INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

University of Alberta

Analysis of the Sequence of An Insertion-Like Element from *Mycobacterium avium*

by

Xiaoling Puyang

A thesis to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

Department of Medical Microbiology and Immunology

Edmonton, Alberta
Fall 1997

University of Alberta

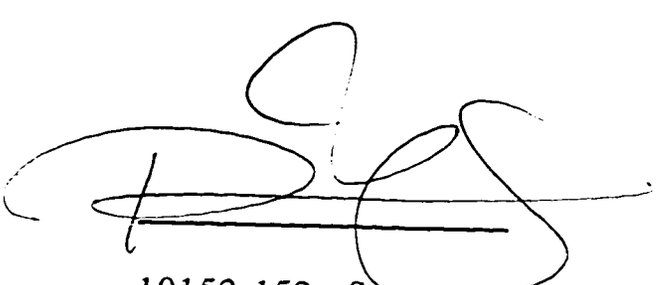
Library Release Form

Name of Author: Xiaoling Puyang

Title of Thesis: Analysis of the Sequence of An Insertion-Like Element from *Mycobacterium avium*

Degree: Master of Science

Year This Degree Granted: 1997



10152-152 Street
Edmonton, Alberta
T5P 1X9

Sept. 30, 97

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Analysis of the Sequence of An Insertion-Like Element from *Mycobacterium avium* submitted by Xiaoling Puyang in partial fulfillment of the requirement for the degree of Master of Science.

Dr. Dennis Y. Kunimoto
(Supervisor)

Dr. A. Mark Joffe

Dr. James A. Talbot

Dr. Laura S. Frost

Dr. Mark S. Peppler

Sept. 24, 97

For my grandmother, who passed away when I just started my graduate studies.

It was her who showed me the beauty of the world.

Abstract

An insertion sequence (IS) designated ISN was isolated and characterized from a *Mycobacterium avium* clinical strain. ISN was found by hybridization with the pMB22/S12 probe from IS900 of *Mycobacterium paratuberculosis*. ISN is 1418 bp in size with 65% G+C content. ISN has a high degree of homology with IS900-related elements. Similar to IS900-related elements, ISN has neither terminal inverted repeats nor flanking direct repeats. Analysis of three insertion sites for ISN in the *M. avium* strain and the corresponding potential insertion site from two *M. avium* strains indicates a consensus insertion sequence of $CATGCN_{(4-5)}TCCTN_{(2)}G$ for ISN insertion. In the three clones examined, ISN has the same orientation with respect to this target site. ISN has two major open reading frames. ORF1179 encodes a predicted protein of 393 amino acids. ORF930, on the complementary strand of ORF1179, encodes a protein of 310 amino acids. The Shine-Dalgarno sequence for ORF930 is partially located in the flanking region, similar to other IS900-related elements.

The analysis of the comparable features of IS and the variable occurrence in related organisms is useful for studying the evolution of these elements and their hosts.

Acknowledgments

I would like to give my heartfelt thanks to my supervisor, Dr. Kunimoto, for his guidance, advice and patience throughout my studies, most of all, for his caring and understanding.

I would also like to thank Dr. Joffe and Dr. Talbot for their time and advice as my committee members.

I would like to thank: Dr. Dequan Kong for his helpful suggestions, technical training and friendship; Amy Ng for her technical assistance; Dr. Zhongming Ge for his kind help during the last stage of my studies; San Vinh for his photographic assistance; and, the staff in the office of MMID for their help and patience during my studies.

Special thanks go to all my friends from MMID, Dr. Dovochi's group and off campus, too many to be named, who have always been generous and friendly in supporting and helping me in different ways. They were always there, any time, whenever I needed.

It is hard to find the words to express my thanks to my dear husband, Xiaogang Gao, and my lovely son, Yang Gao, for always loving, understanding and supporting me, and all the sacrifices they made for me.

It would not have been possible to complete my studies without the love and encouragement of my parents, my in-laws and my family in China.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIDS	acquired immunodeficiency syndrome
ATP	adenosine triphosphate
bp	basepair
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
dNTP	deoxynucleoside triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytosine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddTTP	dideoxythymidine triphosphate
dITP	2′-deoxyinosine-5′ triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
G+C	guanine + cytosine
GCG	Genetics Computer Group
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulonic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
IS	insertion sequence
Kb	kilobase
kDa	kilodalton
LB	Luria Bertani

MAC	*Mycobacterium avium* complex
MCS	multiple cloning site
ORF	open reading frame
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
RFLP	restriction fragment length polymorphism
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
S-D	Shine-Dalgarno
SDS	sodium dodecyl sulphate
TAE	tris acetate EDTA buffer
TBE	tris borate EDTA buffer
TE	Tris EDTA buffer
TEMED	N, N, N′, N′-tetramethylethylenediamine
Tm	melting temperature
Tn	transposon
Tris	tris (hydroxymethyl) aminomethane
UV	ultraviolet
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

I. Introduction

1. General Properties of Mycobacteria

The name *Mycobacterium*, meaning fungus-bacterium, was first introduced by Lehman and Neumann in 1896 (Grange, 1988). Mycobacteria are rod-shaped, aerobic, nonsporulating, gram-positive bacteria, though they are not easily stainable using this method. However, they share a distinctive staining property commonly described as Acid-Fast. Due to their lipid-rich cell walls, these bacteria resist decolorization by acidified organic solvents containing mineral acids after being stained with phenol-containing dyes. These complex cell walls also contribute to the slow growth rate, virulence, intracellular survival and pathogenicity for mycobacteria (Grange, 1988)

At present, about 41 species of mycobacteria have been recognized (Grange, 1988). In addition to the classic tubercle bacillus (*Mycobacterium tuberculosis*) and leprosy bacillus (*Mycobacterium leprae*), other mycobacteria called 'atypical' were divided initially by Runyon into four groups according to their growth rate and pigmentation:

I. photochromogens (pigment formed in the light)

II. scotochromogens (pigment formed in the dark)

III. non-chromogens

IV. rapid growers

The tubercle bacilli and leprae bacilli are pathogenic organisms that cause chronic diseases producing lesions of the granulomatous

type. Most of the atypical mycobacteria are found in the environment and are harmless. Only some of them are medically significant. Nevertheless, the importance of atypical mycobacteria as opportunistic pathogens in immunocompromised individuals is increasing (Inderlied *et al.*, 1993).

2. *Mycobacterium avium* Complex (MAC)

(1) Classification

Mycobacteria included in the MAC are classified in Runyon group III, which are non-pigmented and grow extremely slowly. However, MAC is a group of very closely related strains whose classification is still somewhat confusing. Based on seroagglutination analysis, MAC is classified into 28 serovars included in two species: *M. avium* and *M. intracellulare*, which are phenotypically similar (Inderlied *et al.*, 1993). However, the difference between *M. avium* and *M. intracellulare* has been well established by using either restriction fragment length polymorphism (RFLP) patterns or DNA hybridization to probes, e.g. Gen-Probe (McFadden *et al.*, 1987; Musial *et al.*, 1988; Picken *et al.*, 1988).

Currently, several molecular tools have been applied with various degrees of success. DNA-DNA hybridization studies and numerical taxonomy analysis have demonstrated that two other specific pathogens are also closely related to *M. avium* (Thorel *et al.*, 1990). The first one is *Mycobacterium paratuberculosis*, the causative agent of Johne's disease or paratuberculosis (chronic enteritis) in ruminants. It has also been implicated in Crohn's disease (chronic regional enteritis) in humans (Cocito *et al.*, 1994; Fidler *et al.*, 1994).

The other, wood-pigeon mycobacteria, can cause chronic enteritis in animals and has been shown to cause Johne's disease in some experimental studies (Moss *et al.*, 1992). Thorel et al. (Thorel *et al.*, 1990) have proposed three subspecies of *M. avium* on the basis of genetic and phenotypic studies: *M. avium* subsp. *avium*, *M. avium* subsp. *paratuberculosis*, and *M. avium* subsp. *silvaticum* (wood-pigeon bacillus).

rRNA sequences have also been used to evaluate the genetic relatedness of various bacterial genera and species. Since the variable region of 16S rRNA sequences is conserved at the species level, the sequence specificity in the 16S rRNA is sufficient for species identification (Woese, 1987). By comparing the aligned 16S rRNA sequences from a number of mycobacteria, the sequences specific to most mycobacterial species have been defined (Rogall *et al.*, 1990).

More recently, specific insertion sequences (IS) have been found in MAC. This has led to another classification of *M. avium* strains on the basis of their possession of a related IS element (see more details later in this chapter).

Other techniques, such as restriction fragment length polymorphism (RFLP), restriction analysis of diverse genetic regions and pulsed-field gel electrophoresis (PFGE) have been used for various levels of species, subspecies, and strain characterization within the MAC (Arbeit *et al.*, 1993; Hampson *et al.*, 1989; Mazurek *et al.*, 1993; McFadden *et al.*, 1987).

(2) Epidemiology

MAC organisms are ubiquitous in the environment. They have

been recovered from soil, water, house dust and other environmental sources (Inderlied *et al.*, 1993). They are considered to be of low pathogenicity and colonizers that rarely cause disease. However, an increasing number of MAC infections have been reported during the past two decades. Worldwide, these organisms are currently the most prevalent causes of the 'opportunistic' mycobacterial diseases in humans, occurring predominantly in northern temperate areas (Grange, 1988; Inderlied *et al.*, 1993). Studies have shown that 98% of MAC strains isolated from AIDS patients are *M. avium*, compared with only 60% isolated from patients without AIDS. The remaining strains are *M. intracellulare* (Guthertz *et al.*, 1989). The reasons for the close association between MAC and AIDS are not known.

So far, sources of infection and routes of transmission of MAC to humans have not been well defined. Studies have suggested that environmental sources of water constitute the greatest risk of exposure for humans, but the modes of acquisition are still not clear (Wendt *et al.*, 1980).

M. avium is an important cause of disease in poultry and swine, and is the most significant causal agent of tuberculosis-like disease in birds. The excretion of this organism from its animal hosts presents a potential source of infection for both humans and animals. The range of wild bird movement facilitates the dissemination of *M. avium* (Meissner and Anz, 1977). However the direct transmission of *M. avium* from animals to humans is rare and human-to-human transmission has never been reported (McFadden *et al.*, 1992).

(3) Pathogenesis and Diseases

The first case of human disease caused by *M. avium* was reported

in 1943 in Minnesota (Inderlied *et al.*, 1993). However, the mechanisms of pathogenesis of the MAC are not well understood. It has been suggested that MAC infection is due to activation of a subclinical endogenous infection (McFadden *et al.*, 1992). However strains frequently isolated from AIDS patients with disseminated disease are not commonly found in the stools of healthy persons (Hampson *et al.*, 1989). Current evidence points to the intestinal tract as the primary route of *M. avium* infection in AIDS patients (Gray and Rabeneck, 1989; Klatt *et al.*, 1987). The respiratory tract seems to be a secondary and significantly less frequent portal of entry (Jacobson *et al.*, 1991; Knapp *et al.*, 1987). Although, asymptomatic respiratory and intestinal colonization with *M. avium* can be seen in healthy people, the development of focal or disseminated disease is rare. Ingestion of mycobacteria in water or food can lead to colonization of the intestinal tract (Collins, 1989; Malpother and Sanger, 1984). In the human intestinal lumen, the bacteria bind to enterocytes and probably M cells and quickly penetrate the intestinal epithelial cells before translocating into the lamina propria. The bacteria can colonize Peyer's patches and are eventually localized in the liver and spleen as well as found circulating in the blood (Inderlied *et al.*, 1993).

In non-immunocompromised patients, MAC organism usually causes cavitary pulmonary disease. Cervical lymphadenitis, chronic osteomyelitis, renal and skin infections also occur but disseminated infection of MAC is extremely unusual (Inderlied *et al.*, 1993).

Studies indicate that more than 90% of AIDS patients with symptomatic MAC infection exhibit evidence of disseminated mycobacterial disease affecting multiple organs (Benson and Ellner,

1993). This is usually accompanied by continuous, generally high-grade mycobacteremia characterized by intermittent fever, sweats, weakness, anorexia and weight loss (Inderlied *et al.*, 1993). MAC localized diseases are seen less often in AIDS patients, although some patients may develop focal pulmonary infection without evidence of dissemination (Modilevsky *et al.*, 1989; Wallace *et al.*, 1990).

The treatment for MAC infection requires multiple drugs administered simultaneously since development of resistance to antibiotics is very common (Inderlied *et al.*, 1993). The recommended length of therapy is 18 to 24 months. Even after cultures become negative, continued therapy is still needed for several months to one year (Inderlied *et al.*, 1993). For patients with AIDS, studies indicate that MAC contributes to early mortality and that prolonged survival is associated with antimycobacterial treatment (Horsburgh *et al.*, 1991). However the optimal length of treatment is unknown. Prolonged treatment may prevent recurrence of bacteremia. Prophylaxis reduces the chance of MAC infection (Inderlied *et al.*, 1993).

(4) MAC Genetics

A. General Features

The mycobacterial genome consists of a single length of DNA in the form of a closed loop. The genome molecular weight is in the range of 2.8×10^9 to 4.5×10^9 dalton. The size of *M. avium* genome is 3.5×10^6 bp and that of *M. paratuberculosis* is 4.4×10^6 to 4.7×10^6 bp (Clark-Curtiss, 1990; Cocito *et al.*, 1994). Most mycobacterial genomes have a guanine (G) plus cytosine (C) content of about 64 to 70 mol%. The base composition of *M. avium* DNA is 66 to 70% G + C, 66 to 67%

for *M. paratuberculosis* and 65 to 69% for *M. intracellulare* (Baess and Mansa, 1978; Cocito *et al.*, 1994; Imaeda, 1985; Imaeda *et al.*, 1988). Plasmids are common in MAC, but attempts to clearly define the biological significance of plasmids in *M. avium* strains have not yet been successful (Martin *et al.*, 1990).

B. Insertion Sequence

IS elements are small, mobile genetic elements encoding genes which are essential for transposition. Bacterial IS elements were discovered during early investigations of the molecular genetics of gene expression in *Escherichia coli* (*E. coli*) and bacteriophage lambda (Galas and Chandler, 1989). They can cause insertional mutations, chromosome rearrangements and altered expression of genes near their sites of insertion. They transpose not only themselves but also their flanking DNA sequences and can thereby spread antibiotic resistance among bacterial species. They have also gained special importance as genetic tools, since they can be used to insert or delete DNA in susceptible hosts. IS elements have been found in a wide variety of prokaryotes including mycobacteria.

a. General Features of IS Elements

Insertion sequences vary in size. They are usually 0.8 to 2.5 Kb in length, except IS gamma delta of *E. coli* F which is about 6 Kb (Galas and Chandler, 1989). DNA sequence analysis has shown that almost all IS elements contain a perfect or nearly perfect inverted repeat involving approximately 9-40 bp at their termini (Galas and Chandler, 1989; Kleckner, 1981). This inverted repeat appears to be essential for transposition for some IS elements. However, some IS elements, for instance IS1000 from *Thermus thermophilus* (Ashby and Bergquist, 1990), IS492 from *Pseudomonas atlantica* (Bartlett

and Silverman, 1989), IS900 from *M. paratuberculosis* (Green *et al.*, 1989) and IS901 from *M. avium* (Kunze *et al.*, 1991) lack terminal inverted repeats. This class of IS elements may use a different method for their transposition from those with inverted repeats, but the exact mechanism is not clear.

IS elements contain at least one major open reading frame (ORF) encoding the transposase. This ORF extends almost the entire length of the element but some IS elements also contain one or more short overlapping ORFs on the opposite strand (Galas and Chandler, 1989; Kleckner, 1981). After transposition, almost all IS elements generate short direct repeats of target sequences at the point of insertion. This direct repeat is presumably created from staggered cleavage and the filling of gaps formed during the joining of the element and target DNA.

Numerous IS elements have been identified in the genomes, plasmids and bacteriophages of a wide range of bacteria genera and species. The copy number of IS elements can be different for various IS elements and can differ considerably from strain to strain for the same IS element. For example, IS1 in some *E. coli* and *Shigella* strains can range from one to a few hundred copies per genome (Ohtsubo *et al.*, 1981; Sawyer *et al.*, 1987). They are particularly frequent as components of natural plasmids in which they are often associated with genes responsible for antibiotic resistance.

b. Genetic Effects of IS Elements

In addition to their transposition ability, which can cause spontaneous mutations or inactivate the function of the target gene, IS elements also mediate a variety of DNA rearrangements.

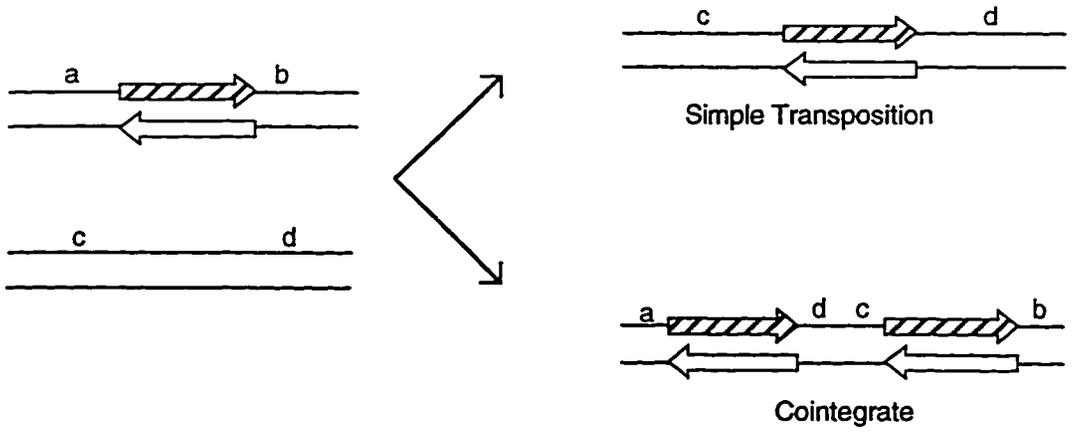
Insertion can generate deletions of flanking DNA, while the

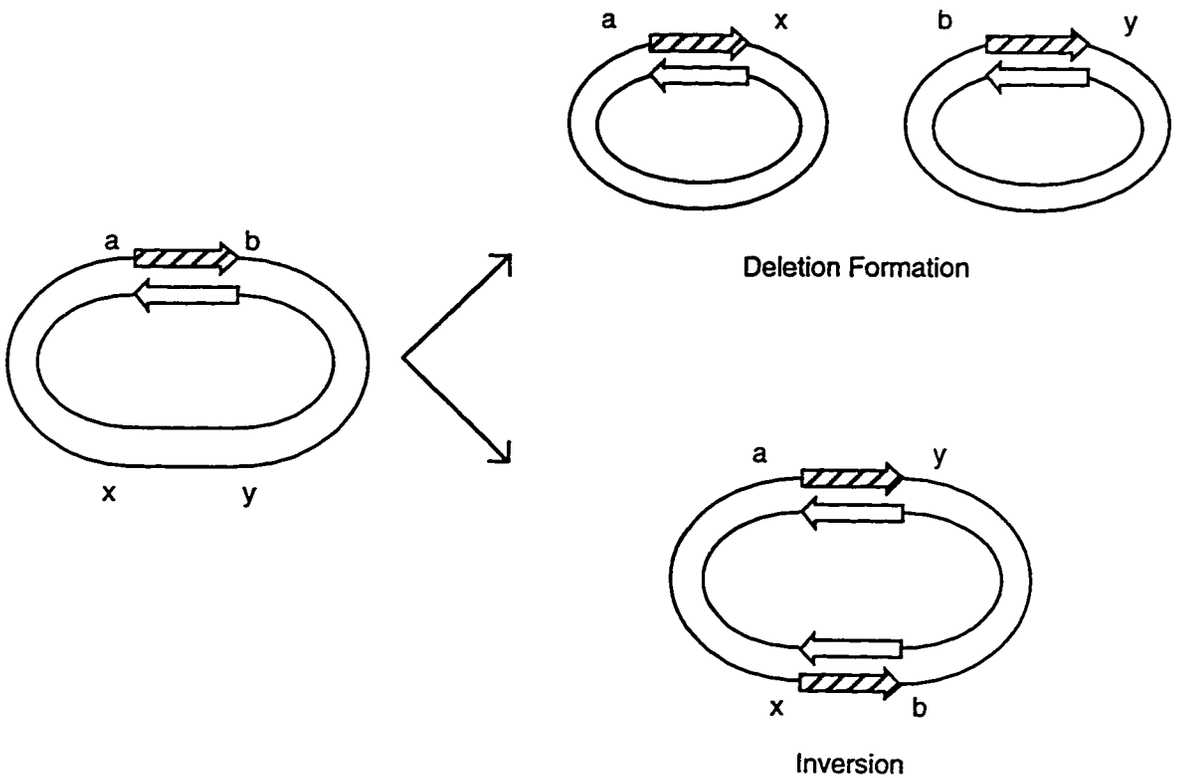
element remains intact. It also can invert a neighboring segment of DNA. This inversion of DNA is accompanied by the appearance of a second copy of the insertion sequence flanking the segment in inverted orientation with respect to the first. Cointegration is another property of IS elements. The resulting cointegrate structure is composed of donor and recipient replicons separated at each junction by a single, directly repeated copy of the transposable element (Galas and Chandler, 1989; Iida *et al.*, 1983; Kleckner, 1981) (Fig. 1).

Not all IS elements exhibit all these capabilities. However, the variety of genetic characteristics have made transposable elements a rich source of experimental material for studying recombination, gene expression and horizontal transmission of genes. They can be used as insertional mutagens since the insertion of a transposable element into a gene may lead to inactivation of the gene. The mutants therefore can be isolated by selecting for the eliminated or altered phenotype of interest. They are also widely used as molecular tags to facilitate genetic mapping, to recognize genes, to analyze gene organization and to isolate genes.

In several pathogenic species, such as *Salmonella*, *Shigella*, *Bordetella* and *Yersinia*, transposition elements have been utilized for the isolation of virulence genes via insertional mutagenesis (Isberg and Falkow, 1985). Weiss (Weiss *et al.*, 1983) used the transposon Tn5 to study virulence factors in *Bordetella pertussis*. Insertions of Tn5 into the chromosome were isolated by selecting for kanamycin resistance after introduction of a replication-defective plasmid that harbored the transposon and the kanamycin resistance gene. Each kanamycin-resistant colony was the result of a transposon-induced mutation. These colonies were then screened for reduced expression

Figure. 1. DNA rearrangements mediated by transposition. (⇄) denotes the transposable element. (a b) denotes the donor DNA and (c d) denotes the recipient DNA. The top part of the figure represents intermolecular transposition which results in either simple insertion or cointegrate formation. The bottom part of the figure represents intramolecular transposition which may lead to adjacent deletion or inversion.


a
b
c
d
c
d
Simple Transposition
a
d
c
b
Cointegrate


a
x
b
y
Deletion Formation
a
b
x
y
a
y
x
b
Inversion

of haemolysin, pertussis toxin, FHA and adenylate cyclase. Analysis of these insertional mutations therefore permitted the isolation of specific genes that could be mapped physically and characterized genetically in a detailed manner.

In addition to disrupting gene function, IS elements may also have strong polar effects on the expression of downstream genes. However, the activation of downstream genes observed for some IS elements is possibly due to either the presence of outwardly directed promoters within the elements or the formation of new promoters on insertion (Galas and Chandler, 1989).

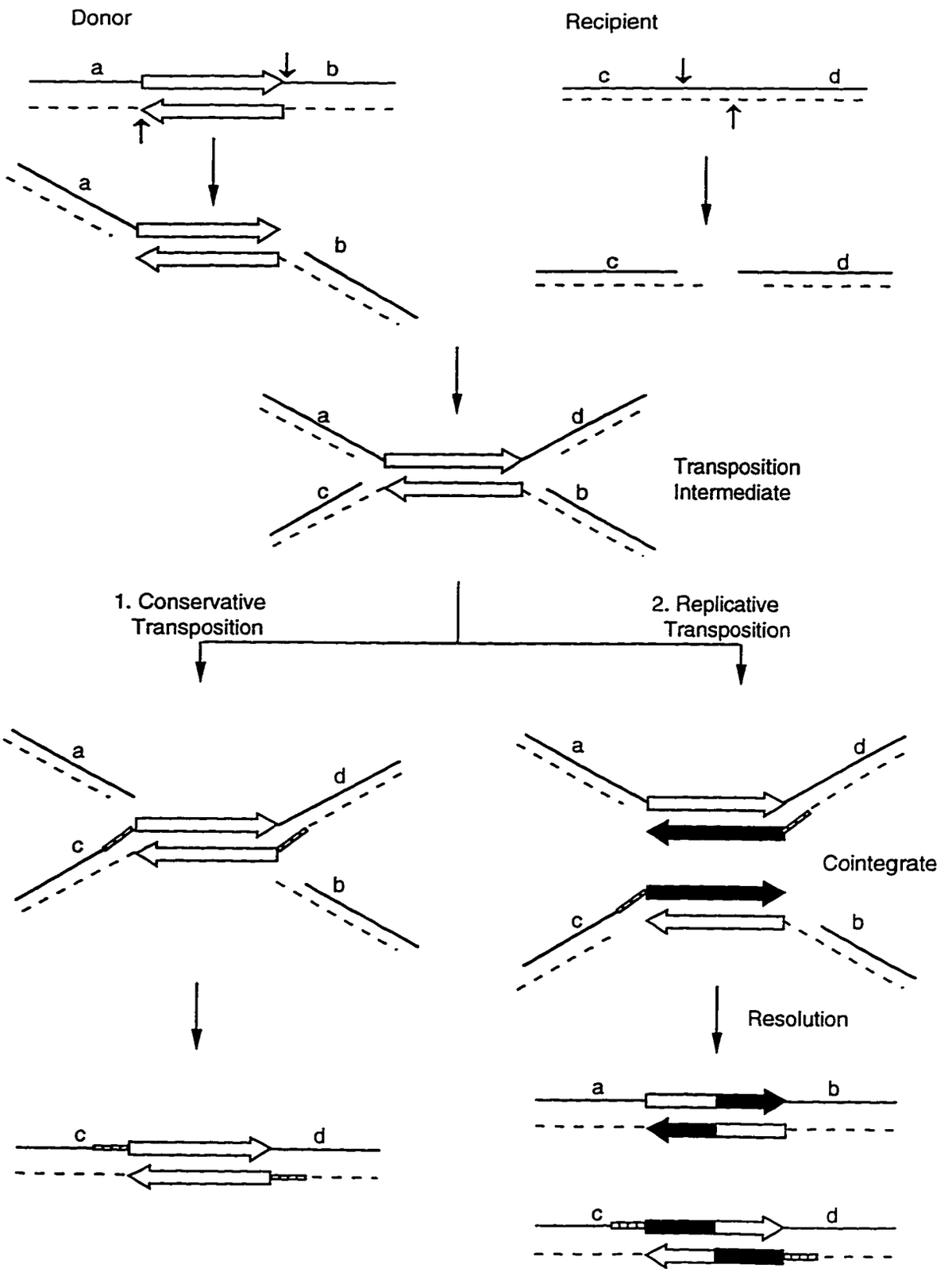
Some insertion DNA sequences can be used to construct a promoter controlled expression system. For example, with an adjacent potential promoter (P_{AN}) and Shine-Dalgarno (S-D) sequence, the ORF2 of IS900 was fused to the *lacZ* gene and inserted into the replicative shuttle vector pRR3. *M. smegmatis* and *M. bovis* BCG that were transformed with this plasmid exhibited β-galactosidase activity. Immunization of mice with the recombinant *M. bovis* BCG resulted in the detection of antibodies against β-galactosidase encoded by the *lacZ* gene (Murray *et al.*, 1992). This demonstrated that β-galactosidase was expressed suggesting that this system may be useful for the expression of foreign genes in *M. bovis* BCG.

c. Transpositional Mechanisms

There are two different mechanisms of transposition: conservative and replicative (Berg *et al.*, 1988; Iida *et al.*, 1983; Shapiro, 1979). The model for these two mechanisms is presented in Figure 2.

In conservative transposition, the transposable element inserts into its new site on target DNA without replication of the element.

Figure 2. The model of two mechanisms of transposition according to Shapiro (Shapiro, 1979). Adapted from McAdam (McAdam *et al.*, 1994) and Martin (Martin *et al.*, 1990). The transposable element (⇨) is cleaved at its 3′ ends. A staggered cleavage on target DNA results in 5′ overhanged ends. Joining the 3′ ends of the element to the 5′ ends of the target forms a transposition intermediate. This intermediate structure can be resolved in two ways: (1) cleavage of the 5′ ends of the element and replication of the target site (▭) leads to a simple insertion of the element into target DNA; (2) replication of the target DNA and the element (➡) leads to the formation of a conintegrate which can be resolved either by a site specific resolvase or by the homologous recombination pathway of the host.


Donor
Recipient
a
b
c
d
Transposition
Intermediate
1. Conservative
Transposition
2. Replicative
Transposition
Cointegrate
Resolution

The ends of donor DNA are not rejoined and the remains are exonucleolytically degraded. In replicative transposition, cuts on just one of the two strands are made at each end of the element. The free ends of the element are joined to the target, and the remaining 3′ ends prime replication across the element. This leads to duplication of the element and the formation of a cointegrate consisting of the complete donor molecule and the target DNAs containing an extra copy of the element joined by direct repeats. This cointegrate can be resolved either by a site specific resolvase or by homologous recombination mediated by the host, resulting in the separation of the donor molecule and the recipient molecule (containing the transposed element).

Neither extensive DNA sequence homology nor general recombination function is needed for transposition (Berg *et al.*, 1988). Transposition requires interaction of element-encoded transposases with sequences at transposable element ends. The element ends have been demonstrated to be essential for transposition (Galas and Chandler, 1989). In some IS elements the ends are recognized and acted on by transposases (Derbyshire and Grindley, 1992; Ichikawa *et al.*, 1987; New *et al.*, 1988). To study the ends of IS1, a set of mutated ends was tested for the ability to bind transposase InsA and for transposition activity. They found that there were two distinct functional domains in the ends of IS1: one contained the specificity determinants for recognition by InsA and one determined the rate of some steps in the transposition process other than InsA binding (Galas and Chandler, 1989). The IS element insertion specificity may be due to recognition of particular sequences or conformation in target DNAs. Transposition is regulated by both the IS element and

host factors. These could include a negative regulatory protein encoded in the IS, transcription termination factor Rho within an IS, and DNA polymerase I, DNA gyrase, Dam methylase and Dna A protein from the host (Galas and Chandler, 1989).

C. IS Elements in MAC

To date, six IS elements have been isolated from MAC. These elements, along with some IS900-related elements from other bacteria, are summarized in Table 1.

IS900, the first IS element characterized in mycobacteria, was identified from the clone pMB22. This clone was derived from a genomic library of a *M. paratuberculosis* strain isolated from the tissue of a patient with Crohn's disease (Green *et al.*, 1989). IS900 was shown to be highly specific to *M. paratuberculosis* (Moss *et al.*, 1991; Moss *et al.*, 1992b; Vary *et al.*, 1990).

Upon screening DNA from *M. avium* strains by Southern blot analysis using the clone pMB22 as a probe, some strains of *M. avium* were found to produce a complex banding pattern. Subsequently, IS901 was identified and found in most *M. avium* animal strains (Kunze *et al.*, 1991; Kunze *et al.*, 1992; Nishimori *et al.*, 1995) and IS902 was identified from *M. avium* subsp. *silvaticum* (Moss *et al.*, 1992a). Since IS901 and IS902 are virtually identical, only IS901 will be considered here (see below).

Recently, two more IS elements, IS1110 and IS1245, have been identified from *M. avium* and added into the GenBank database. IS1110 was discovered during a study of plasmids in AIDS-derived *M. avium* strains. IS1110 is a 1457 bp element lacking terminal inverted repeats and is related to IS900 and IS901. IS1110 was detected in some *M. avium* isolates and was considered as a highly

Table 1. IS elements in MAC and some IS900-related elements in other bacteria

Sequence name	Host	Size (bp)	Direct repeats (bp)	Inverted repeats (bp)	References
IS900	*M. paratuberculosis*	1451	-	-	Green *et al.*, 1989
IS901	*M. avium*	1472	-	-	Kunze *et al.*, 1991
IS902	*M. avium*	1470	-	-	Moss *et al.*, 1992
IS1110	*M. avium*	1457	5	-	Hernandez Perez *et al.*, 1994
IS1245	*M. avium*	1313	-	15*	Guerrero *et al.*, 1995
IS1141	*M. intracellulare*	1588	4-5	23	Genbank (unpublished)
IS110	*Streptomyces coelicolor*	1550	-	15*	Bruton and Chater, 1987
IS116	*S. clavuligerus*	1421	-	-	Leskiw *et al.*, 1990
IS117	*S. coelicolor*	2527			Henderson *et al.*, 1990
IS492	*Pseudomonas atlantica*	1202	5	-	Bartlett and Silverman, 1989

* indicates imperfect inverted repeats

mobile genetic element since the transposition events of IS1110 can be detected in random colonies without any selection pressure (Hernandez Perez *et al.*, 1994).

IS1245 was identified from an *M. avium* genomic fragment. It is a 1313 bp element with two imperfect inverted repeats and one open reading frame. IS1245 can be detected in a variety of *M. avium* isolates, but human isolates characteristically were found to have multiple copies and a greater diversity of RFLP patterns (Guerrero *et al.*, 1995).

In *M. intracellulare*, IS1141 has been identified (GenBank entry only, unpublished). IS1141 is 1588 bp long, containing 23 bp inverted repeats at its ends. Although IS1141 only appears in some *M. intracellulare* strains, its transposition may be associated with colonial variation.

IS900, together with IS901, forms one of the best studied groups of mycobacterial repeated DNA sequences. Since these two IS elements are also closely related to my study, I will focus on these two elements.

D. IS900

IS900 has a size of 1451 bp. It lacks terminal inverted repeats and flanking direct repeats. An analysis of the sequences flanking IS900 in three different clones and one unoccupied locus from *M. avium* revealed a specific recognition sequence of 5′ CATG$(N)_{4-6}$*CNCCTT 3′ (the asterisk corresponds to the insertion site) for IS900 insertion. In the three clones examined, the element has the same orientation with respect to this recognition sequence.

IS900 showed an overall homology of 52% with IS110, an IS element identified in *Streptomyces coelicolor* (Bruton and Chater,

1987; Chater *et al.*, 1985). No homologies with other insertion sequences were observed. The strongest homology between IS900 and IS110 occurs at their 3′ end, at both the DNA and the amino acid level. From the unique sequence at the 5′ region of IS900, species specific probes have been generated by either PCR amplification or subcloning of IS900; e.g. subclone pMB22/S12, which is located at position 163-414 of the 5′ end of IS900 (El-Zaatari, 1994; Vary *et al.*, 1990). These DNA probes were reported specifically distinguish *M. paratuberculosis* from a wide range of other organisms including members of MAC (Frank and Cook, 1996; Moss *et al.*, 1991; Moss *et al.*, 1992b).

DNA sequence analysis found that IS900 has 66% G + C content, similar to that of its host. There is a major ORF of 1197 nucleotides (ORF1197) corresponding to a protein of 399 amino acids. The predicted GUG initiation codon is preceded by an S-D sequence. In vitro, under the control of the exogenous promoters, transcription and translation of ORF1197 revealed expression of a protein, p43, which corresponded to the predicted molecular mass of the ORF1197 and was presumed to function as a transposase (Tizard *et al.*, 1992).

Recently, a second ORF (ORF2) from IS900 has been identified (Doran *et al.*, 1994; Murray *et al.*, 1992). ORF2 is in the opposite direction to the ORF1197. However, the transcriptional and translational start signals of this ORF are identified outside of the IS900 element. The S-D sequence is downstream of the initial codon and is adjacent to the complementary sequence of the 3′ end of the IS900 element. This S-D sequence (GAGGAA) is highly homologous to the flanking sequences of IS900. A promoter sequence P_{AN} was located outside the 3′ end of IS900. ORF2 expression was detected

by fusion of P_{AN} and ORF2 with the *lacZ* gene. This ORF2 protein may have a role in the increased pathogenicity of *M. paratuberculosis* associated with iron uptake (Doran *et al.*, 1994).

England (England *et al.*, 1991) has successfully demonstrated that IS900 can facilitate stable integration of a foreign gene into mycobacteria. This was done by constructing an artificial transposon with two copies of the insertion sequence IS900 flanking a kanamycin resistance gene into a non-mycobacterial replicating vector. Constructs were then transformed into mycobacteria by electroporation and the kanamycin-resistant transformants were selected. Analysis of these transformants indicates that IS900 can transpose involving both simple insertion and cointegrate formation.

E. IS901

IS901 was found in pathogenic strains of *M. avium*. It is 1472 bp in size. Although the homology between IS901 and IS900 is only 60%, they share similar properties. Like IS900, IS901 has no terminal inverted repeats or flanking direct repeats. It has a potential insertion site of $CATN_{(7)}$*TTCCNTTC and inserts in the same orientation with respect to the target site in two clones which have been checked. IS901 has 62% G + C content. It contains a major ORF for expression of a protein of 401 amino acids presumed to function as a transposase.

An ORF2 was also identified by Doran et al. (Doran *et al.*, 1994) with characteristics similar to those of IS900. The ORF2 of IS901 is located on the complementary strand to its transposase gene. This ORF2 is 1431 bp in size, starting with a GUG initiation codon and preceded by a potential S-D sequence, GAGGA.

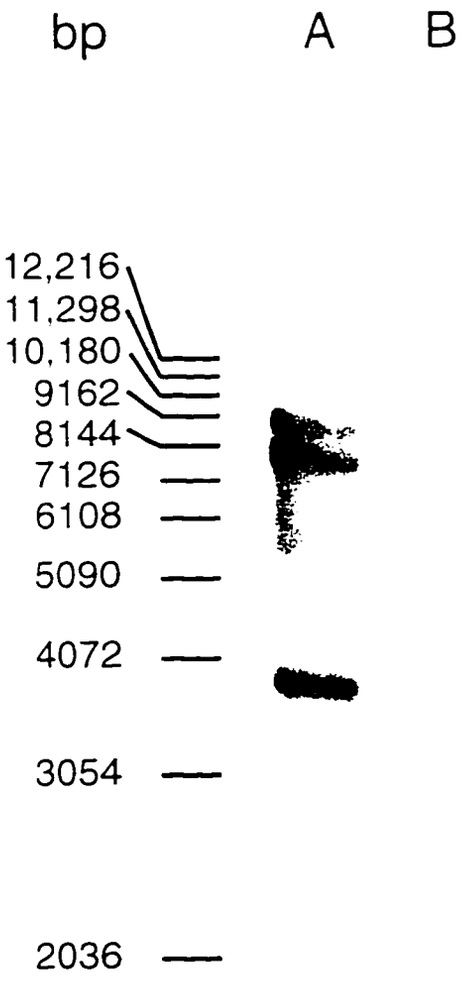
IS901 has been used to classify *M. avium* into two subtypes: *M.*

avium type A/I strains possessing IS901 and *M. avium* type A strains lacking IS901. *M. avium* type A/I appears to be primarily an animal and bird pathogen. It is rarely isolated from humans, and so far has not been identified in any MAC strains from AIDS patients. *M. avium* type A is primarily an environmental organism, and is the predominant *M. avium* type causing human disease, particularly in AIDS patients (Kunze *et al.*, 1992; McFadden *et al.*, 1992). The reasons for the markedly different disease associations between these two types of *M. avium* remain unclear, although an in vivo study of the proliferation of *M. avium* in mice implicated IS901 with increased pathogenicity (Kunze *et al.*, 1991).

3. Objectives of This Project

This study was based on the observations of other individuals in Dr. Kunimoto's lab. As previously described, the subclone pMB22/S12 of IS900 was reported to be relatively specific to *M. paratuberculosis*. In Dr. Kunimoto's lab, a summer student used pMB22/S12 as a probe to screen 66 human MAC isolates, two of which hybridized with this probe. Figure 3 contains my own Southern blot and hybridization confirming this observation. Strain one was obtained from the lymph node of an immunocompetent child. This strain was identified as *M. avium* by a Gen-Probe test with a *M. avium* probe (GEN-PROBE, Inc., San Diego, CA). Strain two was from the sputum of a 73 year old man with pneumonia and congestive heart failure, and was identified as *M. intracellulare* by a Gen-Probe test with a *M. intracellulare* probe (GEN-PROBE, Inc., San Diego, CA). Using pMB22/S12 to probe *Bam*HI-digested DNA, strain

Figure 3. *Bam*HI-digested genomic DNA of strain one and strain two probed with radiolabeled pMB22/S12. Strain one shows three hybridizing bands at 3.8 Kb, 8 Kb and 9 Kb (lane A) and strain two shows one hybridizing band at 10 Kb (lane B).


bp
A
B
12,216
11,298
10,180
9162
8144
7126
6108
5090
4072
3054
2036

one showed three positive hybridizing bands at 3.8 Kb, 8 Kb and 9 Kb, whereas strain two showed one hybridizing band at 10 Kb. From strain one, a 1 Kb fragment from the 3.8 Kb hybridizing band had been cloned as pPP1.0 and sequenced. A search of Genbank with the NIH BLAST program revealed that the 1 Kb fragment contained an area that has 82% homology with the first 500 bp of IS900.

Based on these results, i.e. the appearance of more than one copy in a MAC isolate and the partial sequence with a high degree of homology to IS900, it was suggested that these two MAC isolates may contain IS900-like element(s). Therefore my project was to complete the identification of this element by further cloning, sequencing and data analysis (In this paper this IS900-like element was named ISN).

II. Material and Methods

1. Bacterial Strains, Plasmids and Phage

Clinical MAC isolates were received from the Mycobacteriology Department of the Provincial Laboratory of Public Health for Northern Alberta, University of Alberta.

E. coli strains DH5α and TG2 were obtained from Gibco/BRL Life Technologies (Bethesda, Maryland)

DNA cloning vectors used in this study were pBluescript II SK+/- and pBluescript II KS+/- (Stratagene Cloning Systems, La Jolla, CA). The helper phage used was R408 kindly given by Dr. John Elliott's lab.

2. Media and Growth Conditions

Mycobacterial strains were grown on Middlebrook 7H10 agar slants at 37°C in 10% CO_2 for 4-6 weeks.

E. coli strains were grown in either broth or on solid media. The broth used was Terrific broth (TB), Luria-Bertani broth (LB) and 2 × YT/MT broth (Sambrook *et al.*, 1989). The solid media was made from LB with 1.5% agar added. All *E. coli* cultures were maintained at 37°C for different periods as experiments required.

3. Isolation of DNA

(1) Chromosomal DNA extraction

With visible colonies on solid media, mycobacteria were scraped with a loop and transferred into 14 ml polypropylene tubes with 0.5-1.0 ml of siliconized and dried 40 mesh glass beads. Tubes were vortexed vigorously for 6-10 minutes. 600 µl of TE buffer (10 mM Tris and 1 mM EDTA, pH 8.0), and 1 ml of phenol was added and mixed well. The solution was then centrifuged at 1,000 × g for 1 minute. The supernatant was transferred to a microcentrifuge tube (Eppendorf) and centrifuged again at 14,000 × g for 3 minutes. The aqueous layer was transferred to a new Eppendorf tube and an equal volume of phenol was added. The tube was shaken well and centrifuged at 14,000 × g for 5 minutes. The aqueous phase was re-extracted 3 to 5 times with a 1:1 phenol-chloroform mixture until the interphase was clear. Then the aqueous phase was extracted once with 1 volume of chloroform. To the extracted aqueous phase, 1/10 volume of 3.0 M sodium acetate (pH 5.2) and two volumes of -20°C absolute ethanol were added. After chilling in a -70°C freezer for at least 30 minutes, the tube was centrifuged at 14,000 × g at 4°C for 10 minutes. The supernatant was removed and the pellet was washed with cold 70% ethanol twice. The pellet was then dried and dissolved in 50-100 µl of TE buffer containing 1 µl of 10 mg/ml RNase A (Boehringer Mannheim, Laval, Quebec).

(2) Miniprep of Plasmid DNA

Plasmid DNA was prepared by an alkaline denaturation mini-preparation procedure, which is a modification of the method of Sambrook (Sambrook *et al.*, 1989). An isolated bacterial colony was

incubated in 4 ml of TB supplemented with ampicillin on a roller-wheel at 37°C overnight. The culture was then transferred and centrifuged in an Eppendorf tube. The supernatant was removed and the cell pellet was resuspended by vortexing in 160 µl of Solution I (50 mM glucose, 25 mM Tris-Cl [pH 8.0], 10 mM EDTA [pH 8.0]). 320 µl of fresh Solution II (1% SDS, 0.2 N NaOH) was added and the tube was inverted 5 to 10 times, then incubated on ice for 5 minutes. After that, 240 µl of Solution III (5 M potassium acetate in glacial acetic acid) was added. The tube was inverted until the solution was thoroughly mixed and then incubated on ice for 5 to 10 minutes. The tube was then centrifuged at 14,000 × g at 4°C for 5 minutes. The supernatant was transferred to a new Eppendorf tube, and precipitated with 1 volume of isopropanol. The pelleted DNA was washed with cold 70% ethanol and air dried. The dried DNA pellet was redissolved in 350 µl TE buffer containing 2 µl of 10 mg/ml RNase A. After the tube was incubated in a 37°C waterbath for one hour, 300 µl of TE buffer was added to bring the total volume to about 700 µl. An equal volume of phenol was added and the tube was shaken briefly then centrifuged at 14,000 × g at room temperature for 5 minutes. The aqueous layer was extracted with equal volumes of 1:1 phenol-chloroform and then chloroform. The extracted aqueous layer was then mixed with 1/10 volume of 3 M sodium acetate (pH 5.2) and 2 volumes of cold absolute ethanol. After storing at - 70°C for 30 minutes, the DNA was collected, washed and dried. The DNA pellet was dissolved in 50 µl of TE buffer. The DNA concentration was measured with a TKO 100 Mini-Fluorometer (Hoefer Scientific Instruments, San Francisco, CA).

(3) Large-Scale Preparation of Plasmid DNA

A modified large-scale alkaline denaturation procedure was used in this study (Sambrook *et al.*, 1989). A single colony was cultured in 4 ml of broth on a roller wheel at 37°C to late log phase. Then, 500 ml of TB broth supplemented with ampicillin was inoculated with 1 ml of a late exponential growth phase culture and incubated at 37°C overnight. The culture was centrifuged in a centrifuge bottle at 4,000 × g at 4°C for 15 minutes. The cell pellet was resuspended well in 10 ml of lysis buffer (50 mM glucose, 25 mM Tris [pH 8.0], 10 mM EDTA [pH 8.0], 5 mg/ml lysozyme) and maintained on ice for 10-20 minutes. 20 ml of freshly prepared 0.2 N NaOH, 1% SDS was added and the swirled mixture was incubated at room temperature for 5 minutes. 15 ml of ice cold 7.5 M ammonium acetate (pH 7.5) was then added. The solution was mixed thoroughly by swirling and rotating, and stored on ice for 5 minutes. Following incubation, the bacterial lysate was centrifuged at 12,000 × g at 4°C for 10 minutes and the supernatant was transferred into a 50 ml conical tube containing 25 µl of RNase A (10 mg/ml). After one hour incubation at 37°C, one volume of 1:1 phenol-chloroform was added. Following 2-3 minutes of vigorous shaking, the mixture was centrifuged at 4,000 × g for 5 minutes. The supernatant was transferred to a fresh 50 ml conical tube and precipitated with one volume of cold isopropanol at -20°C for at least 30 minutes. The DNA was collected by centrifugation at 12, 000 × g at 4°C for 30 minutes and rinsed with ice cold 70% ethanol twice. The pellet was dried under vacuum for 10-15 minutes, and resuspended in 1.5 ml TE buffer and 3 ml of CsCl stock (1.2 gm/ml H_2O [wt/vol]). The solution was mixed and

transferred to a VTi 75 tube (Beckman). 50 µl of ethidium bromide (10 mg/ml) was added. The tube was then filled with 2:1 CsCl stock-distilled water and heat sealed. Density gradient centrifugation was carried out in a 75 Ti fixed angle rotor (Beckman) in a Beckman L5-75 ultracentrifuge at 55,000 rpm at 19°C overnight. Using a 16G needle and a 5 ml syringe the plasmid DNA band (the lower band) was pulled out and transferred into a 14 ml conical polypropylene tube. Ethidium bromide was removed by extracting the plasmid solution with water-saturated n-butanol 4-6 times. The aqueous phase was precipitated with 70% ethanol at -20°C for 20 minutes. DNA was pelleted at 8,000 × g at 4°C for 30 minutes and rinsed with 70% ethanol. After the pellet was resuspended in 400 µl of TE buffer in an Eppendorf tube, 1/10 volume of 3 M sodium acetate (pH 5.3) and 1 ml of cold 95% ethanol were added. The mixture was stored at -20°C for 30 minutes and then centrifuged at 14,000 × g at 4°C for 10 minutes. The DNA pellet was washed, dried and dissolved in 200 µl of TE buffer, and the DNA concentration was determined.

4. Restriction Endonuclease Digestion of DNA

The enzymes used in this study were obtained from either New England Biolabs (Beverly, MA) or Gibco BRL Life Technologies (Grand Island, NY). The amount of enzyme used was adjusted depending on the amount of DNA to be digested. The digestion buffer and temperature were as recommended by the manufacturers.

5. Agarose Gel Electrophoresis

Ultra pure agarose (Gibco BRL) was used at 0.8%-1.0% in 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA). The agarose was heat dissolved. After the dissolved agarose was cooled and ethidium bromide was added, the gel was then poured into a plastic gel-casting tray. The solid gel was submerged in 1 × TAE buffer and the DNA samples were loaded. The gel was electrophoresed using constant voltage at room temperature. Bands of DNA were visualized by illumination with long-wavelength (302 nm) ultraviolet light. Photographs were taken by an Imager Video-Capture System (Appligene).

Bands with DNA to be used in further experiments were excised from the agarose gel with a clean safety razor blade and placed in an Eppendorf tube. DNA was then purified using the Geneclean II Kit (BIO 101 Inc., Vista California). Three volumes of NaI (6 M) was added to the weighed gel slice and kept at 55°C for 5-10 minutes. After the agarose gel was completely dissolved, 10 µl of Genclean glassmilk (a silica matrix) was added and placed on ice for 5-10 minutes. Then the matrix and the bound DNA were pelleted by centrifugation for 10-20 seconds. The pellet was washed with NEW WASH (a solution of NaCl, Tris, EDTA and ethanol) 3 times. After the pellet was dried, the bound DNA was eluted with 20 µl of TE buffer or distilled water at 55°C for 5 minutes. The tube was centrifuged for 30 seconds to pellet the silica matrix. The eluted DNA was stored at -20°C until used.

6. Radioactive Labeling of DNA

In this study DNA probes were labeled by the random primer labeling method described by Feinberg and Vogelstein (Feinberg and Vogelstein, 1984). Approximately 50 ng of DNA to be labeled was denatured by heating at 95-100°C for 2-3 minutes and rapidly chilled on ice, followed by the addition of 10 µl of mixture containing unlabeled dATP, dTTP and dGTP, HEPES, hexadeoxy-ribonucleotides, TE buffer, BSA, 5 µl of [α-^{32}P]dCTP and 5U of Klenow enzyme (New England Biolabs) in a 50 µl volume of reaction. The reaction was incubated at 37°C for 1-5 hours. The labeled probe was purified with a Sephadex G-50 (Pharmacia, Uppsala, Sweden) drip column to remove unincorporated labeled nucleotides. For purification, the Sephadex G-50 slurry was added into a 1 ml syringe which had a circular disc of glassfibre filter placed in the bottom. The Sephadex G-50 column was washed with 300 µl of TE buffer twice. The 50 µl of labeled sample was loaded into the column and the tube was rinsed with 250 µl of TE buffer which was also loaded into the column and the eluate discarded. Then 300 µl of TE buffer was added to the column twice and 300 µl eluates were separately collected in two Eppendorf tubes. 1 µl of eluate from each tube was counted in 5 ml of scintillation fluid in a Beckman LS6800 scintillation counter. The probe was stored at -20°C until further use.

7. Southern Blot Hybridization

DNA to be screened was digested, loaded on an agarose gel and

electrophoresis was performed. The DNA banding patterns and the 1 Kb molecular weight markers (Gibco BRL) were recorded with the Imager Video-Capture System (Appligene). The position of the molecular weight markers was copied by marking on a transparency. The gel was immersed in a denaturation solution (0.5 M NaOH, 1 M NaCl) and gently agitated for 30 minutes. The denaturing solution was decanted and a neutralization solution was added (3 M NaCl, 0.5 M Tris, pH7.0) for another 30 minutes with gentle shaking. The gel was then placed on 2-3 sheets of Whatman 3MM paper (Fisher Scientific, Pittsburgh, PA) soaked in 10 × SSC (1.5 M NaCl, 150 mM sodium citrate, pH7.0) on top of a piece of Cling Wrap. The Cling Wrap was then folded to cover the four edges of the 3MM paper. A piece of Hybond™-N nylon membrane (Amersham Life Science Inc., Arlington Heights, IL) was cut to the same size as the gel to be blotted and placed on the top of the gel. Air bubbles were pressed out by rolling a 20 ml Pasteur pipet on the membrane. Two more pieces of 3MM paper were placed on the top and then a 3-4 inch layer of paper towels. This stack was pressed with a 1kg weight. The transfer was carried out for 10-12 hours. The membrane was marked with the position of the gel wells. The DNA on the membrane was fixed by cross-linking in the UV Stratalinker™ 2400 (Stratagene) and the membrane was allowed to be air dried. The blot was then subjected to prehybridization by soaking in 10 ml of hybridization solution (50% [v/v] formamide, 50 µg/ml single stranded salmon sperm DNA, 5 × SSC, 10 × Denhardt's, 5 µM EDTA, 0.5% SDS and 0.1 g/ml Dextran) in a hybridization tube and the tube was rotated at 42°C for 1 hour. After adding the ^{32}P labeled probe, the hybridization mixture was continually rotated at 42°C overnight. Then the

hybridization solution was decanted into a radioactive waste container. The blot was washed with 0.5 × SSC, 0.2% SDS wash solution at 55°C for 30 minutes 2-3 times and checked for counts by using a Geiger-Mueller counter (Ludlum Measurements Inc., Sweetwater, TX). The blot was wrapped in Saran wrap and exposed to X-ray film (Kodak or Fuji) in a cassette holder.

8. Ligation and Transformation

The desired DNA and a vector were digested with restriction enzyme(s). When digestion resulted in the same termini on both ends of linearized vector, alkaline phosphatase was used to remove the 5′ phosphate to reduce self ligation. This dephosphorylation was carried out by incubation of the vector with calf intestinal alkaline phosphatase (CIAP) (GIBCO BRL Life Technologies, Inc.) at 37°C for 30 minutes. The DNA fragment and the vector to be ligated were recovered by gel electrophoresis and DNA extraction. Ligation was carried out usually with a 1:3 ratio of vector and insert, and 1U of T4 DNA ligase (GIBCO BRL Life Technologies, Inc.). The ligation reaction was carried out at 16°C overnight. Half of the ligation mixture was used for transformation.

The competent cells used for transformation were prepared by a modified $CaCl_2$ procedure (Sambrook *et al.*, 1989). Cells were grown in 2.5 ml 2 × YT/MT broth on a rolling wheel at 37°C overnight. 0.5-1 ml of this overnight culture was used to inoculate 500 ml of 2 × YT/MT and grown with shaking at 37°C until the culture's OD_{600} was read at 0.5. Then the culture was chilled on ice for 10 minutes and centrifuged in a prechilled centrifuge bottle at 1,500 ×g at 4°C for 10

minutes. The supernatant was decanted and the cell pellet was resuspended gently in half of the volume (250 ml) of prechilled 50 mM $CaCl_2$, 20 mM HEPES solution. The cells were pelleted and again suspended in 1/10 of the original volume (50 ml) of prechilled solution containing 50 mM $CaCl_2$, 20 mM HEPES and 5% (V/V) glycerol. Cells were dispensed in prechilled Eppendorf tubes and frozen on dry ice and stored at -70°C for further use.

Prior to transformation, competent cells were thawed on ice. 100 µl of cells was transferred to a prechilled 14 ml polystyrene tube. Ligated DNA was added and gently mixed. After 30 minutes on ice, the transformation mixture was heat-shocked in a 42°C water bath for 90 seconds, followed by 3 minutes on ice. Then 800 µl of LB broth was added to the reaction and incubated on a roller wheel at 37°C for 45 minutes. About 100 µl of the transformation culture was plated on appropriate selective media with 4 µl of IPTG (isopropylthio-β-D-galactoside) (200 mg/ml) and 40 µl of X-gal (20 mg/ml) spread on the surface. The plates were incubated at 37°C overnight.

9. Genomic DNA Cloning

M. avium genomic DNA was digested with restriction endonuclease and DNA fragments were fractionated by electrophoresis on agarose gels. Fragments of the desired size were isolated from the gel using a Geneclean II Kit (Bio 101 Inc.) as described previously, then ligated into the pBluescript vector. Recombinant plasmids were transformed into *E. coli* DH5α.

10. Colony Screening

Colony screening was used to select the colony with the desired insert. Using sterilized toothpicks, colonies were transferred from an experiment plate and inoculated onto duplicate antibiotic plates, one of which had a Hybond™-N nylon membrane disc (Amersham Life Science Inc.) on the surface. The two plates were incubated at 37°C overnight. The plate without the nylon membrane was kept as a template and the nylon membrane was used for hybridization with a ^{32}P-labeled probe. The membrane was first soaked in 10 ml of 0.5 M NaOH at room temperature for 5 minutes to lyse the cells. After removing excess fluid, the membrane was soaked in 10 ml of solution containing 1.5 M NaCl and 0.5 M Tris, pH7.4 for 5 minutes. The membrane was then soaked in 10 ml of 1.5 M NaCl and 2 × SSC for another 5 minutes. The filter was then dried, UV cross-linked and hybridized with a radiolabeled probe. The autoradiograph of the hybridized filter was aligned with the template plate. Hybridizing colonies were picked and grown for plasmid preparation.

11. DNA Sequencing

(1) Sequencing Primers

T3/T7 primers which flank the multiple cloning site (MCS) of pBluescript vectors were used to sequence the insert of pBluescript recombinants. Some specific primers used in this study were selected from newly determined sequences. The oligonucleotides used for sequencing and/or PCR in this study are listed in Table 2.

Table 2. Primers for sequencing and/or PCR

Name	Nucleotide sequence	Location	Supply
T3	5′ AAT TAA CCC TCA CTA AAG GG 3′	from T3 and T7 bacteriophage promoters flanking the MCS of pBluescript vectors.	Commercially available
T7	5′ GTA ATA CGA CTC ACT ATA GGG C 3′		
DKU62	5′ AGA GTT TGA TCC TGG CTC AG 3′	corresponding to *E. coli* 16S rRNA positions 8 to 28 and 614 to 633 respectively.	DNA Synthesis Laboratory, University of Alberta
DKU92	5′ CAC GC(T/C) CAC AGT TAA GC(T/C) GT 3′		
DKU117	5′ CCG CGG ATT CTC CGT CC 3′	located on flanking regions of ISN in pEH5.0. DKU117 is 76 bases from the 5′ and DKU118 is 87 bases from the 3′ end of ISN.	DNA Synthesis Laboratory, University of Alberta
DKU118	5′ GCC CTG GCG TTC CTA TGC C 3′		
DKU119	5′ ACT TGC TGG CCA CCT TCC 3′	located on flanking regions of ISN in pB3.8. DKU119 is 52 bases from the 5′ and DKU120 is 78 bases from the 3′ end of ISN.	DNA Synthesis Laboratory, University of Alberta
DKU120	5′ GCA TCA TCT TCG GCC ACC 3′		

(2) Template Preparation

In this study, an Exonuclease III (Exo III) based nested deletion strategy was applied, using the procedure developed by Henikoff (Henikoff, 1984). In brief, the plasmid with the insert DNA to be sequenced was digested with two different restriction enzymes: one generated a 3′ overhang near the primer binding site and another generated a 5′ overhang or blunt end adjacent to the insert. This was followed by digestion with Exo III (New England Biolabs) which specifically digests insert DNA from 5′ overhang or blunt end restriction sites and the primer binding site is protected from Exo III digestion by the 3′ overhang restriction site. Samples of the Exo III digestion were removed at 30 second intervals into the tubes containing S1 nuclease (Boehringer Mannheim, Laval, Quebec), which was used to remove the single-stranded tails remaining. After neutralization and heat inactivation of the S1 nuclease, Klenow DNA polymerase was added to blunt the fragment ends. These vector-containing deleted fragments were circularized by ligation. The ligation mixture was used directly to transform competent cells. Thus, different regions of the insert fragment were brought close to the primer site for sequencing. In this way, two nested sets of deletions were generated from both ends of the insert DNA. One set of nested deletions was prepared for double strand DNA sequencing by plasmid mini-preparation or large-scale preparation. Another set was prepared for single-stranded DNA sequencing.

For single strand template DNA preparation, the mini-prepped DNA was transformed in the F' *E. coli* TG2 cells and plated. One single colony of the F' strain was picked into 2 ml 2 × YT/MT with ampicillin (100 µg/ml) and incubated at 37°C for 4 hours to

overnight. 0.2 ml of the resulting culture was transferred to 2 new fresh tubes containing 2 ml 2 x YT/MT without ampicillin and grown at 37°C for one hour, followed by superinfection with 0.2 ml of M13 helper phage R408 and grown at 37°C for 6 to a maximum of 8 hours. The cultures were then transferred to 2-3 Eppendorf tubes and pelleted at 4°C. The supernatant was immediately transferred to a new tube avoiding any pellet. Subsequent to precipitation with PEG (8000) at a final concentration of 10%, DNA was washed with cold 70% ethanol, treated with RNase, extracted with phenol, phenol-chloroform (1:1 [vol:vol]) and chloroform, and precipitated with 3 M sodium acetate acid (pH 5.2) and ethanol. The pellet was washed with cold 70% ethanol twice, dried and resuspended in 50 µl of distilled water. 3-4 µl of SS DNA was electrophoresed with helper phage in an adjacent well on a 0.7% agarose gel to determine yield and purity, and the rest of the SS DNA was stored at -70°C.

(3) Sequencing

DNA was sequenced using the dideoxy chain termination method of Sanger with either the Sequenase™ Version 2.0 DNA Sequencing Kit, (Amersham Life Science, United States Biochemical, Cleveland, Ohio) or the Taq DNA polymerase cycle sequencing developed by Murray and Craxton (Craxton, 1991; Murray, 1989).

For sequencing with the Sequenase™ Version 2.0 DNA Sequencing Kit, 1 µg of DNA for single stranded DNA template and 3-5 µg of DNA for plasmid template DNA were required. For sequencing double-stranded DNA template, alkaline-denaturation of template DNA was carried out. Briefly, template DNA was denatured by 0.2 N NaOH and heated in boiling water for 5-10 minutes followed by ethanol and

sodium acetate precipitation. The following procedure was the same for sequencing single stranded and double stranded template DNA. 1-2 pmol of sequencing primer was annealed to the single stranded DNA or denatured double stranded DNA in a total volume of 10 µl with sequenase buffer (40 mM Tris-HCl [pH 7.5], 20 mM $MgCl_2$, 50 mM NaCl). The annealing reaction was incubated at 65°C for 3 minutes then cooled slowly to room temperature. This was followed by labeling the reaction with [α-^{35}S]dATP (New England Nuclear, Boston, Massachusetts). To the annealing reaction, 1 µl of 0.1 M DTT, 2 µl of 1 × labeling mixture (5 × labeling mixture contained 7.5 µM of each dGTP, dCTP, dTTP), 1 µl of [α-^{35}S]dATP and 2 µl of 8-fold diluted Sequenase™ Version 2.0 were added. The labeling mixture was incubated at room temperature for 2-5 minutes prior to transfer of 3.5 µl of labeling reaction to each termination well containing 2.5 µl ddGTP, ddCTP, ddTTP and ddATP termination mixtures respectively (80 µM of each 4 dNTPs, 50 mM NaCl and 8 µM of ddGTP, ddCTP, ddTTP and ddATP respectively). After incubation at 37°C for 5 minutes, the reactions were stopped by adding 4 µl of stop solution (95% [v/v] formamide, 20 mM EDTA, 0.05% [w/v] bromophenol blue, 0.05% [w/v] xylene cyanol FF). The reaction was heated at 95-100°C for 3 minutes before loading onto a sequencing gel.

For cycle sequencing, about 1 pmol of sequencing primer was used for 5′ end labeling with [γ-^{33}P]ATP by T4 polynucleotide kinase (Boehringer Mannheim, Laval, Quebec). The labeling reaction was carried out in a volume of 5 µl with 1 µl of sequencing primer, 1 µl of 5 × kinase buffer (300 mM Tris-HCl [pH 7.8], 50 mM $MgCl_2$, 1 M KCl), 1 µl of [γ-^{33}P]ATP (2 pmol) and 1 µl of 1U/µl T4 polymerase kinase.

The reaction mixture was incubated at 37°C for 10 minutes and then at 55°C for 5 minutes to terminate the reaction. The quality and the quantity of template DNA required for cycle sequencing was less than that required in a typical noncycling reaction. 0.5 μg or less of template DNA was used for a cycle sequencing reaction. To 5 μl of end-labeled primer, 26 μl of template DNA and distilled water, 4.5 μl of 10 × Taq sequencing buffer (300 mM Tris-HCl [pH 9.0], 50 mM $MgCl_2$, 300 mM KCl), and 1 μl of 5U/μl Taq DNA polymerase (GIBCO BRL) were added. 8 μl of this mixture was transferred into each of the four sequencing reaction tubes, which contained 2 μl of ddGTP, ddCTP, ddTTP and ddATP termination mixtures (100 μM each dNTP and 0.2 to 2 mM ddGTP, ddCTP, ddTTP and ddATP respectively). 2 drops of mineral oil was dispensed into each reaction tube before starting the temperature cycling program. The temperatures for cycle sequencing with T3, T7 primers were 94°C for 40 seconds for denaturing, 55°C for 30 seconds for annealing, and 72°C for 1 minute for extension/termination for 30 cycles. Reactions were terminated by adding 5 μl of stop solution (95% [v/v] formamide, 10 mM EDTA [pH8.0], 0.1% [w/v] bromophenol blue, 0.1% [w/v] xylene cyanol). The reaction tubes were heated at 95-100°C for 5 minutes prior to loading onto a sequencing gel.

To solve compression problems, dGTP substitutes, dITP and 7-deaza-dGTP, were used. The reaction conditions were identical to those described above.

Direct sequencing of PCR products was also applied. PCR products were recovered from agarose gels with a Geneclean II kit in 12 μl of distilled water. To this 12 μl DNA template, 15 pmol of primer was

added. The mixture was heated in boiling water for 8 minutes and immediately frozen on dry ice. For each reaction, 1 µl of 5 × label mixture, 1 µl of 0.1 M DTT, 1 µl of [α-^{35}S]dATP, 3U of Sequenase™ Version 2.0 and 1 µl of distilled water were added to the primer template mixture which had been rapidly thawed to room temperature. Then the reaction was incubated at room temperature for 2-5 minutes, and 4.2 µl of the reaction was aliquoted to the 4 termination tubes containing 2.5 µl ddGTP, ddCTP, ddTTP and ddATP termination respectively. After incubation at room temperature for 5 minutes and an additional 5 minutes at 37°C, the reaction was terminated by adding 5 µl of formamide stop solution. The reaction was heated at 95-100°C for 3-5 minutes before loading the sequencing gel.

Polyacrylamide gels were used for electrophoresis of sequencing products. This gel was prepared with 6% of 19:1 of acrylamide/bis-acrylamide mixture, 7 M urea and 1 × TBE buffer (90 mM Tris, 90 mM boric acid, 2 mM EDTA). About 60 ml of gel was polymerized with 700 µl of 10% ammonium persulfate and 30 µl of TEMED. The gel plate was assembled with the electrophoresis apparatus (Owl Scientific Plastic Inc., Cambridge, MA) in a vertical position. After the sequencing sample was loaded, the gel was run in 1 × TBE buffer at constant power (40W) for the time required to achieve optimal resolution of the sequence. After the electrophoresis was completed, the gel mold was removed from the apparatus. The siliconized plate was then gently removed and the gel was lifted by placing a piece of Whatman 3MM paper on top of the gel with gentle pressing. The gel was then covered with Saran Wrap and dried in a gel dryer (Hoefer

Scientific Instruments, San Francisco, CA) with gel side up. The gel was dried under partial vacuum at 80°C for 2-3 hours. Saran Wrap was peeled off and the dried gel was exposed to X-ray film at room temperature overnight. The autoradiograph was developed and the sequence was read manually.

12. PCR Technique

The DNA to be amplified was diluted to 1 ng/µl. One nanogram of DNA was mixed with 49 µl of a prepared reaction mixture containing 5 µl of 10 × PCR buffer (200 mM Tris-HCl [pH 8.4], 500 mM KCl), 1.5 µl of 50 mM $MgCl_2$, 1 µl of 10 mM dNTP, 0.5 µM of each of the oligonucleotide primers and 0.5 µl of 5 U/µl Taq polymerase. The mixture was covered with 50 µl of mineral oil. The reaction was subjected to an initial denaturation step at 94°C for 1 minute, annealing at 55°C for 1 minute and extension at 72°C for 1 minute for 30 cycles using the Techne PHC-2 thermocycler (TECHNE Corporation). Annealing temperature was estimated using the general rule Tm-5°C, where Tm is approximated with the formula, $2°C \times [N_{A/T}] + 4°C \times [N_{C/G}]$, where G, C, T and A correspond to the four different bases (Innis and Gelfand, 1990).

13. PCR Colony Screening

PCR was used for quick screening of the different size inserts in cloning transformants. Colonies to be screened were picked with sterile toothpicks and inoculated onto a marked antibiotic plate, and

incubated at 37°C overnight. The colony was touched by a toothpick and removed to an Eppendorf tube containing 50 µl of distilled water. The remaining bacteria on the plate was kept as a template for further plasmid preparation. The tube was heated in boiling water for 5 minutes and centrifuged at 14,000 × g for 1 minute. 2-5 µl of supernatant was used as template DNA for PCR amplification. The PCR product was electrophoresed on an agarose gel to determine its size. The transformants with correct-sized inserts were grown for plasmid preparation.

III. Results

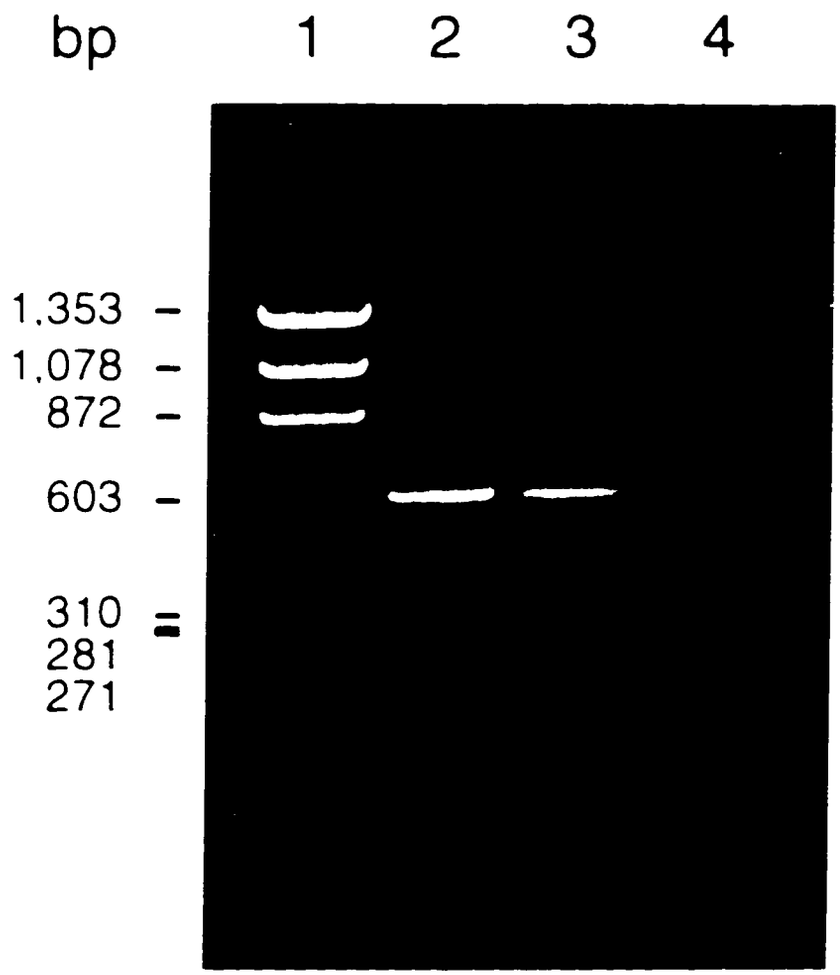
1. Strain Identification by 16S rRNA

Strain one and strain two are human isolates from the Provincial Lab of Northern Alberta. The BACTEC system (Roberts *et al.*, 1991) was used for culturing the mycobacteria and biochemical identification was used to confirm the strains as MAC group. Furthermore strain one was positive by the Gen-Probe *M. avium* probe and strain two was positive by the Gen-Probe *M. intracellulare* probe. To further confirm the identification of these two strains, 16S rRNA sequencing was used.

rRNA is an essential constituent of prokaryotic and eukaryotic ribosomes. The 16S rRNA molecule is highly conserved, with rare sequence changes in certain positions. However, the location of these changes is specific to the group or species in which they occur (Fox *et al.*, 1980; Woese *et al.*, 1983). In mycobacteria, the sequences of the 16S rRNA genes comprises about 1500 nucleotides. The nucleotide sequences specific for a range of Mycobacterium species have been defined by computer-assisted sequence comparisons. Most mycobacteria can be identified at position 123-273 (numbering system for *E. coli*) (Rogall *et al.*, 1990). Therefore appropriate primers for this region were used to amplify a DNA fragment followed by direct sequence determination.

With primers DKU62 and DKU92 (Table 2), fragments of approximately 600 bp were amplified from both strain one and strain two (Fig. 4) followed by direct sequencing. The sequence

Figure 4. PCR amplification of the 600 bp fragment of the gene coding for 16S rRNA from strain one (lane 2) and strain two (lane 3). *Hae*III digested ϕ x 174 DNA marker is in lane 1 and lane 4 is the negative control.


bp
1
2
3
4
1,353
1,078
872
603
310
281
271

results were then compared with previously characterized mycobacterial 16S rRNA sequences (Rogall et al., 1990). Strain one was found to possess a 16S rRNA sequence almost identical to the published reference sequence of *M. avium* and strain two possessed a 16S rRNA sequence similar to the reference sequence of *M. intracellulare* (Fig. 5). Therefore strain one was confirmed as *M. avium* and strain two as a probable *M. intracellulare*.

2. Restriction Map of the 10 Kb Fragment in Strain Two

Since strain two appeared to be *M. intracellulare*, it was suspected that the 10 Kb fragment which hybridized with pBM22/S12 in strain two could be another IS900-related sequence. From this 10 Kb *Bam*HI fragment, a 2.5 Kb *Pst*I fragment which hybridized with pMB22/S12 was cloned. The restriction map of this 2.5 Kb fragment (Fig. 6) is completely different from that of strain one (Fig. 7). This indicated that there may be two different IS900-like fragments in these two organisms. Since strain one contained three hybridizing bands and part of the sequence from strain one had shown high degree of homology to IS900, strain one was chosen for further study.

3. Characterization of the 3.8 Kb Hybridizing Fragment from Strain One

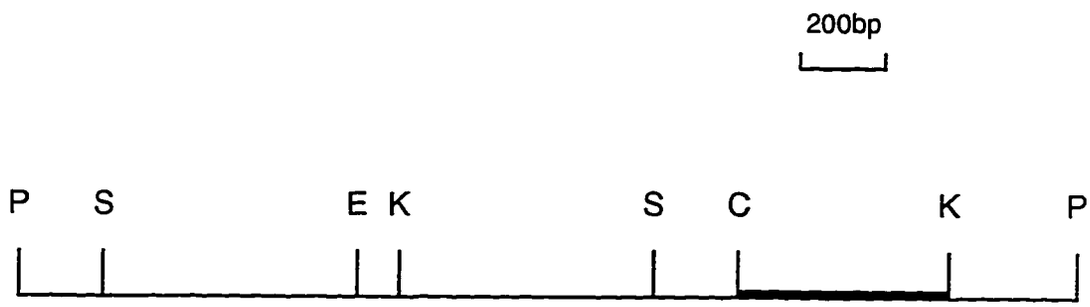
(1) Restriction Enzyme Digestion and Hybridization

As mentioned in the first chapter, *Bam*HI-digested chromosomal DNA of strain one hybridized with the ^{32}P-labeled pMB22/S12 probe

Figure 5. Alignment of published 16S rRNA reference sequences of *M. avium* and *M. intracellulare* (Rogall *et al.*, 1990). with the 16S rRNA sequences of strain one and strain two. *M. avium* was used as the reference sequence. The first nucleotide in the figure corresponds to *E. coli* 16S rRNA position 123 (Brosius *et al.*, 1978). The bars indicate deletions. Nucleotides different from those of *M. avium* are indicated. The underlined nucleotides are the specific positions for *M. avium* and *M. intracellulare.*

```
          130          140          150          160
           •            •            •            •
CGT GGG CAA TCT ACC CTG CAC TTC GGG ATA AGC CTG GGA   M. avium
... ... ... ... G.. ... ... ... ... ... ... ... ...  M. intracellulare
... ... ... ... G.. ... ... ... ... ... ... ... ...  Strain one
... ... ... ... G.. ... ... ... ... ... ... .C. ...  Strain two

          170          180          190          200
           •            •            •            •
AAC TGG GTC TAA TAC CGG ATA GG- ACC TCA AGA CGC ATG    M. avium
... ... ... ... ... ... ... ... ... .TT ..G ... ...  M. intracellulare
... ... ... ... ... ... ... ... ... ... ... ... ...  Strain one
... ... ... ... ... ... ... ... ... .TT ... ... ...  Strain two

          210          220          230
           •            •            •
--T CT- TCT GGT GGA AAG C-- TTT --- TGC               M. avium
... ... .TA ... ... ... ... ... ... ...               M. intracellulare
... ... ... ... ... ... ... ... ... ...               Strain one
... ... .T. ... ... ... ... ... ... ...               Strain two
```

Figure 6. Restriction map of 2.5 Kb *Pst*I fragment from strain two. C, *Cla*I; E, *Eco*RI; K, *Kpn*I; P, *Pst*I; S, *Sac*I. The thick solid line represents the region that hybridized with pMB22/S12.



showed three bands, 3.8 Kb, 8 Kb and 9 Kb. The 3.8 Kb fragment had been cloned into pUC19 by previous work of Dr. Kunimoto's lab. This clone was named as pB3.8. The restriction map of pB3.8 was determined (Fig. 7). Subsequently, fragments from pB3.8 were subcloned. pB3.8 DNA was digested with *Bam*HI, *Kpn*I, *Pst*I, *Xho*I individually or in combination and ligated to similarly digested pBluescript II vectors. The inserts of the resultant subclones were then ^{32}P-labelled and used as probes to hybridize the *Bam*HI-digested chromosomal DNA of strain one to locate the ISN on pB3.8. If the DNA showed three hybridizing bands, the corresponding probe was then considered as a part of ISN. It was found that the left end of ISN was located approximately at 0.8 Kb and the right end was at 2.8 Kb of pB3.8. From the left side, a 1 Kb fragment had been cloned (pPP1.0) and sequenced previously by others in our lab. To sequence the right side of this region, a 2 Kb *Bam*HI-*Pst*I fragment was cloned into a pBluescript II vector. The resultant clone was named pBP2.0.

(2) Sequencing pBP2.0 and Primary Analysis

To sequence both strands of pPB2.0, two nested deletion sets, pBX and pAE, were produced. The polymerase chain reaction was used for screening the transformants for the presence of the appropriate size inserts, as well as confirming the preservation of priming sites for sequencing (Fig. 8). The pBX deletion set was subject to double stranded DNA sequencing and the pAE set was determined by single stranded DNA sequencing. Figure 9 outlines the strategy used to sequence ISN in pB3.8. Computer analysis of the sequence of pPB2.0 combined with previous sequence data indicated that there is high degree of homology between the fragment in pB3.8 and IS900. This

Figure 7. Restriction map of pB3.8. B, *Bam*HI; C, *Cla*I; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sac*I; X, *Xho*I. The arrows indicate the possible two ends of ISN. From pB3.8, a 2 Kb *Bam*HI-*Pst*I fragment (in thick solid line) was cloned for sequencing.

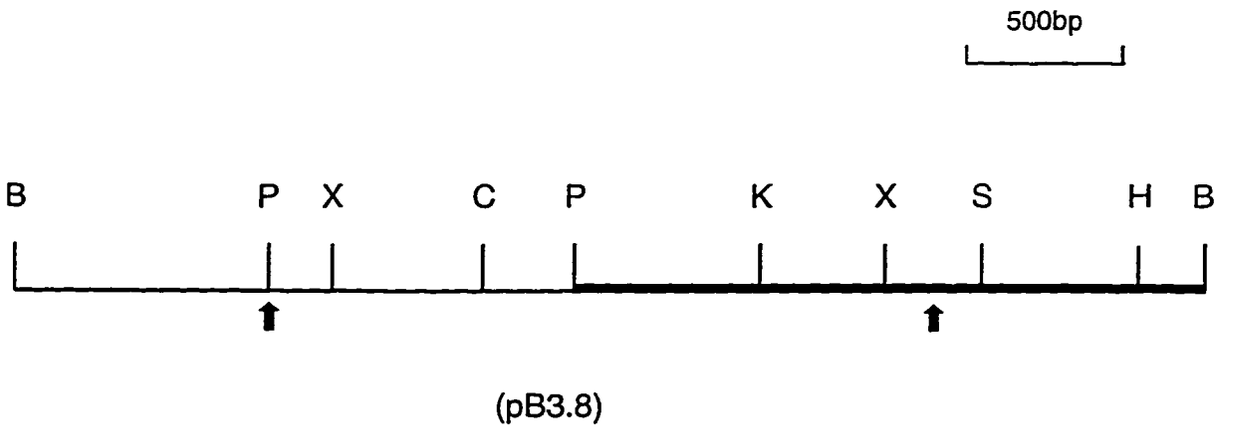
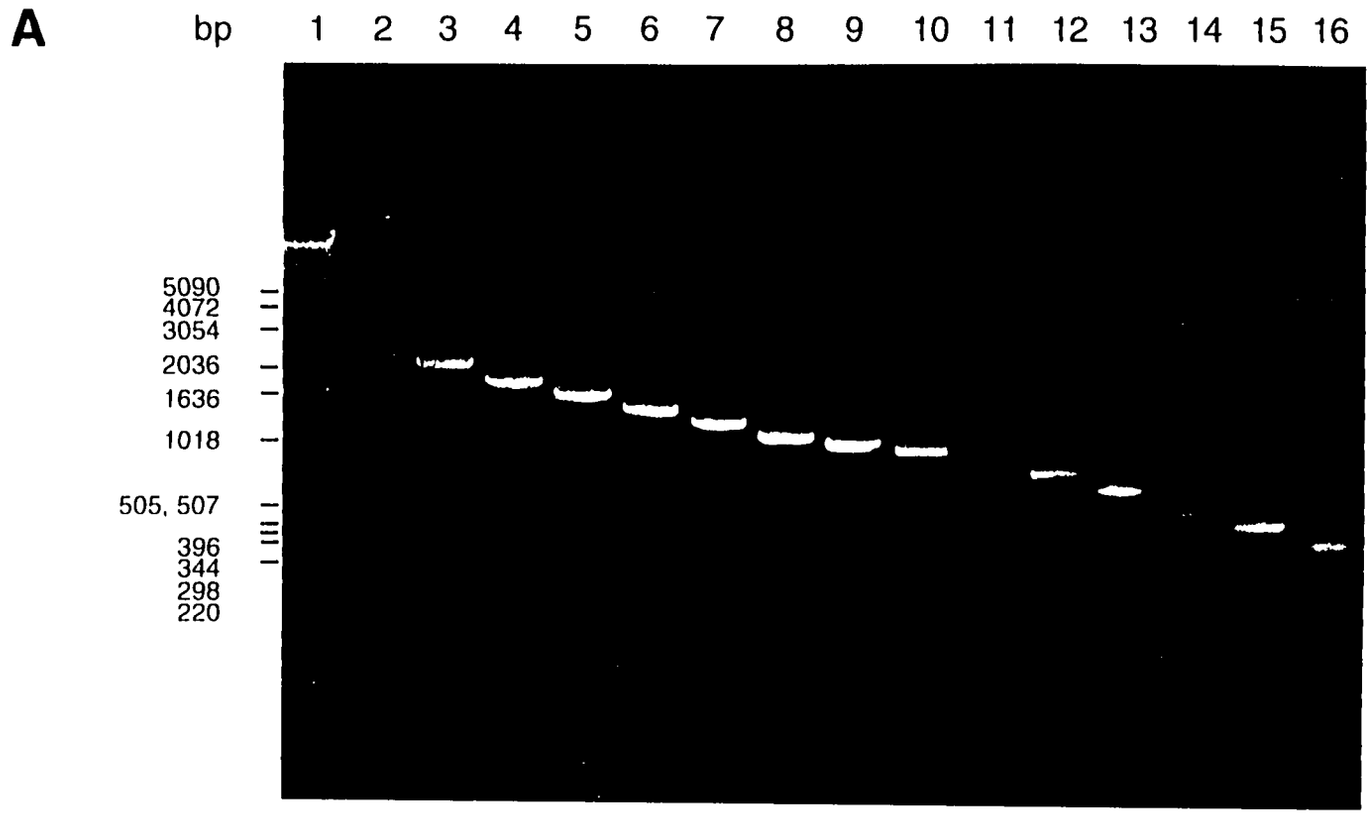


Figure 8. By PCR screening, two sets of clones, pBX (A) and pAE (B), with nested deletion fragments were obtained for sequencing. (A) lane 1, 1 Kb DNA ladder; lane2, the fragment amplified from clone pBP2.0; lane 3 to 14, the fragments amplified from pBX clones. (B) lane 1, 1 Kb DNA ladder; lane 2, the fragment amplified from pBP2.0; lane 3 to 16, the fragments amplified from pAE clones.


A
bp
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
5090
4072
3054
2036
1636
1018
505, 507
396
344
298
220

B

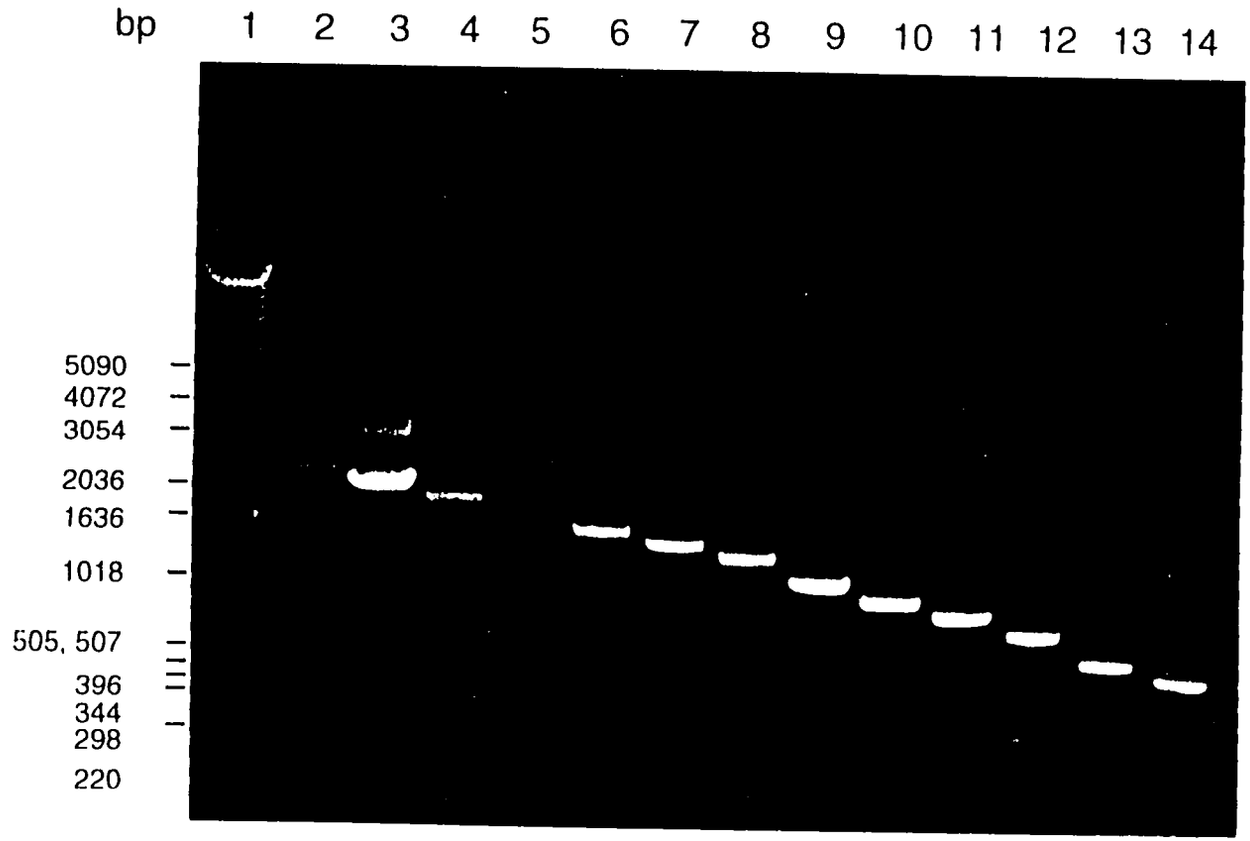
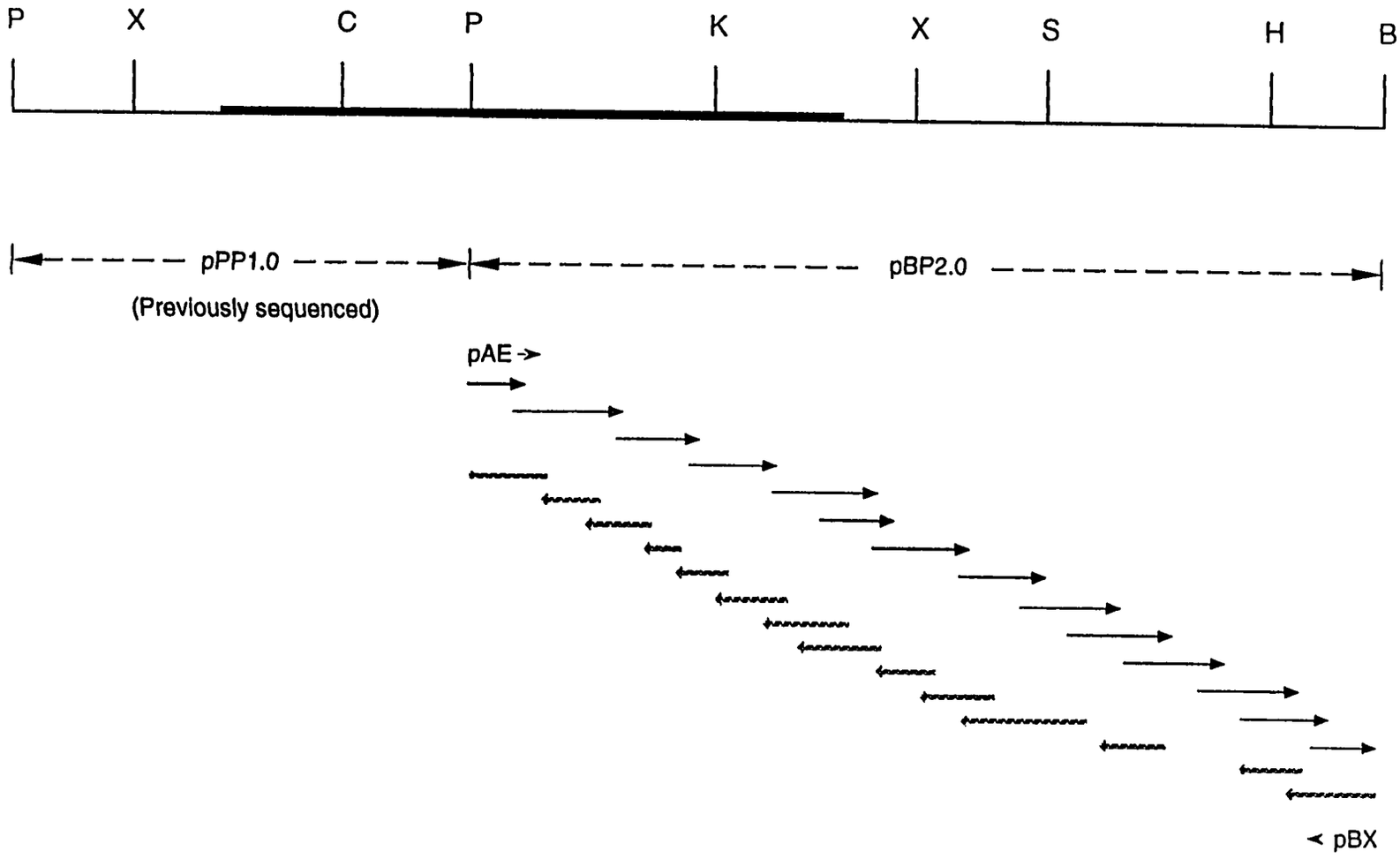

bp
1
2
3
4
5
6
7
8
9
10
11
12
13
14
5090
4072
3054
2036
1636
1018
505, 507
396
344
298
220

Figure 9. The restriction map and the strategy used to sequence ISN in pB3.8. The thick solid line denotes ISN and the thinner line denotes the flanking regions. pPP1.0 was previously sequenced by Dr. Kunimoto's laboratory and pPB2.0 was sequenced by using two sets of deletions, pBX (⟵) and pAE (→). B, *Bam*HI; C, *Cla*I; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sac*I; X, *Xho*I.


P
X
C
P
K
X
S
H
B
pPP1.0
(Previously sequenced)
pBP2.0
pAE
pBX

fragment seems to be an IS900-related element. No significant inverted repeats or direct repeats were found. Therefore from pB3.8 alone it was not possible to identify the exact end points of ISN. Further clones were required to confirm the presence of ISN as well as to determine the points at which ISN diverges from the genome of *M. avium.*

4. Cloning and Characterization of pB8.0 and pEH5.0

*Bam*HI-digested strain one had three bands when hybridized with pMB22/S12. In addition to the 3.8 Kb band from pB3.8, two other bands of 8 Kb and 9 Kb needed to be cloned to identify the divergence points of ISN. Strain one chromosomal DNA was digested with *Bam*HI and subjected to electrophoresis on a 0.8% TAE gel. The DNA fragments corresponding to 7-10 Kb were excised. The DNA was purified and ligated with *Bam*H1 digested and dephosphorylated pBluescript II vector. Several hundred white, ampicillin-resistant colonies were selected and screened with the ^{32}P-labelled pMB22/S12 probe by colony hybridization. As a result, the 8 Kb fragment was isolated, named pB8.0. However cloning of the 9 Kb fragment failed. To make this 9 Kb *Bam*HI fragment easier to clone, strain one genomic DNA was digested with *Bam*HI, *Cla*I, *Eco*RI *Hind*III, *Kpn*I, *Pst*I, *Sac*I and *Xho*I restriction enzymes individually or in combination, and then subjected to Southern blot and hybridization. Finally, from the 9 Kb *Bam*HI fragment, a 5 Kb *Eco*RI-*Hind*III fragment was defined and cloned into pBluescript II vector as pEH5.0.

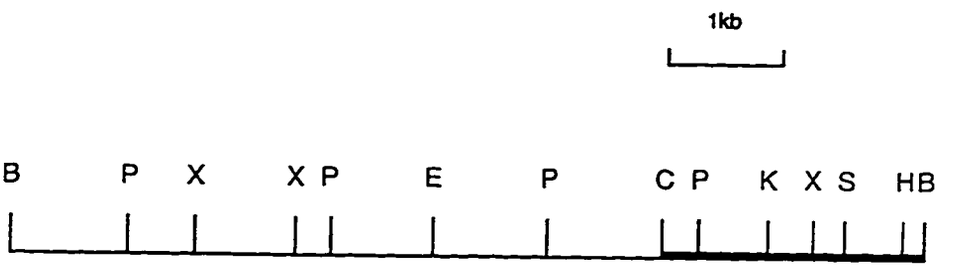
Restriction mapping of pB8.0 and pEH5.0 was carried out by

digesting these two clones with the restriction enzymes which corresponded to the enzymes used for mapping pB3.8 (Fig. 10). Comparing the restriction site profiles of pB3.8, pB8.0 and pEH5.0 indicated that these three clones contained the same restriction mapped fragment of about 1.3 Kb. Further subcloning and sequencing of pB8.0 and pEH5.0 was subsequently performed. The results from comparison of the three clones' sequence data indicated that these three clones contain the same fragment, ISN. The divergence points of ISN for the three clones were determined (Fig. 11).

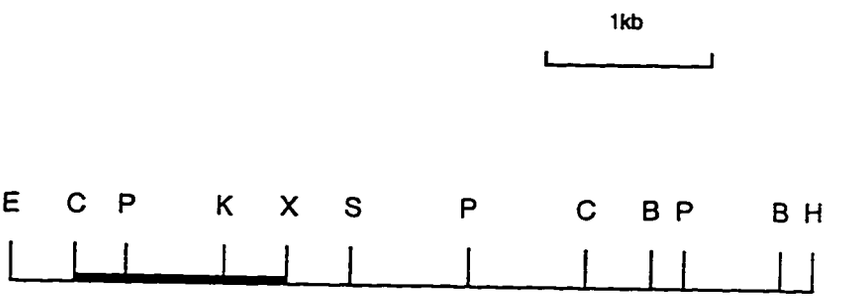
5. Determination of the Ends of ISN and Its Possible Insertion Site in *M. avium*.

To further confirm the termini of ISN, two more *M. avium* strains, *M. avium* A and *M. avium* B, which did not contain IS900 related elements were examined for the equivalent insertion locus. From sequences flanking ISN in PB3.8 and pEH5.0, two pairs of primers, DKU119/DKU120 and DKU117/DKU118, were selected respectively (Table 2). From *M. avium* A and *M. avium* B genomic DNA, the potential insertion sites were amplified with DKU117/DKU118 primers by PCR. These 150 bp PCR products were directly sequenced. Three insertion sites from pB3.8, pB8.0 and pEH5.0, and two potential insertion sites from *M. avium* A and *M. avium* B indicated that there is a consensus sequence CATGC within 10 bases flanking the 5′ terminus and a consensus sequence $TCCTN_{(2)}G$ flanking the 3′ terminus of ISN (Fig. 11). This suggests that $CATGCN_{(4-5)}TCCTN_{(2)}G$ represents an insertion site for ISN. In these three clones examined,

Figure 10. Restriction maps of pB8.0 and pEH5.0. B, *Bam*HI; C, *Cla*I; *Eco*RI H; *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sac*I; X, *Xho*I. The thick solid line indicates the segment with the same restriction enzyme site profile as that in pB3.8.

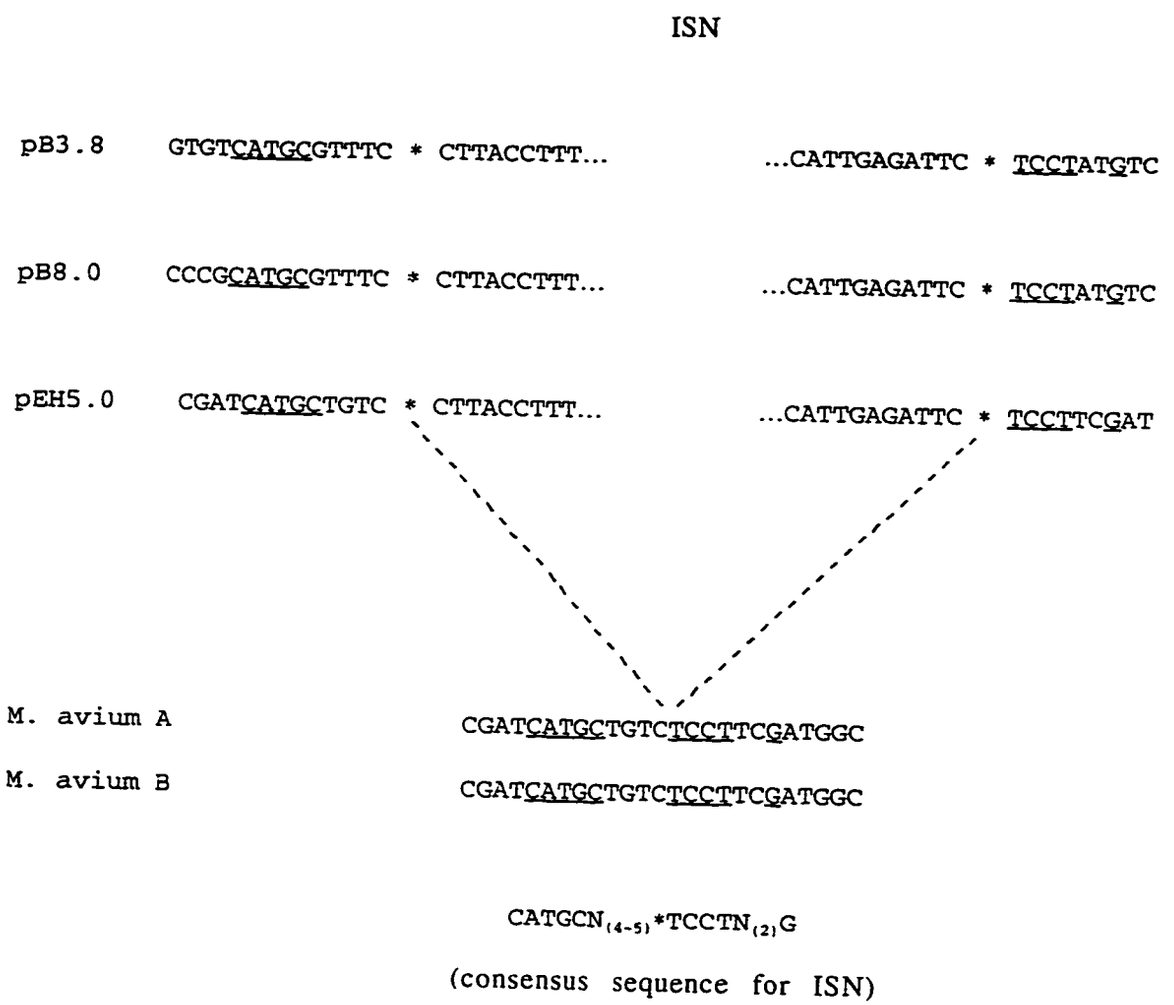


(pB8.0)



(pEH5.0)

Figure 11. The sequences of three insertion sites for ISN in *M. avium*, as determined from clones pB3.8, pB8.0 and pEH5.0, are shown above. Below are the sequences of a potential insertion site in two ISN free *M. avium* strains. Conserved nucleotides are underlined. (*) denotes the element-insertion site junction.



ISN has the same orientation with respect to this consensus sequence. This consensus insertion site is similar to the putative insertion site sequences described for IS900-related elements (Fig. 12). This suggests that similar short sequences might be involved in the transposition of these elements.

6. ISN Sequence Analysis

The nucleotide sequence of ISN is shown in Figure 13. It is 1418 bp and the G + C content is 65%. Analysis with the program STRIDER 1.2 of the sequences adjacent to the 5′ and 3′ termini revealed that, in common with IS900 and IS901, ISN has neither terminal inverted repeats nor flanking direct repeats.

DNA sequence analysis also revealed that there are two major and two short ORFs in ISN (Fig. 14). On the top strand in Fig. 14 there is one major ORF starting at position 172 and ending with a TGA codon at position 1398. In this same reading frame several possible GTG and ATG start codons were found. Among these various possibilities, initiation is only likely to take place with a GTG codon at nucleotide 220 on the basis that it is adjacent to a recognizable S-D sequence (GGAGG) from nucleotides 211 to 215. This is similar to IS900 and other related IS elements. This ORF of 1179 nucleotides (ORF1179) encodes a protein of 393 amino acids with an expected Mr of 44 kDa.

It is clear that most of the genes in bacteria have similar sequences at position -10 and -35 (Dale and Patki, 1990). It is noted that within the first 180 nucleotides of ISN and upstream of ORF1179 there is a -35-like region with the sequence TTGAGA and a -10-like

Figure 12. Comparison of the putative insertion sequence of ISN, IS900, IS901, IS902 and IS116. Conserved nucleotides are underlined for each IS element. *, insertion position of each of the respective IS elements. (1), this study; (2), (Green *et al.*, 1989); (3), (Moss *et al.*, 1992); (4), (Kunze *et al.*, 1991); (5), (Leskiw *et al.*, 1990).

elements/clones	putative insertion site sequences
ISN/pB3.8[1]	GTGTCATGCGTTTC*TCCTATGTC
ISN/pB8.0[1]	CCCGCATGCGTTTC*TCCTATGTC
ISN/pEH5.0[1]	CGATCATGCTGTC*TCCTTCGAT
IS900/pMB22[2]	TGGTCATGTGGTGT*CTCCTTCGCG
IS900/pMB15[2]	AACGACATGTGTT*CCCCTTACGC
IS900/pMB55[2]	ATGGTCATGGTGG*CCCCTTGGCA
IS900/pMBJ2[3]	CATGACGA*CTCCTTG
IS900/PMBJ3[3]	CATGTGG*CTCCTTC
IS901/PUS410[4]	CATGCGCTGA*TTCCTTTCAG
IS901/PUS411[4]	CATTGGTGGC*TTCCGTTCGG
IS902/PZM22[3]	CATGATCAATT*CCTTTC
IS902/PZM25[3]	CATTTACAGT*CCTTTC
IS116/PIJ702[5]	CATGGTCGG*TCTCCTGGT

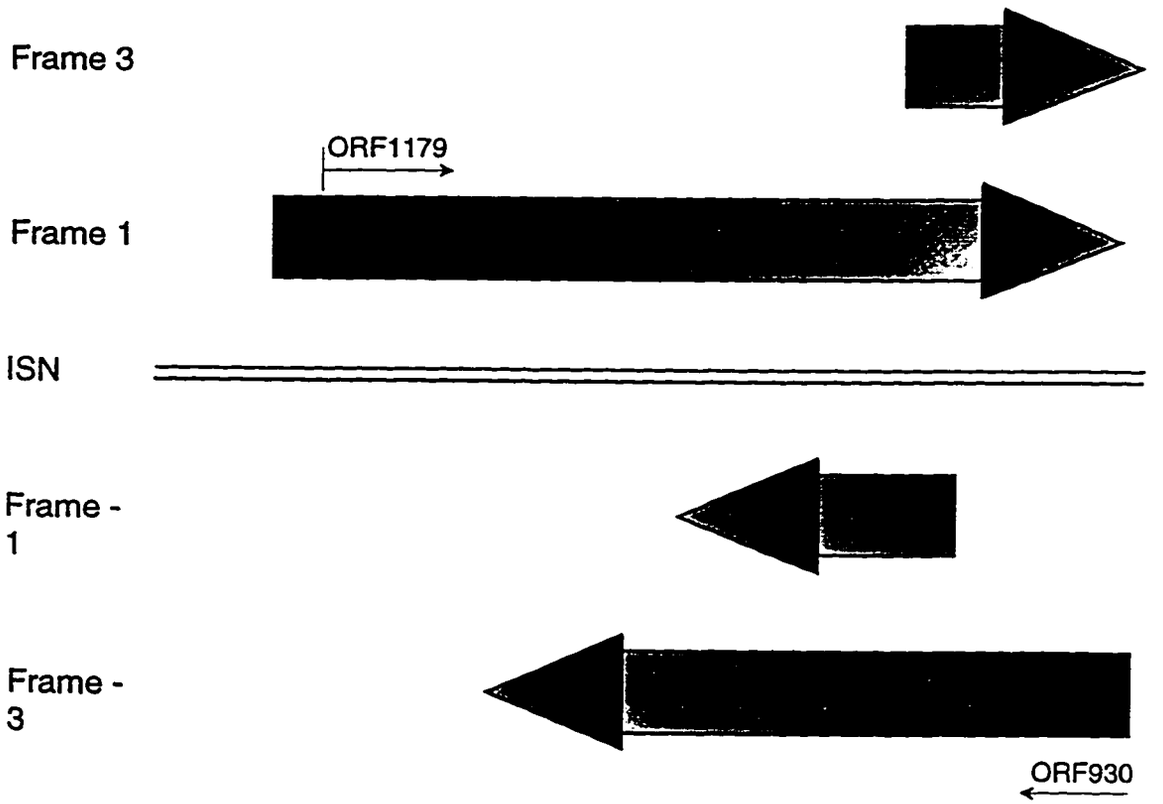
Figure 13. The nucleotide sequence of ISN. For ORF1179, the possible S-D and the amino acid sequences are shown. The -35 and -10 regions are also indicated. On the complementary strand, a potential S-D sequence (sequence in bracket is ISN immediate flanking sequence), the initial codon AUG and the stop codon UAG for ORF2 are underlined. The amino acid sequence for ORF2 is not shown.

```
TACCTTTCTTGCAGGGTGGTTGTTGCCCTCGGCCGTACGTTCGAACTGCCAGGACGTCGG
ATGGAAAGAACGTCCCACCAACAACGGGAGCCGGCATGCAAGCTTGACGGTCCTGCAGCC
                    20                  40                  60

TATGCTTCATGCGTTGCGGTGTTCATGCGAGGAGATTGGCCGCCCGACGTCCGCGACGAC
ATACGAAGTACGCAACGCCACAAGTACGCTCCTCTAACCGGCGGGCTGCAGGCGCTGCTG
                    80                  100                 120

          -35                       -10
TCGACCGCTAATTGAGAGATGCGATTGATCGCTGTGTAAGGACACGCCGGCGTGGTCGTC
AGCTGGCGATTAACTCTCTACGCTAACTAGCGACACATTCCTGTGCGGCCGCACCAGCAG
                    140                 160                 180

                                    S-D     V  S  Q  Q  V  W  A
TGCTGGGTTGATAGGGATGCCAATGATCACGGAGGTGCTGTGAGCCAACAGGTCTGGGCC
ACGACCCAACTATCCCTACGGTTACTAGTGCCTCCACGACACTCGGTTGTCCAGACCCGG
                    200                 220                 240

G  V  D  A  G  S  P  T  T  I  A  W  L  S  M  R  R  V  S  D
GGTGTCGATGCCGGAAGTCCGACCACCATTGCGTGGCTATCGATGCGGAGGGTCAGCGAC
CCACAGCTACGGCCTTCAGGCTGGTGGTAACGCACCGATAGCTACGCCTCCCAGTCGCTG
                    260                 280                 300

C  C  Q  T  G  S  Q  R  R  N  R  T  A  G  A  D  R  R  S  S
TGCTGTCAGACGGGTAGCCAACGACGAAACCGCACTGCTGGAGCTGATCGACGCAGTAGC
ACGACAGTCTGCCCATCGGTTGCTGCTTTGGCGTGACGACCTCGACTAGCTGCGTCATCG
                    320                 340                 360

S  N  P  T  R  E  V  T  W  A  I  D  L  N  A  G  W  C  R  A
TCCAATCCGACGCGTGAAGTTACCTGGGCGATCGACCTGAACGCCGGTTGGTGCCGCGCT
AGGTTAGGCTGCGCACTTCAATGGACCCGCTAGCTGGACTTGCGGCCAACCACGGCGCGA
                    380                 400                 420

A  D  H  L  A  H  A  A  E  Q  R  A  L  H  P  R  P  Q  D  L
GCTGATCACCTTGCTCATGCCGCCGAGCAGCGTGCTCTACATCCCCGGCCGCAGGATCTA
CGACTAGTGGAACGAGTACGGCGGCTCGTCGCACGAGATGTAGGGGCCGGCGTCCTAGAT
                    440                 460                 480

P  R  S  A  G  Y  R  G  D  G  K  S  D  A  K  D  A  A  V  I
CCACGCTCGGCCGGCTACCGTGGCGACGGCAAAAGCGACGCCAAGGACGCCGCCGTCATC
GGTGCGAGCCGGCCGATGGCACCGCTGCCGTTTTCGCTGCGGTTCCTGCGGCGGCAGTAG
                    500                 520                 540

A  D  Q  A  R  M  R  R  D  L  Q  P  L  R  P  G  D  D  I  A
GCCGACCAAGCTCGGATGCGCCGCGACCTGCAGCCGCTGCGACCCGGTGATGACATCGCG
CGGCTGGTTCGAGCCTACGCGGCGCTGGACGTCGGCGACGCTGGGCCACTACTGTAGCGC
                    560                 580                 600

V  E  L  R  I  L  T  A  G  A  P  T  W  S  S  D  R  T  R  V
GTCGAGCTGCGCATCCTGACCGCTGGCGCACCGACCTGGTCGTCCGATCGCACCCGGGTG
CAGCTCGACGCGTAGGACTGGCGACCGCGTGGCTGGACCAGCAGGCTAGCGTGGGCCCAC
                    620                 640                 660
```

```
I  N  R  L  R  A  Q  L  L  E  Y  F  P  A  L  E  R  G  F  D
ATCAACCGGCTGCGTGCCCAGCTGCTTGAGTACTTCCCCGCCCTGGAACGCGGCTTCGAT
TAGTTGGCCGACGCACGGGTCGACGAACTCATGAAGGGGCGGGACCTTGCGCCGAAGCTA
                  680                 700                 720

Y  S  A  S  K  A  A  L  I  L  L  T  G  Y  Q  T  P  D  G  C
TACAGCGCCAGCAAGGCCGCGCTGATCCTGTTGACCGGCTACCAAACCCCCGACGGCTGC
ATGTCGCGGTCGTTCCGGCGCGACTAGGACAACTGGCCGATGGTTTGGGGGCTGCCGACG
                  740                 760                 780

A  R  R  A  A  R  L  T  A  W  L  H  K  R  K  A  R  N  A  S
GCGCGCCGCGCTGCCCGCCTGACAGCTTGGCTGCATAAACGCAAGGCCCGCAACGCCTCT
CGCGCGGCGCGACGGGCGGACTGTCGAACCGACGTATTTGCGTTCCGGGCGTTGCGGAGA
                  800                 820                 840

A  V  A  A  T  A  I  E  A  R  Q  R  Q  H  T  T  V  P  G  Q
GCTGTCGCGGCCACGGCTATCGAGGCCCGCCAGCGCCAACACACCACCGTGCCCGGCCAA
CGACAGCGCCGGTGCCGATAGCTCCGGGCGGTCGCGGTTGTGTGGTGGCACGGGCCGGTT
                  860                 880                 900

N  L  A  A  A  M  V  A  R  L  A  K  E  V  M  T  L  D  T  E
AACCTCGCCGCCGCCATGGTCGCCCGCCTGGCTAAGGAGGTGATGACCCTCGACACCGAA
TTGGAGCGGCGGCGGTACCAGCGGGCGGACCGATTCCTCCACTACTGGGAGCTGTGGCTT
                  920                 940                 960

I  A  E  T  D  T  M  I  E  D  R  F  R  R  H  R  H  A  E  I
ATCGCAGAAACCGACACGATGATCGAGGACCGATTTCGCCGCCACCGCCACGCCGAGATC
TAGCGTCTTTGGCTGTGCTACTAGCTCCTGGCTAAAGCGGCGGTGGCGGTGCGGCTCTAG
                  980                 1000                1020

L  L  S  M  P  G  F  G  V  I  L  G  A  E  F  L  A  A  T  A
CTCCTGAGCATGCCCGGCTTCGGCGTCATACTCGGCGCCGAGTTCCTCGCCGCCACCGCA
GAGGACTCGTACGGGCCGAAGCCGCAGTATGAGCCGCGGCTCAAGGAGCGGCGGTGGCGT
                  1040                1060                1080

H  E  R  F  R  L  R  R  S  S  C  R  R  V  G  L  A  P  V  P
CATGAGCGCTTTCGACTCCGTCGATCGTCTTGCCGGCGTGTCGGGCTGGCCCCGGTACCG
GTACTCGCGAAAGCTGAGGCAGCTAGCAGAACGGCCGCACAGCCCGACCGGGGCCATGGC
                  1100                1120                1140

R  D  S  G  R  I  S  G  N  L  K  R  P  R  R  Y  D  R  R  L
CGGGATTCCGGGCGCATCAGCGGCAATCTCAAACGCCCCCGCCGCTACGACCGACGACTG
GCCCTAAGGCCCGCGTAGTCGCCGTTAGAGTTTGCGGGGGCGGCGATGCTGGCTGCTGAC
                  1160                1180                1200

L  R  A  C  Y  L  S  A  Q  I  A  I  R  T  D  P  A  S  R  T
CTGCGAGCCTGTTACCTGTCCGCCCAAATCGCCATCCGTACCGACCCCGCATCCCGGACC
GACGCTCGGACAATGGACAGGCGGGTTTAGCGGTAGGCATGGCTGGGGCGTAGGGCCTGG
                  1220                1240                1260

Y  Y  N  R  K  R  A  E  G  K  T  H  T  Q  S  I  L  A  L  A
TACTACAACCGCAAAAGAGCCGAAGGCAAAACCCACACCCAATCCATCCTCGCCCTGGCC
ATGATGTTGGCGTTTTCTCGGCTTCCGTTTTGGGTGTGGGTTAGGTAGGAGCGGGACCGG
                  1280                1300                1320
```

```
R  R  R  L  N  V  L  W  A  M  L  R  D  H  A  V  Y  Q  P  A
CGCCGGCGCCTCAACGTCCTGTGGGCCATGCTGCGCGACCACGCTGTCTACCAACCCGCA
GCGGCCGCGGAGTTGCAGGACACCCGGTACGACGCGCTGGTGCGACAGATGGTTGGGCGT
                  1340                1360                1380

T  T  T  A  A  A
ACCACTACTGCGGCGGCTTGACAACGTCATTGAGATTC(TCCT)
TGGTGATGACGCCGCCGAACTGTTGCAGTAACTCTAAG(AGGA)
                  1400      ← ORF2          S-D
```

Figure 14. Open reading frames in ISN. The ORF1179 and ORF930 are indicated.



region with the sequence TGTAAG. These two sequences are in good agreement with the *E. coli* promoter consensus sequences of TTGACA and TATAAT respectively (Dale and Patki, 1990). Conceivably, this region could promote the transcription of ORF1179, although the significance of this region has not yet been determined at present.

The sequence homologies between ISN and IS900, IS901, IS1110, IS110 and IS116 were determined by sequence alignment with Wisconsin GCG software (Table 3). The DNA sequence comparisons show that ISN has 82% homology with IS900, and has variable homology of 53%-62% with other related IS elements. No significant homologies were found with IS elements from *E. coli* or other mycobacterial species. Comparison of the amino acid sequences of the major ORFs from IS900, IS901, IS1110, IS110 and IS116 shows that ORF1179 has 63% of the residues identical with that of IS900, and 30-42% identity with the others.

CLUSTAL alignment of the major ORFs of these elements showed that the 3′ carboxy terminus of the predicted proteins have higher similarities than the 5′ ends (Fig. 15). Of note, all IS elements contain the conserved motif K-DAKDA, which has been found in reverse transcriptase enzymes (Xiong and Eickbush, 1988).

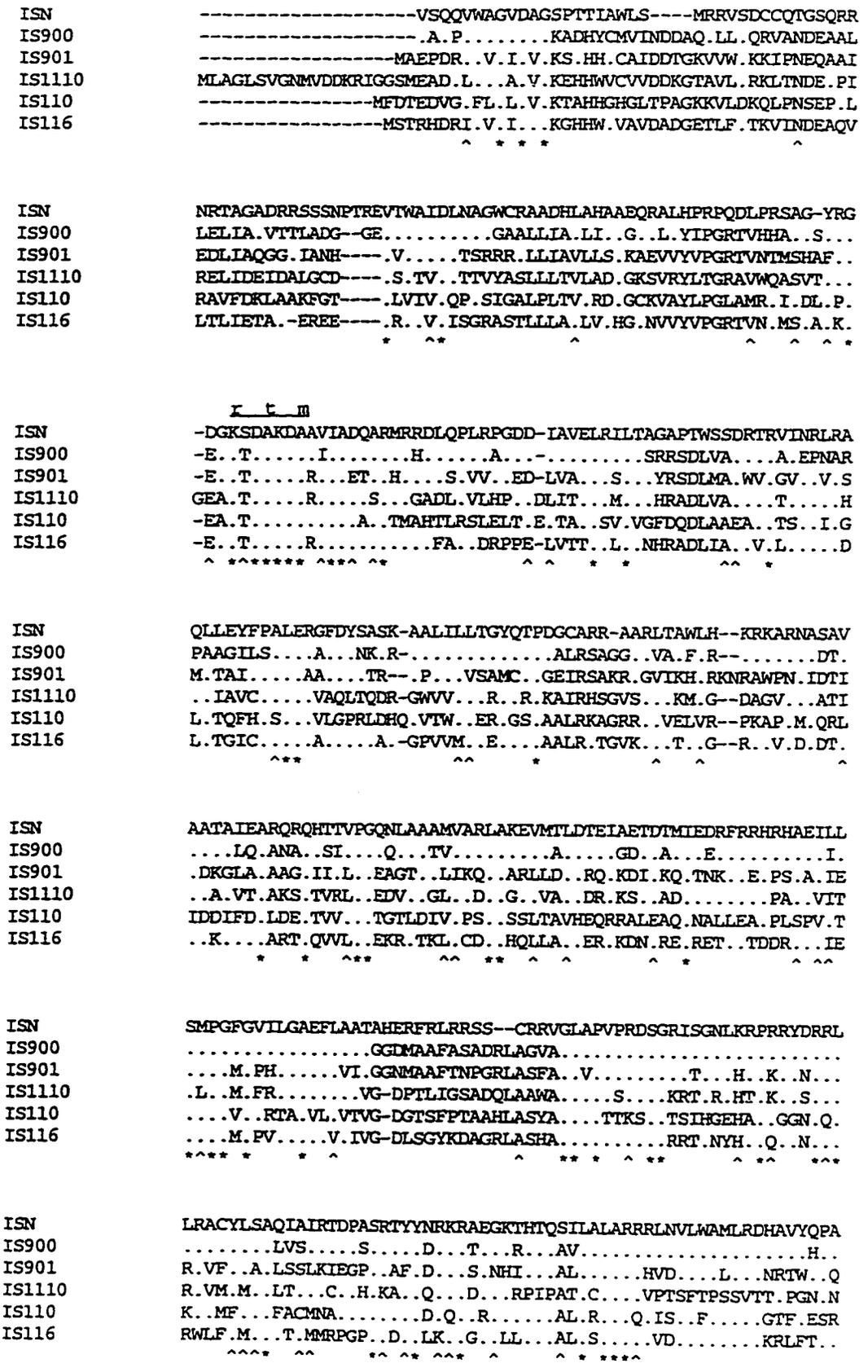
Similar to IS900 and IS902, ISN has another long open reading frame, ORF930, on the complementary strand of the ORF1179. ORF930 starts at the very 3′ end of ISN at nucleotide 1410 and goes to 481. Its ATG start codon is preceded by a S-D sequence AGGAGA which is partially located in the flanking region (Fig. 11; Fig. 13). This ORF930 encodes a predicted protein of 310 amino acids.

Comparison of the amino acid sequence of ORF2 shows that ORF930 of ISN has 53% of residues identical to ORF2 of IS900 and

Table 3. DNA sequence homologies and amino acid sequence homologies between ISN and IS900, IS901, IS1110, IS110 and IS116.

	Homologies with ISN (%)	
	Nucleotide sequence	Amino acid sequence
IS900	82	63
IS901	59	40
IS1110	61	40
IS110	53	30
IS116	62	42

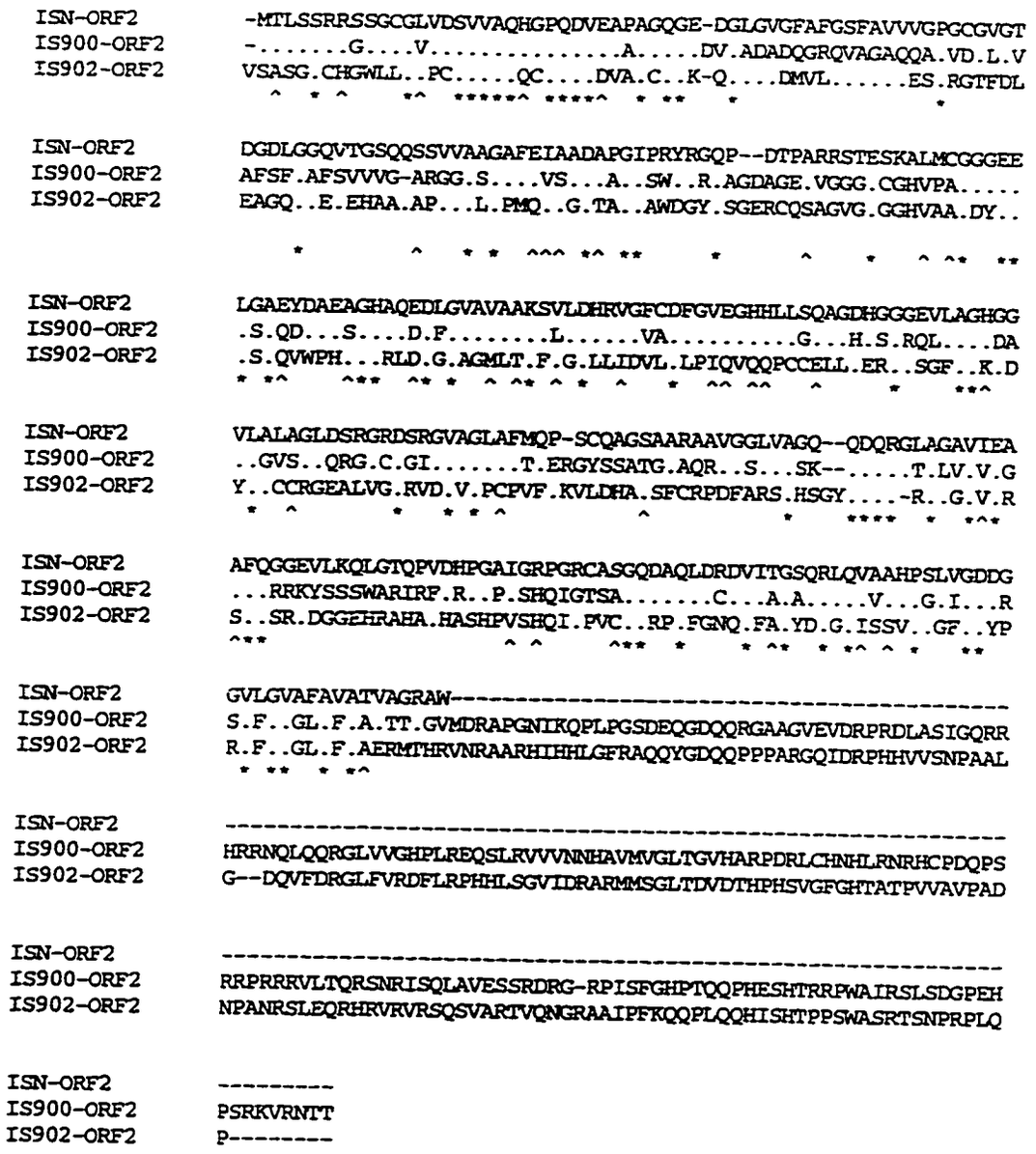
Figure 15. CLUSTAL alignment of the putative transposases of ISN, IS900, IS901, IS1110, IS110 and IS116. For each sequence, (.) denotes residues identical with ISN and (-) denotes gaps and deletions comparing with ISN. (*) indicates the residues present in all six sequences are identical and (^) indicates residues conserved in these aligned sequences. Sequence homology with a reverse transcriptase motif is indicated (rtm).


ISN
IS900
IS901
IS1110
IS110
IS116
r t m

```
ISN             TTTAAA----------------------------------------------
IS900           ......----------------------------------------------
IS901           QPTV..A---------------------------------------------
IS1110          HPRSPSRRLDIFIESPFAIYALVSGVGGPAVGAIVLAVVVDDDDDKKAFGGR
IS110           MPAGVELAA-------------------------------------------
IS116           PFVTQTA---------------------------------------------
```

31% identity with ORF2 of IS902. However, computer searches of ORF930 in Genbank failed to reveal any significant similarity to other known sequences. CLUSTAL alignment of the ORF2s from ISN, IS900 and IS902 exhibits higher homology at their 5′ ends (Fig. 16). Compared to the other two ORF2s, ORF930 is approximately 170-180 amino acids shorter.

Figure 16. CLUSTAL alignment of the amino acid sequence of ORF2. (.) denotes residues identical with ISN and (-) represents the gaps and deletions in the sequences during alignment. (*) indicates residues presented in all three sequences are identical and (^) indicates residues are conserved in these three sequences.


ISN-ORF2 -MTLSSRRSSGCGLVDSVVAQHGPQDVEAPAGQGE-DGLGVGFAFGSFAVVVGPGCGVGT
IS900-ORF2 -.....G....V.....A.....DV.ADADQGRQVAGAQQA.VD.L.V
IS902-ORF2 VSASG.CHGWLL..PC.....QC....DVA.C..K-Q....DMVL.....ES.RGTFDL
ISN-ORF2 DGDLGGQVTGSQQSSVVAAGAFEIAADAPGIPRYRGQP--DTPARRSTESKALMCGGGEE
IS900-ORF2 AFSF.AFSVVVG-ARGG.S....VS...A..SW..R.AGDAGE.VGGG.CGHVPA.....
IS902-ORF2 EAGQ..E.EHAA.AP...L.PMQ..G.TA..AWDGY.SGERCQSAGVG.GGHVAA.DY..
ISN-ORF2 LGAEYDAEAGHAQEDLGVAVAAKSVLDHRVGFCDFGVEGHHLLSQAGDHGGGEVLAGHGG
IS900-ORF2 .S.QD...S....D.F.....L.....VA.....G...H.S.RQL....DA
IS902-ORF2 .S.QVWPH...RLD.G.AGMLT.F.G.LLIDVL.LPIQVQQPCCELL.ER..SGF..K.D
ISN-ORF2 VLALAGLDSRGRDSRGVAGLAFMQP-SCQAGSAARAAVGGLVAGQ--QDQRGLAGAVIEA
IS900-ORF2 ..GVS..QRG.C.GI.....T.ERGYSSATG.AQR..S...SK--.....T.LV.V.G
IS902-ORF2 Y..CCRGEALVG.RVD.V.PCPVF.KVLDHA.SFCRPDFARS.HSGY....-R..G.V.R
ISN-ORF2 AFQGGEVLKQLGTQPVDHPGAIGRPGRCASGQDAQLDRDVITGSQRLQVAAHPSLVGDDG
IS900-ORF2 ...RRKYSSSWARIRF.R..P.SHQIGTSA.....C...A.A.....V...G.I...R
IS902-ORF2 S..SR.DGGEHRAHA.HASHPVSHQI.PVC..RP.FGNQ.FA.YD.G.ISSV..GF..YP
ISN-ORF2 GVLGVAFAVATVAGRAW----------
IS900-ORF2 S.F..GL.F.A.TT.GVMDRAPGNIKQPLPGSDEQGDQQRGAAGVEVDRPRDLASIGQRR
IS902-ORF2 R.F..GL.F.AERMTHRVNRAARHIHHLGFRAQQYGDQQPPPARGQIDRPHHVVSNPAAL
ISN-ORF2 ----------
IS900-ORF2 HRRNQLQQRGLVVGHPLREQSLRVVVNNHAVMVGLTGVHARPDRLCHNHLRNRHCPDQPS
IS902-ORF2 G--DQVFDRGLFVRDFLRPHHLSGVIDRARMMSGLTDVDTHPHSVGFGHTATPVVAVPAD
ISN-ORF2 ----------
IS900-ORF2 RRPRRRVLTQRSNRISQLAVESSRDRG-RPISFGHPTQQPHESHTRRPWAIRSLSDGPEH
IS902-ORF2 NPANRSLEQRHRVRVRSQSVARTVQNGRAAIPFKQQPLQQHISHTPPSWASRTSNPRPLQ
ISN-ORF2 ---------
IS900-ORF2 PSRKVRNTT
IS902-ORF2 P--------

IV. Discussion

A DNA element that hybridized to pMB22/S12 of IS900 was found in a *M. avium* isolate from a non-AIDS patient. As pMB22/S12 was reported to be specific for *M. paratuberculosis*, it was suspected that this DNA element represented a new IS element. Southern blot analysis showed that there are three copies of this element dispersed throughout the genome. This element was proposed to be a novel insertion sequence of the IS900 family in *M. avium* and is designated ISN in this paper.

To identify ISN from this *M. avium* isolate, the region suspected of containing ISN and its flanking regions in pB3.8 were sequenced. Sequence analysis indicated that ISN is IS900-like element and lacks significant inverted repeats and direct repeats. To locate the divergence points of ISN, two more clones containing ISN (pB8.0 and pEH5.0) were isolated and their terminal sequences were determined. The exact ends of ISN and its insertion site in *M. avium* were identified by examining the sequences at ISN's divergence points in three clones and the sequences of potential insertion site from ISN-free *M. avium* strains. Computer analysis revealed that the sequence of ISN has a high degree of homology with that of the well studied IS900. The structural similarities and the sequence homology between ISN and IS900 provide strong supportive evidence for ISN being an insertion-like element.

ISN is 1418 bp in length. The overall G + C content is 65% which is similar to that found in the host genomic DNA (about 62-70%), and differs markedly from that of *E. coli* IS elements (about 45%). This G

+ C content feature of ISN is similar to that of IS element IS900, IS901, and IS110 which also have G + C contents comparable to that of their hosts *M. paratuberculosis*, *M. avium* and *Streptomyces* respectively (Brutonand Chater, 1987; Green *et al.*, 1989; Kunze *et al.*, 1991). This suggests that the IS elements of all these organisms may be related.

The sequence analysis shows that in contrast to most other insertional elements, ISN does not possess either terminal repeats or flanking direct repeats. This absence is characteristic of most previously characterized IS900-related elements. The sequence analysis of the flanking regions of clones pB3.8, pB8.0 and pEH5.0, and the potential insertion sites in two *M. avium* isolates indicates an insertion site specificity. The consensus insertion site for ISN is 5′ $CATGCN_{4-5}$ * $TCCTN_2G$ 3′ (the asterisk denotes site of insertion) and in three clones the element inserts in the same orientation with respect to this target sequence, which is similar to the insertion site sequences described for IS900-related elements.

In most IS elements, the inverted repeats are believed to provide the recognition and binding sites for the elements' transposase protein (Berg *et al.*, 1988; Galas and Chandler, 1989; Grindley, 1985). Several studies of different ISs have demonstrated that the ends are essential components of the transposition apparatus. The Tn3 transposase has been shown to recognize and bind to the ends of Tn3 (New *et al.*, 1988), and Grindley and Wiater found that the Tn1000 transposase, which is similar to that of Tn3, can recognize and bind to the 39 bp inverted repeats at the ends of Tn1000 (Grindley and Wiater, 1988). In IS50 (Berg *et al.*, 1988) and IS903 (Derbyshire and Grindley, 1992), the inverted repeat ends are composed of two

functional domains which are critical for transposase binding. However, the reason why the recognition and binding sites for the transposase in many IS elements exist as inverted repeats remains unclear. It has also been observed that Tn21 and Tn1721 can continue to transpose, even when one end is completely deleted, although at a substantially reduced frequency (Grindley, 1985).

Most IS elements generate direct repeats at their insertion sites during their transposition. The evidence indicates that the direct repeat is not part of the preexisting sequence of the transposed DNA but, rather, is a consequence of the mechanism of insertion. The direct repeat is produced from the staggered cut of the targeted sequence followed by repair upon transposition. These features exist in most transposons and IS elements investigated so far. One exception is IS492 which lacks terminal inverted repeats but does have a 5 bp direct repeat (Bartlett and Silverman, 1989). Tn554 also lacks both inverted and direct terminal repeats, and yet it is extremely site-specific (Murphy, 1988). Similarly, several IS900-related elements from *M. paratuberculosis*, *M. avium* and *Streptomyces* do not have either inverted repeats and/or direct repeats, but they do have a similar consensus insertion site. These IS elements suggest that the inverted repeat is not essential for all transposase binding and that target staggered-cut is not the only model for transposition.

Analysis of ISN in three clones reveals that they have the same orientation with respect to their consensus insertion site. This specific insertion suggests that the flanking sequence plays an important role during the transposition. Examination of 13 insertion site sequences from ISN and other IS900-related elements (Fig.11)

revealed that they all have CAT at the left insertion site and CCT at the right insertion site. These two 3-nucleotide elements were separated by 6-9 nucleotides in consensus insertion sequences. It was also noted that the same 3-nucleotide elements also appear in their IS sequences with the internal CAT at the right side, 7-9 nucleotides away from the CCT in the right flanking region, and the internal CCT at the left side, 5-11 nucleotides away from the CAT in the left flanking region. These short sequences might be important for transposase recognition. The presence of nearby regions of limited homology with the ends of ISs has often been suggested to play a role in site selection (Galas and Chandler, 1989). It is assumed that a transposase can make double-stranded cuts of the transposable element at both its ends. The excised duplex segment is integrated into the target by ligating the extended single strands of the target DNA to the appropriate strands at the transposable element ends (Grindley, 1985). Morisato & Kleckner found that Tn10 transposase itself can make double-stranded cuts (Grindley, 1985). Their study supports the cut-and-paste process. Therefore, it is conceivable that, like class I restriction endonucleases, transposase in ISN and IS900-related IS elements may recognize and cut the CAT-CCT asymmetric sequences at both ends of transposable elements in the donor and the consensus insertion site in the recipient, and then ligate them in the same order. If these cleavages are made by blunt-end cutting followed by the cut and-paste process, then simple-insertion occurs. This transposition will not generate target duplication and the IS will always be in the same orientation with their consensus site.

England et al. (England *et al.*, 1991), by analyzing integration of

IS900 into several genomic sites, demonstrated that transposition in IS900 involves simple insertion as well as a replicative mechanism. It was proposed that many transposons and IS elements can use both replicative and nonreplicative mechanisms to different extents. To date, the mechanisms for transposition of IS900 family elements are not completely understood. Nevertheless, the insertion specificity of these IS elements may reflect base sequence recognition by their transposition proteins. Several studies suggest that the topology of target DNA is also important for some transposable elements (Berg *et al.*, 1988; Galas and Chandler, 1989). Therefore this recognition specificity may depend not only on the specific nucleotides but also the local DNA conformation.

The homology between ISN and IS900, IS901, IS110, IS116 and IS1110 at the nucleotide level was determined by sequence alignment. The alignment shows ISN has the highest homology with IS900 (82%), followed by IS116 (62%), IS1110 (61%) IS901 (59%), and IS110 (53%) The greatest homology exists at their 3′ ends. No significant homology was observed between ISN and IS elements from *E. coli* or from other mycobacterial species.

The alignment of the amino acid sequence of the major ORFs in ISN and in each of the IS900 family elements also shows a high degree of homology between ISN and IS900 (63%), followed by IS116 (42%), IS901 (40%), IS1110 (40%) and IS110 (30%). These high homologies indicate that ISN is closely related to IS900. The similarities between the predicted protein of ORF1179 in ISN and those specified by IS900-related elements suggests that these proteins may have comparable functions. The amino acid sequence of ORF1179 also has the conserved motif K-DAKDA which was found in

other ORFs of IS900 related elements (Kunze *et al.*, 1991; Leskiw *et al.*, 1990) and this motif has similarity to the reverse transcriptase motif (Xiong and Eickbush, 1988), although the role of this motif in transposition is not clear at present. Similar to many other IS elements, functional studies of the products of the ISN ORFs have not yet been performed.

ORF1179 might encode a protein which is responsible for transposition of ISN. Since ISN contains -10 and -35-like regions upstream of the ORF1179 and these regions are acceptably close to the presumed translational S-D sequence preceding the GTG initiation codon, it would be expected that ORF1179 can be expressed under the control of the promoter within the ISN. This structural feature is also noticed in IS900. It has been shown that ORF1197 of IS900 has been expressed in *E. coli* under the control of an exogenous promoter. The protein product, p43, has a molecular mass of 44.5 kDa corresponding to the predicted molecular mass of ORF1197 (Tizard *et al.*, 1992). Whether a promoter within IS900 drives the expression of p43 in *M. paratuberculosis* has not been determined. Further studies of the protein p43 could facilitate investigation of the mechanism of transposition of IS900 and IS900-related elements.

Most IS elements only possess a single, large ORF which encodes for a transposase protein. Some IS elements, however, have additional ORF(s) on both strands (Galas and Chandler, 1989). Further analysis of the IS900 and IS902 sequences revealed the presence of a second ORF, both designated ORF2 which run on the complementary strand of the putative transposase genes (Doran *et al.*, 1994; Murray *et al.*, 1992). In the case of IS900, the S-D sequence and UGA termination codon for ORF2 are encoded by the sequence that flanks

IS900. In IS902, the sequence required for translation of ORF2 is contained entirely within IS902.

As in IS900 and IS902, ISN also contains another ORF, ORF930, on the complementary strand to the putative transposase gene. This ORF is predicted to encode a protein of 310 amino acids. Its translational signals are formed from ISN and its flanking region (Fig. 13), but it does not seem to have any upstream transcriptional signals.

It appears that the expression of ORF2 on these elements requires sequences adjacent to and outside the element for transcription or both transcription and translation signals. This phenomenon was also observed in another related element IS116, which was isolated from *Streptomyces clavuligerus* (Leskiw *et al.*, 1990). IS116 has a putative sequence which is similar to that of IS900. Our analysis of IS116 indicates that there is also an ORF2 on the complementary strand of the transposase gene. This ORF2 is 1233 bp with an AUG start codon at nucleotide 1415 and potentially encodes a protein of 411 amino acids. Seven nucleotides away from the initiation codon, a 5′ ACCAGGAGA 3′ sequence in the flanking region provides a potential S-D sequence. It is not clear if there are transcriptional signals in the adjacent sequence, although the sequences required for translation of this ORF2 are contained in the IS116 element and its flanking sequence.

Doran (Doran *et al.*, 1994) studied ORF2 sequences in IS900 and IS902 and found that both were possibly capable of encoding a protein. Both ORF2 potential coding regions are very large, over 900 bp in size. An analysis of codon usage in the ORF2 amino acid sequences reveals a bias that is consistent with other mycobacterial genes (Dale and Patki, 1990). Furthermore, Murray et al. have

mapped a promoter sequence, P_{AN}, adjacent to ORF2 on an IS900 element in *M. paratuberculosis* (Murray *et al.*, 1992). A DNA fragment containing P_{AN} and part of ORF2, was fused to the *lacZ* gene and subsequently expressed in *M. bovis* BCG. This indicates that in *M. paratuberculosis* at least one ORF2 sequence of IS900 possesses an upstream promoter capable of driving expression of a polypeptide encoded by this ORF. Recently, in *M. paratuberculosis*, the expression of this ORF2 has been detected at the level of transcription by reverse transcription-PCR, and the translation products have also been detected by two specific antibodies in western blots of protein extracts from *M. paratuberculosis* (Doran *et al.*, 1997).

Although it is unusual that the signals required for gene expression of an IS element might be obtained from an adjacent sequence, it is not only found in IS900-family elements, but also in other IS elements, including IS elements in *Agrobacterium vitis*, *Rhizobium fredii*, *Bradyrhizobium japonium*, *Pseudomonas* sp. and *Shigella sonnei*. These IS elements all create a TAG stop codon for their transposases by inserting into a 5′ CTAG 3′ target sequence (Fournier *et al.*, 1993; Kleckner, 1981). As further analysis of the putative insertion sites for IS900-related elements showed (Fig. 12) the complementary strands of those insertion site sequences do indeed compose a S-D and AUG initiation codon-like structure. Combining the results from the analysis of insertion site sequences and the studies of IS900 ORF2, it is reasonable to assume that these IS elements are "in search" of the translation and transcription signals needed to obtain an active element or to express genes encoded by IS themselves. This may also explain their unusual insertion site specificity and the orientation of the element with

respect to its target site.

Previous studies of other IS elements also suggest that some IS elements have outwardly-directed promoter or partial promoter sequences which are present within the ends of ISs (Galas and Chandler, 1989). Upon their transposition, these IS elements are able to initiate transcription directed outward into flanking DNA and activate downstream genes. In the case of IS900-related elements, the inward transcription of IS genes could be driven by an adjacent promoter in the flanking region of insertion. This may not only activate an IS gene, but also, simultaneously interrupt the host gene function either by occupying the promoter of the host gene or by the internal transcription terminator which is carried by the IS element.

The significance of the ORF2s remains unclear. It is proposed that ORF2 in IS900-related elements is not essential for transposition and it may only play a regulatory role as a repressor in transposition, since IS1110 does not have such an ORF2 (Hernandez Perez *et al.*, 1994). Doran et al. noticed that in ORF2 on IS900 and IS902, there are ten amino acid sequences, which are similar to a highly conserved domain in eight proteins, all involved in various transport mechanisms in both prokaryotic and eukaryotic organisms. It is proposed that the proteins encoded by ORF2 on IS900 and IS902 may perform a conserved function associated with the iron transport mechanism in *M. paratuberculosis* and *M. avium* subsp. *silvaticum* (Doran *et al.*, 1994).

Comparison of the amino acid sequence of ORF930 with those of IS900 and IS902 revealed an overall 50% identity with IS900 and 33% identity with IS902, although ORF930 is approximately 170-180 amino acids smaller and lacks the ten conserved amino acid motif

which was found in ORF2 of IS900 and IS902. This degree of sequence conservation suggests a degree of protein structural and functional conservation, which may result in certain similar properties of these elements or may be an indication of the evolution of these elements.

Unlike other related IS elements, ISN seems to be uncommon in *M. avium*. Of 66 MAC isolates screened, only one *M. avium* strain contained the ISN element and only 3 copies were present, whereas IS900 is found in multiple copies (12-15 copies) in almost all *M. paratuberculosis* strains with little polymorphism (McFadden *et al.*, 1990); IS901 has about 6-9 copies in *M. avium* strains with a high frequency of appearance in animal strains (89%), a low rate in non-AIDS patient strains and is lacking in AIDS patient and environmental isolates (Kunze *et al.*, 1992). IS1110 was recently found as another IS900 family element from an AIDS patient isolate. Although it is not commonly found in *M. avium* isolates, IS1110 is reported to have a significant degree of mobility, with more diverse RFLP patterns than those seen with IS900 and IS901 (Hernandez Perez *et al.*, 1994). So far, the distribution of ISN in different sources of *M. avium* isolates and its polymorphism has not been investigated. Nevertheless, the analysis of these elements indicate that although there are many sequence structure similarities, they still have some unique features. These differences might be due to their host and the IS elements themselves.

IS elements are widespread among prokaryotes. IS elements can move between phage genomes, plasmids and chromosomes, and transfer from one species/strain to another (Iida *et al.*, 1983). The source of these IS900 family elements is unknown. They may

originally have been carried on a phage, since IS110 was found as an occasional passenger on the temperate *Streptomyces* phage C31 (Brutonand Chater, 1987; Green *et al.*, 1989). However, evidence shows that these elements may also come from plasmids. For example, IS1110 was isolated from the plasmid pLR20 of *M. avium* strain LR541 when an increase in the size of this plasmid was observed (Hernandez Perez *et al.*, 1994). IS116, too, was first observed following the introduction of the *Streptomyces* plasmid pIJ702 into protoplasts prepared from *Streptomyces clavuligerus* spores. Then the IS116 was isolated from plasmid pIJ702 in one transformant (Leskiw *et al.*, 1990). Plasmids have been commonly found in strains of *M. avium* complex. There is not an obvious connection found between plasmids and virulence in MAC, although it was reported that these plasmids may encode some metal resistance functions (Crawford and Falkinham, 1990). It may reasonably be supposed that these plasmids also play an important role in the distribution of IS elements among MAC organisms.

Recently, more and more IS900-like elements have been identified from mycobacteria and other bacterial genera, such as IS110 from *Streptomyces coelicolor*, IS117 from *Streptomyces coelicolor* A3(2) (Henderson *et al.*, 1990), IS116 from *Streptomyces clavuligerus*, IS492 from *Pseudomonas atlantica*, IS901/902 from *M. avium*, IS1110 from *M. avium* and our ISN, from *M. avium*. This suggests that *M. avium* may carry more undiscovered IS elements related to the IS900 family. As a new member of this family, IS1110 was isolated from *M. avium* LR541 strain. Initial tests using a probe derived from IS1110 seemed to indicate that IS1110 is not commonly detected in *M. avium* with only two strains positive (both

from AIDS patients). However, prolonged exposure of the blot found that several strains have weak hybridization bands with different polymorphism. This suggests that there are yet other IS900-like elements carried in *M. avium* isolates. As mentioned before, workers in Dr. Kunimoto's lab noticed that one *M. intracellulare* strain may carry another IS900-related element as well. It seems there is extensive variation among members of the IS900 family. This element has been found in many bacterial genera and more have yet to be identified. Therefore this class of IS element may be far more widely distributed than is currently realized.

Many IS elements possess comparable features and distribute in related organisms. These features appear to make significant contributions to the evolution of their hosts. The elements' comparative analysis can shed light on bacterial evolution. IS900 related elements have several characteristics that make them of great interest for phylogenetic and taxonomic studies. Analyzing the sequence variation among these elements allows us to speculate about the evolution of these elements and their hosts.

As indicated by Kunze (Kunze *et al.*, 1991), IS900 and IS901 might have been acquired by separate events after the speciation of their hosts, *M. paratuberculosis* and *M. avium*. The sequence analysis showed that there is 18% difference at the DNA level between ISN and IS900. This level of diversity is much higher than the genome sequence divergence estimated between *M. avium* and *M. paratuberculosis* (< 2%) (McFadden *et al.*, 1992). The DNA diversity among the IS elements of *M. avium* is even higher. The sequence comparison between ISN and other IS900-related elements presents another example that the acquisition of these elements was

subsequent to the divergence of the host species. As mentioned before, these elements might originally have been carried on plasmids or phages. With an association of these elements with such vectors, there may be great potential for horizontal transfer of IS900-like elements both within and among strains of MAC and other distantly related organisms. This hypothesis is consistent with previous studies of the evolution of the IS1 element among different species of enteric bacteria (Bisercic and Ochman, 1993; Lawrence *et al.*, 1992).

The distribution of IS900 family elements in *Mycobacterium*, *Streptomyces* and other bacteria, and the variable occurrence in *M. avium* strains may be evidence of horizontal transfer after the divergence of these species. This implies that horizontal transfer is common but that the rate of transfer may be relatively low so that there is sufficient time for IS element to accumulate substantial genetic divergence. The ISN's relatively fewer copies and low frequency of occurrence in *M. avium* probably reflects the more recent invasion of this element in *M. avium* strains.

IS elements significantly contribute to host DNA rearrangements, epidemiological study and as tools for genetic research. Various studies show that IS900 family elements can be used or have the potential to be used for various mycobacterial research. For example, IS900 was used to develop primers and probes to specifically detect *M. paratuberculosis*; the polymorphism of IS1110 in *M. avium* isolates suggests this element could be a valuable epidemiological tool for studying *M. avium* infections; the association of IS901 with virulence in mice suggests that IS901 could be a useful tool for studying the genetics of the virulence of mycobacteria; and, the

artificial transposon made from IS900 makes it possible to introduce foreign genes into mycobacteria. Since, at present, we do not know much about ISN, the potential value of this IS element still needs further study. Nevertheless, the results of ISN analysis adds more data to the IS900 family, which already occupies a central position in many areas of mycobacterial research.

V. Bibliography

Arbeit, R., Slutsky, A., Barber, T., Maslow, J., Niemczyk, S., Falkinham III, J. and O'Connor, G. (1993) Genetic diversity among strains of *Mycobacterium avium* causing monoclonal and polyclonal bacteremia in patients with AIDS. *Journal of Infectious Diseases* **167**:1384-1390.

Ashby, M. K. and Bergquist, P. L. (1990) Cloning and sequence of IS1000, a putative insertion sequence from *Thermus thermophilus* HB8. *Plasmid* **24**:1-11.

Baess, I. and Mansa, B. (1978) Determination of genome size and base ratio on deoxyribonucleic acid from mycobacteria. *Acta pathologica et microbiologica scandinavica-section B, microbiology* **86**:309-312.

Bartlett, D. H. and Silverman, M. (1989) Nucleotide sequence of IS492, a novel insertion sequence causing variation in extracellular polysaccharide production in the marine bacterium *Pseudomonas atlantics*. *Journal of Bacteriology* **171**:1763-1766.

Benson, C. A. AND Ellner, J. J. (1993) *Mycobacterium avium* complex infection and AIDS: advances in theory and practice. *Clinical Infectious Diseases* **17**:7-20.

Berg, D. E., Kazic, T., Phadnis, S., Dodson, K. W. and Lodge, J. K. (1988) Mechanism and regulation of transposition. In *Transposition*, edited by Kingsman, A. J. New York, N.Y.: Cambridge University Press.

Bisercic, M. and Ochman, H. (1993) The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium*. *Journal of Bacteriology* **175**:7863-7868.

Bruton, C. J. and Chater, K. F. (1987) Nucleotide sequence of IS110, an insertion sequence of *Streptomyces coelicolor* A3(2). *Nucleic Acids Research* **15**:7053-7065.

Chater, K. F., Bruton, C. J., Foster, S. G. and Tobek, I. (1985) Physical and genetic analysis of IS110, a transposable element of *Streptomyces coelicolor*. *Molecular & General Genetics* **200**:235-239.

Clark-Curtiss, J. E. (1990) Genome structure of mycobacteria. In *Molecular Biology of the Mycobacteria*, edited by McFadden, J. J. London: Academic Press.

Cocito, C., Gilot, P., Coene, M., De Kesel, M., Poupart, P. and Vannuffel, P. (1994) Paratuberculosis. *Clinical Microbiology Reviews* **7** (3):328-345.

Collins, F. M. (1989) Mycobacterial disease, immunosuppression, and acquired immunodeficiency syndrome. *Clinical Microbiology Reviews* **2**:360-367.

Crawford, J. T. and Falkinham, J. O. I. (1990) Plasmids of the *Mycobacterium avium* complex. In *Molecular Biology of the Mycobacteria*, edited by McFadden, J. J. London: Academic Press.

Craxton, M. (1991) *Methods: a companion to methods in enzymology.*

Dale, J. W. and Patki, A. (1990) Mycobacterial gene expression and regulation. In *Molecular Biology of the Mycobacteria*, edited by McFadden, J. J. London: Academic Press.

Derbyshire, K. and Grindley, N. D. F. (1992) Binding of the IS903 transposase to its inverted repeat in vitro. *The EMBO Journal* **11**:3449-3455.

Doran, T., Tizard, M., Millar, D., Ford, J., Sumar, N., Loughlin, M. and Hermon-Taylor, J. (1997) IS900 targets translation initiation signals in *Mycobacterium avium* subsp. *paratuberculosis* to facilitate expression of its *hed* gene. *Microbiology* **143**:547-552.

Doran, T. J., Davies, J. K., Radford, A. J. and Hodgson, A. L. (1994) Putative functional domain within ORF2 on the mycobacterium insertion sequences IS900 and IS902. *Immunology & Cell Biology* **72**:427-434.

El-Zaatari, F. A. K., Naser, S. A., Engstrand, L., Hachem, C. Y. and Graham, D. Y. (1994) Identification and characterization of *Mycobacterium paratuberculosis* recombinant proteins expressed in *E. coli. Current Microbiology* **29**:177-184.

England, P. M., Wall, S. and McFadden, J. (1991) IS900-promoted stable integration of a foreign gene into mycobacteria. *Molecular Microbiology* **5**:2047-2052.

Feinberg, A. P. and Vogelstein, B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Analytical Biochemistry* **137**:266-7.

Fidler, H. M., Thurrell, W., McI Jonson, N., Rook, G. A. W. and McFadden, J. J. (1994) Specific detection of *Mycobacterium paratuberculosis* DNA associated with granulomatous tissue in Crohn's disease. *Gut* **35**:506-510.

Fournier, p., Paulus, F. and Otten, L. (1993) IS870 requires a 5′-CTAG-3′ target sequence to generate the stop codon for its large ORF1. *Journal of Bacteriology* **175**:3151-3160.

Fox, G. E., Stackebrandt, E., Hespell, R. B. G., J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bossen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. and Woese, C. R. (1980) The phylogeny of prokaryotes. *Science* **209**:457-463.

Frank, T. S. and Cook, S. M. (1996) Analysis of paraffin sections of Crohn's disease for *Mycobacterium paratuberculosis* using polymerase chain reaction. *Modern Pathology* **9**:32-35.

Galas, D. J. and Chandler, M. (1989) Bacterial insertion sequences. In *Mobile DNA*, edited by Berg, D. E. Washington, D.C.: American Society for Microbiology.

Grange, J. M. (1988) *Mycobacteria and Human Disease*. London:

Edward Arnold.

Gray, J. R. and Rabeneck, L. (1989) Atypical mycobacterial infection of the gastrointestinal tract in AIDS patients. *American Journal of Gastroenterology* **84**:1521-1524.

Green, E. P., Tizard, M. L. V., Moss, M. T., Thompson, J., Winterbourne, D. J., McFadden, J. J. and Hermon-Taylor, J. (1989) Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. *Nucleic Acids Research* **17**:9063-9073.

Grindley, N. and Wiater, L. A. (1988) γδ transposase and integration host factor bind cooperatively at both ends of γδ. *EMBO Journal* **7**:1907-1912.

Grindley, N. D. F. (1985) Transpositional recombination in prokaryotes. *Annual Review of Biochemistry* **54**:863-896.

Guerrero, C., Bernasconi, C., Burki, D., Bodmer, T. and Telenti, A. (1995) A novel insertion element from *Mycobacterium avium*, IS1245, is a specific target for analysis of strain relatedness. *Journal of Clinical Microbiology* **33**:304-307.

Guthertz, L. S., Damsker, B., Bottone, E. J., Ford, E. G., Midura, T. F. and Janda, J. M. (1989) *Mycobacterium avium* and *Mycobacterium intracellulare* infections in patients with and without AIDS. *Journal of Infectious Diseases* **160**:1037-1041.

Hampson, S. J., Thompson, J., Moss, M. T., Portaels, F., Green, E. P., Hermon-Taylor, J. and McFadden, J. J. (1989) DNA probes demonstrate a single highly conserved strain of *Mycobacterium avium* infecting AIDS patients. *Lancet* **1**:65-68.

Henderson, D. J., Brolle, D.-F., Kieser, T., Melton, R. E. and Hopwood, D. A. (1990) Transposition of IS117 (the *Streptomyces coelicolor* A3(2) mini-circle) to and from a cloned target site and into secondary chromosomal sites. *Molecular & General Genetics* **224**:65-71.

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**:351.

Hernandez Perez, M., Fomukong, N. G., Hellyer, T., Brown, I. N. and Dale, J. W. (1994) Characterization of IS1110, a highly mobile genetic element from *Mycobacterium avium*. *Molecular Microbiology* **12**:717-724.

Horsburgh, C. R., Jr., Havlik, J. A., Ellis, D. A., Kennedy, E., Fann, S. A., DuBois, R. E. and Thompson, S. E. (1991) Survival of patients with acquired immune deficiency syndrome and disseminated *Mycobacterium avium* complex infection with and without antimycobacterial chemotherapy. *American Review of Respiratory Disease* **144**:557-559.

Ichikawa, H., Ikeda, K., Wishart, W. L. and Ohtsubo, E. (1987) Specific binding of transposase to terminal inverted repeats of transposable elements. *Proceedings of the National Academy of Sciences, USA*

84:8220-8224.

Iida, S., Meyer, J. and Arber, W. (1983) Prokaryotic IS elements. In *Mobile Genetic Elements*, edited by Shapiro, J. A. New York: Academic Press.

Imaeda, T. (1985) Deoxyribonucleic acid relatedness among selected strains of *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium bovis* BCG, *Mycobacterium microti*, and *Mycobacterium africanum*. *International Journal of Systematic Bacteriology* **35**:147-150.

Imaeda, T., Broslawski, G. and Imaeda S. (1988) Genomic relatedness among mycobacterial species by nonisotopic blot hybridization. *International Journal of Systematic Bacteriology* **38**:151-156.

Inderlied, C. B., Kemper, C. A. and Bermudez, L. E. (1993) The *mycobacterium avium* complex. *Clinical Microbiology Reviews* **6**:266-310.

Innis, M. A. and Gelfand, D. H. (1990) Optimization of PCRs. In *PCR protocols: a guide to methods and applications*, edited by Innis, M. A. New York: Academic Press.

Isberg, R. R. and Falkow, S. (1985) Genetic analysis of bacterial virulence determinants in *Bordetella pertussis* and the pathogenic *Yersinia*. *Current Topics in Microbiology and Immunology* **118**:1-11.

Jacobson, M. A., Hopewell, D. M., Yajko, W. K., Hadley, W. K., Lazarus, E., Mohanty, P. K., Modin, G. W., Feigal, D. W., Cusick, P. S. and Sande, M. A. (1991) Natural history of disseminated *Mycobacterium avium* complex infection in AIDS. *Journal of Infectious Diseases* **164**:994-998.

Klatt, E. C., Jensen, D. F. and Meyer, P. R. (1987) Pathology of *Mycobacterium avium-intracellulare* infection in acquired immunodeficiency syndrome. *Human Pathology* **18**:709-714.

Kleckner, N. (1981) Transposable elements in prokaryotes. *Annual Review of Genetics* **15**:341-404.

Knapp, A., Stern, G. A. and Hood, C. I. (1987) Mycobacterium avium-intracellulare corneal ulcer. *Cornea* **6**:175-180.

Kunze, Z. M., Portales, F. and McFadden, J. J. (1992) Biologically distinct subtypes of *Mycobacterium avium* differ in possession of Insertion sequence IS901. *Journal of Clinical Microbiology* **30**:2366-2372.

Kunze, Z. M., Wall, S., Appelberg, R., Silva, M. T., Portaels, F. and McFadden, J. J. (1991) IS901, a new member of a widespread class of atypical insertion sequences, is associated with pathogenicity in *Mycobacterium avium*. *Molecular Microbiology* **5**:2265-2272.

Lawrence, J. G., Ochman, H. and Hartl, D. L. (1992) The evolution of insertion sequences within enteric bacteria. *Genetics* **131**:9-20.

Leskiw, B. K., Mevarech, M., Barritt, L. S., Jensen, S. E., Henderson, D. J., Hopwood, D. A., Bruton, C. J. and Chater, K. F. (1990) Discovery of an insertion sequence, IS116, from *Streptomyces clavuligerus* and its relatedness to other transposable elements from actinomycetes. *Journal of General Microbiology* **136**:1251-1258.

Malpother, M. E. and Sanger, J. G. (1984) In vitro interaction of *Mycobacterium avium* with intestinal epithelial cells. *Infection and Immunity* **45**:67-73.

Martin, C., Ranes, M. and Gicquel, B. (1990) Plasmids, antibiotic resistance, and mobile genetic elements in mycobacteria. In *Molecular Biology of the Mycobacteria*, edited by McFadden, J. J. London: Academic Press.

Mazurek, G. H., Hartman, S., Zhang, Y., Brown, B. A., Hector, J. S. R., Murphy, D. and Wallace, R. J., Jr. (1993) Large DNA restriction fragment polymorphism in the *Mycobacterium avium-M. intracellulare* complex: a potential epidemiologic tool. *Journal of Clinical Microbiology* **31**:390-394.

McFadden, J., Kunze, Z. and Seechurn, P. (1990) DNA probes for detection and identification. In *Molecular Biology of the Mycobacteria*, edited by McFadden, J. J. London: Academic Press.

McFadden, J. J., Butcher, P. D., Thompson, J., Chlodini, R. and Hermon-Taylor, J. (1987) The use of DNA probes identifying restriction-fragment-length polymorphisms to examine the *Mycobacterium*

avium complex. *Molecular Microbiology* **1**:283-291.

McFadden, J. J., Kunze, Z. M., Portaels, F., Labrousse, V. and Rastogi, N. (1992) Epidemiological and genetic markers, virulence factors and intracellulare growth of *Mycobacterium avium* in AIDS. *Research in Microbiology* **143**:423-430.

Meissner, G. and Anz, W. (1977) Sources of *Mycobacterium avium* complex infection resulting in human diseases. *American Review of Respiratory Disease* **116**:1057-1064.

Modilevsky, T., Sattler, F. R. and Barnes, P. F. (1989) Mycobacterial disease in patients with human immunodeficiency virus infection. *Archives of Internal Medicine* **149**:2201-2205.

Moss, M. T., Green, E. P., Tizard, M. L., Malik, Z. P. and Hermon-Taylor, J. (1991) Specific detection of *Mycobacterium paratuberculosis* by DNA hybridization with a fragment of the insertion element IS900. *Gut* **32**:395-398.

Moss, M. T., Malik, Z. P., Tizard, M. L. V., Green, E. P., Sanderson, J. D. and Hermon-Taylor, J. (1992a) IS*902*, an insertion element of the chronic-enteritis-causing *Mycobacterium avium* subsp. *silvaticum*. *Journal of General Microbiology* **138**:139-145.

Moss, M. T., Sanderson, J. D., Tizard, M. L. V., Hermon-Taylor, J., El-Zaatari, F. A. K., Markesich, D. C. and Graham, D. Y. (1992b) Polymerase chain reaction detection of *Mycobacterium*

paratuberculosis and *Mycobacterium avium* subsp *silvaticum* in long term cultures from Crohn's disease and control tissues. *Gut* **33**:1209-1213.

Murphy, E. (1988) Transposable elements in *Staphylococcus*. In *Transposition*, edited by Kingsman, A. J. New York, N.Y.: Cambridge University Press.

Murray, A., Winter, N., Lagranderie, M., Hill, D. F., Rauzier, J., Timm, J., Leclerc, C., Moriarty, K. M., Gheorghiu, M. and Gicquel, B. (1992) Expression of *Escherichia coli* beta-galactosidase in *Mycobacterium bovis* BCG using an expression system isolated from *Mycobacterium paratuberculosis* which induced humoral and cellular immune responses. *Molecular Microbiology* **6**:3331-3342.

Murray, V. (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Research* **17**:8889.

Musial, C. E., Tice, L. S., Stockman, L. and Roberts, G. D. (1988) Identification of mycobacteria from culture by using the Gen-Probe rapid diagnostic system for *Mycobacterium avium* complex and *Mycobacterium tuberculosis* complex. *Journal of Clinical Microbiology*:2120-2123.

New, J. H., Eggleston, A. K. and Fennewald, M. (1988) Binding of Tn3 transposase to the inverted repeats of Tn3. *Journal of Molecular Biology* **201**:589-599.

Nishimori, K., Eguchi, M., Nakaoka, Y., Onodera, Y., Ito, T. and Tanaka, K. (1995) Distribution of IS901 in strains of *Mycobacterium avium* complex from swine by using IS901-detacting primers that discriminate between *M. avium* and *Mycobacterium intracellulare*. *Journal of Clinical Microbiology* **33**:2102-2106.

Ohtsubo, H., Nyman, K., Doroszkiewicz, W. and Ohtsubo, E. (1981) Multiple copies of iso-insertion sequences of IS1 in *Shigella dysenteriae* chromosome. *Nature (London)* **292**:640-643.

Picken, R. N., Tsang, A. Y. and Yang, H. L. (1988) Speciation of organisms within the *Mycobacterium avium-Mycobacterium intracellulare-Mycobacterium scrofulaceum* (MAIS) complex based on restriction fragment length polymorphisms. *Molecular & Cellular Probes* **2**:289-304.

Roberts, G. D., Koneman, E. W. and Kim, Y. K. (1991) Mycobacterium. In *Manual of Clinical Microbiology, fifth edition*, edited by Balows, A. Washington, D.C.: American Society for Microbiology.

Rogall, T., Flohr, T. and Böttger, E. C. (1990) Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. *Journal of General Microbiology* **136**:1915-1920.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning a Laboratory Manual*. second ed. New York: Cold Spring Harbor Laboratory Press.

Sawyer, S. A., Dykhuizen, D. E., DuBose, R. F., Green, L., Mutangadura Mhlanga, T., Wolczyk, D. F. and Hartl, D. L. (1987) Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. *Genetics* **115**:51-63.

Shapiro, J. A. (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proceedings of the National Academy of Sciences, USA* **76**:1933-1937.

Thorel, M. F., Krichevsky, M. and Levy Frebault, V. V. (1990) Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. *International Journal of Systematic Bacteriology* **40**:254-260.

Tizard, M. L. V., Moss, M. T., Sanderson, J. D., Austen, B. M. and Hermon-Taylor, J. (1992) p43, the protein product of the atypical insertion sequence IS900, is expressed in *Mycobacterium paratuberculosis*. *Journal of General Microbiology* **138**:1729-1736.

Vary, P. H., Andersen, P. R., Green, E., Hermon-Taylor, J. and McFadden, J. J. (1990) Use of highly specific DNA probes and the polymerase chain reaction to detect *Mycobacterium paratuberculosis* in Johne's disease. *Journal of Clinical Microbiology* **28**:933-937.

Wallace, R. J., Jr., O'Brien, R., Glassroth, J., Raleigh, J. and Dutt, A. (1990) Diagnosis and treatment of disease caused by nontuberculosis mycobacteria. *American Review of Respiratory Disease* **142**:940-953.

Weiss, A. A., Hewlett, E. L., Myers, G. D. and Falkow, S. (1983) Pertussis toxin and extracytoplasmic adenylate cyclase as virulence factors. *Infection and Immunity* **42**:33-41.

Wendt, S. L., George, K. L., Parker, B. C., Gruft, H. and Falkinham, J. O. I. (1980) Epidemiology of infection by nontuberculous mycobacteria. *American Review of Respiratory Disease* **122**:259-263.

Woese, C. R. (1987) Bacterial evolution. *Microbiological Reviews* **51**:221-271.

Woese, C. R., Gutell, R., Gupta, R. and Noller, H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiological Reviews* **47**:621-669.

Xiong, Y. and Eickbush, T. H. (1988) Similarity of reverse transcriptase-like sequences of virus, transposable elements and mitochondrial introns. *Molecular Biology & Evolution* **5**:675-690.