some indication that RT remains complexed with msDNA at the 3' ends of the RNA and DNA strands (Lampson *et al.*, 1991b). Therefore, it may be that the msDNA-RT complex, in conjunction with mismatch repair proteins, is directed to either preexisting chromosomal nicks, or specific chromosomal target sequences in which the msDNA-protein complex

Figure 5.6 **Model for insertion of *msd* repeats.** Two similar, but slightly different models, for generation of multiple repeats of the *msd* gene and for transposition of full length retrons are presented here. (A) *msd* repeats may be generated by directly using the DNA portion of msDNA as a template at a chromosomal break. The thick double line represents chromosomal DNA. Reverse transcriptase (stippled oval) associated with msDNA binds cellular mismatch repair proteins (lined circles) at the bulge regions in the stem and is directed to a chromosomal break, or the msDNA protein complex creates a break at a particular target site. Using the 3'-OH as a primer, subsequent polymerase activity (open circle), which may be either reverse transcriptase or DNA Pol I, then copies a region of the msDNA to "repair" the chromosomal break (dashed line). Second strand synthesis (arrow) is moderated by either RT or by a host polymerase, which results in an insertion of *msd* sequences (dashed lines).


association of msDNA with
accessory proteins and target DNA
DNA polymerization
second strand synthesis
double stranded, integrated
form of the *msd* sequence

subsequently generates a nick. Once the msDNA-protein complex becomes associated with the insertion site, polymerase activity, using the 3'-OH of the nick as a primer and msDNA as a template, fills in the gap to generate a partial *msd* repeat. The polymerase activity could be provided by either DNA polymerase I or reverse transcriptase. Subsequently, second strand synthesis occurs and completes the insertion of the *msd* sequences. It should be pointed out that this mechanism is an attempt to explain how the *msd* repeats are generated and is not sufficient to account for insertion of full length, single copy retron elements.

Regardless of the mode of dissemination, repetitive elements exert a great deal of control over genomic processes including: gene regulation at both the levels of transcription (Bachellier *et al.*, 1994) and translation (Stern *et al.*, 1988), genome rearrangements (Gilson *et al.*, 1984), and genome organization (Bachellier *et al.*, 1993). Thus it is equally likely that the msDNA repeat elements reported here may have similar functions in the evolution and generation of genetic diversity of the host bacterium. Interestingly, the *Nannocystis* subgroup of myxobacteria is believed to be evolving at a faster rate than the other myxobacteria, and is the only bacterial group which has been demonstrated to contain multiple copies of the *msd* gene. As noted above, msDNAs with mismatches in the stem region are capable of binding proteins involved in mismatch repair, and over expression of those msDNAs leads to an increase in mutation frequency in *E. coli* (Maas *et al.*, 1994). The question that remains to be answered is whether msDNA, either by sequestering key repair proteins or by disruption of important cellular genes by insertion of the *msd* repeat has increased the mutation rate of this genus, or alternatively, does a high rate of mutation in the *Nannocystis* allow for generation of the *msd* repeats?

In conclusion, this report presents data that provide further evidence that retron-encoded msDNAs may be mobile elements, and is the first report of multiple, incomplete copies of retrons in the bacterial chromosome. Additionally, insertion and generation of msDNA repeats may be mutagenic and potentially important factors in affecting the evolution of prokaryotes. The mechanism for generation of the short repeats and their potential presence in other bacteria are intriguing questions deserving attention which may yield exciting clues to the function and origin of retron elements in bacteria.

Summary

It appears to me that there are two areas that need to be focused on to answer the above questions about retrons and msDNA. The two are not mutually exclusive, and one could easily see how the two areas fit together, but until more specific data are available about either msDNA or retrons, they can be treated separately. First, there is the question of mobility or transposition of whole retron elements. This is a popular theory to explain the presence of retrons; they are selfish DNAs, responsible for their own propagation. This theory fits nicely with reported data if one concentrates on the *E. coli* retrons. Therefore, I would suggest directly assaying for transposition of a retron in *E. coli*. The availability of a variety of constructs, vectors, selective markers, and the relative simplicity of growing *E. coli* makes it the most obvious organism to use for these studies. Since this was the subject of my preliminary exam, which included an outline of some experimental methods for testing this hypothesis, I shall not dwell on it here.

In light of the recent discovery of short *msd* repeats in *N. exedens* Na e434, it would be interesting to screen a collection of *E. coli* for similar repeats, or the presence of other partial retron sequences, which may indicate loss due to purifying selection, or aborted transposition, similar to SINE elements. Since RNA-associated msDNA may be an artifact, a study designed to detect the "unbranched" forms described by Lim may reveal these msDNAs are present in a significant portion of the proteobacteria and could provide clues as to a function for the satellite molecule. One way to screen for these would be based on using the polymerase chain reaction and degenerate primers based on the conserved amino acid domains of the RT. This approach may also detect retrons which contain RT, but do not make msDNA However, this method has been problematic and has not yet yielded consistent, reliable results.

For retrons in the myxobacteria, an evolutionary comparison of the reverse transcriptase (RT) genes from Na e434 and Mx162, similar to the one reported here, may give some clues to the origin of RT in the myxobacteria. Most likely this study would indicate that RT was present in an ancestor common to all the myxobacteria. Finally, it would be interesting to look at *M. xanthus* strains with and without retron Mx65 to compare flanking sequences to look for a possible target site for the retron to insert into; this would be similar to the current study of repeats in *N. exedens*.

The second area of investigation would concentrate on msDNA and the generation of the *msd* repeats. Initial studies would concentrate on whether or not msDNA actually

binds proteins and if so, which ones, with a specific interest in mismatch repair proteins, DNA pol I, and combinations thereof. Protein binding by msDNA could be tested using a gel shift assay. Additionally, it would be interesting to see if the msDNA-protein complex is capable of binding double-stranded DNA and/or chromosomal DNA. I would also propose to assay the potential for msDNA or the msDNA-RT complex to cleave or nick double-stranded DNA or chromosome. This could be tested by engineering the target site, if one can be identified, into a plasmid vector, incubating the super-coiled plasmid and msDNA-protein complex for a brief period, and assaying for either open-circular forms of the plasmid (single-strand nick), or for linear forms (double-strand cleavage). Before the model for generation of the *msd* repeats can be tested, an experimental system needs to be developed. This would first require demonstration that generation of these repeats is an active, continuing process. In short, one would need to directly observe formation of a new repeat element. Potentially, this could be done using *N. exedens* Na e434 as the experimental system, by serial plating of *N. exedens* Na e434 for several generations, and screening for changes in the hybridization pattern. Since genetic experiments in the myxobacteria is hampered by a lack of vectors and reporter systems, it may be better to focus on *E. coli* strains. The *E. coli* strains which contain retrons that have been demonstrated to be mutagenic would be useful tools for this study (Maas *et al.*, 1994). By examining many different repeats, a consensus sequence for the insertion site might be generated, which could be engineered onto a plasmid to directly assay for repeat generation. For example, if the insertion site can be engineered into β-galactosidase gene, then insertion of the repeat might block β-galactosidase activity and insertions could subsequently be screened for by blue-white selection. These experiments could be carried out in a variety of *E. coli* hosts, including strains with mutations in the recombination pathway, under many different conditions, such as temperature stress or ultraviolet light exposure.

Maas *et al.* (1994) demonstrated that overexpression of msDNA with a bulge region was mutagenic. Interestingly, this bulge contains a "GATC" which is the target sequence for methylase. Also, the Na e434 msDNA A has a similar "GATC" sequence in a bulge region. Therefore, the possibility that msDNA can be methylated in the stem region should be addressed as should the possibility that the "GATC" sequence is required for the observed mutagenic effects. The first approach to this question is to alter the "GATC" site and determine if the altered msDNA is still mutagenic. If it is, then the methylase activity is likely not involved. If, however, altering this site reduces or eliminates the frameshift

mutations normally observed, then restriction endonuclease digestion of the msDNA with the restriction endonucleases *Dpn*II (blocked by methylation) and *Sau*3AI (not blocked by methylation) may help elucidate if this site is methylated on the msDNA.

As for the *Nannocystis* strains, I would test the reversion rate of Lac^+ reversion of a Lac^- marker to see if a high rate of Lac^- to Lac^+ revertants are observed, which might indicate these strains are highly error prone, and may be defective for one or more repair systems. Finally, I would be curious to see if processed msDNA can serve as a primer for reverse transcriptase or for DNA polymerase I, which would address a possible "function" of msDNA. This could be tested *in vitro* by creating a template with sequences complimentary to the msDNA, and incubation of the template in the presence of purified RT and ^{32}P-labeled nucleotide.

Acknowledgments

The authors would like to thank Justin "Batman" McCain and Julia Roberts for help in cloning of the msDNA repeat elements. Also, thanks go to Hans Reichenbach and Larry Shimkets for the *Nannocystis* strains and the *Melittangium lichenicola* respectively.

References

Altshul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic Local Alignment Search Tool. J. Mol. Biol. **215:**403-410.

Applegate, B. M., U. Matrubutham, J. Sansevrino, and G. S. Sayler. 1995. Biodegradation Genes as Marker Genes in Microbial Populations, p. 6.1.8. *In* A. D. L. Akkermans, J. D. van Elas, and F. J. de Bruijn (ed.), Molecular Microbiology Ecology Manual. Kluwer Academic Publisher, Kluwer, Netherlands.

Applegate, B. M., R. Stapleton, W. Johnston, L. Calicotte, and G. S. Sayler. 1995. submitted for publication. .

Bachellier, S., D. Perrin, M. Hofnung, and E. Gilson. 1993. Bacterial Interspersed Mosaic Elements (BIMEs) are Present in the Genome of Klebsiella. Mol. Microbiol. **7:**537-544.

Bachellier, S., W. Saurin, D. Perrin, M. Hofnung, and E. Gilson. 1994. Structural and Functional Diversity Among Bacterial Interspersed Mosaic Elements (BIMEs). Mol. Microbiol. **12:**61-70.

Baltimore, D. 1970. Viral Dependent RNA Polymerase. Nature (London). **226:**1205-1210.

Banfalvi, Z., A., Nieuwkoop, M. Schell, L. Besl, and G. Stacey. 1988. Regulation of *nod* gene expression in *Bradyrhizobium japonicum.* Mol. Gen. Genet. **214:**420-424.

Beltran, P., S. A. Plock, N. H. Smith, T. S. Whittam, D. C. Old, and R. K. Selander. 1991. Reference Collection of Strains of the *Salmonella typhimurium* Complex from Natural Populations. J. Gen. Microbiol. **173:**601-606.

Birnboim, H. C., and J. Doly. 1979. A Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid DNA. NAR. **7:**1513-1523.

Blackburn, E. H. 1993. Telomerase, p. 411-424. *In* A. M. Skalka and S. P. Goff (ed.), Reverse Transcriptase. Cold Spring Harbor Labratory Press, Cold Spring Harbor.

Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink. 1985. Ty Elements Transpose Through an RNA Intermediate. Cell. **40:**491-500.

Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow. 1977. Construction and Characterization of New Cloning Vehicles: II. A New Multipurpose Cloning System. Gene. **2:**95-113.

Boone, L. R., and A. M. Skalka. 1993. Strand Displacement Synthesis by Reverse Transcriptase. Cold Spring Harbor Labratory Press, Cold Spring Harbor.

Bradford, M. M. 1976. A Rapid and Sensitive Method for the Quantification of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye-Binding. Anal. Biochem. **72:**248-254.

Chapman, K. B., and J. D. Boeke. 1991. Isolation and Characterization of the Gene Encoding Yeast Debranching Enzyme. Cell. **65:**483-492.

Cheng, Y. L., L. V. Kalman, and D. Kaiser. 1994. The *dsg* Gene of *Myxococcus xanthus* Encodes a Protein Similar to Translation Initiation Factor IF3. J. Bacteriol. **176:**1427-1433.

Chomyzynski, P., and N. Sacchi. 1987. Single-Step Method for RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. Anal. Biochem. **162:**156-159.

Church, G. M., and W. Gilbert. 1984. Genomic Sequencing. Proc. Natl. Acad. Sci., USA. **81:**1991-1995.

Devereux, J., P. Haeberli, and O. Smithies. 1984. A Comprehensive Set of Sequence Analysis Programs for Vax. Nucl. Acids Res. **12:**387-395.

Dhundale, A., T. Furuichi, M. Inouye, and S. Inouye. 1988. Mutations that Affect Production of Branched RNA-Linked msDNA in *Myxococcus xanthus*. J. Bacteriol. **170:** 5620-5624.

Dhundale, A., M. Inouye, and S. Inouye. 1988. A New Species of Multicopy Single-Stranded DNA from *Myxococcus xanthus* with Conserved Structural Features. J. Biol. Chem. **263:**9055-9058.

Dhundale, A., B. C. Lampson, T. Furuichi, M. Inouye, and S. Inouye. 1987. Structure of msDNA from *Myxococcus xanthus*: Evidence for a Long, Self-Annealing RNA Precursor for the Covalently Linked, Branched RNA. Cell. **51:**1105-1112.

Dhundale, A. R., T. Furuichi, S. Inouye, and M. Inouye. 1985. Distribution of Multicopy Single-Stranded DNA Among Myxobacteria and Related Species. J. Bacteriol. **164:**914-917.

Di Nocera, P. P., C. Contursi, and G. Minchiotti. 1994. LINE-Related Elements in *Drosophila melanogaster*. Genetica. **94:**173-180.

Dombroski, B. A., A. F. Scott, and H. Kazazian Jr. 1993. Two Additional Retrotransposons Isolated From a Human L1 Subfamily that Contains an Active Retrotransposable Element. Proc. Natl. Acad. Sci., USA. **90:**6513-6517.

Downard, J. S., and D. R. Zusman. 1985. Differential Expression of Protein S Genes During *Myxococcus xanthus* Development. J. Bacteriol. **161:**1146-1155.

Dworkin, M. 1973. Cell-Cell Interactions of the Myxobacteria. Symp. Soc. Gen. Microbiol. **23:**125-147.

Dworkin, M. 1991. Cell-Cell Interactions of the Myxobacteria, p. 179-216. *In* M. Dworkin (ed.), Microbial Cell-Cell Interactions. ASM Press, Washington, D. C.

Eickbush, T. H. 1992. Transposing Without Ends: The NON-LTR Retrotransposable Elements. The New Biologist. **4:**430-440.

Feinberg, A. P., and B. Volgstein. 1983. A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity. Anal. Biochem. **132:**6-13.

Ferat, J.-L., and F. Michel. 1993. Group II Self-Splicing Introns in Bacteria. Nature (London). **364:**358-361.

Flavell, A. J. 1995. Retroelements, Reverse Transcriptase, and Evolution. Comp. Biochem. Physiol. **110B:**3-15.

Furuichi, T., A. Dhundale, M. Inouye, and S. Inouye. 1987. Branched RNA Covalently Linked to the 5' End of a Single-Stranded DNA in *Stigmatella aurantiaca*: Structure of msDNA. Cell. **48:**47-53.

Furuichi, T., S. Inouye, and M. Inouye. 1987. Biosynthesis and Structure of Stable Branched RNA Covalently Linked to the 5' End of Multicopy Single-Stranded DNA of *Stigmatella aurantiaca.* Cell. **48:**55-62.

Ghosal, D., I.-S. You, and I. C. Gunsalus. 1987. Nucleotide Sequence and Expression of Gene *nah*H of Plasmid NAH7 and Homology to *xyl*E of TOL pWWO. Gene. **55:**19-28.

Gilson, E., J.-M. Clement, D. Brutlag, and M. Hofnung. 1984. A Family of Dispersed Repetitive Extragenic Palindromic DNA sequences in *E. coli.* EMBO J. **3:**1417-1421.

Gilson, E., D. Perrin, and M. Hofnung. 1990. DNA Polymerase I and a Protein Complex Bind Specifically to *E. coli* Palindromic Unit Highly Repetitive DNA: Implications for Bacterial Chromosome Organization. Nuc. Acid Res. **18:**3941-3952.

Harayama, S., M. Kok, and E. L. Neidle. 1992. Functional and Evolutionary Relationships Among Diverse Oxygenases. Ann. Rev. Microbiol. **46:**565-601.

Herzer, P. J., S. Inouye, and M. Inouye. 1992. Retron-Ec107 is Inserted into the *Escherichia coli* Genome by Replacing a Palindromic 34 bp Intergenic Sequence. Mol. Micro. **6:**345-354.

Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic Distribution of Branched RNA-Linked Multicopy Single-Stranded DNA Among Natural Isolates of *E. coli.* J. Bacteriol. **172:**6175-6181.

Hsu, M.-Y., S. G. Eagle, M. Inouye, and S. Inouye. 1992. Cell-Free Synthesis of the Branched RNA-Linked msDNA from Retron-Ec67 of *Escherichia coli.* J. Biol. Chem. **267:**13823-13829.

Hsu, M.-Y., S. Inouye, and M. Inouye. 1989. Structural Requirements of the RNA Precursor for the Biosynthesis of the Branched RNA-Linked Multicopy Single-Stranded DNA of *Myxococcus xanthus.* J. Biol. Chem. **264:**6214-6219.

Hsu, M.-Y., C. Xu, M. Inouye, and S. Inouye. 1992. Similarity Between the *Myxococcus xanthus* and *Stigmatella aurantiaca* Reverse Transcriptase Genes Associated with Multicopy, Single-Stranded DNA. J. Bacteriol. **174:**2384-2387.

Hulton, C. S. J., C. F. Higgins, and P. M. Sharp. 1991. ERIC Sequences: a Novel Family of Repetitivve Elements in the Genomes of *Escherichia coli*, *Salmonella typhimurium*, and Other Enterobacteria. Mol. Micribiol. **5:**825-834.

Inouye, M., and S. Inouye. 1992. Retrons and Multicopy, Single-Stranded DNA. J. Bacteriol. **174:**2419-2424.

Inouye, S. 1990. Cloning and DNA Sequence of the Gene Coding for the Major Sigma Factor From *Myxococcus xanthus*. J. Bacteriol. **172:**80-85.

Inouye, S., P. Herzer, and M. Inouye. 1990. Two Independent Retrons with Highly Diverse Reverse Transcriptases in *Myxococcus xanthus*. Proc. Natl. Acad. Sci. USA. **87:**942-945.

Inouye, S., M.-Y. Hsu, S. Eagle, and M. Inouye. 1989. Reverse Transcriptase Associated with the Biosynthesis of the Branched RNA-Linked msDNA in *Myxococcus xanthus*. Cell. **56:**709-717.

Inouye, S., and M. Inouye. 1995. Structure and Evolution of Bacterial Reverse Transcriptase. Virus Genes. **11:**(in press).

Inouye, S., M. G. Sunshine, E. W. Six, and M. Inouye. 1991. Retronphage ϕR73: An *E. coli* Phage that Contains a Retroelement and Integrates into a tRNA Gene. Science. **252:**969-9971.

Joyce, G. F. 1989. RNA Evolution and the Origins of Life. Nature (London). **338:**217-224.

Kawaguchi, T., P. J. Herzer, M. Inouye, and S. Inouye. 1992. Sequence Diversity of the 1.3 kb Retron (retron-Ec107) Among Three Distinct Phylogenetic Groups of *Escherichia coli*. Mol. Micro. **6:**355-361.

Kehrmeyer, S. R., B. M. Applegate, H. C. Pinkart, D. B. Hedrick, D. C. White, and G. S. Sayler. 1995. Combined Lipid/DNA Extraction Method for Environmental Samples. J. Microb. Meth. **in press.**

Komano, T., T. Franceschini, and S. Inouye. 1987. Identification of a Vegetative Promoter in *Myxococcus xanthus*: a Promoter That Has Homology to Histones. J. Mol. Biol. **196:**517-524.

Lampson, B. C. 1993. Retron Elements of the Myxobacteria, p. 109-128. *In* M. Dworkin and D. Kaiser (ed.), Myxobacteria II. ASM Press, Washington, D.C.

Lampson, B. C., M. Inouye, and S. Inouye. 1989. Reverse Transcriptase with Concomitant Ribonuclease H Activity in the Cell-Free Synthesis of Branched RNA-Linked msDNA of *Myxococcus xanthus*. Cell. **56:**701-707.

Lampson, B. C., M. Inouye, and S. Inouye. 1991. Survey of Multicopy Single-Stranded DNAs and Reverse Transcriptase Genes Among Natural Isolates of *Myxococcus xanthus*. J. Bacteriol. **173:**5363-5370.

Lampson, B. C., S. Inouye, and M. Inouye. 1991. msDNA of Bacteria. Prog. Nuc. Acid Res. **40:**1-24.

Lampson, B. C., J. Sun, M.-Y. Hsu, J. Vallejo-Ramirez, S. Inouye, and M. Inouye. 1989. Reverse Transcriptase in a Clinical Isolate of *Escerichia coli*: Production of Branched RNA-Linked msDNA. Science. **243:**1033-1038.

Lampson, B. C., M. Viswanathan, M. Inouye, and S. Inouye. 1990. Reverse Transcriptase from *Escherichia coli* Exists as a Complex with msDNA and is Able to Synthesize Double-Stranded DNA. J. Biol. Chem. **265:**8490-8496.

Li, S., B.-U. Lee, and L. J. Shimkets. 1992. *csg*A Expression Entrains *Myxococcus xanthus* Development. Gene Develop. **6:**401-410.

Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A New Method for Estimating Synonymous and Nonsynonymous Rates of Nucleotide Substitution Considering the Relative Likelihood of Nucleotide and Codon Changes. Mol. Biol. Evol. **2:**150-174.

Lim, D. 1995. Analysis of a Retron Ec86 and Ec67 Insertion Site in *Escherichia coli*. Plasmid. **34:**58-61.

Lim, D. 1992. Structure and Biosynthesis of Unbranched Multicopy Single-Stranded DNA by Reverse Transcriptase in a Clinical *Escherichia coli* Isolate. Mol. Micro. **6:**3531-3542.

Lim, D. 1991. Structure of Two Retrons of *Escherichia coli* and Their Common Chromosomal Insertion Site. Mol. Micro. **5:**1863-1872.

Lim, D., T. A. T. Gomes, and W. K. W. K. Maas. 1990. Distribution of msDNA Among Serotypes of Enteropathogenic *Escherichia coli* Strains. Mol. Microbiol. **4:**1711-1714.

Lim, D., and W. K. Maas. 1989. Reverse Transcriptase-Dependent Synthesis of a Covalently Linked, Branched DNA-RNA Compound in *E. coli* B. Cell. **56:**891-904.

Luan, D. D., M. H. Korman, J. L. Jacubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is Primed by a Nick at the Chromosomal Target Site: A Mechanism for NON-LTR Retrotransposition. Cell. **72:**595-605.

Maas, W. K., C. Wang, T. Lima, G. Zubay, and D. Lim. 1994. Multicopy Single-Stranded DNAs With Mismatched Base Pairs are Mutagenic in *Escherichia coli*. Mol. Microbiol. **14:**437-441.

Maniatis, T., J. Sambrook, and E. F. Fritsch (ed.). 1989. Molecular Cloning: a Laboratory Maunal. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Maxam, A., and W. Gilbert. 1980. Sequencing End-Labeled DNA With Base-Specific Chemical Cleavages. Meth. Enzymol. **65:**499-525.

McClure, M. A. 1991. Evolution of Retroposons by Acquisition or Deletion of Retrovirus-Like Genes. Mol. Biol. Evol. **8:**835-856.

McCurdy, H. D. 1969. Studies on the Taxonomy of the *Myxobacterales*. I. Record of Canadian Isolates and Survey of Methods. Can. J. Microbiol. **15:**1453-1461.

Mirochnitchenko, O., S. Inouye, and M. Inouye. 1994. Production of Single-Stranded DNA in Mamalian Cells by Means of a Bacterial Retron. J. Biol. Chem. **269:**2380-2383.

Miyata, S., A. Oshima, S. Inouye, and M. Inouye. 1992. *In vivo* Production of a Stable Single-Stranded cDNA in *Saccharomyces cerevisiae* by Means of a Bacterial Retron. Proc. Natl. Acad. Sci. USA. **89:**5735-5739.

Moran, J. V., C. M. Wernette, K. L. Mecklenburg, R. A. Butow, and P. S. Perlman. 1992. Intron 5α of the COXI Gene of Yeast Mitochondrial DNA is a Mobile Group I Intron. Nuc. Acid Res. **20:**4060-4076.

Ochman, H., and A. C. Wilson. 1987. Evolution in Bacteria: Evidence for a Universal Substitution Rate in Cellular Genomes. J. Mol. Evol. **26:**74-86.

Ogram, A., G. S. Sayler, and T. Barkay. 1987. The Extraction and Purification of Microbial DNA from Sediments. J. Microbiol. Meth. **7:**57-66.

Oppenheim, A. B., K. E. Rudd, I. Mendelson, and D. Teff. 1993. Integration Host Factor Binds to a Unique Class of Complex Extragenic Repetitive DNA Sequences in *Escherichia coli*. Mol. Microbiol. **10:**113-122.

Oshima, K., R. Koishi, M. Matsuo, and N. Okada. 1993. Several Short Interspersed Repetitive Elements (SINEs) in Distant Species May Have Originated from a Common Ancestral Retrovirus: Characterization of a Squid SINE and a Possible Mechanism for Generation of tRNA-Derived retroposons. Proc. Natl. Acad. Sci. USA. **90:**6260-6264.

Reichenbach, H., and M. Dworkin. 1992. The Myxobacteria, p. 3416-3487. *In* A. Barlows (ed.), The Prokaryotes, second ed. Springer-Verlag, New York.

Rice, S. A., J. Bieber, J.-Y. Chun, G. Stacey, and B. C. Lampson. 1993. Diversity of Retron Elements in a Population of Rhizobia and Other Gram-Negative Bacteria. J. Bacteriol. **175:**4250-4254.

Rice, S. A., and B. C. Lampson. 1995. Bacterial Reverse Transcriptase and msDNA. Virus Genes. **11:**(in press).

Rice, S. A., and B. C. Lampson. 1995. Phylogenetic Comparison of Retron Elements Among the Myxobacteria: Evidence for Vertical Inheritance. J. Bacteriol. **177:**37-45.

Ringleberg, D., and D. C. White. 1992. DOE-WSRS Rhizosphere PLFA/Sterol Analysis. Final Report MARX11-SK808V. University of Tennessee.

Rosseló-Mora, R. A., J. Lalucat, and E. García-Valdés. 1994. Comparative Biochemical and Genetic Analysis of Naphthalene Degradation Among *Pseudomonas stutzeri* Strains. Appl. Environ. Microbiol. **60:**966-972.

Sanger, F., S. Nicklen, and R. Coulson. 1977. DNA Sequencing With Chain-Terminating Inhibitors. Proc. Natl. Acad. Sci. USA. **74:**5463-5467.

Sanseverino, J., B. M. Applegate, J. M. Henry King, and G. S. Sayler. 1993. Plasmid-Mediated Mineralization of Naphthalene, Phenanthracene, and Anthracene. Appl. Environ. Microbiol. **59:**1931-1937.

Sawyer, S. A., D. E. Dykhuizen, R. F. DuBose, L. Green, T. Mutangadura-Mhlango, D. F. Wolczyk, and D. L. Hartl. 1987. Distribution and Abundance of Insertion Sequences Among Natural Isolates of *Escherichia coli.* Genetics. **115:**51-63.

Sayler, G. S., J. Flemming, B. M. Applegate, and C. Werner. 1991. Nucleic Acid Extraction and Analysis: Detecting Genes and Their Activity in the Environment, p. 237-257. *In* E. M. Wellington and J. D. Van Elas (ed.), Genetic Interactions Between Microorganisms in the Environment.

Selander, R. K., and N. H. Smith. 1990. Molecular Population Genetics of Salmonella. Rev. Med. Microbiol.219-228. **1:**Rev. Med. Microbiol. 1:219-228.

Sharp, P. M. 1991. Determinants of DNA Sequence Divergence Between *Escherichia coli* and *Salmonella typhimurium*: Codon Usage, Map Position, and Concerted Evolution. J. Mol. Evol. **33:**23-33.

Shimada, M., H. Hosaka, H. Takaku, J. S. Smith, M. J. Roth, S. Inouye, and M. Inouye. 1994. Specificity of Priming Reaction of HIV-1 Reverse Transcriptase 2'-OH or 3'OH. J. Biol. Chem. **269:**3925-3927.

Shimamoto, T., M.-Y. Hsu, S. Inouye, and M. Inouye. 1993. Reverse Transcriptases from Bacterial Retrons Require Specific Secondary Structures at the 5'-End of the Template for the cDNA Priming Reaction. J. Biol. Chem. **1993:**2684-2692.

Shimamoto, T., M. Inouye, and S. Inouye. 1995. The Formation of the 2',5'-Phosphodiester Linkage in the cDNA Priming Reaction by Bacterial Reverse Transcriptase in a Cell-Free System. J. Biol. Chem. **270:**581-588.

Shimamoto, T., M. Shimada, M. Inouye, and S. Inouye. 1995. The Role of Ribonuclease H in Multicopy, Single-Stranded DNA Synthesis in Retron-Ec73 and Retron-Ec107 of *Escherichia coli.* J. Bacteriol. **177:**264-267.

Shimkets, L. 1993. The Myxobacterial Genome, p. 85-107. *In* M. Dworkin and K. Kaiser (ed.), Myxobacteria II. American Society for Microbiology, Washington, D.C.

Shimkets, L., and C. R. Woese. 1992. A Phylogenetic Analysis of the Myxobacteria: Basis for Their Classification. Proc. Natl. Acad. Sci. USA. **89:**9459-9463.

Simon, J. M., J. D. Osslund, R. Saunders, B. D. Ensley, S. Suggs, A. Harcourt, W. C. Suen, D. L. Cruden, D. J. Gibson, and G. Zylstra. 1993. Sequences of Genes Encoding Naphthalene Dioxygenase in *Pseudomonas putida* G7 and *Pseudomonas* sp. NCBI 9816-4. Gene. **127:**31-37.

Singh, B. N. 1947. Myxobacteria in Soils and Composts; Their Distribution, Number, and Lytic Action on Bacteria. J. Gen. Microbiol. **1:**1-10.

Stahl, D. A., B. Flesher, H. R. Mansfield, and L. Montgomery. 1988. Use of Phylogenetically Based Hybridization Probes for Studies of Ruminal Microbial Ecology. Appl. Environ. Microbiol. **54:**1079-1084.

Stern, M. J., E. Prossnitz, and G. Ferro-Iuzzi Ames. 1988. Role of the Intercistronic Region in Post-Transcriptional Control of Gene Expression in the Histidine Transport Operon of *Salmonella typhimurium*: Involvement of REP Sequences. Mol. Microbiol. **2:**141-152.

Sun, J., P. J. Herzer, M. P. Weinstein, B. C. Lampson, S. Inouye, and M. Inouye. 1989. Extensive Diversity of Branched-RNA-Linked Multicopy, Single-Stranded DNAs in Clinical Strains of *Escherichia coli.* Proc. Natl. Acad. Sci. USA. **86:**7208-7212.

Sun, J., M. Inouye, and S. Inouye. 1991. Association of a Retroelement with a P4-Like Cryptic Prophage (Retronphage ϕR73) Integrated into the Selencystyl tRNA Gene of *Escherichia coli.* J. Bacteriol. **173:**4171-4181.

Suzuki, M., T. Hayakawa, J. P. Shaw, M. Rekik, and S. Harayama. 1991. Primary Structure of Xylene Monooxygenase: Similarities to and Differences from the Alkane Hydroxylation System. J. Bacteriol. **173:**1690-1695.

Temin, H. 1989. Retrons in Bacteria. Nature (London). **339:**254-255.

Temin, H. M., and S. Mizutani. 1970. RNA-Dependent DNA Polymerase in Virions of Rous Sarcoma Virus. Nature (London). **226:**1211-1213.

Viswanathan, M., M. Inouye, and S. Inouye. 1989. *Myxococcus xanthus* msDNA Mx162 Exists as a Complex with Proteins. J. Biol. Chem. **264:**13665-13671.

Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon. 1985. Nucleotide Sequence of the AIDS Virus, LAV. Cell. **40:**9-17.

Wakefield, J. K., S. A. Jablonski, and C. D. Morrow. 1992. In Vitro Enzymatic Activity of Human Immunodeficiency Virus Type I Reverse Transcriptase Mutants in the Highly Conserved YMDD Amino Acid Motif Correlates with the Infectious Potential of the Proviral Genome. J. Virol. **66:**6806-6812.

Waldrop, M. M. 1989. Did Life Really Start Out in an RNA World. Science. **246:**1248-1249.

Wang, G.-H., and C. Seeger. 1992. The Reverse Transcriptase of Hepatitis B Virus Acts as a Protein Primer for Viral DNA Synthesis. Cell. **71:**663-670.

Wilson, C. A., S. S. Carlson, and T. J. White. 1977. Biochemical Evolution. Ann. Rev. Biochem. **46:**573-639.

Woese, C. R. 1987. Bacterial Evolution. Microbiol. Rev. **51:**221-271.

Xiong, Y., and T. H. Eickbush. 1990. Origin and Evolution of Retroelements Based Upon Their Reverse Transcriptase Sequences. EMBO J. **9:**3353-3362.

Yang, Y., and G. Ames. 1988. DNA Gyrase Binds to the Family of Prokaryotic Repetitive Extragenic Palindromic Sequences. Proc. Natl. Acad. Sci., USA. **85:**8850-8854.

Yanisch-Perron, C., J. Viera, and J. Messing. 1985. Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequences of the M13mp18 and pUC19 Vectors. Gene. **33:**103-119.

Yee, T., T. Furuichi, S. Inouye, and M. Inouye. 1984. Multicopy Single-Stranded DNA Isolated from a Gram-Negative Bacterium, *Myxococcus xanthus*. Cell. **38:**203-209.

Yen, K.-M. 1988. Genetics of Naphthalene Catabolism in *Pseudomonads*. CRC Crit. Rev. Microbiol. **15:**247-267.

Yen, K.-M., and I. C. Gunsalus. 1985. Regulation of Naphthalene Catabolic Genes of Plasmid NAH7. J. Bacteriol. **162:**1008-1013.

You, I.-S., D. Ghosal, and I. Gunsalus. 1991. Nucleotide Sequence Analysis of the *Pseudomonas putida* PpG7 Salicylate Hydroxylase Gene (*nah*G) and its 3'-Flanking Region. Biochem. **30:**1635-1641.

Zimmerly, S., H. Guo, P. S. Perlman, and A. M. Lambowitz. 1995. Group II Intron Mobility Occurs by Target DNA-Primed Reverse Transcription. Cell. **82:**545-554.

Zylstra, G. J., and D. T. Gibson. 1989. Toluene Degradation by *Pseudomonas putida* F1. J. Biol. Chem. **264:**14940-14946.

Vita

Scott Rice was born in SanDiego, CA on February 26, 1967, the second child of an initiate naval family. He subsequently relocated to Va. Beach, Va. where the ocean was warmer and there he attended public schools. After graduating from Princess Anne High School in 1985 with no particular degree in anything, he began college at Virginia Polytechnic Institute and State University (Va. Tech). During his years there, Mr. Rice received an excellent education and had a lot of fun along the way. In May 1989, the future Dr. Rice graduated from Va. Tech with a Bachelor of Science degree in Biology. The result of many inspirational faculty, such as Jack Webster, Jack Cranford, George Simmons, Robert Benoit, and spirited biological discussions with his good friends, especially Matt and Jehu, not to mention a severe lack of job prospects, encouraged the young Mr. Rice to enroll in graduate school. In the Fall of 1989, he began graduate school at the University of North Carolina, Charlotte, and graduated with a Master of Science degree in Biology in December 1991. Despite a cold and high pressure Masters thesis, our hero pressed onward and upward. In August 1991, Scott began work towards his Ph.D. at another institution of higher education perversely obsessed with the colour orange, known as the University of Tennessee, Knoxville. The degree of Doctor of Philosophy, in Microbiology was conferred upon him in May 1996, despite having finished several months earlier. Presently, he is living somewhere in the outback with his wife and several dingos where they run a successful commercial koala ranch and do science on the side.