slfo

의학박사 학위논문

Molecular taxonomic study of *Mycobacterium yongonense*

Mycobacterium yongonense 균주의 분자분류학적 연구

2016 년 8 월

서울대학교 대학원

의과학과 의과학전공

김 병 준

A thesis of the Degree of Doctor of Philosophy

Mycobacterium yongonense 균주의 분자분류학적 연구

Molecular taxonomic study of *Mycobacterium yongonense*

August 2016

The Department of Biomedical Sciences,
Seoul National University
College of Medicine
Byoung Jun Kim

Mycobacterium yongonense

균주의 분자 분류학적 연구

지도교수 김 범 준

이 논문을 의학박사 학위논문으로 제출함

2016 년 8 월

서울대학교 대학원
의과학과 의과학전공
김 병 준

김병준의 의학박사 학위논문을 인준함
2016 년 8 월

위 원 장	(인)
부위원장	(인)
위 원	(인)
위 원	(인)
위 원	(인)

Molecular taxonomic study of *Mycobacterium yongonense*

by
Byoung Jun Kim

A thesis submitted to the Department of Biomedical Sciences in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Medical Science at Seoul National University College of Medicine

August 2016

Approved by Thesis Committee:

Professor _____Chairman
Professor _____Vice chairman
Professor _____
Professor _____
Professor _____

ABSTRACT

Introduction: Members of *Mycobacterium avium* complex (MAC) are the most frequently isolated non-tuberculous mycobacteria (NTM). Traditionally, MAC complex includes two species of *M. avium* and *M. intracellulare* strains. Recently, using the partial three independent chronometer genes, *hsp65*, ITS1 and 16S rRNA, clinical isolated *M. intracelllulare* strains from Korean patients were separated into 5 groups (INT-1 to -5) due to their genetic diversities.

Methods: A novel strain 05-1390, which is related with *M. intracellulare*, especially INT-5 group, was isolated from a Korean pulmonary patient and proposed to a novel species (*M. yongonense* DSM 45126^T) by biochemical and phylogenetic analyses. This strain is phylogenetically related to *M. intracellulare*, but has a distinct RNA polymerase β-subunit gene (*rpoB*) sequence that is identical to that of *M. parascrofulaceum*, suggesting the acquisition of the *rpoB* gene via a potential lateral gene transfer (LGT) event. To gain better insight into the possibility of LGT event mechanisms in the *M. yongonense*, the complete genome sequence of *M. yongonense* was determined by four types of sequencing methods. Also, to determine the exact taxonomic status of the *M. yongonense* and other INT-5 strains (MOTT-36Y and MOTT-H4Y), genome-based phylogenetic analysis and IS-elements identification were conducted. To this end, genome sequences of the two INT-5 strains were compared with *M. intracellulare* ATCC 13950^T and *M. yongonense* DSM 45126^T. Multilocus sequence typing (MLST) of 35 target

genes and single nucleotide polymorphism (SNP) analyses were conducted to compare the relationship between the two INT-5 strains and *M. yongonense* DSM 45126^{T}. And using the *M. yongonense* genome information, a novel IS-element was identified and applied to develop a novel diagnostic method based on a real-time PCR technique. To figure out the mechanism of LGT events in the genome of *M. yongonense* from *M. parascrofulaceum*, putative LGT regions in *M. yongonense* strains were identified by comparative genomic analysis and the regions were examined by SimPlot and BootScan analysis. Also, in the genome of *M. yongonense* strains, DNA mismatch repair genes (*mutS4A* and *mutS4B*) were identified, and examined the ability in recombination of DNA mismatch repair genes. To this end, recombinant *M. smegmatis* harboring DNA mismatch repair gene was constructed and partial *rpoB* sequences (684 bp) from rifampin resistant *M. tuberculosis* was introduced in the recombinant *M. smegmatis*. After that, the ability of resistance to rifampin and recombination frequencies of recombinant *M. smegmatis* were examined.

Results: A novel strain, 05-1390 was similar with *M. intracellulare* in biochemical traits. Phylogenetic analysis based on ITS1 and the *hsp65* gene indicated that strain 05-1390 was closely related to *M. intracellulare* ATCC 13950^{T}, but the *rpoB* gene sequence showed that it is closely related to *M. parascrofulaceum* ATCC BAA-614^{T}. With these results, strain 05-1390 was proposed to a novel species as *M. yongonense* DSM 45126^{T}. This strain have a circular chromosome (5,521,023 bp; GenBank accession no. of CP003347), a circular plasmid (122,976 bp; GenBank accession no. of JQ657805), and a

linear plasmid (18,089 bp; GenBank accession no. of JQ657806). The genome based MLST and SNP analyses showed that the two INT-5 strains and *M. yongonense* DSM 45126^{T} were closely related than other *M. intracellulare* strains. Also, a novel IS-elements, IS*Myo2* (2,387 bp), belonging to the IS*21* family were identified in the genome of *M. yongonense* DSM 45126^{T} and two INT-5 strains. The developed real-time PCR diagnostic method targeting IS*Myo2* could detect only *M. yongonense* strains including the two INT-5 strains. In the genome of *M. yongonense* DSM 45126^{T}, two putative LGT regions with high sequence similarity with *M. parascrofulaceum* were identified**.** The first region is *rpoBC* operon (OEM_44170~44190) and the second region is containing ORFs from OEM_08030 to OEM_08590 (57 ORFs). SimPlot and BootScan analyses supported that these regions were laterally transferred from *M. parascrofulaceum*. Also, the recombinant *M. smegmatis* harboring DNA mismatch repair gene showed potential of resistance to rifampin after indroducing the partial *rpoB* sequences related with rifampin resistance.

Conclusions: The genome-based phylogenetic analysis showed that *M. yongonense* is taxonomically related with INT-5 strains (MOTT-36Y and MOTT-H4Y). So, previously described taxonomic status of the two INT-5 strains should be revised into *M. yongonense* strains. *M. yongonense* strains could be divided into two distinct genotypes. Type I genotype have the distinct *rpoBC* operon which was laterally transferred from *M. parascrofulaceum*, and Type II genotype have the similar *rpoBC* operon with *M. intracellulare* strains. The result of LGT events could be a criteria for

differentiation between *M. yongonense* Type I and Type II genotypes, and the LGT events might be facilitated by introduction of DNA mismatch repair genes from the genus of *Acidothermus* species to *M. yongonense* strains.

* This work is published in Internationl Journal of Systematic and Evolutionary Microbiology (Kim BJ, Math RK, Jeon CO, Yu HK, Park YG, Kook YH, et al. *Mycobacterium yongonense* sp. nov., a slow-growing non-chromogenic species closely related to *Mycobacterium intracellulare*. Int J Syst Evol Microbiol. 2013; 63(Pt 1):192-9), PLoS One (Kim BJ, Hong SH, Kook YH, and Kim BJ. Molecular evidence of lateral gene transfer in *rpoB* gene of *Mycobacterium yongonense* strains via multilocus sequence analysis. PLoS One. 2013;8(1):e51846.; Kim BJ, Kim BR, Lee SY, Kim GN, Kook YH, and Kim BJ. Molecular Taxonomic Evidence for Two Distinct Genotypes of *Mycobacterium yongonense* via Genome-Based Phylogenetic Analysis. PLoS One. 2016; 11(3):e0152703), Genome Announcements (Kim BJ, Kim BR, Lee SY, Seok SH, Kook YH, and Kim BJ. Whole-Genome Sequence of a Novel Species, *Mycobacterium yongonense* DSM 45126^{T}. Genome Announc. 2013; 1(4)), and BMC Genomics (Kim BJ, Kim K, Kim BR, Kook YH, and Kim BJ. Identification of IS*Myo2*, a novel insertion sequence element of IS*21* family and its diagnostic potential for detection of *Mycobacterium yongonense*. BMC Genomics. 2015; 16(1):794.).

Keywords: *Mycobacterium yongonense*, *M. intracellulare* INT-5 group, genome sequence, comparative genome analysis, multi-locus sequence typing (MLST), single nucleotide polymorphism (SNP), RNA polymerase β-subunit

gene (*rpoB*), lateral gene transfer (LGT), DNA mismatch repair gene.

Student number: 2010-30606.

CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF ABBREVIATIONS

ADC: albumin dextrose catalase

AFLP: amplified fragment length polymorphism

AMC: Asan medical center

Ami: Amikacin

ATCC: American Type Culture Collection

CDS: coding sequence

Cef: Cefoxitin

Cip: Ciprofloxacin

CIP: Collection de I'Institut Pasteur

Cla: Clarithromycin

Cqs: Quantification cycles

DDH: DNA-DNA hybridization

Dox: Doxycycline

EMB: ethambutol

HPLC: High-performance liquid chromatography

hsp65: heat shock protein 65 kD

HWM: High-molecular-weight standard

Imi: Imipenem

IR: inverted repeat

IRL: left inverted repeat

IRR: right inverted repeat

IS: insertion sequence

ITS1: internal transcribed spacer 1

KCTC: Korean Collection for Type Cultures

KIT: Korean Institute of Tuberculosis

LGT: lateral gene transfer

LMW: Low-molecular-weight standard

MAC: *Mycobacterium avium* complex

MALDI-TOF: Matrix-Associated Laser Desorption Ionization-Time of Flight

MLSA: multilocus sequence analysis

MLST: Multilocus sequence typing

Mox: Moxifloxacin

NICEM: National Instrumentation Center for Environmental Management

NTM: non-tuberculous mycobacteria

OADC: oleic acid albumin dextrose catalase

OD: optical density

ORF: open reading frame

PCR: polymerase chain reaction

PFGE: pulsed-field gel electrophoresis

PGAAP: Prokaryotic Genomes Automatic Annotation Pipeline

PNB: p-nitrobenzoate

QIA: The Animal and Plant Quarantine Agency

RAPD: random amplified polymorphic DNA

Rif: Rifampin

***rpoB*:** RNA polymerase β-subunit gene

SEM: scanning electron microscopy

SNP: single nucleotide polymorphism

SNUMC: Seoul National University College of Medicine

Sul: Sulfamethoxazole

TBSA: Tuberculostearic acid

TCH: thiophene-2-carboxylic acid hydrazide

Tms: melting temperatures

Tob: Tobramycin

VNTR: variable number of tandem repeat

GENERAL INTRODUCTION

Within the genus *Mycobacterium*, close to 200 mycobacterial species have now been described. Among the mycobacterial species, prominent pathogens such as *Mycobacterium tuberculosis* complex, *M. leprae* and *M. ulcerans* are exist. In addition, numerous species of environmental mycobacteria called non-tuberculous mycobacteria (NTM) are responsible for various kinds of diseases.

Members of *M. avium* complex (MAC) are the most frequently isolated NTM in global as well as in Korea [1-3]. Traditionally, MAC includes two species, *M. avium* and *M. intracellulare* [4-6]. Based on phenotypic and genetic characteristics, four subspecies of *M. avium* have been proposed: *M. avium* subsp. *avium*, *M. avium* subsp. *paratuberculosis*, *M. avium* subsp. *silvaticum* [7], and *M. avium* subsp. *hominissuis* [8]. Also, several species which were closely related with *M. intracellulare* have been described by many researchers: *M. bouchedurhonense*, *M. colombiense* [9], *M. vulneris* (formerly sequevar MAC-Q) [10], *M. arosiense* [11], *M. marseillense*, *M. timonense* [12] and *M. chimaera* (formerly sequevar MAC-A) [13].

As described above, members of the MAC are the most frequently isolated NTM in Korea. However, unique to Korea, *M. intracellulare* has been isolated more frequently than *M. avium* strains. According to recent paper, there exists diversities in *M. intracellulare* related clinical isolates due to their genetic heterogeneity in Korea. The *M. intracellulare* strains were grouped into five distinct groups (INT-1 to INT-5, the type strain of *M. intracellulare* is grouped into INT-2) using the combinational

sequence analysis of the partial *hsp65*, internal transcribed spacer (ITS), and 16S rRNA sequences [14].

Recently, rapid developments in molecular biology have resulted in the use of various methods for the identification and differentiation of species of the genus *Mycobacterium*. These include pulsed-field gel electrophoresis (PFGE) [15, 16], insertion sequences (IS-elements such as IS*900*, IS*901*, IS*902* and IS*1245* in MAC complex) based typing [17-19], variable number of tandem repeat (VNTR) analysis [20-22], random amplified polymorphic DNA (RAPD) analysis [23], amplified fragment length polymorphism (AFLP) analysis [24, 25], multilocus sequence typing (MLST) analysis [26-28], and single nucleotide polymorphism (SNP) typing analysis [29]. However, the discriminatory power of these techniques for the NTM has not been fully examined. Among these methods, MLST has also the potential to become a reference method, although its wider use is hampered by limited access, high costs of the sequencing facilities. Also, the discriminatory power of most of the markers not fully established in each NTM species. This in turn may lead to misinterpretation of genotyping results.

In this situation, it is noteworthy that the advances in whole-genome sequencing technologies and with the result, there are available many whole-genome sequecnes of the genus *Mycobacterium* including clinical strains from around the world (ftp://ftp.ncbi.nih.gov/genomes/). These genome information could applied to various analysis such as phylogeography, strain diversity and evolutionary studies. In the case of members of MAC complex, genomes of three species (*M. avium*, *M. colombiense*, *M. intracellulare*) are available. Especially, in the case of *M. intracellulare*, the genome of type strain (ATCC 13950^{T}, INT-2) [30] and four clinical isolates in Korea

(MOTT-02, INT-2; MOTT-64, INT-1; MOTT-36Y, INT-5; MOTT-H4Y, INT-5) were sequenced and the database is available [31-34].

In this study, I focused on a novel mycobacterial species, *M. yongonense* (strain 05-1390) which is phenotypically and genetically related with *M. intracellulare*, especially INT-5 group. First, I described and proposed the *M. yongonense* as a novel strain by biochemical and phylogenetical analysis based results. In this procedure, I found that the RNA polymerase β-subunit gene (*rpoB*) of *M. yongonense* is distinct from *M. intracellulare*, but identical to that of *M. parascrofualceum*, which is a distantly related scotochromogen. So, I hypothesized about the possibility of lateral gene transfer (LGT) event in some genes or loci from *M. parascrofulaceum* to *M. intracellulare* INT-5 strain and the result, a novel species, *M. yongonense* is present now. To prove the presence of the LGT event in the *rpoB* gene of the *M. yongonense* strains, multilocus sequence analysis using the full *rpoB* seqeucnes and partial five target sequences (16S rRNA, *hsp65*, *dnaJ*, *recA*, and *sodA*) was applied to *M. yongonense* strains (DSM 45126^T, MOTT-12 and MOTT-27), *M. intracellulare* strains (ATCC 13950^T, INT-2; MOTT-02, INT-2) and *M. parascrofulaceum* strains (MOTT-01 and ATCC BAA-614^T) for the first study. Second, to determine the exact taxonomic relationship between *M. yongonense* and *M. intracellulare* INT-5 genotype, I planned to obtain the whole genome sequence of *M. yongonense*, and carried out the genome-based phylogenetic analysis. To do this, genome sequences of the two INT-5 strains, MOTT-H4Y and MOTT-36Y were compared with *M. intracellulare* ATCC 13950^T and *M. yongonense* DSM 45126^T. Also together, a novel insertion sequence elements (IS-elements) were identified in the genome of *M. yongonense* and developed a diagnostic method based on a real-time PCR technique. Finally, to find

genetic differences between *M. yongonense* and INT-5 strains, I performed a comparative genome analysis using the genome sequences of *M. yongonense* strains (DSM 45126^T, MOTT-12 and MOTT-27), INT-5 strains (MOTT-36Y and MOTT-H4Y) and *M. parascrofulaceum* (ATCC BAA-614^T). To this end, putative LGT regions were identified in the genome of *M. yongonense* strains by BLAST analysis and the regions were examined by SimPlot and BootScan analysis. Also, among the non-mycobacterial ORFs, DNA mismatch repair gene was identified. To prove the hypothesis that the DNA mismatch repair genes in the genome of *M. yongonense* strains could facilitate the homologous recombination event under the harsh environment, I generated a recombinant *M. smegmatis* strain harboring DNA mismatch repair gene and cultured on 7H10 agar plate containing rifampin after introducing the partial *M. tuberculosis rpoB* sequences which are related to rifampin resistance. After that the colonies which could resistant to rifampin were selected and their *rpoB* sequences determined and examined the frequency of homologous recombination.

CHAPTER 1

Mycobacterium yongonense sp. nov., a novel slow growing nonchromogenic species closely related to *Mycobacterium intracellulare*

INTRODUCTION

Among the slow-growing mycobacteria responsible for opportunistic infections, members of *Mycobacterium avium* complex (MAC) are the non-tuberculous mycobacteria (NTM) most frequently isolated in clinical settings [6, 35, 36]. Traditionally, MAC includes two species, *M. avium* and *Mycobacterium intracellulare* [4-6]. Recent, advances in molecular taxonomy have fueled the identification of novel species within MAC [9-13, 37]. Details on genetic separations between closely related strains of *M. intracellulare* have been disclosed by many researchers. Seven MAC or *M. intracellulare*-related species, *Mycobacterium bouchedurhonense*, *Mycobacterium colombiense* [9], *Mycobacterium vulneris* (formerly sequevar MAC-Q) [10], *Mycobacterium arosiense* [11], *Mycobacterium marseillense*, *Mycobacterium timonense* [12], and *Mycobacterium chimaera* (formerly sequevar MAC-A) [13], have recently been described.

As in other countries, MAC is the most frequently isolated NTM in Korea. However, unique to Korea, *M. intracellulare* has been isolated more frequently than *M. avium* [3, 38, 39]. Recently, a paper was reported about diversity of *M. intracellulare* related strains showing genetic heterogeneity with one another in Korea, distinct from other countries; the strains can be divided into a total of five distinct groups using the sequence analysis of *hsp65*, internal transcribed spacer and 16S rRNA genes [14].

Generally, the informative genes associated with the central dogma of bacteria, such as the 16S rRNA gene or the RNA polymerase β-subunit gene (*rpoB*), have been reported to be recalcitrant to lateral gene transfer (LGT) events. However, the LGT

events of informative genes within the genus *Mycobacterium* have been disclosed in two recent reports. One report described the potential LGT event of the *rpoB* gene between three groups of strains belonging to *M. abscessus* (*M. abscessus sensu stricto*, *M. massiliense* and *M. bolletii*) [40]; the other report described the potential LGT event of the 16S rRNA gene between *Mycobacterium franklinii* and *Mycobacterium chelonae* [41].

In the present study, first, I described a novel clinical isolate that is closely related to *M. intracellulare* via the polyphasic taxonomic approach using a combination of biochemical tests and molecular analyses targeting the 16S rRNA [42, 43], *hsp65* [44], internal transcribed spacer 1 (ITS1) [14], and *rpoB* [45, 46] gene sequences. Second, I discovered the epidemiologic features of *M. yongonense* from an infection cohort previously identified as *M. intracellulare*. And finally, I proposed the presence of the LGT event in the *rpoB* gene of the *M. yongonense* strains via multilocus sequence analysis (MLSA).

MATERIALS AND METHODS

1. Mycobacterial strains and culture conditions

The strain 05-1390^T (*M. yongonense* DSM 45126^T), used in the stud y was one of the 'difficult-to-identify' isolates from a patient with pulmonary symptoms, which was submitted to the Korean Institute of Tuberculosis (KIT) by mycobacteriology laboratories in Korea in 2005. A sputa sample decontaminated by 1 % NaOH was centrifuged at 13,000 rpm for 15 min. The sedimented sample was resuspended in 1.5 ml of phosphate buffer (pH 6.8) for inoculation into 7H9 broth with ADC and streaked onto 7H10 agar with OADC for a period of 2 weeks. Other strains, for used as control groups, *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T were from KIT, and used in this study (Table 1-1).

Also, for MLSA assay, seven mycobacteria strains, including three reference strains (*M. intracellulare* ATCC 13950^T, *M. parascrofulaceum* ATCC BAA-614^T, and *M. yongonense* DSM 45126^T) and four clinical isolates (MOTT-01, MOTT-02, MOTT-12 and MOTT-27) were used (Table 1-1). Of the four clinical isolates, one (MOTT-01) was identified as *M. parascrofulaceum*, one (MOTT-02) as *M. intracellulare* and two (MOTT-12 and MOTT-27) as *M. yongonense*, using a combination of the *hsp65* and *rpoB* sequence based analyses. These strains were recovered from -70 ℃ storage in a deep-freezer by sub-culturing into Middlebrook 7H9 broth media at an incubation temperature of 37 ℃.

2. Biochemical tests

The morphology and size of the 05-1390^T strain were observed by scanning electron microscopy (SEM) using a JSM-840A JEOL electron microscope (JEOL Ltd, Tokyo, Japan). The phenotypic characteristics of the strain 05-1390^T and the type or reference strains, *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T strains were analyzed and compared. Also, the two *M. yongonense* clinical isolates (MOTT-12 and MOTT-27), their biochemical test profiles were compared with those of three type reference strains: *M. intracellulare* ATCC 13950^T, *M. yongonense* DSM 45126^T and *M. parascrofulaceum* ATCC BAA-614^T. The colony morphology, pigmentation in the dark, photo-induction and growth at different temperatures, of 25 °C, 37 °C, and 45 °C were tested on 7H10 agar plates with OADC during a six-week incubation period. Acid-alcohol-fastness was examined by Ziehl-Neelsen and auramine O staining. The biochemical characteristics, of niacin accumulation, nitrate reductase, arylsulfatase on days 3 and 14, heat-stable catalase (pH 7, 68 °C), tellurite reductase, Tween 80 hydrolysis, urease and pyrazinamidase were tested [47]. Inhibition tests including tolerance to thiophene-2-carboxylic acid hydrazide (TCH), p-nitrobenzoate (PNB), 5% sodium chloride, ethambutol (EMB), and picric acid were performed, and the ability to grow on MacConkey agar without crystal violet was examined. Growth in 7H9 broth with ADC adjusted to pH 5.5 and 7.0 was examined. In addition, antimicrobial susceptibility was determined by the agar proportion method in the 7H10 medium [47]. The production of alkaline phosphatase, esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, and crystine arylamidase was also analyzed with the API ZYM kit (bioMerieux) following the recommendations of the manufacturer.

Table 1-1. Mycobacterial strains used in this study.

Strains	Source	Sp. or Str. [a]
M. chimaera	JCM 14737^T	*M. chimaera*
M. colombiense	JCM 16228^T	*M. colombiense*
M. intracellulare	ATCC 13950^T	INT-2
M. yongonense	DSM 45126^T	05-1390, INT-5
M. parascrofulaceum	ATCC BAA-614^T	*M. parascrofulaceum*
MOTT-01	AMC [b]	*M. parascrofulaceum*
MOTT-02	AMC	INT-2
MOTT-12	AMC	INT-5
MOTT-27	AMC	INT-5

[a] Separation into species or strain level was performed by combination of hsp65, ITS-1 and 16S rRNA sequences based analyses [14].

[b] AMC, Asan medical center.

3. Fatty acid methyl esters analysis

Fatty acid methyl esters were obtained from biomass of strain 05-1390^T, with the manner described by Minnikin (1988) [48]. Extracted sample was separated by gas chromatography (GC; Model 5898A, Hewlett Packard) and analyzed using the Sherlock Microbial ID System (MIS).

4. High-performance liquid chromatography (HPLC) analysis of mycolic acids

HPLC was used for the analysis of mycolic acids of 05-1390^T and other strains (*M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T) as described by Butler *et al* [49]. The low- and high-molecular mass standards (Ribi ImmunoChem) were added for peak identification. The Microbial Identification

system (MIDI Inc.) and HPLC mycobacterium library (available at http://www.MycobacToscana.it) were used to identify and compare mycolic acid patterns.

5. Matrix-Associated Laser Desorption Ionization-Time of Flight (MALDI-TOF) mass spectrometry anaylsis

For a lipid analysis, lipids were extracted with $CHCl_3$/CH_3OH (1:1 v/v, adding 0.5 µl of 2,5-dihydroxybenzoic acid) from 30ml of 7H9 broth cultures of 05-1390^T, *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T. The samples were subject to matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Voyager DE-STR MALDI-TOF instrument (Perseptive Biosystems) equipped with a pulse nitrogen laser emitting at 337 nm as previously reported [50].

6. DNA extraction and polymerase chain reaction (PCR)

Genomic DNA of each mycobacterial strain was extracted by the bead-beater phenol extraction method [51]. Briefly, a loopful of colonies from cultured strains was suspended in 200 µl of TEN buffer (10 mM Tris-HCl, 1 mM EDTA and 100 mM NaCl; pH 8.0), placed in a 2.0-ml screw-cap microcentrifuge tube filled with 200 µl slurry of glass beads (diameter, 106 µm; Sigma-Aldrich, Co., St. Louis, MO, USA) and 200 µl of phenol/chloroform/isoamyl alcohol (PCI, 25:24:1, saturated with 10 mM Tris and 1 mM EDTA, pH 8.0) (Sigma-Aldrich, Co., St. Louis, MO, USA). To disrupt the bacterial cell wall, the tube containing glass beads mixture was oscillated on a Mini-Bead Beater (Biospec Products, Bartlesville, OK, USA) for 30 sec, and to

separate the phases, centrifuged for 15 min with 13,000 rpm at 4 °C. After the aqueous phase (~300 μl) was transferred into another tube, 45 μl of 3 M sodium acetate (pH 5.2) and 600 μl of ice-cold ethanol were added. To precipitate the DNA, the mixture was kept at -20 °C for 10 min. After centrifugation at 13,000 rpm for 15 min, the DNA pellet was washed with 70 % ethanol, dissolved in 50 μl of TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH 8.0).

To obtain a total of four target sequences, 16S rRNA (1,395 bp), *hsp65* (441 bp), *rpoB* (701 bp) and internal transcribed spacer region 1 (ITS1, 298 bp) were directly determined after PCR amplification using purified genomic DNA as a template, as previously described [ref]. In brief, the template DNA (50 ng) and 20 pmol of each primer were added to a PCR mixture tube (AccuPower PCR PreMix; Bioneer, Daejeon, South Korea) containing one unit of Taq DNA polymerase, 250 μM of deoxynucleotide triphosphate, 10 mM Tris-HCl (pH 8.3), 10 mM KCl, 1.5 mM $MgCl_2$, and gel loading dye. The final volume was adjusted to 20 μl with distilled water, and the reaction mixture was then amplified as previously described [4, 14, 44-46] using a model 9700 Thermocycler (Perkin-Elmer Cetus). The sequences of the primers for the amplification are shown in Table 1-2. The 16S rRNA gene sequence of strain 05-1390^T determined in this study was compared with those in the GenBank database using the BLAST program (http://www.ncbi.nlm.nih.gov/blast/).

Also, in order to verify the LGT of the *rpoB* gene between the *M. parascrofulaceum* and the three *M. yongonense* strains (MOTT-12, MOTT-27, and DSM 45126^T), the full *rpoB* gene sequences (3,447 or 3,450 bp) and sequences from five other targets [16S rRNA (1,383 or 1,395 bp), *hsp65* (603 bp), *dnaJ* (192 bp), *recA* (1,053 bp), and *sodA* (501 bp)] of the four clinical and three reference stains were analyzed. The PCR

amplification and sequence analysis of the *rpoB* (partial and complete), 16S rRNA, *hsp65*, *dnaJ*, *recA*, and *sodA* genes were performed as described previously [14, 51-54]. A total of six primer sets were used for the amplification of the full *rpoB* gene (3,447 or 3,450 bp) sequence. The locations and sequences of the primers for the *rpoB* amplification are shown in Figure 1-1 and Table 1-2, respectively. These primer sets were designed using the whole genome sequence database of *M. intracellulare* ATCC 13950^{T} (GenBank no. ZP_05227774) and *M. avium* 104 (GenBank no. NC_008595). The sequences of the primers for the amplification and sequencing of the *rpoB* (partial and complete), 16S rRNA, *hsp65*, *dnaJ*, *recA*, and *sodA* genes are also shown in Table 1-2.

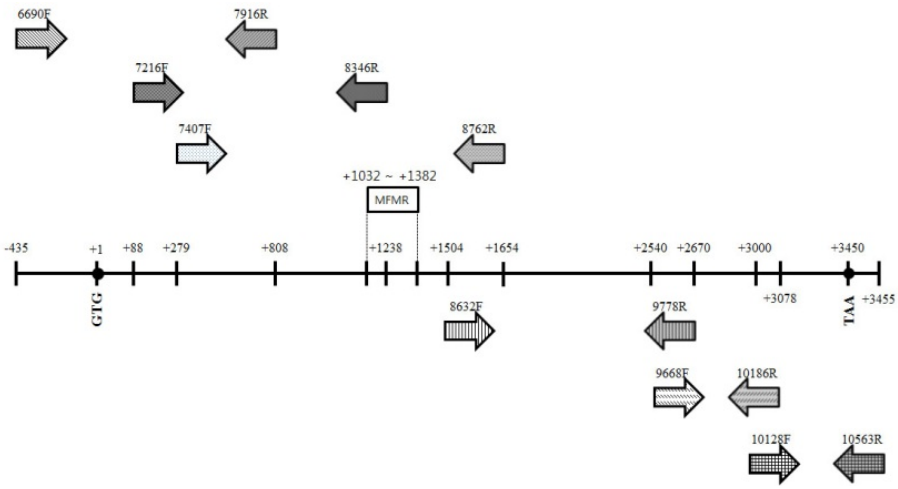


Figure 1-1. Locations of primers used for amplification of the full *rpoB* (3,450 bp) gene sequence in this study. All the primer sequences are shown in Table 1-2.

Table 1-2. The primer sets used for PCR amplification and sequencing in this study.

Genes (Aligned sequence length)	Primer	Primer sequences (5' to 3')	Amplicon size (bp)	Used study
16S rRNA (1,383 or 1,385 bp)	285	GAG AGT TTG ATC CTG GCT CA	1,525	1, 2 [a]
	261	AAG GAG GTG ATC CAG CCG CA		
hsp65 (401 bp)	Tb11	ACC AAC GAT GGT GTG TCC AT	441	1
	Tb12	CTT GTC GAA CCG CAT ACC CT		
rpoB (701 bp)	MycoF	GGC AAG GTC ACC CCG AAG GG	711	1
	MycoR	AGC GGC TGC TGG GTG ATC ATC		
ITS-1 (298 bp)	ITS1 F	TTG TAC AAC ACC GCC CGT C	439	1
	ITS1 R	TCT CGA TGC CAA GGC ATC		
rpoB (3,447 or 3,450 bp)	6690 F	CTT CGT GTA ACT CGG CGG CCC	1,247	2
	7916 R	AGA ACA GGT TCT CCA GCA GGG		
	7216 F	TTC GAG TGG TTG ATC GGC TCG	1,151	2
	8346 R	AAC TGG GAG AGC TGG CTG GTG		
	7407 F	CAA GGA CAT GAC GTA CGC GGC	1,376	2
	8762 R	AGT CCA CCT CGG ACG ACG GCA		
	8632 F	ATC CAC TAC CTG ACC GCC GAC	1,167	2
	9778 R	GAT GAT GTC CAC CGG CGT GCC		
	9668 F	TGG CCC AGA AGC GCA AGA TCT	539	2
	10186 R	CTG CTG GGT GAT CAT CGA GTA		
	10128 F	CAT CAT GAA GCT GCA CCA CCT	456	2
	10563 R	AGT GAT TAA GCC AGG TCC TCA		
rpoB (306 bp)	MF	CGA CCA CTT CGG CAA CCG	351	2
	MR	TCG ATC GGG CAC ATC CGG		
hsp65 (603 bp)	HspF3	ATC GCC AAG GAG ATC GAG CT	644	2
	HspR4	AAG GTG CCG CGG ATC TTG TT		
dnaJ (192 bp)	dnaJ F	GGG TGA CGC G(G/A)C ATG GCC CA	236	2
	dnaJ R	CGG GTT TCG TCG TAC TCC TT		
sodA (501 bp)	SodlgF	GAA GGA ATC TCG TGG CTG AAT AC	547	2
	SodlgR	AGT CGG CCT TGA CGT TCT TGT AC		
recA (1,053 bp)	RecF1	GGT GTT CGN CTA NTG TGG TG	703	2
	RecR1	AGC TGG TTG ATG AAG ATY GC		
	RecG1	CTS GAR ATC GCC GAC ATG CTG	596	2
	RecR2	TTG ATC TTC TTC TCG ATC TC		
	Rec3288F	CAA GCA GGC CGA GTT CGA CAT C	319	2
	Rec3575R	AGG ATC CTG CGC CTG CTC		

[a] 1, Identification and description of 05-1390^T as *M. yongonense*; 2, MLSA assay.

7. Sequence analysis

The four determined target sequences (*rpoB*, 16S rRNA, *hsp65* and ITS1) of the strain 05-1390^T and 6 target gene sequences [*rpoB* (partial and complete), 16S rRNA, *hsp65*, *dnaJ*, *recA*, and *sodA* genes] for MLSA anaylsis were used for multiple alignments with other mycobacterial reference strains using the multiple-alignment algorithm of the MegAlign program as previously described [51, 53]. Phylogenetic trees based on the four independent gene and six MLSA target gene sequences were constructed with the neighbor-joining [55] and maximum parsimony [56] methods. Evolutionary distance matrices were generated according to the Jukes-Cantor model [57]. The MEGA 4.0 program [58] was used for all phylogenetic analyses, and construction of the neighbor-joining trees was evaluated in each case by bootstrap analysis based on 1,000 replicates [59].

8. Nucleotide accession numbers

The sequences of the partial 16S rRNA (1,395 bp), *hsp65* (441 bp), *rpoB* (701 bp) and ITS1 (298 bp) genes from 05-1390^T strain and the sequences of the seven target genes [*hsp65* gene (603 bp), 16S rRNA (1,383 or 1,395 bp), the *rpoB* gene (306 bp), the full *rpoB* gene (3,447 or 3,450 bp), *dnaJ* (192 bp), *recA* (1,053 bp), and *sodA* (501 bp)] from a total of seven strains, including three reference (*M. intracellulare* ATCC 13950^T, *M. parascrofulaceum* ATCC BAA-614^T, and *M. yongonense* DSM 45126^T) and four clinical (MOTT-01, MOTT-02, MOTT-12, and MOTT-27) strains were deposited in the GenBank database; the GenBank numbers are presented in Table 1-3.

Table 1-3. GenBank accession numbers

Strains	*rpoB* (306 bp)	*rpoB* (701 bp)	*rpoB* (3,450 bp)	*hsp65* (401 bp)	*hsp65* (603 bp)	16S rRNA (1,385 bp)	*dnaJ* (192 bp)	*recA* (1,053 bp)	*sodA* (501 bp)	ITS-1 (439 bp)
M. yongonense	JF271806	JF738053	JF271806	JN605801	JF271809	JF271817	JQ937025	JQ937024	JQ937023	JF738054
M. intracellulare	AF057472	-	JQ411539	-	JF271810	JF271818	JQ411546	JQ411525	JQ411518	-
M. parascrofulaceum	JF271829	-	JQ411538	-	JF271807	JF271815	JQ411545	JQ411524	JQ411517	-
MOTT-01	JF271835	-	JQ411533	-	JF271814	JF271822	JQ411540	JQ411519	JQ411512	-
MOTT-02	JF271830	-	JQ411534	-	JF271808	JF271816	JQ411541	JQ411520	JQ411513	-
MOTT-12	JF271833	-	JQ411535	-	JF271812	JF271820	JQ411542	JQ411521	JQ411514	-
MOTT-27	JF271832	-	JQ411536	-	FJ849777	JF271819	JQ411543	JQ411522	JQ411515	-

RESULTS

Morphology and biochemical characteristics of the strain 05-1390^T

Acid-fast and generally rod-shaped bacteria with frequently curved bacilli (0.5-3μm length) having no spores or filaments were observed by microscopy and SEM (Figure 1-2). The optimal growth temperature was 37 °C; however, no growth was observed at 45 °C. On Middlebrook 7H10 agar plates, mature colonies were detected in more than seven days. The growth rate of this strain is similar to that of *M. intracellulare* ATCC 13950^T. The growth of primarily smooth and non-pigmented colonies was observed on Middlebrook 7H10 agar plates under both dark and photo-induction conditions. However, some yellow colonies were also found. This strain showed tolerance to 10 mg TCH ml^{-1} and 500 mg PNB ml^{-1}, but no growth on media containing 5% NaCl, picric acid, ethambutol and on MacConkey agar plate without crystal violet. Negative responses were observed for niacin accumulation, urease activity, nitrate reductase and tween 80 hydrolysis (within 5 days). In addition, the activities of catalase, arylsulfatase (3 and 14 days), tellurite reductase and pyrazinamidase showed positive responses. The biochemical characteristics of strain 05-1390^T were similar to those of *M. intracellulare* ATCC 13950^T. Lipase (C14) and valine arylamidase tests were negative while alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase and cysteine arylamidase tests were positive. These enzymatic activities were similar to *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T, apart from the cysteine arylamidase activity. However, strain 05-1390^T showed positive growth in 7H9 broth with ADC adjusted to pH 5.5, which may be

used for differentiation between strain 05-1390^T and *M. intracellulare* ATCC 13950^T. Details from a comparison of the biochemical profiles and cultural characteristics of 05-1390^T, *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T are shown in Table 1-4. The results of the antimicrobial susceptibility test showed that strain 05-1390^T was more susceptible to nearly all of the antibiotics, except for amikacin (8 μg ml^{-1}) and sulfamethoxazole (32 μg ml^{-1}), compared with *M. intracellulare* ATCC 13950^T. These differences in antimicrobial susceptibility may be useful for clinical diagnoses. The results of the antibiotic susceptibility test are shown in Table 1-5.

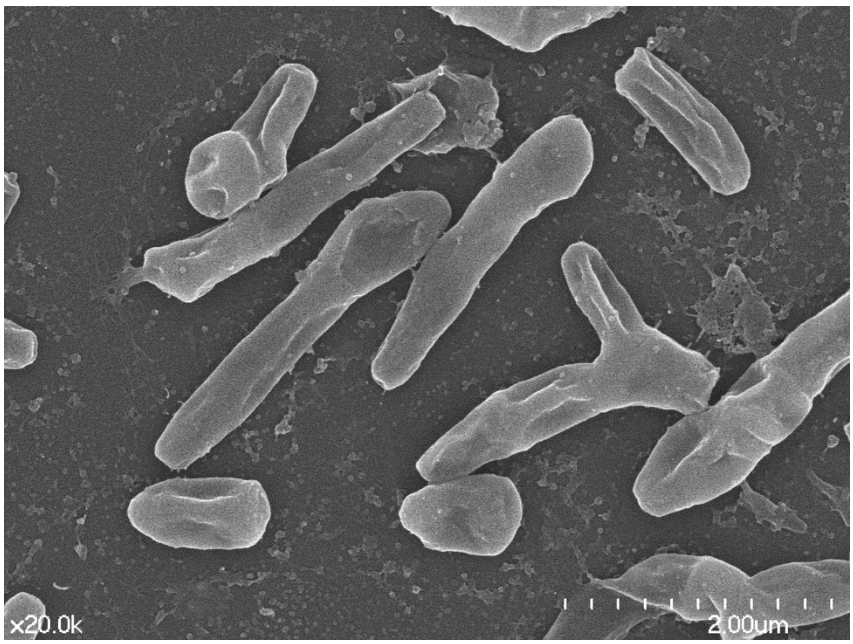


Figure 1-2. Scanning electron micrograph (SEM) image of the strain 05-1390^T. SEM image showed that the morphology – long rod shaped and 3~5 μm in length – of the strain 05-1390^T.

Table 1-4. Cultural and biochemical characteristics that differentiate 05-1390^T and *M. intracellulare*-related strains: 1, 05-1390^T; 2, *M. intracellulare* ATCC 13950^T; 3, *M. chimaera* JCM 14737^T; 4, *M. colombiense* JCM 16228^T. +++, Strong growth; ++, good growth, +, positive/growth; −, negative/no growth; ±, variable. All strains showed negative results in niacin, picric acid and tween hydrolysis (< 5 days) tests, and positive results in pyrazineamidase and catalase tests.

Characteristics	**1**	**2**	**3**	**4**
Growth on 7H10 agar plate at:				
25 ℃	+	+	+	+
37 ℃	++	++	+	+
45 ℃	–	–	–	–
Growth detectable after:				
< 7 days	–	–	–	–
> 7 days	++	++	++	+
Morphology *	SWY	IWY	SWY	IWY
Pigmentation †	N	N	N	N
Nitrate reductase	–	–	–	–
Arylsulfatase				
3 days	±	±	–	–
14 days	+	+	+	±
Tellurite reductase	+	+	–	+
Tween hydrolysis				
> 10 days	±	–	±	+
Urease	–	–	–	+
Growth with:				
10 mg TCH ml $^{-1}$	++	+	++	+
500 mg PNB ml -1	+	+	±	–
5% NaCl	–	–	–	–
Ethambutol	–	–	±	–
7H9 broth (pH 5.5)	+	–	+	+
Growth on:				
MacConkey agar	–	–	–	–
API ZYM kit				
Alkaline phosphate	+	+	+	+
Esterase (C4)	+	+	+	+
Esterase lipase (C8)	+	+	+	+
Lipase (C14)	–	–	–	–
Leucine arylamidase	+	+	+	+
Valine arylamidase	–	–	–	–
Crystine arylamidase	+	–	–	±

* S, smooth; I, Intermediate; Y, Yellow; W, White.

† N, Non-photochromogenic

Table 1-5. Comparison of the antibiotic susceptibility between the strain 05-1390^T and *M. intracellulare* ATCC 13950^T. Ami, Amikacin; Cef, Cefoxitin; Cip, Ciprofloxacin; Cla, Clarithromycin; Dox, Doxycycline; Imi, Imipenem; Mox, Moxifloxacin; Rif, Rifampin; Sul, Sulfamethoxazole; Tob, Tobramycin.

Strains	MIC (μg/ml)									
	Ami	Cef	Cip	Cla	Dox	Imi	Mox	Rif	Sul	Tob
05-1390^T	8	32	0.5	≤0.5	≤0.25	4	≤0.125	≤0.125	32	≤0.25
M. intracellulare	4	64	1	≤0.5	16	≥64	2	0.25	16	8

Fatty acid methyl esters analysis

When using the MI7H10 3.80 database and the MYCO6 method, the ID result showed that 05-1390^T belonged in the *M. avium/intracellulare* group (0.531). The predominant fatty acids of 05-1390^T were $C_{16:0}$ (32.56%) and $C_{18:1}$ ω-9-cis (19.84%). The fatty acid pattern was composed of unbranched saturated and unsaturated fatty acid esters with chains of carbon atoms of various lengths. Tuberculostearic acid ($C_{18:0}$ 10-ME TBSA; 5.33%) was also detected. The detailed fatty acid patterns of 05-1390^T are shown in supplementary Table 1-6.

Table 1-6. Composition of fatty acid methyl esters derived from biomass of 05-1390^T. Examples of abbreviation – 16:0, hexadecanoic acid (palmitic acid); 18:1 cis-9, cis-9-octadecanoic acid (oleic acid); 18:0 10-ME TBSA, 10-methylocatadecanoic acid (tuberculostearic acid).

Fatty acids	Composition (%) [a]
14:0	6.27
16:1 ω-9-cis	1.24
16:1 ω-7-cis	2.61
16:1 ω-6-cis	8.47
16:0	32.56
17:1 ω-7-cis/18-OH/ 17:1 ω-6-cis/17:0 cyclopropane	2.22
18:1 ω-9-cis	19.84
18:1 ω-7-cis	1.30
18:0	1.98
18:0 10-ME TBSA	5.33
20:0-ALC 18.838 ECL/20-OH/ 19:0 cyclopro ω-10-cis/19:0 cyclopro ω-8-cis	16.59

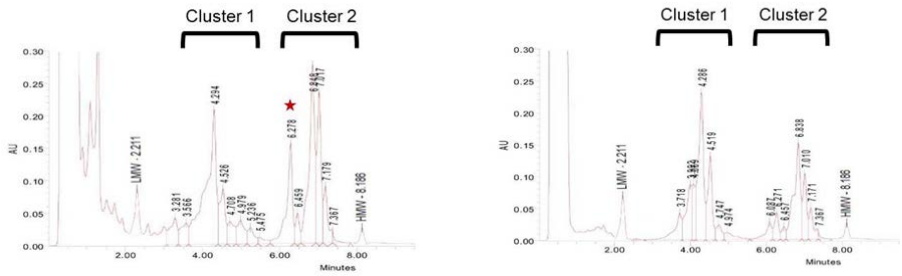
[a], The values are average of the results from independently quardruplicated tests.

HPLC analysis of mycolic acids

The HPLC profiles of strain 05-1390^T showed two clusters of peaks, one early and one late, similar to the other mycobacteria compared here. However, a prominent peak around 6.2 to 6.4 min allowed for differentiation of strain 05-1390^T from *M.*

intracellulare ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T (Figure 1-3).

(A) 05-1390^T (B) *M. intracellulare* ATCC 13950^T



(C) *M. chimaera* JCM 14737^T (D) *M. colombiense* JCM 16228^T

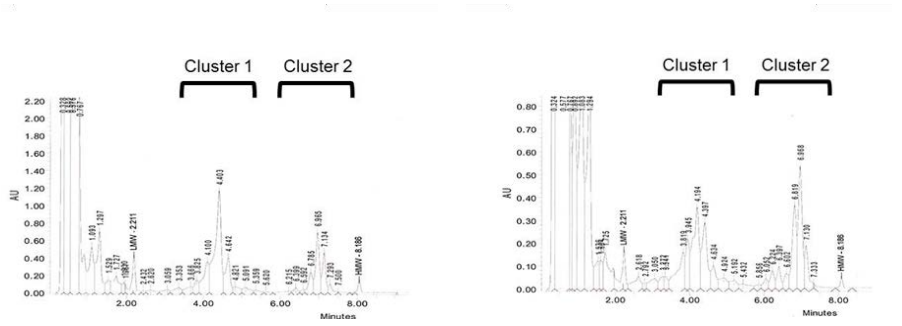
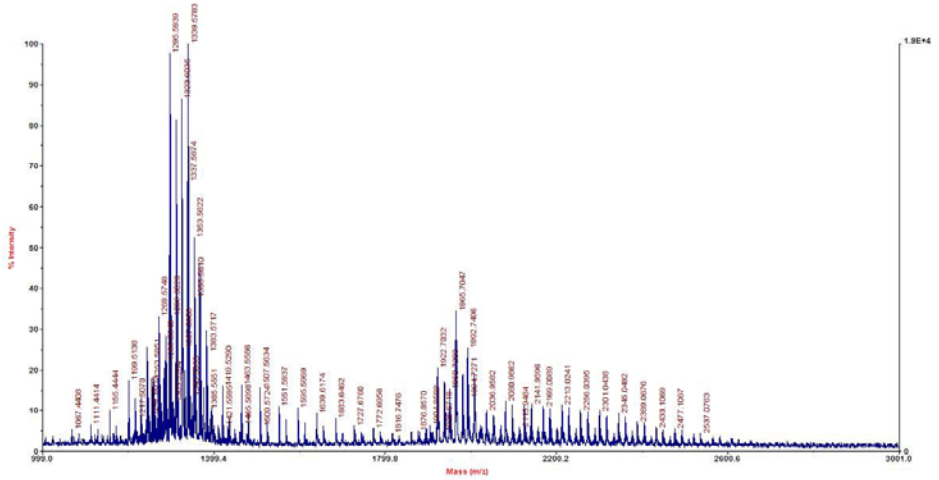


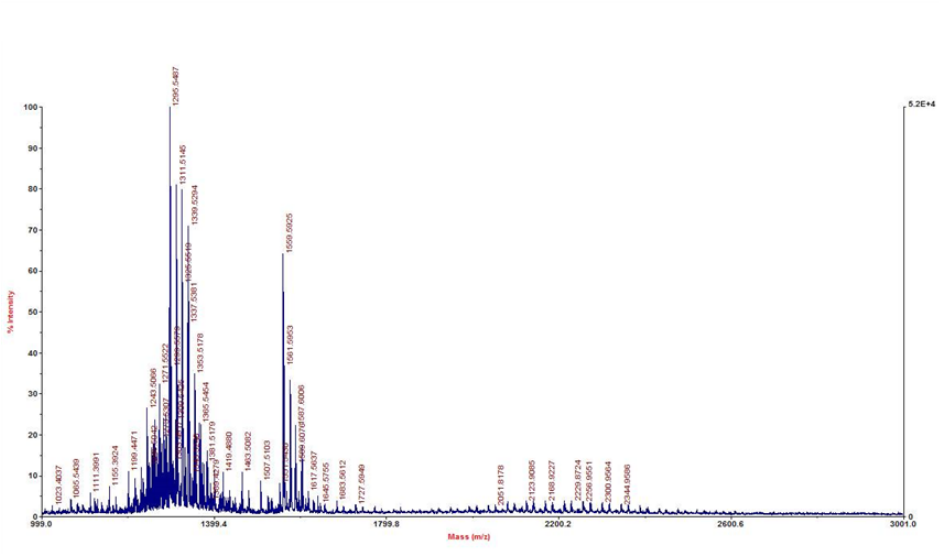
Figure 1-3. HPLC profiles of purified mycolic acids. Comparison of mycolic acid profiles of (A) strain 05-1390^T, (B) *M. intracellulare* ATCC 13950^T, (C) *M. chimaera* JCM 14737^T, and (D) *M. colombiense* JCM 16228^T species obtained from HPLC analysis. The relative retention time is indicated for each peak. LMW, Low-molecular-weight standard; HWM, High-molecular-weight standard. The asterisk means a 05-1390^T-specific HPLC peak.

MALDI-TOF MS analysis

The MALDI-TOF MS profiles of the lipids of the strain 05-1390^T, *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, and *M. colombiense* JCM 16228^T showed different patterns. Each mass spectrum exhibited a cluster of peaks ranging from *m/z* 1199 to *m/z* 1390, with major peak at *m/z* 1295.6 and *m/z* 1339.6 (05-1390^T), *m/z* 1295.5 (*M. intracellulare* ATCC 13950^T), *m/z* 1309.6 (*M. chimaera* JCM 14737^T), and *m/z* 1295.6 (*M. colombiense* JCM 16228^T). However, 05-1390^T showed a second cluster of peaks from *m/z* 1922 to *m/z* 2036 with a major peak at *m/z* 1965.7 that was not detected from *M. intracellulare* ATCC 13950^T, *M. chimaera* JCM 14737^T, or *M. colombiense* JCM 16228^T (Figure 1-4). The unique MALDI-TOF MS profiles of lipids from 05-1390^T suggest that the strain differed from other types of *M. intracellulare*-related mycobacterial strains.

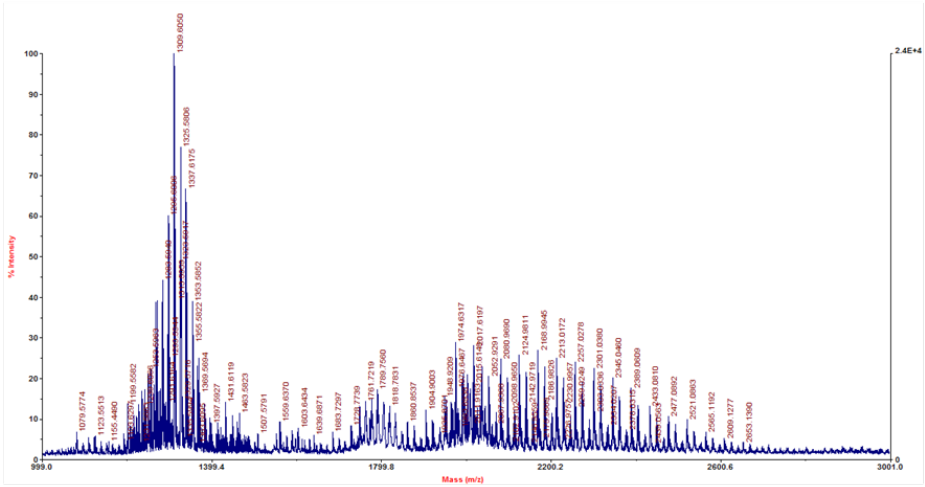
(A) 05-1390^{T}





(B) *M. intracellulare* ATCC 13950^{T}

(C) *M. chimaera* JCM 14737^{T}



(D) *M. colombiense* JCM 16228^{T}

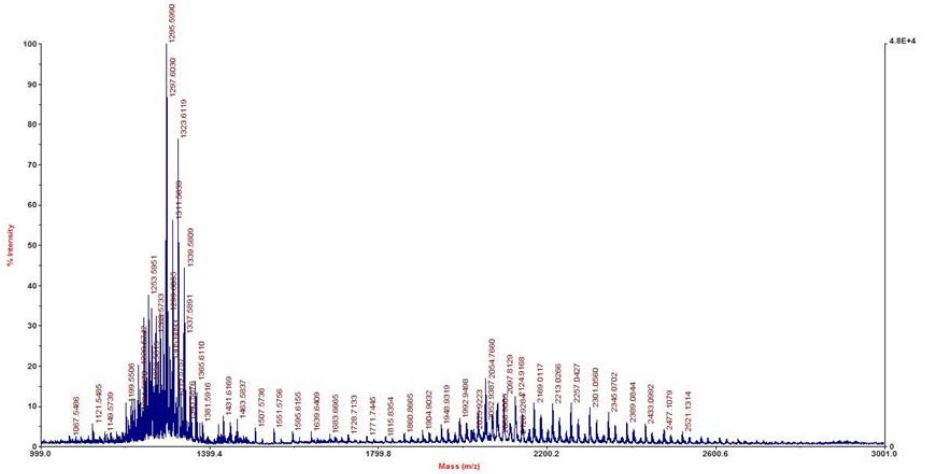


Figure 1-4. Matirx-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectra of lipids from 05-1390^{T}, *M. intracellulare* ATCC 13950^{T}, *M. chimaera* JCM 14737^{T}, and *M. colombiense* JCM 16228^{T}.

Phylogenetic analysis based on sequenced four independent gene sequences

The BLAST results of the 16S rRNA gene sequence (1,385 bp) of strain 05-1390^T showed the closest match (100 %) with the *Mycobacterium avium* complex strain NLA000800397 (GenBank No. EU815938), which was recently isolated in Tanzania [60]. A neighbor-joining tree based on the 16S rRNA gene sequences of the *Mycobacterium* species also indicated the closest relationship between strain 05-1390^T and the *M. intracellulare* related species, *M. marseillense* strain 5356591^T and *M. timonense* strain 5351974^T (Figure 1-5A). A stretch of 1,385 bps from *M. marseillense* strain 5356591^T showed no differences when compared to the 16S rRNA gene sequence of strain 05-1390^T. In a comparison of the 16S rRNA gene sequences of *M. timonense* strain 5351974^T, and *M. chimaera* DSM 44623^T, those of strain 05-1390^T differed by only 1-bp (99.9 % sequence similarity) and 2-bp substitutions (99.9 % sequence similarity), respectively. Compared to the 16S rRNA gene sequence of *M. intracellulare* ATCC 13950^T, 05-1390^T differed by 3-bp substitutions (99.8%). All of the sequence polymorphisms of 05-1390^T, *M. intracellulare* ATCC 13950^T, *M. chimaera* DSM 44623^T, and the *M. timonense* strain 5351974^T were found only in the hypervariable regions, A and B, of the 16S rRNA gene (Figure 1-6).

Phylogenetic trees based on the partial *hsp65* sequences (441 bp) also supported the grouping of strain 05-1390^T within the *M. intracellulare*-related species group, as shown in the 16S rRNA-based tree. However, in this case, the 05-1390^T strain was clustered with the *M. intracellulare* ATCC 13950^T. Strain 05-1390^T shared highest *hsp65* gene sequence similarity with *M. intracellulare* ATCC 13950^T (99.3%) (Figure

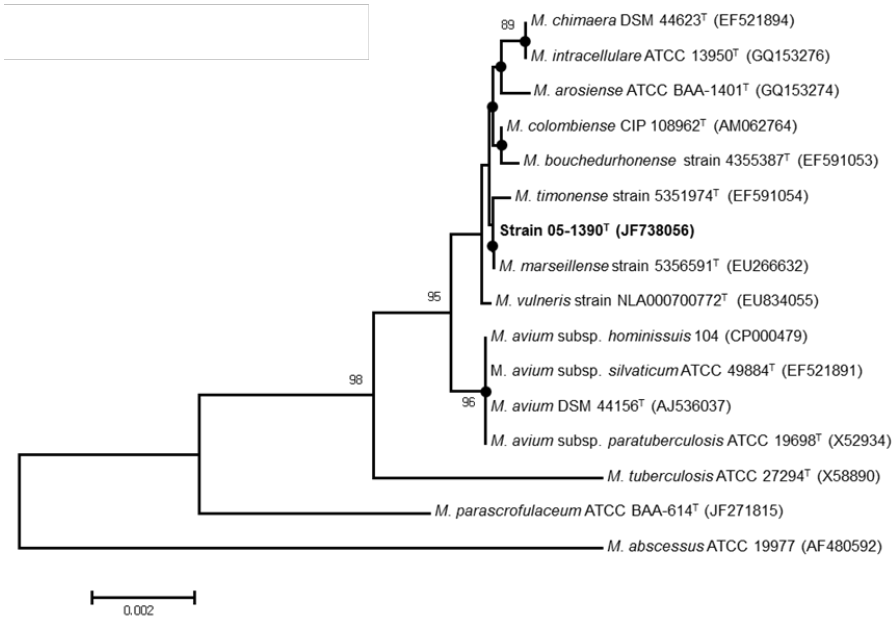
1-5B). In deduced Hsp65 amino acid sequences (134 a. a.), 100 % sequence similarity was found between MAC strains and strain 05-1390^T

Recently, phylogenetic analysis based on the partial *rpoB* gene sequence (701 bp) has proven to be valuable for use in species differentiation between MAC strains. A sequence similarity of 99.3 % was suggested as the criterion of species separation between MAC strains [46]. Therefore, to determine whether strain 05-1390^T meets this standard as a novel MAC species, the partial *rpoB* gene sequence of strain 05-1390^T was subjected to the *rpoB*-based phylogenetic analysis. Surprisingly, the *rpoB* gene sequence of strain 05-1390^T showed more than 6 % divergence compared with those of strains within MAC. Comparison with the GenBank database using the BLAST program showed that the *rpoB* gene sequence of strain 05-1390^T showed the best match (99.4% sequence similarity) with those of *M. parascrofulaceum* ATCC BAA-614^T (GenBank accession no., NZ_ADNV01000269). In fact, when we performed 701-bp *rpoB* gene sequence analysis of *M. parascrofulaceum* ATCC BAA-614^T, our *rpoB* phylogenetic data showed that strain 05-1390^T formed a tight cluster with *M. parascrofulaceum*, but not with other MAC strains, which is strongly supported by high bootstrap values (Figure 1-5C). Sequence analysis of 234 deduced RpoB amino acids showed the strain 05-1390^T had 100 % similarity with *M. parascrofulaceum*, but only 94% similarity with other MAC strains (data not shown), strongly supporting the above notion. These data suggest a lateral gene transfer event of the *rpoB* gene from *M. parascrofulaceum* to the 05-1390^T genome.

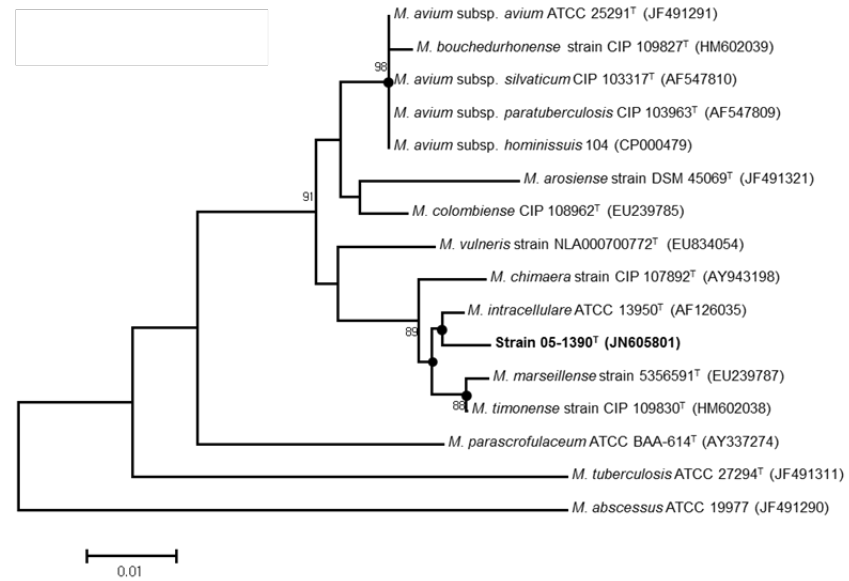
ITS1 has also been reported to be a good genetic marker for phylogenetic separation between MAC members [61]. Our phylogenetic result showed that the ITS1 sequence of strain 05-1390^T was identical to those of the MAC-C genotype (Figure 1-5D).

Also, phylogenetic analysis based on the concatenation (2,827 bp) of the four genes [*hsp65* (441 bp) + *rpoB* (701 bp) + ITS1 (280 bp)+ 16S rRNA (1,395 bp)] was performed, showing that strain 05-1390^{T} formed an independent branch apart from clusters belonging to MAC strains, and the topology of which was strongly supported by high bootstrapping values (Figure 1-5E).

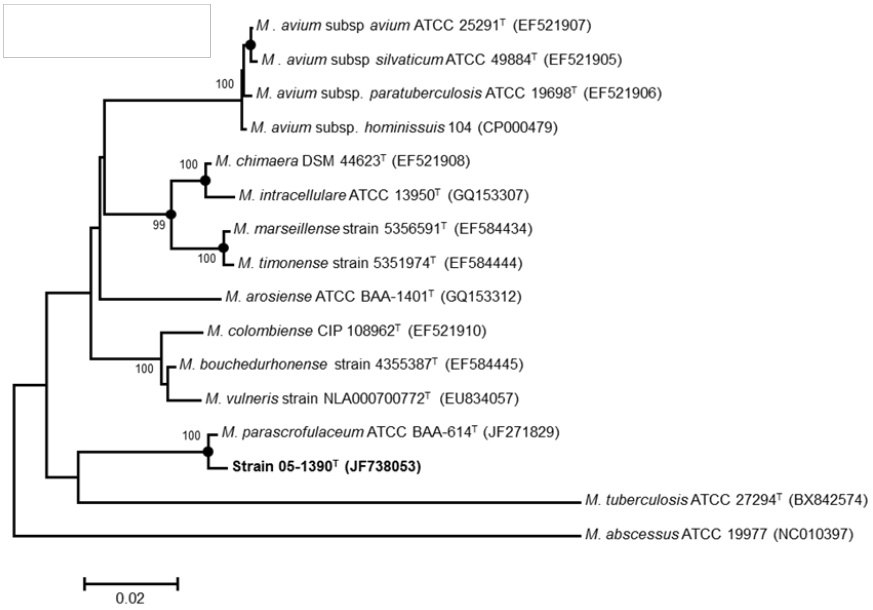
(A) 16S rRNA gene tree



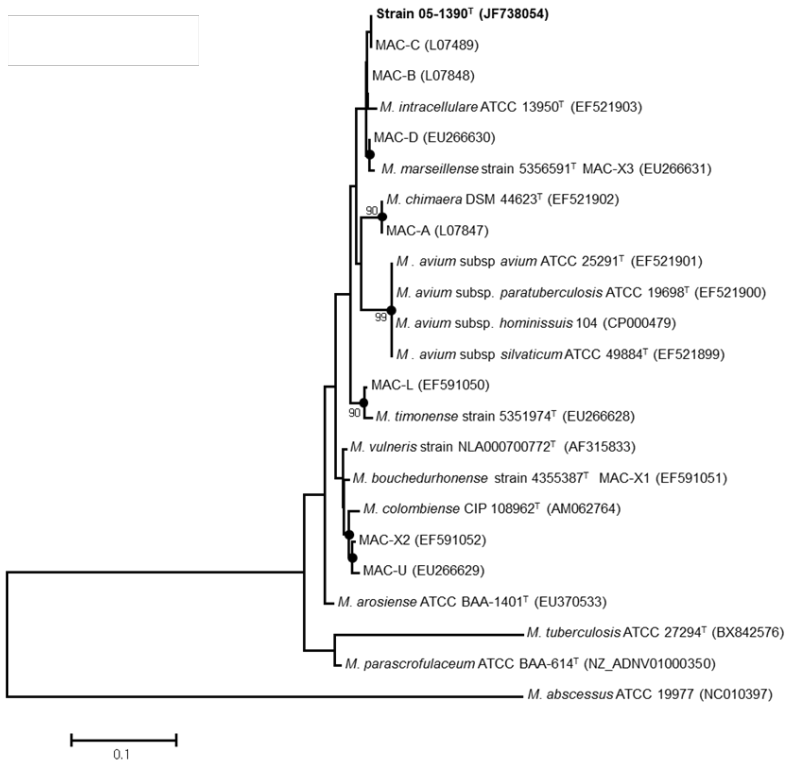
(B) *hsp65* gene tree



(C) *rpoB* gene tree



(D) ITS 1 gene tree



(E) Concatenated tree

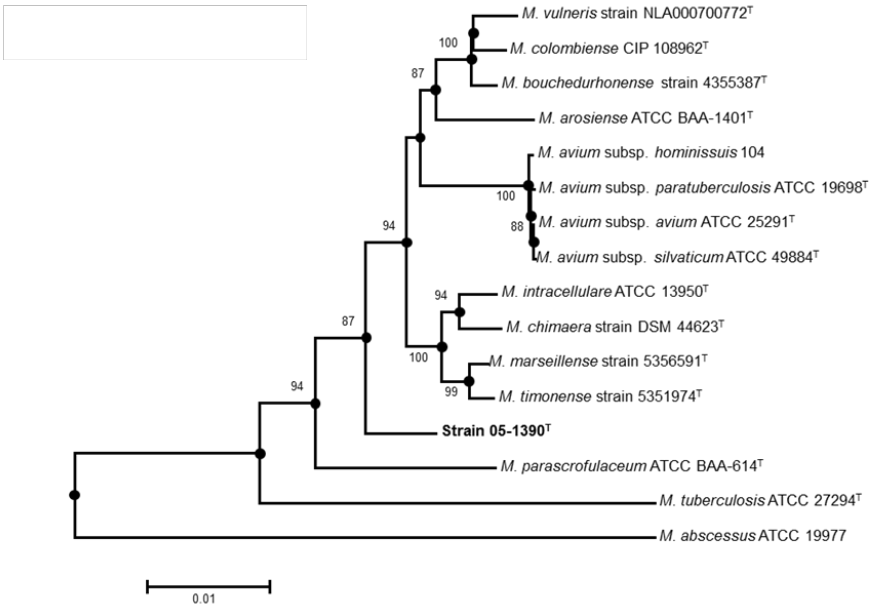


Figure 1-5. Phylogenetic relationships of strain 05-1390^T and other *Mycobacterium* species. (A) 16S rRNA gene (1,385 bp), (B) *hsp65* gene (401 bp), (C) *rpoB* gene (701 bp), (D) internal transcribed spacer (ITS1) (298 bp), and (E) concatenated four genes (16S rRNA, *hsp65*, *rpoB*, and ITS1; 2,768 to 2,785 bp) based phylogenetic trees. These trees were constructed using the neighbor-joining method. The bootstrap values were calculated from 1,000 replications. Bootstrap values of < 50% are not shown. The solid circles indicate that the corresponding groups were supported in the maximum-parsimony analysis. *M. abscessus* ATCC 19977 (NC_010397) was used as

an outgroup in all trees. The bars indicate numbers of substitutions per nucleotide position.

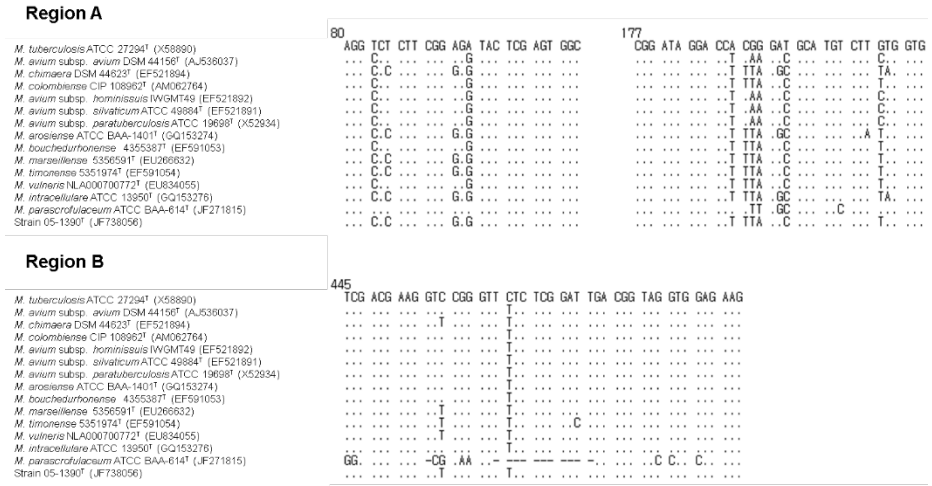


Figure 1-6. Alignments of the hypervariable regions A and B of the 16S rRNA gene sequences of strain 05-1390^{T} and other *M. avium* complex-related strains. Nucleotide positions are according to the 16S rRNA gene sequence of *Escherichia coli*. Only base pairs that differ from *M. tuberculosis* are shown. Except for the sequence of 05-1390^{T}, others were retrieved from GenBank database.

Description of *Mycobacterium yongonense* sp. nov.

Mycobacterium yongonense (yon.gon.en´se. M.L. neut. adj. *yongonense*, indicating yongon-dong, the location of the department performing the taxonomic research on the species).

The organism generally shows a rod-shaped morphology, with frequent curved acid-alcohol fast bacilli. Spores and filaments are not present. The optimal growth temperature is 37 °C. It can grow at 25 °C but not at 45 °C. On the Middlebrook 7H10 agar medium, mature colonies develop in more than seven days. Colonies grown on Middlebrook 7H10 agar reveal a smooth morphology and no pigmentation under both dark and photo-induced conditions. This strain showed tolerance to 10 mg/ml TCH and 500 mg/ml PNB but showed no growth in media containing 5% NaCl and positive or variable responses for arylsulfatase, tellurite reductase, and tween 80 hydrolysis. This strain showed negative responses for niacin accumulation, urease activity, and nitrate reductase activity. Strain 05-1390^T showed positive growth on 7H9 broth with ADC adjusted to pH 5.5, which may be used for differentiation between strain 05-1390^T and *M. intracellulare*. The MALDI-TOF MS profiles of lipids from 05-1390^T exhibited two clusters of peaks, the first ranging from *m/z* 1199 to *m/z* 1390 with major peaks at *m/z* 1295.6 and *m/z* 1339.6 and the second ranging from *m/z* 1922 to *m/z* 2036 with a major peak at *m/z* 1965.7. Genetically, a phylogenetic analysis based on the 16S rRNA and *hsp65* sequences showed that strain 05-1390^T was closely related to *M. intracellulare* within the slowly growing mycobacteria cluster. The ITS1 sequences are identical to those of sequevar MAC-C. Uniquely, an *rpoB* gene-based analysis indicated that it is distantly related to MAC members. Sequences of the 16S rRNA (JF738056), *hsp65* (JN605801), ITS1 (JF738054) and *rpoB* genes (JF738053) of strain 05-1390^T have been deposited in GenBank. The type strain is 05-1390^T (= DSM 45126^T = KCTC 19555^T), isolated from human sputum in Seoul, Korea.

Characterization of the phenotypic traits of the two *M. yongonense* clinical strains (MOTT-12 and MOTT-27) based on conventional biochemical tests.

The conventional taxonomic approaches based on biochemical traits demonstrated that all strains shared similar growth patterns. Pigmentation is known to be the most pronounced difference between *M. intracellulare* and *M. parascrofulaceum*; the former is a nonphotochromogen; however, the latter is a scotochromogen [62]. The two *M. yongonense* clinical strains in the current study (MOTT-12 and MOTT-27) proved to be nonchromogens, suggesting that they are phenotypically closer to *M. intracellulare,* rather than *M. parascrofulaceum.* However, the differences in some biochemical traits such as nitrate reductase, arylsulfatase and tellurite reductase were found between *M. yongonense* DSM 45126^T and the two clinical strains (MOTT-12 and MOTT-27) (Table 1-7).

Table 1-7. Comparison of phenetic and biochemical characteristics between *M. yongonense* DSM 45126^T, MOTT-12, MOTT-27, *M. parascrofulaceum* ATCC BAA-614^T and *M. intracellulare* ATCC 13950^T. 1, *M. yongonense* DSM 45126^T; 2, MOTT-12; 3, MOTT-27; 4, *M. parascrofulaceum* ATCC BAA-614^T; 5, *M. intracellulare* ATCC 13950^T All strains showed negative results in niacin accumulation test, and positive results in heat stable catalase test.

Characteristics	**1**	**2**	**3**	**4**	**5**
Growth on 7H10 agar plate at:					
25 ℃	+	+	+	–	+
37 ℃	++	++	+	+++	++
45 ℃	–	–	–	–	–
Growth detectable after:					
< 7 days	–	–	–	–	–
> 7 days	++	+++	+++	+++	++
Morphology *	SWY	SWY	IWY	SWY	IWY
Pigmentation †	N	N	N	S	N
Nitrate reductase	–	±	±	–	–
Arylsulfatase					

3 days	±	–	–	–	±
14 days	+	–	–	–	+
Tellurite reductase	+	–	–	–	+
Tween hydrolysis					
> 10 days	±	–	–	–	–
Urease	–	–	–	+	–
Growth with:					
10 mg TCH ml $^{-1}$	++	+++	+	+++	+
500 mg PNB ml -1	+	++	+	++	+
5% NaCl	–	–	–	–	–
Growth on:					
MacConkey agar	–	–	–	–	–
Picric acid	–	–	–	–	–

* S, smooth; I, Intermediate; Y, Yellow; W, White.

† N, Non-photochromogenic

Molecular taxonomy of the three *M. yongonense* isolates (MOTT-12, MOTT-27, and DSM 45126^T) via phylogenetic analysis based on full *rpoB* sequences.

In order to prove the hypothesis that there may have been an LGT event for the *rpoB* gene between *M. yongonense* and *M. parascrofulaceum*, the full *rpoB* sequences of seven strains, including the three *M. yongonense* strains [two clinical strains (MOTT-12 and MOTT-27) and one type strain (DSM 45126^T)], were analyzed. All full length *rpoB* sequences obtained in the current study were verified to be encoded in the proper deduced RpoB amino acids in the *in silico* translation. The phylogenetic analysis based on the full *rpoB* sequences (3,450 bp) demonstrated that the three *M. yongonense* isolates (MOTT-12, MOTT-27, and DSM 45126^T) formed a tight cluster with the *M. parascrofulaceum* strains (ATCC BAA-614^T and MOTT-01) rather than with the *M. intracellulare* strains (ATCC 13950^T and MOTT-02). Also their phylogenetic relationship was supported by a high bootstrap value (100.0; Figure 1-7). The sequence similarity value of the full *rpoB* sequences between the three *M. yongonense* strains and two *M. parascrofulaceum* strains ranged from 99.7 % to 99.8 %, which presented eight to nine base pair mismatches among 3,450 bp. However, the sequence similarity values between the three *M. yongonense* strains and two *M. intracellulare* strains

ranged from 94.7 % to 94.9 %, which presented 181 to 196 bp mismatches from 3,450 bp (Table 1-8). The high similarity value observed between the *M. yongonense* and *M. parascrofulaceum* strains indicates that these two different species share almost identical *rpoB* sequences. Furthermore, the similarity values observed between the *M. yongonense* and *M. intracellulare* strains are lower than that of the cut-off value (97.0 %) for the delineation of bacterial species [63] (Table 1-8).

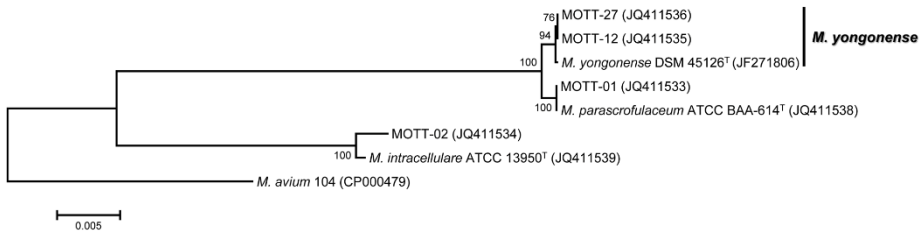


Figure 1-7. Phylogenetic relationships based on the full *rpoB* gene (3,447 or 3,450 bp) sequences. This tree was constructed using the neighbor-joining method. The bootstrap values were calculated from 1,000 replications; bootstrap values of < 50% are not shown.

Table 1-8. Full *rpoB* gene sequence (3,447 and 3,450 bp; right upper side) and concatenated sequence [16S rRNA (1,383 or 1,395 bp) + *hsp65* (603 bp) + *sodA* (501 bp) + *recA* (1,053 bp) + *dnaJ* (192 bp); left down side] similarities between seven mycobacterial strains.

Strains	Sequence similarity (%)							
	MOTT-01	MOTT-02	MOTT-12	MOTT-27	Mint [a]	Myon [b]	Mpara [c]	Mav [d]
MOTT-01		94.6	99.8	99.8	94.8	99.7	100.0	94.6
MOTT-02	94.3		94.7	94.7	99.7	94.7	94.7	95.6
MOTT-12	94.2	99.7		100.0	94.8	100.0	99.8	94.5
MOTT-27	94.2	99.7	100.0		94.8	100.0	99.8	94.5
Mint	94.4	99.9	99.7	99.7		94.9	94.8	95.7
Myon	94.1	99.5	99.6	99.6	99.5		99.7	94.5
Mpara	100.0	94.3	94.2	94.2	94.4	94.1		94.6
Mav	94.2	96.8	96.7	96.7	96.9	96.8	94.2	

[a] Mint, *M. intracellulare* ATCC 13950^T

[b] Myon, *M. yongonense* DSM 45126^T

[c] Mpara, *M. parascrofulaceum* ATCC BAA-614^T

[d] Mav, *M. avium* 104

Phylogenetic analysis based on the 16S rRNA and *hsp65* gene sequences

In order to verify the above hypothesis, a phylogenetic analysis of the three *M. yongonense* strains was performed using two other genes (16S rRNA and *hsp65* genes), which have been used widely for mycobacteria taxonomies and diagnostics [42, 53, 64, 65]. The phylogenetic analysis based on the 16S rRNA sequence (1,383 or 1,395 bp) indicated that the three *M. yongonense* strains belonged to the *M. intracellulare* group, exhibiting a sequence similarity ranging from 99.8% to 100% with two other *M. intracellulare* strains (ATCC 13950^T and MOTT-02; data not shown). The three *M. yongonense* strains exhibited a relatively low level of similarity value (96.8 %) with the *M. parascrofulaceum* strains, which was lower than the universally accepted cut-off value for the 16S rRNA gene (97.0%) for bacteria species delineation (data not shown) [66]. This strongly suggests that the three *M. yongonense* strains are phylogenetically related to *M. intracellulare*.

The *hsp65* gene sequence based methods have been the most widely used methods for mycobacteria taxonomies as alternatives to the 16S rRNA based methods [51, 53]. The three *M. yongonense* strains exhibited some minor variations (99.3 % similarity value with four base pair mismatches of the 603 bp *hsp65* sequences) compared with the other two *M. intracellulare* strains (ATCC 13950^T and MOTT-02). The phylogenetic analysis based on the *hsp65* gene sequence (603 bp) indicated that the three *M. yongonense* strains belonged to the *M. intracellulare* group, rather than to the *M. parascrofulaceum* group, which indicates a low level of sequence similarity value of 94.9 % with the two *M. parascrofulaceum* strains (ATCC BAA-614^T and MOTT-01; data not shown). This also strongly supports their phylogenetic location in *M. intracellulare*.

Phylognentic analysis based on *dnaJ* (192 bp), *recA* (1,053 bp), and *sodA* (501 bp)

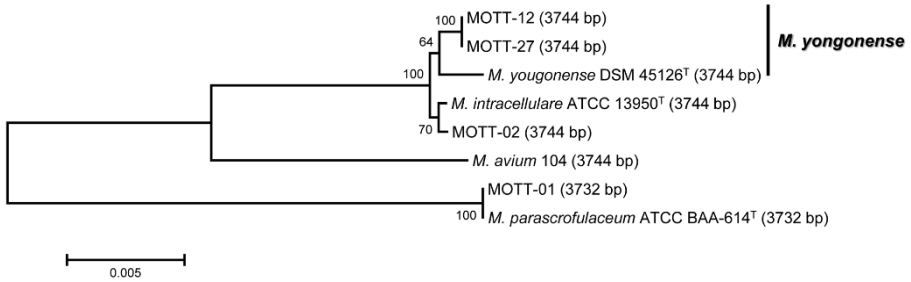
In order to strengthen the above hypothesis, further phylogenetic analyses were performed based on three other genes [*dnaJ* (192 bp), *recA* (1053 bp), and *sodA* (501 bp)], which have been successfully applied in mycobacteria taxonomy [52, 67]. The phylogenetic analyses based on the *dnaJ* gene sequence (192 bp) indicated that the three *M. yongonense* strains belong to the *M. intracellulare* groups and exhibited sequence similarity values of 99.5 % with the other two *M. intracellulare* strains (ATCC 13950^{T}, and MOTT-02). It was also clear that the three *M. yongonense* strains do not belong to the *M. parascrofulaceum* group (similarity value of 93.2 %; data not shown). The phylogenetic analyses based on the *recA* gene sequence (1053 bp) also indicated that the three *M. yongonense* strains belonged to the *M. intracellulare* groups, which exhibited sequence similarity values of 99.4 % to 99.6 % with the other two *M. intracellulare* strains (ATCC 13950^{T} and MOTT-02), rather than with the *M. parascrofulaceum* group (similarity value of 95.4 %; data not shown). The phylogenetic analyses based on the *sodA* gene sequence (501 bp) also indicated that the three *M. yongonense* strains belonged to the *M. intracellulare* groups, which exhibited sequence similarity values of 99.4 % to 99.6 % with the other two *M. intracellulare* strains, rather than with the *M. parascrofulaceum* group (similarity value of 78.8 % to 79.0 %; data not shown). Thus, all phylogenetic analyses based on the other three genes [*dnaJ* (192 bp), *recA* (1053 bp), and *sodA* (501 bp)] also confirmed that the three *M. yongonense* strains with the distinct *rpoB* gene are more closely related to *M. intracellulare* than to *M. parascrofulaceum* as shown in the phylogenetic analyses based on the 16S rRNA and *hsp65* genes.

Phylogenetic analysis based on concatenated sequences of the five MLSA genes and the full *rpoB* gene

Figure 1-8A shows the tree for the seven strains obtained by concatenating the sequences of the five MLSA genes (16S rRNA, *hsp65*, *dnaJ*, *recA*, and *sodA*) (3,732 or 3,744 bp). The tree displays two clearly separated clusters: one for the *M. intracellulare* related strains (three *M. yongonense* and two *M. intracellulare*) and the other for the two *M. parascrofulaceum* strains. A high level of bootstrap values (100 %) was observed for the groupings. The three *M. yongonense* and two *M. intracellulare* strains formed two different branches in one cluster, which indicates their phylogenetic separation. The bootstrap values of both branches were 64% (*M. yongonense*) and 70% (*M. intracellulare*). Although a complete sequence similarity between the two clinical strains (MOTT-12 and MOTT-27) was found, some variations (99.6 % of 3744-bp MLSA sequences) between the clinical strains and the type strain (DSM 45126^T) were found, which indicates that the two clinical strains may be variants of *M. yongonense* DSM 45126^T (Table 1-8). The effect of adding the *rpoB* sequence to the concatenated sequences of the five MLSA genes (7,182 – 7,194 bp) was also studied. The topology of the obtained tree (Figure 1-8B) was radically different from only that constructed from the MLSA gene sequences (Figure 1-8A). The branch of the *M. yongonense* strains forming the same cluster with that of the *M. intracellulare* strains in the MLSA tree were transferred into a cluster belonging to the *M. parascrofulaceum* strains in the MLSA + *rpoB* tree, which was strongly supported with a high level of bootstrap values (100%). The discrepancy observed between the topology structures of both trees

suggests the potential LGT event of the *rpoB* gene from the *M. parascrofulaceum* strain into the *M. yongonense* strain.

(A)



(B)

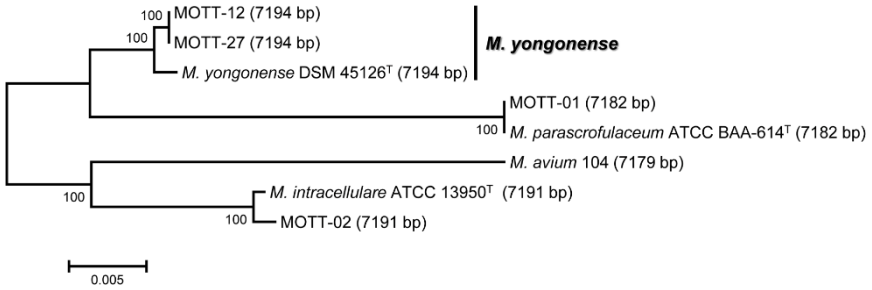


Figure 1-8. Phylogenetic relationships based on concatenated sequences of (A) the five MLSA genes (16S rRNA, *hsp65*, *dnaJ*, *recA* and *sodA*) (3,732 or 3,744 bp) and (B) with the addition of the full *rpoB* gene sequence to the concatenated sequences of the five MLSA genes (7,182 or 7,194 bp). These trees were constructed using the neighbor-joining method. The bootstrap values were calculated from 1,000 replications; bootstrap values of <50% are not shown.

DISCUSSION

In this chapter, I described a slow growing non-chromogenic novel *Mycobacterium* species, strain 05-1390^T. This strain showed similar results with *M. intracellulare* ATCC 13950^T in various biochemical analyses (Table 1-4). However, the strain could grow in the low pH environment and this trait could be used an important criterion to differentiate between the strain 05-1390 and *M. intracellulare*. In the case of HPLC and MALDI-TOF MS analysis using the extracted mycolic acids and lipids, respectively, strain 05-1390 showed unique patterns distinct from those of *M. intracellulare* and other related strains. The 16S rRNA gene sequence (1385 bp) of strain 05-1390^T showed a high degree of similarity to those of the *M. intracellulare* complex, namely *Mycobacterium marseillense* 5351974^T (100 %), *M. intracellulare* ATCC 13950^T (99.8 %) and *Mycobacterium chimaera* DSM 44623^T (99.9 %) (Figures 1-5A and 1-6). Phylogenetic analysis based on internal transcribed spacer 1 (ITS1) and the *hsp65* gene indicated that strain 05-1390 was closely related to *M. intracellulare* ATCC 13950^T, but that it was a distinct phylogenetic entity. Of particular interest, an analysis based on the *rpoB* gene (701 bp) showed that it is closely related to *M. parascrofulaceum* ATCC BAA-614^T (99.4 %), a scotochromogenic strain, rather than to the *M. intracellulare*-related strains. Accumulated results suggested that strain 05-1390 could be differentiated from *M. intracellulare* strains, so the strain proposed as a novel species, *M. yongonense* (=DSM 45126^T=KCTC 19555^T).

To prove the possibility of the LGT events in the *rpoB* gene sequence of *M. yongonense*, MLSA analysis was conducted. The full *rpoB* gene sequence proved useful for the delineation of the bacterial species [63]. A *rpoB* gene sequence similarity of < 97.0 % is reported to be significantly correlated with a DNA-DNA hybridization (DDH) value of < 70%, which is the universal cut-off value for the delineation of a bacterial species [63]. Despite some problems in the bacteria taxonomy, the 16S rRNA gene sequence-based comparisons have been and remain invaluable in describing the prokaryotic diversity; they are indispensable in the delineation of bacterial species [68]. MLSA based concatenated phylogenetic trees with or without full *rpoB* sequence showed that the discrepancy in topology structures of both trees. And this result suggested the potential LGT event in the *rpoB* gene from the *M. parascrofualceum* into the *M. yongonense* (Figure 1-8).

From a clinical perspective, these results emphasize the importance of the MLSA for mycobacteria identification. Currently, the *rpoB* gene has been used widely as a target gene for bacterial identification, particularly for mycobacteria identification [24, 51, 69]. However, the data in this study implies that some strains of *M. yongonense* could be misidentified as *M. parascrofulaceum* when only a single *rpoB* gene is used in the identification or as *M. intracellulare* with use of chronometers other than the *rpoB* gene.

In conclusion, collective consideration of the molecular taxonomic data based on the full *rpoB* and five other genes, which have been used widely for mycobacterial identification has led to the conclusion that the three *M. yongonense* strains with the signature *rpoB* gene have potentially acquired their *rpoB* gene via a very recent LGT event from *M. parascrofulaceum*. However, the details of the LGT events between *M.*

parascrofulaceum and *M. yongonense* strains will be mentioned in the Chapter 3. Furthermore, the data presented here also suggests that the *rpoB* gene analysis alone may have potential for misidentification in mycobacteria diagnostics. Thus, an approach using multi-locus genes should be conducted for mycobacteria identification.

CHAPTER 2

Molecular taxonomic study for two distinct genotypes of *Mycobacterium yongonense* via genome-based phylogenetic analysis

INTRODUCTION

Members of the *Mycobacterium avium* complex (MAC) are the most important nontuberculous mycobacteria (NTM) in terms of clinical and epidemiological aspects [70]. Traditionally, MAC includes two species: *M. avium* and *M. intracellulare* [4-6]. In addition to these two species, recent advances in molecular taxonomy have fueled the identification of novel species within the MAC [9-12, 37].

Our laboratory reported that *M. intracellulare*-related strains from Korean patients showed genetic diversity. This diversity could be used to divide the strains into a total of five distinct groups (INT-1 to -5) via the molecular taxonomic approach using three independent chronometer molecules: *hsp65*, the internal transcribed spacer 1 (ITS1) region and the 16S rRNA sequence [14]. Of these genotypes, the INT-5 strains were distantly related to other genotypes of *M. intracellulare* (INT-1 to -4). In the chapter 1, I introduced a novel species, *M. yongonense*, which is phylogenetically related to *M. intracellulare*. But this species has a distinct RNA polymerase β-subunit gene (*rpoB*) sequence that is identical to that of *M. parascrofulaceum*, suggesting the acquisition of the *rpoB* gene via a potential lateral gene transfer (LGT) event [71]. This strain was clustered into INT-5 group based on the partial gene (*hsp65*, ITS1, and 16S rRNA) sequence analysis [14].

To date, our laboratory have introduced the genome sequences of five *M. intracellulare* strains, one strain of the INT-1 genotype (*M. intracellulare* MOTT-64 [GenBank accession no.,NC_016948]) [32], two strains of the INT-2 genotype (*M. intracellulare* ATCC 13950^{T} and MOTT-02 [GenBank accession no., NC_016946

and NC_016947, respectively]) [30, 33], and two strains of the INT-5 genotype (*M. intracellulare* MOTT-36Y and MOTT-H4Y [GenBank accession no., NC_017904 and AKIG00000000, respectively] [31, 34]. Especially, two INT-5 strains, MOTT-36Y and MOTT-H4Y, showed that they were more closely related to the genome of *M. yongonense* DSM 45126^T than *M. intracellulare* ATCC 13950^T, despite they have *rpoB* sequences identical to *M. intracellulare*, but not *M. parascrofulaceum*.

Recently, Tortoli *et al*. reported pulmonary disease caused by *M. yongonense* strains isolated from patients in Italy [72]. However, this strain notably harbors *rpoB* sequences almost identical to those of *M. intracellulare* but not to those of *M. parascrofulaceum*, suggesting the possibility of the existence of another group of *M. yongonense* strains that were not subject to the LGT event involving the *rpoB* gene from *M. parascrofulaceum*. Furthermore, the potential of its misidentification has recently been proposed [73]. Therefore, the development of a novel diagnostic method for the precies identification of clinical strains of *M. yongonense* is necessary.

Insertion sequence (IS) elements have several traits of great interest in relation to epidemiological evaluations, taxonomic studies and diagnostic purposes. Depending on the degree of mobility and the copy number of IS elements, DNA fingerprints based on Southern blotting and hybridization can be used to infer strain relatedness [74]. For mycobacterial IS elements, it has generally been accepted that genetic rearrangement due to their insertion may frequently be limited to the species or subspecies level [18, 75-77]. The specificity has led to the use of IS elements as markers for mycobacterial diagnosis, such as IS*6110* for the detection of *M. tuberculosis* [76], or IS*900* for the detection of *M. paratuberculosis* [17].

The aims of the present study is to determine the exact taxonomic status of the two INT-5 strains, MOTT-H4Y and MOTT-36Y and *M. yongonense* DSM 45126^{T} strain via genome-based phylogenetic analysis. To this end, I first sequenced the whole genome of *M. yongonense* by various sequencing technologies. Second, using the complete genome sequence of *M. yongonense* and other *M. intracellulare* strains, comparative genome analyses including the comparison of IS elements, multilocus sequencing typing (MLST), and single nucleotide polymorphisms (SNPs) analysis were conducted. Also, I developed a novel real-time PCR method targeting IS-elements for the precise detection of *M. yongonense* strains.

MATERIALS AND METHODS

1. Genome sequencing

The *M. yongonense* DSM 45126^T genome was sequenced using four types of sequencing methods – standard shotgun GS FLX pyrosequencing (770,801 reads), short paired-end GS FLX pyrosequencing (470,728 reads), shotgun clone library Sanger chemistry sequencing (11,211 reads), and fosmid library Sanger chemistry sequencing (822 reads) – to generate scaffolds containing 167 contigs. Sequencing analysis was performed at the National Instrumentation Center for Environmental Management (NICEM, Genome analysis unit) at Seoul National University. A total of 1,253,562 reads were generated, with an average read length of 180 bp, yielding 225,567,303 bp sequences. This represents ~40 x coverage for the estimated 5.5 Mb genome size. All the remaining gaps between contigs were completely filled by ~50 fold Solexa reads and PCR amplifications. Genome annotation was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.gov/genomes/static/Pipeline.html).

2. Genome sequences used in this study

Ten mycobacterial genome sequences, from strains belonging to the M. avium complex (3 *M. intracellulare* strains: ATCC 13950^T, MOTT-02, and MOTT-64; 5 *M. yongonense* strains: DSM 45126^T, MOTT-12, MOTT-27, MOTT-36Y and MOTT-H4Y; one *M. avium* strain: *M. avium* 104; and one *M. parascrofulaceum* strain: *M.*

parascrofulaceum ATCC BAA-614^T) were retrieved from the GenBank database (Table 2-1) and used for comparative genome analysis.

Table 2-1. Whole genome sequences used in this study.

Strains	GenBank No.	Genome size (bp)	G+C ratio	CDS	tRNA	INT-group
M. intracellulare ATCC 13950^T	NC_016946	5,402,402	68.10	5,145	47	INT-2
M. intracellulare MOTT-02	NC_016947	5,409,696	68.10	5,151	47	INT-2
M. intracellulare MOTT-64	NC_016948	5,501,090	68.07	5,251	46	INT-1
M. yongonense DSM 45126^T	NC_020275	5,521,023	67.95	5,222	47	INT-5
M. yongonense MOTT-12	CP015964	5,445,538	68.02	5,157	47	INT-5
M. yongonense MOTT-27	CP015965	5,435,152	68.03	5,041	47	INT-5
M. yongonense MOTT-36Y	NC_017904	5,613,626	67.91	5,128	46	INT-5
M. yongonense MOTT-H4Y	AKIG00000000	5,443,025	68.08	5,020	48	INT-5
M. avium 104	NC_008595	5,475,491	68.99	5,120	46	-
M. parascrofulaceum ATCC BAA-614^T	ADNV00000000	6,564,170	68.5	5,586	47	-

3. Analysis of insertion sequence (IS) elements

Using the sequence information from the mycobacterial genomes, and especially the information on annotated genes, IS elements from each genome were identified. IS elements from *M. yongonense* DSM 45126^T were identified using a default parameter of the IS-finder tool (http://www-is.biotoul.fr) [78] and compared with IS elements from the other mycobacterial genome sequences using the MegAlign (DNASTAR package, http://www.dnastar.com/t-megalign.aspx) [79] and BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) programs. To identify terminal inverted repeat (IR) sequences, upstream and downstream sequences of the targeted *M. yongonense*-

specific IS element were analyzed using the Spectral Repeat Finder tool (http://www.imtech.res.in/raghave/srf/) with default parameters except for a parameter of 'minimum number of copies' (we used this parameter value with 2, the default value was 5) [80].

Also, the identified novel IS element of *M. yongonense* was compared with the IS elements from two other *M. yongonense* strains (MOTT-36Y and MOTT-H4Y) and 16 known IS elements belonging to the IS*21* family. Amino acid or nucleotide sequences were aligned using the ClustalW method in MEGA 4.0 software [58]. Phylogenetic trees were constructed using the neighbor-joining method [55] and maximum composite likelihood (in the case of nucleotide sequences) or number of differences substitution models (in the case of amino acid sequences) in MEGA 4.0 software. Bootstrap values were calculated from 1,000 replications.

4. Complete genome sequence-based phylogenetic analysis

For the phylogenetic analysis of the two INT-5 strains (MOTT-36Y and MOTT-H4Y) and *M. yongonense* DSM 45126^T, their genome sequences were compared with those of other mycobacterial strains. These genome sequences were subjected to whole-genome multiple sequence alignments using the neighbor-joining method [55] by the Mauve Genome Alignments software (http://darlinglab.org/mauve/mauve.html). A phylogenetic tree was generated using the aligned genome sequences and visualized by the TreeViewX program (http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/); additionally, a Venn diagram was constructed by the BLASTCLUST program. The minimum length coverage and identity threshold in BLASTCLUST were 0.9 and 95%, respectively.

5. Phylogenetic analysis based on *rpoB* and 35 selected target gene sequences or single nucleotide polymorphisms (SNPs) of the 35 selected target gene sequences

To analyze the sequence differences among the 3 *M. intracellulare* strains (*M. intracellulare* ATCC 13950^T, *M. intracellulare* MOTT-02, and *M. intracellulare* MOTT-64), 2 INT-5 strains (MOTT-H4Y and MOTT-36Y) and *M. yongonense* DSM 45126^T, the *rpoB* gene and 35 additional gene sequences were selected from the genome sequences. In the selected 35 genes, 9 genes (*argH*, *cya*, *glpK*, *hsp65*, *murC*, *pta*, *recA*, *secA1* and *sodA*) [40, 52, 81, 82] were included for mycobacterial MLST analysis, and other 26 genes were were randomly selected in the housekeeping genes without any standards. The list of chosen genes is as follows: adenylate kinase (*adk*), argininosuccinate lyase (*argH*), chorismate synthase (*aroC*), shikimate 5-dehydrogenase (*aroE*), F0F1 ATP synthase subunit beta (*atpD*), adenylate cyclase (*cya*), cytochrome b6 (*cytB*), ATP-dependent RNA helicase, dead/death box family protein (*deaD*), chromosomal replication initiation protein (*dnaA*), DNA primase (*dnaG*), molecular chaperone DnaK (*dnaK*), chaperone protein (*dnaJ*), 3-oxoacyl-(acyl-carrier-protein) reductase (*fabG*), cell division protein FtsZ (*ftsZ*), fumarate hydratase (*fumC*), malate synthase G (*glcB*), glutamine synthetase type I (*glnA*), glycerol kinase (*glpK*), fructose 1,6-bisphosphatase II (*glpX*), 6-phosphogluconate dehydrogenase (*gnd*), DNA gyrase subunit B (*gyrB*), heat-shock protein 65 kD (*hsp65*), myo-inositol-1-phosphate synthase (*ino1*), NAD-dependent DNA ligase LigA (*ligA*), ATP-dependent DNA ligase (*ligB* and *ligC*), UDP-N-acetylmuramate-L-alanine ligase (*murC*), endonuclease III (*nth*), glucose-6-phosphate isomerase (*pgi*), phosphoglycerate kinase (*pgk*), phosphate acetyltransferase (*pta*), recombinase A

(*recA*), recombination protein F (*recF*), preprotein translocase subunit SecA (*secA1*), and [Mn]-superoxide dismutase (*sodA*) (Table 2-2). The retrieved *rpoB* gene or the concatenated 35 selected gene sequences were multiply aligned using the ClustalW method in the MEGA 4.0 software [58]. Using the multiple alignment matrix, phylogenetic trees were constructed using the neighbor-joining method in the MEGA 4.0 software. The bootstrap values were calculated from 1,000 replications.

SNPs were extracted from the multiple alignments of *rpoB* gene sequences and the 35 selected gene sequences from the 3 *M. intracellulare* strains (*M. intracellulare* ATCC 13950^{T}, *M. intracellulare* MOTT-02, and *M. intracellulare* MOTT-64), 2 *M. intracellulare* INT-5 strains (MOTT-H4Y and MOTT-36Y), *M. yongonense* DSM 45126^{T} and *M. parascrofulaceum* ATCC BAA-614^{T}. Then, the patterns were compared. Additionally, the extracted SNP sequences were concatenated and used to construct a phylogenetic tree as described above.

Table 2-2. Total and *M. yongonense*-group related-SNPs from targeted 35 genes and *rpoB* gene of *M. intracellulare* strains

No.	Genes	Compared nucleotide size (bp)	Total SNPs (n)	*M. yongonense*-group related-SNPs (n)
1	*adk*	534	49	0
2	*argH*	1,431	199	10
3	*aroC*	1,206	134	0
4	*aroE*	888	134	0
5	*atpD*	1,461	164	0
6	*cya*	1,554	207	5
7	*cytB*	1,704	185	0
8	*deaD*	1,704	228	14
9	*dnaA*	1,503	193	11
10	*dnaG*	1,953	250	5
11	*dnaJ*	1,149	124	2
12	*dnaK*	1,860	118	2
13	*fabG*	768	80	0
14	*ftsZ*	1,161	108	0
15	*fumC*	1,407	152	0
16	*glcB*	2,169	183	0
17	*glnA*	1,437	106	1
18	*glpK*	1,527	223	2
19	*glpX*	987	101	0
20	*gnd*	1,521	147	0
21	*gyrB*	2,034	245	4
22	*hsp65*	1,626	103	6
23	*ino1*	1,095	81	2
24	*ligA*	2,082	257	3
25	*ligB*	1,530	167	0
26	*ligC*	1,056	173	3
27	*murC*	1,479	207	0
28	*nth*	792	84	0
29	*pgi*	1,665	203	0
30	*pgk*	1,236	187	0
31	*pta*	2,160	328	3
32	*recA*	1,053	49	0
33	*recF*	1,158	235	18
34	*rpoB*	3,390	185	0
35	*secA1*	2,829	268	3
36	*sodA*	624	118	0

6. Application of SNP analysis to MAC clinical isolates

To confirm the different SNP patterns between the INT-5 strains and other *M. intracellulare* strains (INT-1 or INT-2 strains), five genes of other *M. intracellulare* clinical isolates, which proved to have more *M. yongonense*-group related signature SNPs than others, were amplified and sequenced for further analysis. The five selected genes were *argH*, *dnaA*, *deaD*, *hsp65* and *recF*.

A total of 16 clinical isolates were used in this study. These clinical isolates were collected from the Asan Medical Center (Seoul, Republic of Korea) and Seoul National University Hospital (Division of Pulmonary & Critical Care Medicine, Seoul, Republic of Korea). These strains were identified into genotypes by phylogenetic analysis based on *hsp65*, ITS1 and 16S rRNA gene sequences [14] and *rpoB* sequence analysis [12, 45]. These strains were grouped as follows: five INT-1 strains (Asan 29591, 29778, 36309, 37128, and 37721), five INT-2 strains (Asan 36638, 37016, 38392, 38402, and 38585), and six INT-5 strains (MOTT-12, MOTT-27, MOTT-68Y, MOTT-H4Y, MOTT-36Y and Rhu). For genomic DNA extraction, the clinical isolates were cultured on Middlebrook 7H10 agar plates supplemented with OADC (BD GmbH, Heidelberg, Germany) for 7-10 days in a 5% CO_2 incubator at 37 °C. Genomic DNA was prepared by the bead beater-phenol extraction method as previously described [51].

Genomic DNA from each of the five INT-1 strains (Asan 29591, 29778, 36309, 37128, and 37721), five INT-2 strains (Asan 36638, 37016, 38392, 38402, and 38585), and four INT-5 strains (MOTT-12, MOTT-27, MOTT-68Y and Rhu) was used to amplify the five selected genes. As a positive and a negative control of PCR of 5 genes, genomic DNA of *M. intracellulare* ATCC 13950^T and distilled water were also used.

The five primer sets were as follows: *argH*, argH_F 5' – TGA GCA AGT CCA CCC ATT TC – 3' and argH_R 5' – TGG CGT CGA TGG AGT TGT C – 3'; *dnaA*, dnaA_F 5' – ACG AGC CTC AAC CGC C – 3' and dnaA_R 5' – CTC ACG GCA CAG GTA CAT CG – R'; *deaD*, DEAD_F 5' – GGA ATA CAA GCA GGT GGC ACT – 3' and DEAD_R 5' – GCG TTC GTA GTC CTG GAC CA – 3'; *hsp65*, hspF3 5' – ATC GCC AAG GAG ATC GAG CT – 3' and hspR4 5' – AAG GTG CCG CGG ATC TTG TT – 3' and *recF*, recF_F 5' – GAA ATC CCT GTC TGG CGC – 3' and recF_R 5' – TCA TGC GCG CAT CTC C – 3'. Template DNA and each primer pair (20 pmol) were added to the PCR premix (AccuPower PCR PreMix, Bioneer), and PCR was conducted by subjecting the samples to 5 min at 95 ℃, followed by 30 cycles of 95 ℃ for 30 s, 58-60 ℃ for 30 s, and 72 ℃ for 1 min, and a final extension at 72 ℃ for 5 min. The PCR reaction was performed in a MyCycler (Bio-RAD). The PCR products were detected and purified using the MEGAquick-Spin Total Fragment DNA Purification kit (iNtRON) for direct sequencing. Sequencing reactions were performed using an MJ Research PTC-225 Peltier Thermal Cycler and ABI PRISM BigDye Terminator Cycle Sequencing kits with the AmpliTaq DNA polymerase (FS enzyme, Applied Biosystems) following the manufacturer's protocols. The obtained sequences were aligned using the MegAlign software package (DNASTAR), and then the SNPs in the sequences were analyzed.

7. Real-time PCR analysis

Twenty-eight mycobacterial reference strains and 63 clinical isolates were used for real-time PCR analysis. Twenty-three of the 28 mycobacterial reference strains (with the exception of *M. massiliense* KCTC 19086^{T} and 3 *M. yongonense* strains, DSM

45126^T, MOTT-H4Y, MOTT-36Y and *M. tuberculosis* ATCC 27294^T) were provided by the Korean Institute of Tuberculosis (KIT). In the case of *M. tuberculosis* ATCC 27294^T, genomic DNA of *M. tuberculosis* was provided from the KIT. The *M. massiliense* KCTC 19086^T strain was provided by the Korean Collection for Type Cultures (KCTC), and the three *M. yongonense* strains, DSM 45126^T, MOTT-H4Y and MOTT-36Y, were from Seoul National University College of Medicine (SNUMC). No ethics approval was required for the bacterial isolates in this study.

The extracted genomic DNA from reference or clinical isolated mycobacteria and a set of primers (forward primer, 5′-TTC TTT GCC GGA GAA GAC TTT-3′; reverse primer, 5′-GGG TGG ACC TTG ATC AGC TC-3′) were used for the real-time PCR analysis. Primers were designed to produce a 338-bp IS*Myo2* amplicon (from the 799th to 1136th nucleotide in the *M. yongonense* DSM 45126^T IS*Myo2* sequence) from all strains of *M. yongonense* using Oligo V 6.5 (Molecular Biology Insights).

A 10 μl of reaction mixture was prepared for each sample, with the following components: 1 μl of Taq buffer (including 20 mM $MgCl_2$) supplied together with Ex Taq HS (Takara), 0.25 μM forward primer, 0.25 μM reverse primer, 0.2 mM dNTPs, 0.7 mg/ml BSA (NEB), 0.5 × SYBR Green I (Sigma S9430), 3 % DMSO, 0.25 U of ExTaq HS, and sterile distilled water. The cycling conditions were 2 min at 95 °C and 5 s at 98 °C, followed by 35 or 45 cycles of 10 s at 98 °C, 10 s at 64 °C and 40 s at 72 °C (single acquisition of fluorescence signals). Melting curve analysis was performed as follows: 10 s at 98 °C and 1 min at 70 °C, after which the temperature was increased from 70 °C to 98 °C at a temperature transition rate of 0.2 °C/s, with continuous acquisition of the fluorescence signal. Quantification cycles (Cqs) and amplicon melting temperatures (T_ms) were measured using LightCycler 96 system

software, V1.1. The T_m specificity was verified via duplicate real-time PCR measurements with a panel of reference mycobacterial DNAs. A total of 63 clinical isolates were subsequently tested for the identification of *M. yongonense* species. The detection limit of the real-time PCR assay for *M. yongonense* was tested in duplicate using serially diluted genomic DNA ranging from 10 ng to 10 fg.

RESULTS

General features of *M. yongonense* DSM 45126^T genome sequence

M. yongonense DSM 45126^T genome show it to have a circular DNA of 5,521,023 bp, a circular plasmid of 122,976 bp, and a linear plasmid of 18,089 bp (Figure 2-1). The *M. yongonense* DSM 45126^T genome contains protein-coding genes (5,222 open reading frames [ORFs]) similar to those of *M. intracellulare* ATCC 13950^T (5,145 ORFs) and contains the same number of tRNA genes as *M. intracellulare* ATCC 13950^T (47 tRNA genes). The genome of *M. yongonense* DSM 45126^T has a G+C content of 67.95%, and the two plasmids have G+C contents of 65.92 and 66.69%. The whole genome sequences of *M. yongonense* DSM 45126^T have been deposited at GenBnak under the accession numbers of CP003347 (NC_020275, for chromosomal DNA) and JQ657805 and JQ657806 (for two plasmids).

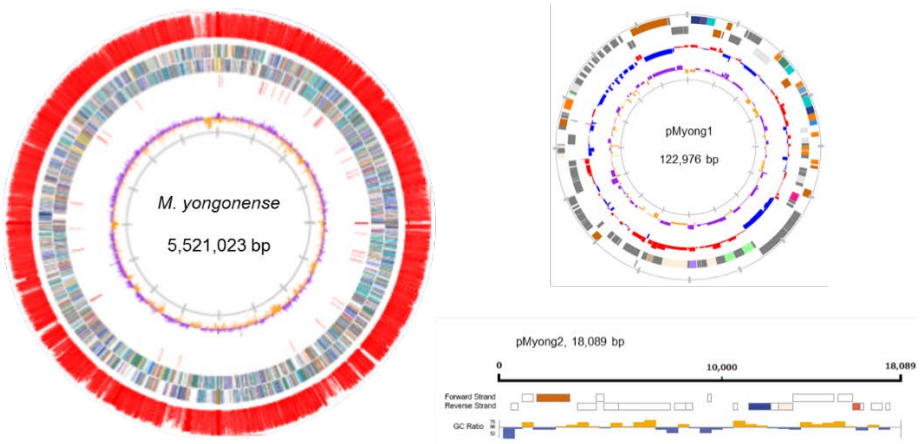


Figure 2-1. Genome contents of *M. yongonense* DSM 45126^{T}. Circular representation of the chromosome in left side. The first circle shows the compared ORFs of *M. intracellulare* ATCC 13950^{T}. The second two circles show forward and reverse genes. The third circle shows tRNA genes (red). The forth circle shows GC ratio (yellow/purple). Circular representation of the 122-kb plasmid, pMyong1 in the right-up side. Linear map representation of the 18-kb plasmid, pMyong2 in the right-down side. Forward and reverse genes and the GC ratio are indicated with the same code as for the chromosome map.

Genome sequence-based phylogenetic analysis of two INT-5 strains, MOTT-H4Y and MOTT-36Y

The phylogenetic relationships between 3 *M. intracellulare* strains (ATCC 13950^T, MOTT-02, and MOTT-64), 2 INT-5 strains (MOTT-H4Y and MOTT-36Y) and one *M. yongonense* strain (DSM 45126^T) were analyzed using the genome sequence information (Figure 2-2). All 6 strains were clustered together in a branch. The strains were separated into two different branches: one including 3 *M. intracellulare* strains (ATCC 13950^T, MOTT-02, and MOTT-64) and the other including *M. yongonense* DSM 45126^T and the two INT-5 strains, MOTT-H4Y and MOTT-36Y. This result indicated that the two INT-5 strains were more closely related to *M. yongonense* DSM 45126^T than the 3 *M. intracellulare* strains (ATCC 13950^T, MOTT-02, and MOTT-64) (Figure 2-2).

To assess the number of genes shared between each genome, we performed a BLASTCLUST analysis on the four genomes (*M. intracellulare* ATCC 13950^T, MOTT-H4Y, and *M. yongonense* DSM 45126^T or *M. intracellulare* ATCC 13950^T, MOTT-36Y, and *M. yongonense* DSM 45126^T). At the level of 95% identity, MOTT-36Y or MOTT-H4Y shared more orthologous coding sequences (CDSs) with *M. yongonense* DSM 45126^T (4,271/5,128 CDSs, 83.3 % or 4,287/5,020 CDSs, 85.4 %, respectively) than *M. intracellulare* ATCC 13950^T (4,101/5,128 CDSs, 80.0 % or 4,052/5,020 CDSs, 80.7 %, respectively) (Figure 2-3).

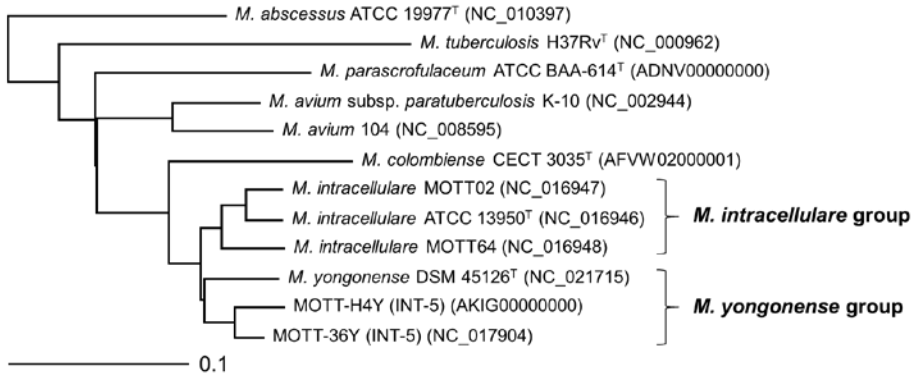


Figure 2-2. Phylogenetic tree based on whole-genome sequences of 3 *Mycobacterium intracellulare* group, 2 INT-5 group, *M. yongonense* and other mycobacterial strains. The tree was calculated using the neighbor-joining method by the Mauve Genome Alignment software and visualized by the TreeViewX program. The bar indicates the number of substitutions per nucleotide position.

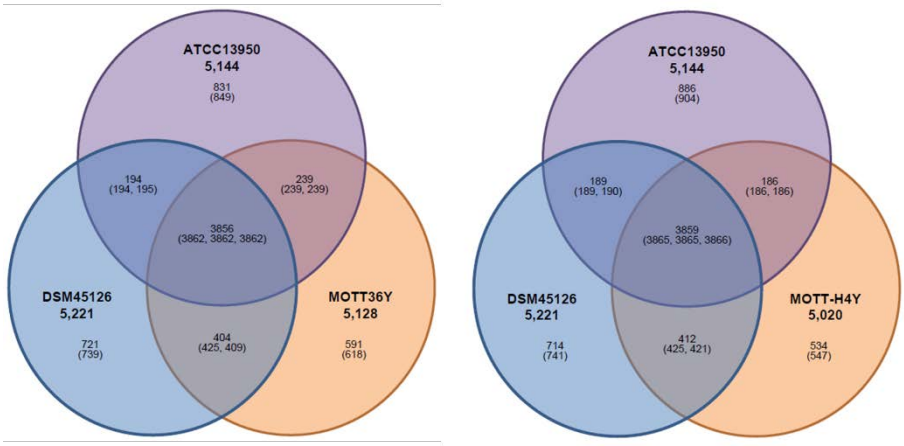
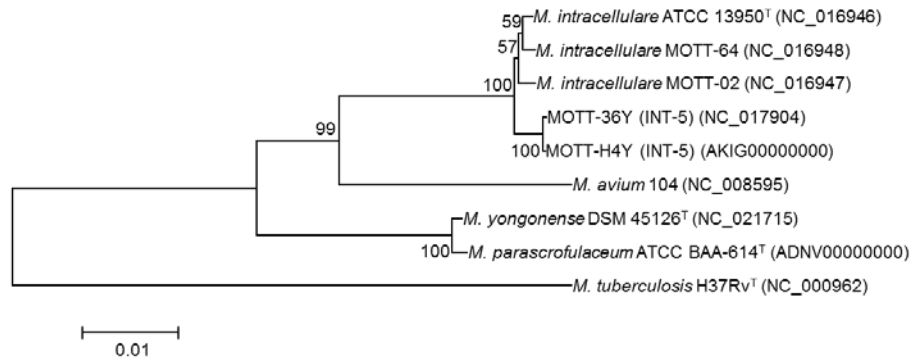


Figure 2-3. Venn diagrams based on genome information of INT-5 strains. Venn diagrams showing orthologous ORFs among four mycobacterial species as determined by BLASTCLUST analysis. Numbers in parenthesis include paralogous ORFs.

Phylogenetic analysis of two INT-5 strains, MOTT-H4Y and MOTT-36Y based on the *rpoB* gene sequences and the sequences of 35 selected target genes

The taxonomic signature of *M. yongonense* was previously reported to be based on the *rpoB* gene sequence. The sequence of this gene is identical to the distantly related species *M. parascrofulaceum*, which enables the separation of the 2 closely related species *M. intracellulare* and *M. yongonense* [71, 83]. Therefore, to obtain the exact taxonomic delineation of the two INT-5 strains we compared their taxonomic location by phylogenetic analysis based on the sequences of *rpoB* and 35 selected target genes. The entire sequences of *rpoB* and the 35 selected genes were retrieved from the genome sequences of 6 mycobacterial strains [3 *M. intracellulare* strains (*M. intracellulare* ATCC 13950^T, MOTT-02, and MOTT-64), 2 INT-5 strains (MOTT-36Y and MOTT-H4Y) and *M. yongonense* DSM 45126^T] (Table 2-1) and subjected to phylogenetic analysis. In the *rpoB* gene (3,375 to 3,462 bp)-based phylogenetic analysis, the two INT-5 strains MOTT-H4Y and MOTT-36Y were clustered into the group including the *M. intracellulare* strains (*M. intracellulare* ATCC 13950^T, MOTT-02, and MOTT-64) and were separated from *M. yongonense* DSM 45126^{T}and *M. parascrofulaceum* ATCC BAA-614^T (Figure 2-4A). However, in the phylogenetic analyses based on the sequences of the 35 selected genes, the two INT-5 strains MOTT-H4Y and MOTT-36Y were clustered into *M. yongonense* DSM 45126^T and separated from the other 3 *M. intracellulare* strains with a high bootstrap value (> 99%), as shown in the genome sequence-based phylogenetic analysis (Figures 2-2 and 2-4B). These results suggest that there may be a distinct *M. yongonense* genotype having an *rpoB* gene sequence that is almost identical to *M. intracellulare*.

(A)



(B)

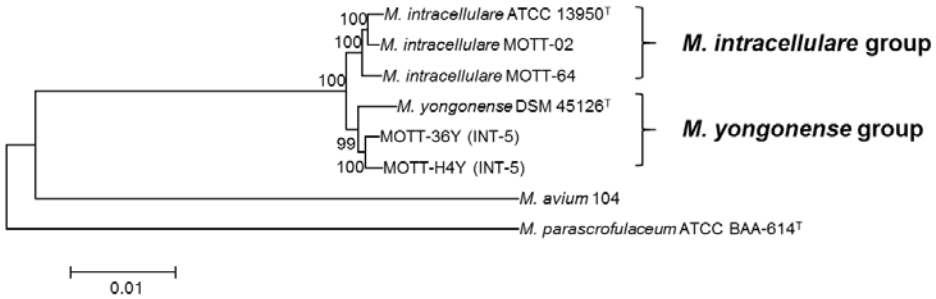


Figure 2-4. Neighbor-joining phylogenetic tree based on the *rpoB* or 35 concatenated genes from 6 *Mycobacterium intracellulare* strains. (A) A tree based on the whole *rpoB* gene sequences from 3 *M. intracellulare*, 2 INT-5 strains, *M. yongonense*, and *M. parascrofulaceum*. (B) A tree based on the 35 concatenated gene sequences from 3 *M. intracellulare*, 2 INT-5 strains, *M. yongonense*, and *M. parascrofulaceum*. The bootstrap values were calculated from 1,000 replications and values <50 % were not shown. The bar indicates the number of substitutions per nucleotide position. *M. tuberculosis* H37Rv and *M. avium* 104 were used as outgroups in the *rpoB* gene- or concatenated genes-based phylogenetic trees, respectively.

Phylogenetic analysis of two INT-5 strains (MOTT-H4Y and MOTT-36Y) based on single nucleotide polymorphisms (SNPs) of the *rpoB* and 35 targeted genes

Multiple alignments of the *rpoB* and 35 gene sequences from the 3 *M. intracellualre* (*M. intracellulare*, MOTT-02 and MOTT-64), 2 INT-5 (MOTT-H4Y and MOTT-36Y), *M. yongonense* and *M. parascrofulaceum* showed that there were *M. yongonense* group- related SNPs in 17 genes [*hsp65* (6 *M. yongonense* group-related SNPs/103 total SNPs), *argH* (10/199), *cya* (5/207), *dnaJ* (2/124), *glpK* (2/223), *pta* (3/328), *recF* (18/235), *secA1* (3/268), *deaD* (14/228), *dnaA* (11/193), *dnaG* (5/250), *dnaK* (2/118), *glnA* (1/106), *gyrB* (4/245), *ino1* (2/81), *ligA* (3/257), and *ligC* (3/173)] (Table 1). Detailed *M. yongonense* group-related SNP signatures are listed in Table 2-3.

In the case of *rpoB* gene, there was no *M. yongonense* group-related SNPs, however, *rpoB* gene of *M. yongonense* shared identical 151 SNPs with that of *M. parascrofulaceum*. A concatenated phylogenetic tree was constructed using the extracted SNP sequences. The tree showed that the two INT-5 strains were clustered into *M. yongonense* DSM 45126^{T} and separated from the other 3 *M. intracellulare* strains based on the phylogenetic analyses of the complete genome sequences and 35 concatenated gene sequences (Figures 2-2, 2-4B and 2-5A).

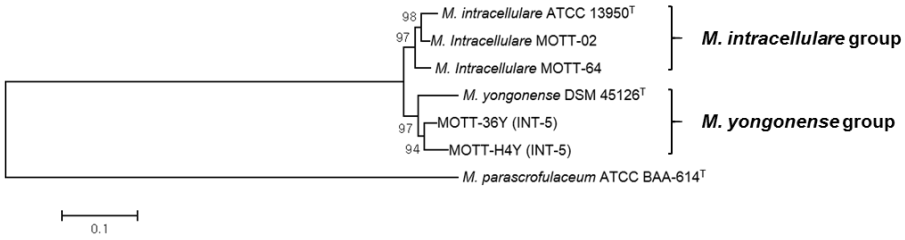
Table 2-3. Details of *M. yongonense* group-related SNP signatures

Genes	*M. yongonense* group-related SNP signatures
argH	C105**T**[a] C132**G** C138**G** C244**T** G303**A** C306**G** C322**T** G339**C** T566**C** C603**A**
cya	C399**G** C414**G** G432**C** C483**G** C504**T**
deaD	G204**C** G210**C** G216**T** G276**A** A315**T** C376**T** C648**T** G840**A** A894**G** T993**C** G1062**C** C1068**G** G1191**A** C1383**T**
dnaA	T222**C** C441**G** G639**A** G651**C** G714**C** G759**A** C921**T** C1035**G** G1080**A** A1326**G** G1341**C**
dnaG	C897**G** G921**T** C1350**G** C1488**T** G1560**T**
dnaJ	C849**T** C1008**T**
dnaK	C1476**T** T1509**C**
glnA	T1434**A**
glpK	T30**C** C723**G**
gyrB	C297**G** G375**C** C660**T** G702**C**
hsp65	G198**A** C555**G** G633**C** C726**T** G1191**C** G1539**C**
ino1	G291**C** G396**A**
ligA	A146**C** C441**G** G1986**A**
ligC	C384**T** G813**A** C933**T**
pta	C1368**T** C1371**T** C1464**T**
recF	C171**T** C249**A** C264**T** C279**T** G336**A** T429**C** A467**G** T534**G** G570**C** C579**T** G586**C** T660**C** G771**T** T796**C** T937**C** G963**A** C1009**T** G1123**A**
secA1	G645**C** T717**G** C1854**G**

All the nucleotide positions were determined from *Mycobacterium intracellulare* ATCC 13950^{T} strain.

[a] Bold characters represent *M. yongonense* group-related SNPs.

(A)



(B)

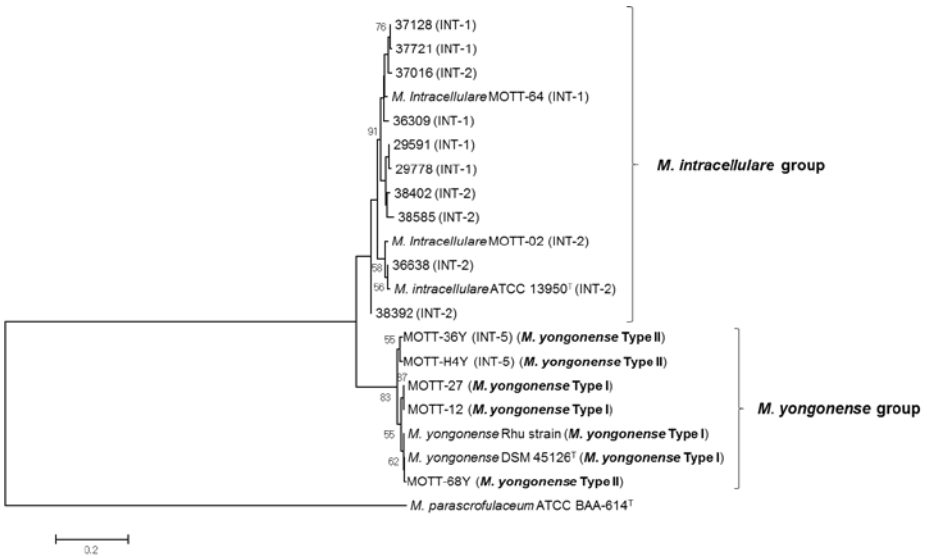


Figure 2-5. Neighbor-joining phylogenetic tree based on concatenated SNPs. (A) A concatenated SNP-based tree from 35 target genes of 3 *M. intracellulare*, 2 INT-5 strains, *M. yongonense*, and *M. parascrofulaceum*. (B) A concatenated SNP-based tree from 5 selected genes (*argH*, *dnaA*, *deaD*, *hsp65* and *recF*) from 3 *M. intracellulare*, 2 INT-5 strains, *M. yongonense*, and *M. parascrofulaceum* and 14 clinical isolate strains. The indicated INT-groupings were assigned in a previous report [14]. The bootstrap values were calculated from 1,000 replications and values <50% were not shown. The bar indicates the number of substitutions per nucleotide position.

Application of the *M. yongonense*-related SNP analysis to MAC clinical isolates

To develop SNP analysis to enable the selective identification of *M. yongonense* strains from the MAC strains, five genes (*argH, deaD, dnaA, hsp65* and *recF*) were selected that possessed a higher number of *M. yongonense* group (*M. yongonense* DSM 45126^T and two INT-5 strains MOTT-H4Y and MOTT-36Y) -related SNPs compared to the other genes. To explore the usefulness of this assay, sequence analysis of the five genes was applied to a total of 14 MAC clinical isolates from different Korean patients [five *M. intracellulare* INT-1 strains (Asan 29591, 29778, 36309, 37128 and 37721), five *M. intracellulare* INT-2 strains (Asan 36638, 37016, 38392, 38402 and 38585), and four INT-5 strains (MOTT-12, MOTT-27, Rhu and MOTT-68Y)] was subjected to phylogenetic analysis.

All four INT-5 strains had 10 *M. yongonense* group-related SNPs in the partial *argH* gene sequence out of 10 *M. yongonense* group-related SNPs (from 105 nt to 657 nt). However, two INT-1 (Asan 29778 and 37721) and two INT-2 group (Asan 37016 and 38392) strains also shared one *M. yongonense* group-related SNP (G339C). All four INT-5 strains had 7 *M. yongonense* group-related SNPs in the partial *dnaA* gene sequence out of 7 *M. yongonense* group-related SNPs (from 627 nt to 1257 nt). However, one INT-2 group strain (Asan 38392) also shared one *M. yongonense* group-related SNP (C921T). All four INT-5 strains had 7 *M. yongonense* group-related SNPs in the partial deaD gene sequence out of 7 *M. yongonense* group-related SNPs (from 588 nt to 1191 nt). However, three INT-2 strains also shared one or four *M. yongonense* group-related SNPs (Asan 36638: C1068G; Asan 38392: C1068G; and Asan 38585: C648T, C681T, G1062C, and C1068G). In the partial *hsp65* gene sequence with 4 *M. yongonense* group-related SNPs (from 192 nt to 726 nt), two INT-

5 group strains (MOTT-12 and MOTT-27) had only one *M. yongonense* group-related SNPs (G198A) and the three INT-1 or INT-2 SNPs (C555, G633 and C726), while the other two strains (Rhu and MOTT-68Y) had 4 *M. yongonense* group-related SNPs. All of the INT-5 group strains had 11 *M. yongonense* group-related SNPs in the partial *recF* gene sequence out of 11 *M. yongonense* group-related SNPs (from 520 nt and 1131 nt). However, one INT-1 (Asan 36309: T660C and G1123A) and one INT-2 strain (Asan 38392: T660C) shared one or two *M. yongonense* group-related SNPs.

The phylogenetic analysis based on the concatenated SNP sequences (395 bp) extracted from the five target genes showed that all four INT-5 strains of *M. yongonense* may share were clearly separated from the other *M. intracellulare* clinical isolates (Figure 2-5B). These results suggested the usefulness of SNP analysis for the taxonomic separation of *M. yongonense* from closely related *M. intracellulare* strains.

Characterization of the IS*Myo2* IS element specific to *M. yongonense*

To select IS elements specific to *M. yongonense* strains for the diagnosis of *M. yongonense*, I compared the distributions of IS elements/transposase sequences between 3 *M. yongonense* strains (DSM 45126^T, MOTT-36Y, and MOTT-H4Y) and 4 other MAC strains (*M. avium* 104, 3 strains of *M. intracellulare*: ATCC 13950^T, MOTT-02, and MOTT-64) via analysis of the seven retrieved mycobacterial genomes (Table 2-1). I identified a total of 56, 40 and 53 IS elements in *M. yongonense* DSM 45126^T, M. yongonense MOTT-36Y, and *M. yongonense* MOTT-H4Y, respectively, using the IS finder program (Table 2-4). In the case of *M. yongonense* DSM 45126^T, 12 types of IS families (IS*5*, IS*21*, IS*30*, IS*110*, IS*256*, IS*481*, IS*607*, IS*630*, IS*1380*, IS*1634*, IS*L3*, and IS*NCY*) were identified in the genome. Through comparison of the

distributions of IS elements among the 7 retrieved mycobacterial genome sequences, seven IS elements (IS*Myo2*, IS*5376*, IS*Mysp3*, IS*Acl1*, IS*Mch6*, IS*Mav2*, and IS*1602*) were identified which found in only the genomes of *M. yongonense* strains. Among these IS elements, I finally targeted an IS*Mt2*- like IS element belonging to the IS*21* family, designated IS*Myo2*, which was found in only the 3 *M. yongonense* strains, and not in other examined strains (Table 2-5).

Table 2-4. Distribution of IS-element/transposase in *M. avium* complex strains

Strains	IS-element/transposase (n)
M. intracellulare ATCC 13950^T	30
M. intracellulare MOTT-02	52
M. intracellulare MOTT-64	36
M. yongonense DSM 45126^T	56
M. yongonense MOTT-36Y	40
M. yongonense MOTT-H4Y	53
M. avium 104	129

Table 2-5. Distribution of IS elements from *M. yongonense* and other *M. intracellulare* strains

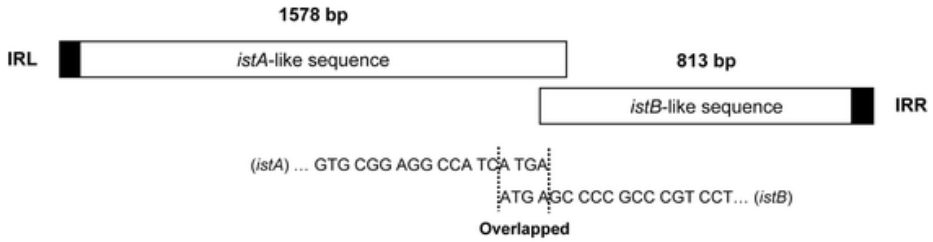
IS family	IS group	IS element	Myon	Mint	MOTT02	MOTT64	MOTT36Y	MOTTH4Y	Mav
IS5	IS427	ISMit1	3	-	-	-	-	-	-
IS21	-	ISMyo2	6	-	-	-	5	4	-
	-	IS5376	2	-	-	-	2	2	-
IS30	IS51	ISMysp3	2	-	-	-	1	1	-
	-	ISAcl1	5	-	-	-	7	5	-
	-	ISLxc3	1	9	5	5	-	1	-
	-	ISMsm8	1	9	5	4	-	-	-
	-	ISPfr16	1	-	-	-	1	-	-
IS110	-	ISCps4	1	-	-	-	-	-	-
	-	ISMch6	1	-	-	-	1	2	-
IS256	-	IS2606	1	-	-	-	1	-	-
	-	ISAnsp14	2	-	-	-	1	-	-
	-	ISOan7	1	-	-	-	-	-	-
	-	ISMgi5	1	-	-	-	-	-	-
	-	ISMlu1	1	-	-	-	1	-	-
	-	ISMysp7	1	-	-	-	9	2	4
	-	ISWch2	1	-	-	-	-	-	-
IS481	-	ISGur11	1	-	-	-	-	-	-
	-	ISMav2	1	-	-	-	2	3	-
	-	ISMva2	2	-	-	-	-	1	-
	-	ISMyma6	1	-	-	-	-	1	-
	-	ISMysp4	3	-	-	-	-	-	-
IS607	-	IS1602	1	-	-	-	1	1	-
IS630	-	ISMsm2	1	-	-	-	2	2	-
	-	ISMsm5	1	-	-	-	-	-	-
IS1380	-	ISMgi1	1	-	-	-	-	-	-
	-	ISMsm10	3	-	-	-	-	-	-
	-	ISMva3	1	-	-	-	1	-	-
IS1634	-	IS1549	2	-	-	-	2	-	-
	-	ISMsm6	1	-	-	-	1	-	-
ISL3	-	IS1207	1	-	-	-	-	-	-
	-	ISMsm4	4	-	-	-	4	-	-
ISNCY	ISDol1	ISRjo3	1	-	-	-	-	-	-

Myon: *M. yongonense* DSM 45126^T, Mint: *M. intracellulare* ATCC 13950^T, MOTT02: *M. intracellulare* MOTT-02, MOTT64: *M. intracellulare* MOTT-64, MOTT36Y: *M. yongonense* MOTT-36Y, MOTTH4Y: *M. yongonense* MOTT-H4Y, Mav: *M. avium* 104.

As shown for other IS elements of the IS*21* family [84], the *M. yongonense*-specific IS elements also exhibit two consecutive open reading frames: a long upstream frame, designated *istA* (1,578-bp), and a shorter downstream frame, designated *istB* (813-bp). *istA*- and *istB*-like sequences overlap between the stop codon of *istA* and the start codon of *istB*. Additionally, upstream and downstream of the two overlapping ORFs, there are three left inverted repeat (IRL) sequences (IRL1: ATGGGACCACCCG, IRL2: ATGGGACCACCTG, IRL3: ATGGGACCACCCG) and two right inverted repeat (IRR) sequences (IRR1: ATGGGACCACCGT, IRR2: ATGGGACCGGTTG) (Figure 2-6).

A BLAST database search at the protein level revealed that *istA*- and *istB*- like sequences from *Saccharopolyspora spinosa* show sequence identities of 76.4 % (396/518) and 76.4 % (204/267), respectively, with those of IS*Myo2*. However, no mycobacterial IS elements matching IS*Myo2* were found (data not shown). Based on these findings, the IS element identified in the *M. yongonense* strains was considered a putative novel IS element belonging to the IS*21* family and was designated IS*Myo2* according to the nomenclature suggested by Mahillon and Chandler [85], and its GenBank accession No. is KP861986.

The copy numbers of IS*Myo2* in the genomes of the 3 *M. yongonense* strains, DSM 45126^{T}, MOTT-36Y, and MOTT-H4Y, were 6, 5 and 4, respectively (Tables 2-5 and 2-6). Exceptionally, in the genome of *M. yongonense* MOTT-36Y, a copy of IS*Myo2* (W7S_12150) was identified, but there was no stop codon between the *istA*- and *istB*-like sequences. Additionally, in the case of *M. yongonense* MOTT-H4Y, a copy of IS*Myo2* included only an *istB*-like sequence (W7U_06705), and no *istA*-like sequences (Table 2-6).



IRL1: ATGGGACCACCCG (-89~-77)
IRL2: ATGGGACCACCTG (-66~-54)
IRL3: ATGGGACCACCCG (-34~-22)

IRR1: ATGGGACCACCGT (reverse complemented) (+74~+86)
IRR2: ATGGGACCGGTTG (reverse complemented) (+49~+61)

Figure 2-6. Schematic representation of the sequence of the *M. yongonense*-specific IS elements IS*Myo2*. For the *istA*-like and *istB*-like sequences, the stop and start codons overlap. Three left inverted repeat (IRL) sequences were found upstream of the *istA*-like sequence, and two right inverted repeat (IRR) sequences were found downstream of *istB*-like sequence.

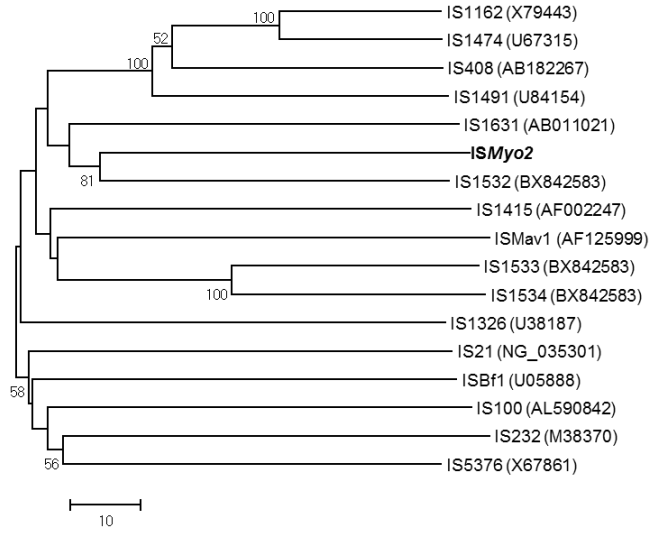
Table 2-6. IS*Myo2* elements in the genomes of *M. yongonense* DSM 45126^T, *M. yongonense* MOTT-36Y and *M. yongonense* MOTT-H4Y

Strains	ORFs	Descriptions	Locations	Orientation	Size (bp)
M. yongonense DSM 45126^T	OEM_10980	*istB*-like	1086916~1087728	(-)	813
(6 copies)	OEM_10990	*istA*-like	1087725~1089302	(-)	1578
	OEM_21160	*istB*-like	2198025~2198837	(-)	813
	OEM_21170	*istA*-like	2198834~2200411	(-)	1578
	OEM_27650	*istA*-like	2924845~2926422	(+)	1578
	OEM_27660	*istB*-like	2926419~2927231	(+)	813
	OEM_30230	*istB*-like	3273171~3273983	(-)	813
	OEM_30240	*istA*-like	3273980~3275557	(-)	1578
	OEM_51540	*istA*-like	5433834~5435411	(+)	1578
	OEM_51550	*istB*-like	5435408~5436220	(+)	813
	OEM_51780	*istB*-like	5463978~5464790	(-)	813
	OEM_51790	*istA*-like	5464787~5466364	(-)	1578
M. yongonense MOTT-36Y	W7S_04860	*istA*-like	974819~976369	(+)	1551
(5 copies)	W7S_04865	*istB*-like	976366~977184	(+)	819
	W7S_08695	*istB*-like	1862284~1863096	(-)	813
	W7S_08700	*istA*-like	1863093~1864631	(-)	1539
	W7S_08730	*istA*-like	1870036~1871574	(+)	1539
	W7S_08735	*istB*-like	1871571~1872383	(+)	813
	W7S_12150	*istA/istB*-like	2654227~2656575	(-)	2349
	W7S_14150	*istA*-like	3100693~3102225	(+)	1533
	W7S_14155	*istB*-like	3102222~3103034	(+)	813
M. yongonense MOTT-H4Y	W7U_00620	*istA*-like	122119~123654	(+)	1536
(4 copies)	W7U_00625	*istB*-like	123651~124463	(+)	813
	W7U_08405	*istA*-like	1846603~1848339	(+)	1737
	W7U_08410	*istB*-like	1848323~1848967	(+)	645
	W7U_09150	*istA*-like	2010206~2011756	(+)	1551
	W7U_09155	*istB*-like	2011753~2012571	(+)	819
	W7U_22700	*istA*-like	4903150~4904700	(+)	1551
	W7U_22705	*istB*-like	4904752~4905516	(+)	765
	W7U_06705	*istB*-like	1488364~1489182	(-)	819

For comparison with other IS elements related to the IS*21* family, 16 additional IS element sequences were retrieved from the GenBank database and compared with the *istA*- and *istB*-like sequences of IS*Myo2*. IS*Myo2* clustered together with IS*1532* from *M. tuberculosis* for two ORFs, though they showed low sequence similarity at the amino acid level (33.5 % for the *istA* sequence and 34.3 % for the *istB* sequence) (Figure 2-7).

The phylogenetic tree based on the *istA*- and *istB*-like sequences from the *M. yongonense* strains showed the presence of four different groups (Figure 2-8). In the case of *M. yongonense* DSM 45126^{T}, the observed IS*Myo2* sequences were highly conserved in the genome. However, those of *M. yongonense* MOTT-36Y or MOTT-H4Y showed variations in the genomes. For the *istA* sequence (1,578 bp), the sequence homologies between the 6 alleles of *M. yongonense* DSM 45126^{T}, the 5 alleles of *M. yongonense* MOTT-36Y, and the 4 alleles of *M. yongonense* MOTT-H4Y ranged from 99.7 to 100.0 %, from 75.6 to 100.0 %, and from 81.1 to 99.9 %, respectively. For the *istB* sequence (813 bp), the sequence homologies between the 6 alleles of *M. yongonense* DSM 45126^{T}, the 5 alleles of *M. yongonense* MOTT-36Y, and the 4 alleles of *M. yongonense* MOTT-H4Y were 100 %, from 79.2 to 100.0 %, and from 84.7 to 100.0 %, respectively.

(A) *istA*-like sequence



(B) *istB*-like sequence

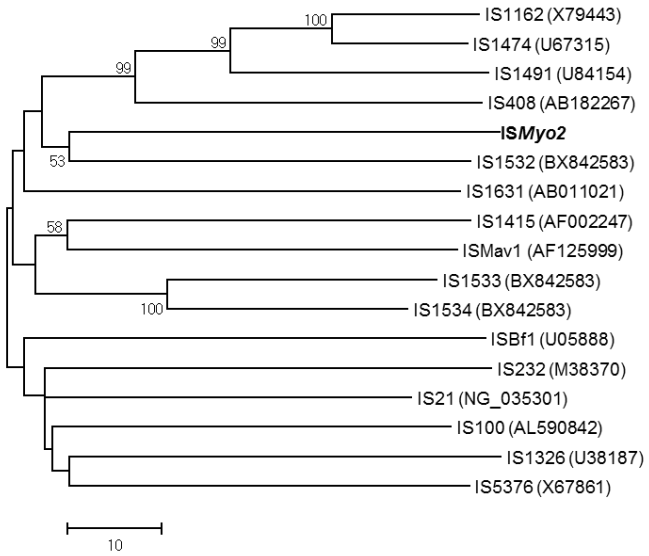
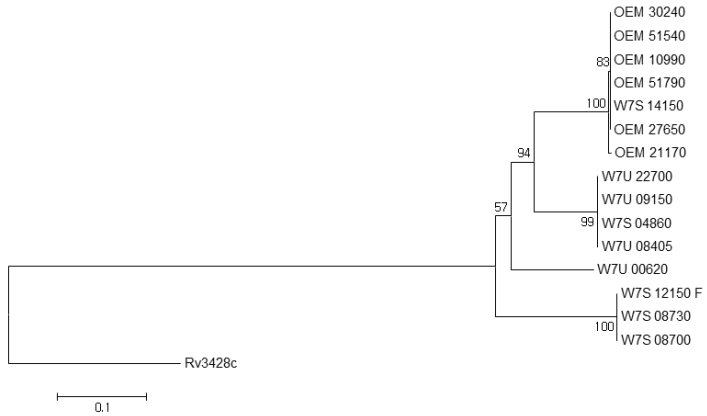


Figure 2-7. Phylogenetic trees based on (A) *istA*-like sequences and (B) *istB*-like sequences from *M. yongonense* DSM 45126^T and 16 other bacteria. The trees were constructed using the neighbor-joining method. Bootstrap values were calculated from

1,000 replications, and values of <50 are not indicated in the trees. The bars indicate the numbers of substitutions per amino acid position.

(A) *istA*-like sequence



(B) *istB*-like sequence

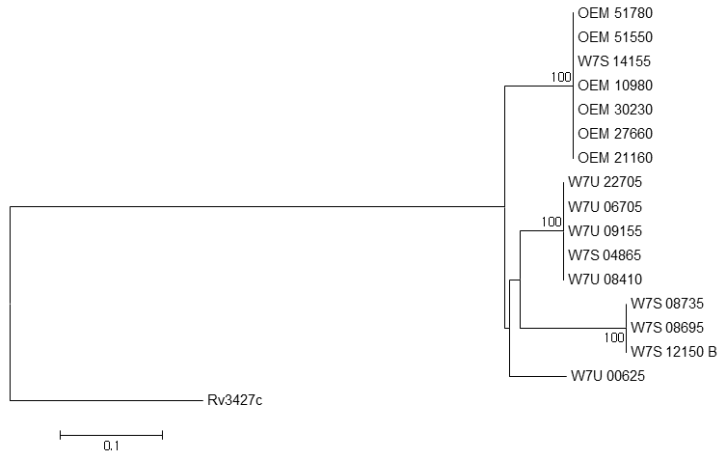


Figure 2-8. Phylogenetic tree based on (A) *istA* -like sequences and (B) *istB* -like sequences from *M. yongonense* DSM 45126^{T}, *M. yongonense* MOTT-36Y and *M. yongonense* MOTT-H4Y. IS-elements of *M. tuberculosis* were used as an outgroup. The trees were constructed using the neighbor-joining method. The bootstrap values were calculated from 1,000 replications and <50 were not indicated. The bars indicate numbers of substitutions per nucleotide position.

Development of a real-time PCR assay targeting IS*Myo2* for the detection of *M. yongonense* strains

To develop appropriate primer sets based on IS*Myo2* sequences for the specific amplification of *M. yongonense* strains, IS*Myo2* sequences from a total of 15 copies from the genomes of 3 *M. yongonense* strains were compared. Finally, I designed a primer set targeting sequences that are conserved in all 15 copies of IS*Myo2*, producing 338-bp amplicons (from the 799th to the 1136th nucleotide of the IS*Myo2* sequence of *M. yongonense* DSM 45126^{T}) (Figure 2-9).

Specificity of the diagnostic assays

The specificity of the real-time PCR assay developed for the identification of *M. yongonenese* was tested in 28 reference strains of mycobacteria via Cq and melting curve analyses. Three *M. yongonense* strains (DSM 45126^{T}, MOTT-36Y, and MOTT-H4Y) were specifically identified through measurement of Cq and T_m (~93 °C), whereas none of the 25 other reference strains showed any detectable Cqs or melting temperatures (Table 2-7, Figure 2-10). The real-time PCR assay was then applied to 63 clinical isolates, including a number of MAC species; again, only the 3 *M. yongonense* strains were identified, and not the 60 other clinical MAC isolates (*M. intracellulare* INT-1: 35 strains, *M. intracellulare* INT-2: 16 strains, *M. intracellulare* INT-3: 1 strain, *M. avium*: 8 strains and *M. chimaera*: 1 strain), based on the measurement of Cqs and specific T_ms (~93 °C) (Table 2-8, Figure 2-11A and B). Agarose gel electrophoresis analysis of the real-time PCR products of 6 *M. yongonense* strains (3 reference and 3 clinical strains) revealed a single electrophoretic band of the predicted size (338 bp) (Figure 2-11C).

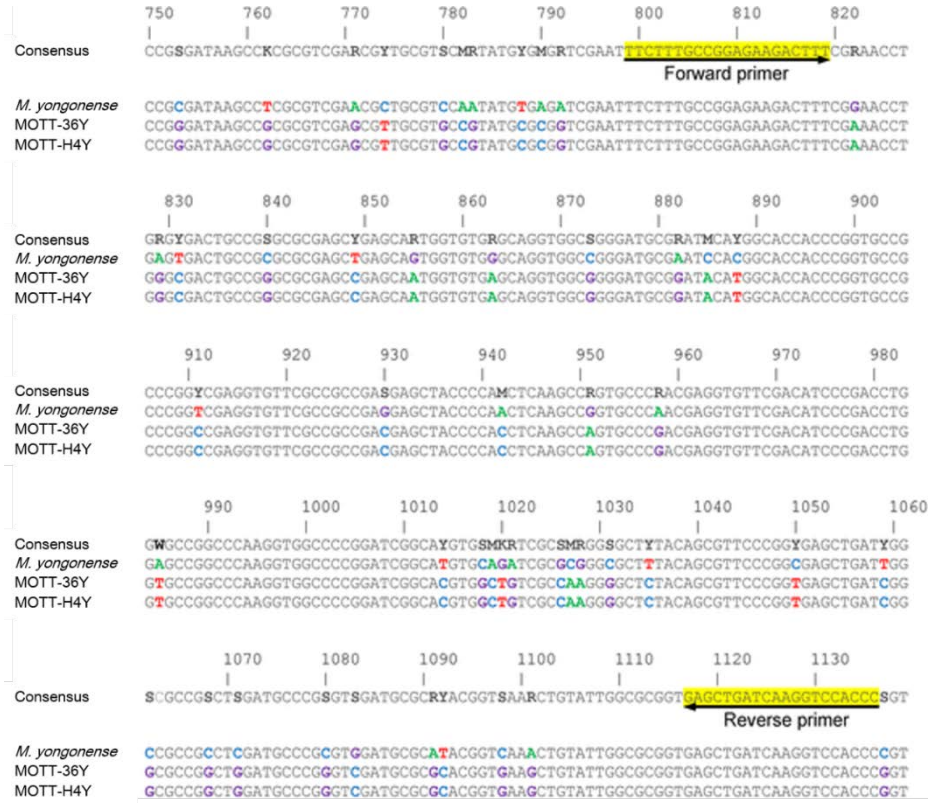


Figure 2-9. Primers designed for the identification of *Myocbacterium yongonense* on the basis of IS*Myo2* sequence alignment for *M. yongonense* strains. Arrows indicate the primer positions. The numbers indicate the nucleotide positions in the *istA*-like sequence of IS*Myo2* of *M. yongonense*. Boldface bases denote the bases that differ from those in the consensus sequence. The strains included in this analysis were as follows: *M. yongonense* DSM 45126^{T}; *M. yongonense* MOTT-36Y; and *M. yongonense* MOTT-H4Y.

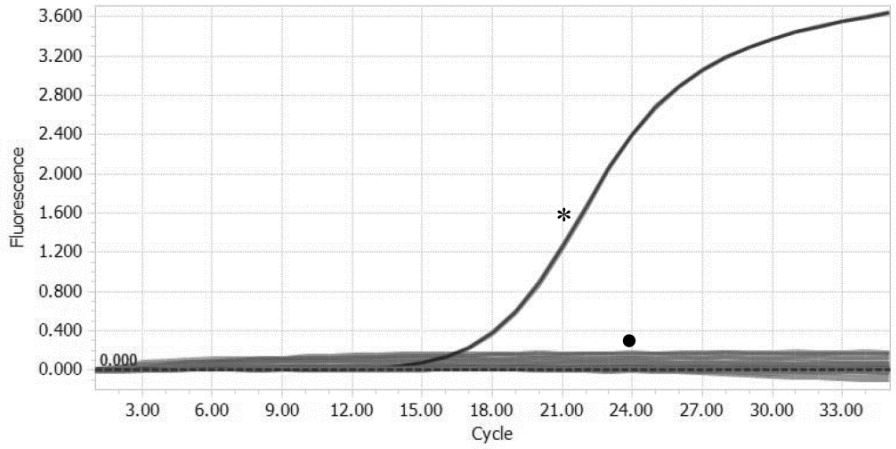
Table 2-7. Mycobacterial reference strains used in the present study and specificity of real-time PCR for the detection of the insertion sequence specific to *Mycobacterium yongonense.*

Species	Strain	Source[a]	Real-time PCR[b]	
			Cq	T_m (℃)
M. abscessus	ATCC 19977^T	ATCC	-	-
M. avium	ATCC 25291^T	ATCC	-	-
M. celatum	ATCC 51131^T	ATCC	-	-
M. chelonae	ATCC 35749^T	ATCC	-	-
M. chimaera	CIP 107892^T	CIP	-	-
M. flavescens	ATCC 14474^T	ATCC	-	-
M. fortuitum	ATCC 6841^T	ATCC	-	-
M. gastri	ATCC 15754^T	ATCC	-	-
M. genavense	ATCC 51233^T	ATCC	-	-
M. gordonae	ATCC 14470^T	ATCC	-	-
M. haemophilum	ATCC 29548^T	ATCC	-	-
M. intracellulare	ATCC 13950^T	ATCC	-	-
M. kansasii	ATCC 12478^T	ATCC	-	-
M. malmoense	ATCC 29571^T	ATCC	-	-
M. massiliense	KCTC 19086	KCTC	-	-
M. marinum	ATCC 29571^T	ATCC	-	-
M. peregrinum	ATCC 27294^T	ATCC	-	-
M. phlei	ATCC 35784	ATCC	-	-
M. scrofulaceum	ATCC 19981^T	ATCC	-	-
M. smegmatis	ATCC 607	ATCC	-	-
M. szulgai	ATCC 35799^T	ATCC	-	-
M. terrae	ATCC 15755^T	ATCC	-	-
M. tuberculosis	ATCC 27294^T	ATCC	-	-
M. ulcerans	ATCC 19423^T	ATCC	-	-
M. xenopi	ATCC 19250^T	ATCC	-	-
M. yongonense	DSM 45126^T	SNUMC	16.79 ± 0.08	93.10 ± 0.08
M. yongonense	MOTT-H4Y	AMC	20.30 ± 0.01	93.20 ± 0.03
M. yongonense	MOTT-36Y	AMC	17.27 ± 0.04	93.21 ± 0.04

[a] ATCC, American Type Culture Collection; CIP, Collection de I'Institut Pasteur; KCTC, Korean Collection for Type Cultures; SNUMC, Seoul National University College of Medicine.

[b] Cqs and T_ms were obtained by duplicate real-time PCR using SYBR Green I as a detection chemistry and melting curve analyses, and the data represent the means ± standard deviations. -, not detectable.

(A)



(B)

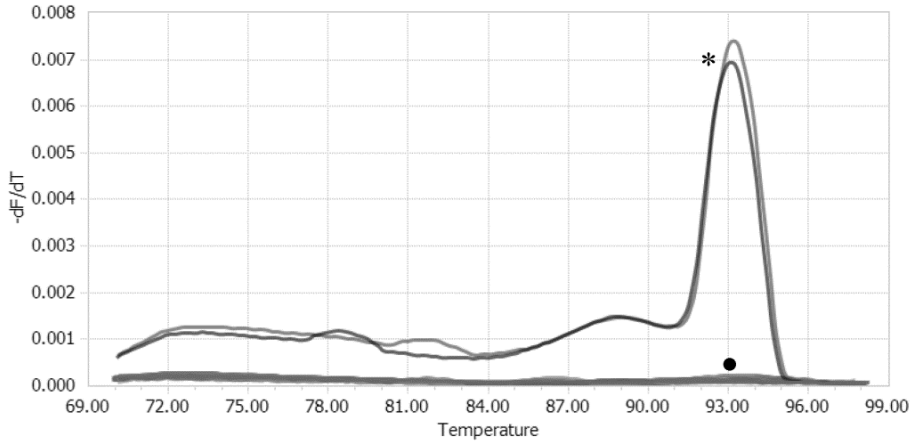


Figure 2-10. Specificity test for the real-time PCR assay developed for the identification of *M. yongonense* with using 28 reference strains of *Mycobacterium* species. *M. yongonense* was specifically identified based on Cq and melting temperature measurements. The tested strains are the same as those listed in Table 2-7 and were tested in duplicate via SYBR Green I real-time PCR. (A) amplification curves. (B) melting curve analysis. * *M. yongonense* DSM 45126^{T}, ・ *Mycobacterium* reference strains.

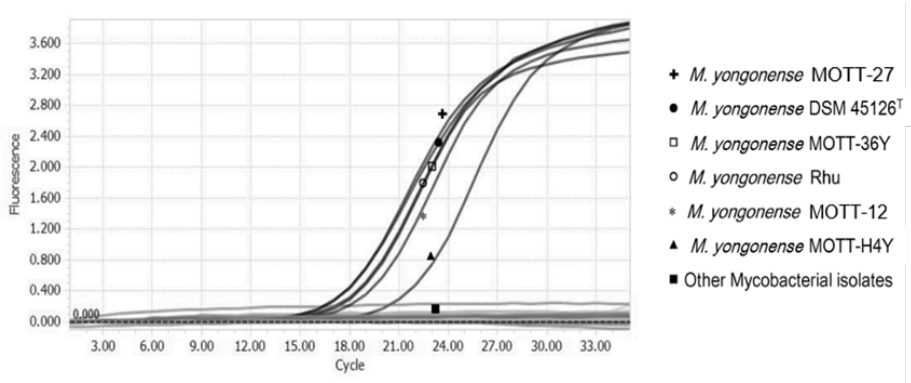
Table 2-8. Identification of the insertion sequence specific to *M. yongonense* among clinical isolates by real-time PCR.

Species	Strain	Source[a]	Group	Real-time PCR[b]	
				Cq	T_m (℃)
M. intracellulare	Asan 29778	AMC	INT 1	-	-
M. intracellulare	Asan 35474	AMC	INT 1	-	-
M. intracellulare	Asan 36005	AMC	INT 1	-	-
M. intracellulare	Asan 36069	AMC	INT 1	-	-
M. intracellulare	Asan 36531	AMC	INT 1	-	-
M. intracellulare	Asan 28764	AMC	INT 2	-	-
M. intracellulare	Asan 36155	AMC	INT 2	-	-
M. intracellulare	Asan 36379	AMC	INT 2	-	-
M. intracellulare	Asan 36795	AMC	INT 2	-	-
M. intracellulare	Asan 36186	AMC	INT 3	-	-
M. intracellulare	Asan 37088	AMC	INT 1	-	-
M. intracellulare	Asan 37349	AMC	INT 1	-	-
M. intracellulare	Asan 37358	AMC	INT 1	-	-
M. intracellulare	Asan 37625	AMC	INT 1	-	-
M. intracellulare	Asan 37128	AMC	INT 1	-	-
M. intracellulare	Asan 37262	AMC	INT 1	-	-
M. intracellulare	Asan 37447	AMC	INT 1	-	-
M. intracellulare	Asan 37635	AMC	INT 1	-	-
M. intracellulare	Asan 37692	AMC	INT 1	-	-
M. intracellulare	Asan 37721	AMC	INT 1	-	-
M. intracellulare	Asan 38013	AMC	INT 1	-	-
M. intracellulare	Asan 38094	AMC	INT 1	-	-
M. intracellulare	Asan 38363	AMC	INT 1	-	-
M. intracellulare	Asan 38397	AMC	INT 1	-	-
M. intracellulare	Asan 38478	AMC	INT 1	-	-
M. intracellulare	Asan 35222	AMC	INT 1	-	-
M. intracellulare	Asan 37183	AMC	INT 1	-	-
M. intracellulare	Asan 35642	AMC	INT 1	-	-
M. intracellulare	Asan 35314	AMC	INT 1	-	-
M. intracellulare	Asan 36526	AMC	INT 1	-	-
M. intracellulare	Asan 36482	AMC	INT 1	-	-
M. intracellulare	Asan 35830	AMC	INT 1	-	-
M. intracellulare	Asan 36993	AMC	INT 1	-	-
M. intracellulare	Asan 35224	AMC	INT 1	-	-
M. intracellulare	Asan 35581	AMC	INT 1	-	-
M. intracellulare	Asan 35701	AMC	INT 1	-	-
M. intracellulare	Asan 36036	AMC	INT 1	-	-
M. intracellulare	Asan 36397	AMC	INT 1	-	-
M. intracellulare	Asan 36485	AMC	INT 1	-	-
M. intracellulare	Asan 37680	AMC	INT 2	-	-
M. intracellulare	Asan 37896	AMC	INT 2	-	-
M. intracellulare	Asan 38189	AMC	INT 2	-	-
M. intracellulare	Asan 38392	AMC	INT 2	-	-
M. intracellulare	Asan 38402	AMC	INT 2	-	-
M. intracellulare	Asan 38495	AMC	INT 2	-	-
M. intracellulare	Asan 38585	AMC	INT 2	-	-
M. intracellulare	Asan 36932	AMC	INT 2	-	-
M. intracellulare	Asan 36456	AMC	INT 2	-	-
M. intracellulare	Asan 36004	AMC	INT 2	-	-
M. intracellulare	Asan 36638	AMC	INT 2	-	-
M. intracellulare	Asan 35909	AMC	INT 2	-	-
M. intracellulare	MOTT-64	AMC	INT 1	-	-
M. avium	QIA-22	QIA		-	-
M. avium	QIA-23	QIA		-	-
M. avium	QIA-24	QIA		-	-
M. avium	QIA-25	QIA		-	-
M. avium	QIA-26	QIA		-	-
M. avium	QIA-27	QIA		-	-
M. avium	QIA-28	QIA		-	-
M. avium	QIA-29	QIA		-	-
M. yongonense	MOTT-27	AMC		16.53 ± 0.16	93.19 ± 0.02
M. yongonense	MOTT-12	AMC		17.83 ± 0.15	93.26 ± 0.11
M. yongonense	Rhu	SNUMC		17.18 ± 0.15	93.20 ± 0.08

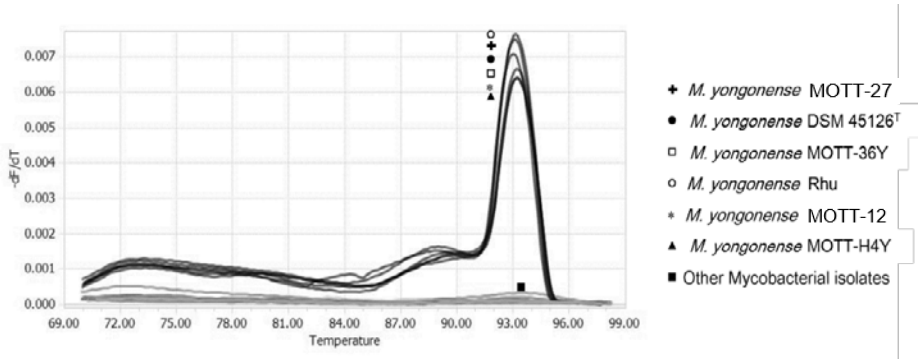
[a] AMC, Asan Medical Center; SNUMC, Seoul National University College of Medicine; QIA, The Animal and Plant Quarantine Agency. [b] Cqs and T_ms were obtained by duplicate real-time PCR using SYBR Green I as a detection chemistry and

melting curve analyses, and the data represent the means ± standard deviations. -, not detectable.

(A)



(B)



(C)

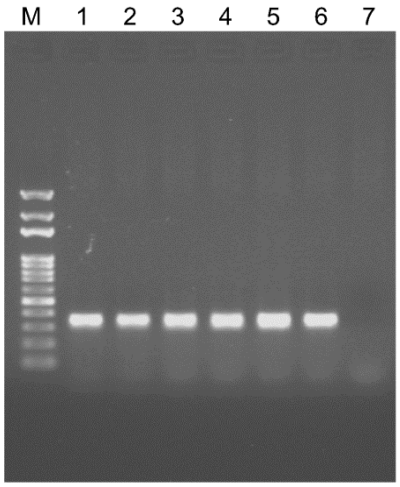


Figure 2-11. Real-time PCR identification of *M. yongonense* from clinical isolates of *Mycobacterium* species. All the *M. yongonense* species among the clinical isolates tested were specifically identified by measurement of their Cqs and melting temperatures. The clinical isolates were tested in duplicate by SYBR Green I real-time PCR. (A) Amplication curves. (B) Melting curve analysis. (C) Agarose gel electrophoresis analysis of real-time PCR products. M, 100 bp DNA marker; 1, *M. yongonense* DSM 45126^{T}; 2, *M. yongonense* MOTT-27; 3, *M. yongonense* MOTT-12; 4, *M. yongonense* MOTT-H4Y; 5, *M. yongonense* MOTT-36Y; 6, *M. yongonense* Rhu; 7, negative control.

Sensitivity of the diagnostic assays

To determine the detection limits of the real-time PCR assay for the detection of *M. yongonense* strains, serially diluted DNA from all 6 strains of the tested *M. yongonense* (3 reference and 3 clinical strains) was subject to real-time PCR. The detection limit of the real-time PCR assay for *M. yongonense* species was 10 fg of genomic DNA in one strain (Rhu) and 0.1 pg in all of the other tested strains (Table 2-9 and Figure 2-12).

Table 2-9. Limit of detection of real-time PCR for the identification of the insertion sequence specific to *M. yongonense*.

Species	Strain	DNA amount	Cq	T_m (℃)
M. yongonense	05-1390	10 ng	13.68 ± 0.12	93.10 ± 0.01
		1 ng	16.95 ± 0.09	93.13 ± 0.11
		100 pg	21.43 ± 0.20	93.13 ± 0.11
		10 pg	26.33 ± 0.09	93.15 ± 0.08
		1 pg	30.83 ± 0.36	93.23 ± 0.01
		100 fg	35.76 ± 0.20	93.22 ± 0.10
		10 fg	-	-
		0 fg	-	-
M. yongonense	MOTT-27	10 ng	13.52 ± 0.16	93.18 ± 0.02
		1 ng	17.24 ± 0.03	93.20 ± 0.14
		100 pg	22.22 ± 0.06	93.28 ± 0.01
		10 pg	27.00 ± 0.06	93.26 ± 0.02
		1 pg	31.18 ± 0.45	93.31 ± 0.02
		100 fg	36.68 ± 0.05	93.39 ± 0.01
		10 fg	-	-
		0 fg	-	-
M. yongonense	MOTT-12	10 ng	14.06 ± 0.06	93.14 ± 0.03
		1 ng	18.61 ± 0.05	93.25 ± 0.01
		100 pg	23.66 ± 0.12	93.22 ± 0.01
		10 pg	28.52 ± 0.25	93.35 ± 0.17
		1 pg	32.62 ± 1.04	93.32 ± 0.04
		100 fg	36.99 ± 0.57	93.37 ± 0.02
		10 fg	-	-
		0 fg	-	-
M. yongonense	MOTT-H4Y	10 ng	16.91 ± 0.01	93.20 ± 0.07
		1 ng	21.49 ± 0.13	93.21 ± 0.08
		100 pg	26.56 ± 0.09	93.25 ± 0.08
		10 pg	31.16 ± 0.27	93.38 ± 0.04
		1 pg	35.27 ± 0.12	93.37 ± 0.17
		100 fg	40.00 ± 0.32	93.35 ± 0.04
		10 fg	-	-
		0 fg	-	-
M. yongonense	MOTT-36Y	10 ng	15.13 ± 0.06	93.10 ± 0.06
		1 ng	17.72 ± 0.16	93.11 ± 0.08
		100 pg	22.14 ± 0.33	93.19 ± 0.08
		10 pg	26.77 ± 0.35	93.24 ± 0.08
		1 pg	31.71 ± 0.35	93.32 ± 0.08
		100 fg	36.59 ± 0.33	93.28 ± 0.02
		10 fg	-	-
		0 fg	-	-
M. yongonense	Rhu	10 ng	13.41 ± 0.23	93.07 ± 0.09
		1 ng	17.31 ± 0.04	93.04 ± 0.06
		100 pg	22.04 ± 0.03	93.13 ± 0.08
		10 pg	26.71 ± 0.38	93.20 ± 0.03
		1 pg	31.19 ± 0.13	93.21 ± 0.03
		100 fg	36.33 ± 0.23	93.21 ± 0.08
		10 fg	40.96 ± 0.08	93.25 ± 0.05
		0 fg	-	-

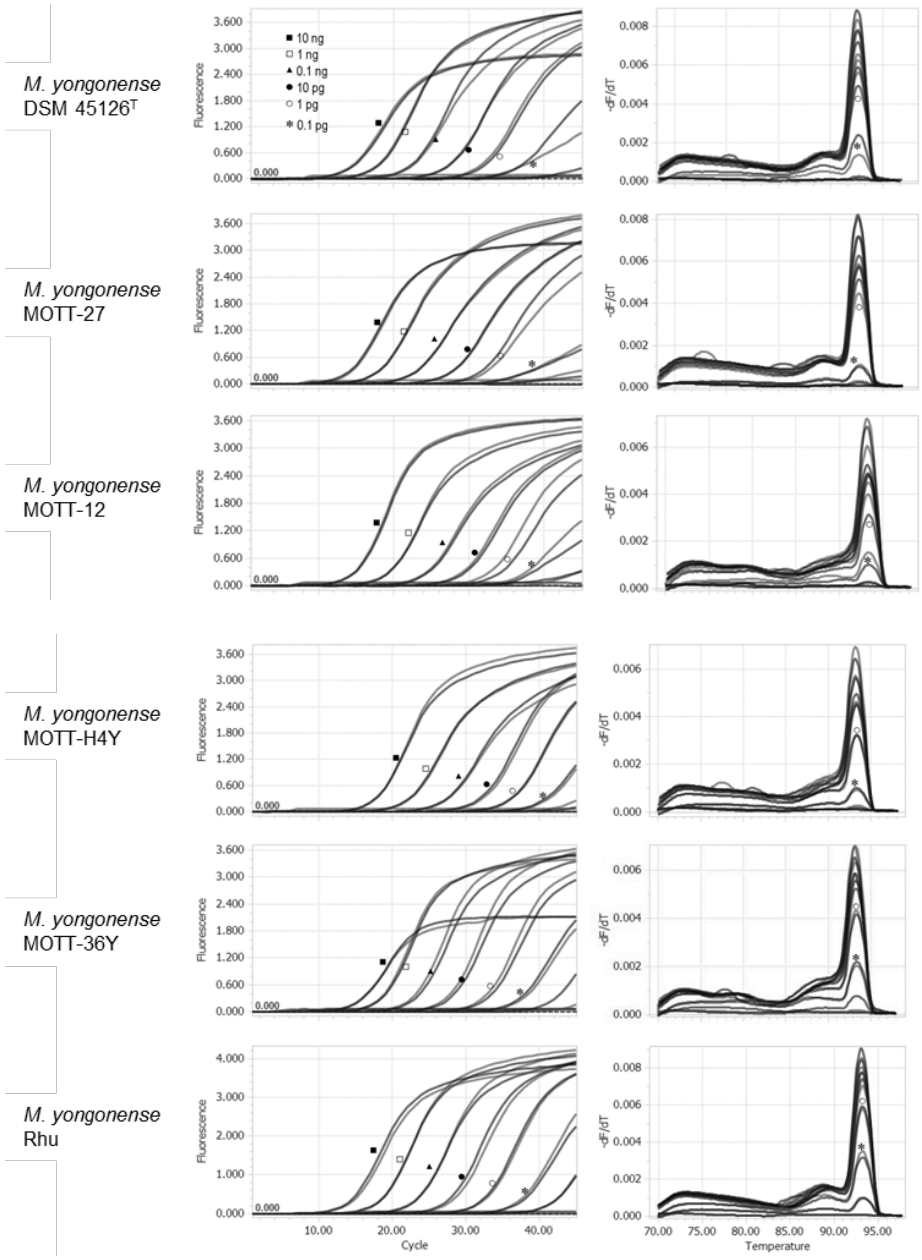


Figure 2-12. Analysis of the detection limits of the real-time PCR assay for the identification of *M. yongonense* species. All of the strains of *M. yongonense* species tested were detected, using as little as 0.1 pg of their genomic DNA.

DISCUSSION

In the present study, phylogenetic analysis based on complete genome sequences, multi-locus sequence typing (MLST) of 35 target genes, and single nucleotide polymorphism (SNP) analysis indicated that the two INT-5 strains, MOTT-H4Y and MOTT-36Y were more closely related to *M. yongonense* DSM 45126^T than the *M. intracellulare* strains. This finding suggests the presence of another distinct genotype in *M. yongonense* that may not have been subjected to the LGT event of *rpoB* from *M. parascrofulaceum*. Therefore, *M. yongonense* could be divided into 2 distinct genotypes: one with the *M. parascrofulaceum rpoB* gene and the other with the *M. intracellulare rpoB* gene, depending on the presence of the LGT event of *rpoB* from *M. parascrofulaceum* (Figures 2-2 and 2-4). Here, I proposed the former and the latter as the *M. yongonense* Type I and Type II genotypes, respectively.

To date, a total of 3 strains (*M. yongonense* DSM 45126^T, MOTT-12 and MOTT-27) belonging to the *M. yongonense* Type I genotype have been introduced via our 2 recent reports [71, 83]. The Rhu strain used in this study was also identified as the Type I genotype by *rpoB* gene analysis (data not shown). In addition to MOTT-H4Y and MOTT-36Y, one additional strain (MOTT-68Y) used in this study was identified as the *M. yongonense* Type II genotype. Although detailed taxonomic proof is needed, *the M. yongonense* strains recently isolated in Italy have the potential to be included in the *M. yongonense* Type II genotype.

LGT is the major mechanism by which bacteria can acquire genetic diversity, guaranteeing their survival under harsh environmental conditions [86, 87]. However, it

is generally accepted that mycobacteria are more resistant to LGT compared to other bacteria, possibly due to the unusually mycolic acid-rich cell wall structure and the relative scarcity of genetic elements such as plasmids and transposable elements [88-90].

Notably, because the *M. yongonense* strains were demonstrated to possess an *rpoB* gene that might have been laterally transferred from the distantly-related scotochromogenic species *M. parascrofulaceum*, these strains have gained increasing importance in the mycobacterial taxonomic fields. One of the noteworthy findings in this study is the identification of a novel genotype of *M. yongonense* without the *rpoB* gene from the LGT event in its genome. A genome comparison study between 3 mycobacterial groups [the *M. yongonense* Type I (subject to the LGT event) and Type II genotypes (without the LGT event) and *M. parascrofulaceum* (gene donor for LGT)] may provide novel insights into our understandings regarding mycobacterial LGT mechanisms.

In the present study, I developed an SNP analysis targeting 5 genes (*argH*, *deaD*, *dnaA*, *hsp65* and *recF*) for the separation of *M. yongonense* from the closely related *M. intracellulare* strains. The concatenated 395-bp SNP-based phylogenetic analysis clearly separated 7 *M. yongonense* strains from 12 closely related *M. intracellulare* strains belonging to the INT-I and INT-2 genotypes, which were the first and the second most prevalent genotypes in Korean patients infected with *M. intracellulare*, respectively, with 83 % bootstrap values (Figure 2-5A). This result suggests the feasibility of this assay for the selective identification of *M. yongonense* strains in clinical settings. Interestingly, this assay could not differentiate 4 Type I (DSM 45126^{T}, MOTT-12, MOTT-27, and Rhu) and 3 Type II strains (MOTT-H4Y, MOTT-36Y and

MOTT-68Y) (Figure 2-5B), suggesting the potential for gene exchanges by LGT events between the 2 genotypes. Notably, a total of 39 *M. yongonense* signature SNPs out of the 395 selected SNPs were found. These SNPs could be used for the development of *M. yongonense*-specific molecular diagnostic methods.

Among the members of the MAC complex, for the diagnosis and epidemiology of 4 *M. avium* subspecies (*M. avium* subsp. *hominissuis*, *M. avium* subsp. *avium*, *M. avium* subsp. *silvaticum* and *M. avium* subsp. *paratuberculosis*), at least four major IS elements have been widely used, namely IS*1245*, IS*1311*, IS*900* and IS*901* [91]. However, despite the increases in global clinical cases and in the species diversity of *M. intracellulare*-related strains, the description of IS elements for their diagnosis and epidemiology has rarely been reported thus far. In this study, I first introduced a novel IS element, IS*Myo2*, specific to *M. yongonense* strains, which belong to the *M. intracellulare*-related species. This element can be used for the diagnosis and epidemiology of these strains, particularly to resolve the current taxonomic need of selectively discriminating *M. yongonense* from other *M. intracellulare* related strains.

In this study, genome analysis using IS finder revealed the presence of a total of 56 IS elements and IS-like elements, consisting of at least 12 types of IS families, in the genome of *M. yongonense* DSM 45126^{T}. Of these IS elements, a total of 7 were shown to be present in only the genomes of 3 *M. yongonense* strains, and were absent in the genomes of 3 *M. intracellulare* strains (Table 2-5). However, considering the copy numbers, BLAST search results, and sequence conservation between IS alleles in a genome, I finally selected IS*Myo2*, belonging to the IS*21* family, as a target IS element for the diagnosis of *M. yongonense* strains.

It is worth noting several characteristic features of IS*Myo2* that make it a useful diagnostic marker for *M. yongonense*. The first is the high specificity of IS*Myo2* for *M. yongonense*. To date, 4 types of IS*21* family members have been described in the genus *Mycobacterium*, three in *M. tuberculosis* (IS*1532*, IS*1533* and IS*1534*) [92] and one in *M. avium* (IS*Mav1*) [91]. Despite belonging to the IS*21* family, the IS*Myo2* element identified in this study shows a very low level of sequence homology with other IS*21* members from mycobacteria, particularly in terms of DNA sequences, suggesting the feasibility of using IS*Myo2* as a diagnostic marker for *M. yongonense*. Furthermore, the application of a real-time PCR assay targeting IS*Myo2* in reference mycobacterial strains demonstrated that the assay was specific to only *M. yongonense* strains (Figure 2-10, Table 2-7). Furthermore, it did not produce any amplicons in any of the 61 examined *M. intracellulare* clinical isolates, which included diverse genotypes such as INT-1, INT-2 and INT-3, which are closely related to *M. yongonense* (Figure 2-11, Table 2-8).

Second, the IS*Myo2*- targeting assay was able to detect 2 types of *M. yongonense* variants. Our genome analysis revealed that IS*Myo2* is also present in the genomes of 2 strains of *M. yongonense*, MOTT-36Y and MOTT-H4Y, which were previously designated as the *M. intracellulare* INT-5 genotype, as well as in genome of *M. yongonense* DSM 45126^{T}. Although two INT-5 strains exhibit *rpoB* sequences identical to that of *M. intracellulare* but not to that of *M. parascrofulaceum*, our phylogenetic analysis based on complete genome sequences showed that these strains and *M. yongonense* DSM 45126^{T} are tightly clustered and are separated from other *M. intracellulare* strains (data not shown), suggesting that they may be members of the same species, *M. yongonense*. Our results also strongly supported the hypothesis

previously put forth by Tortoli *et al*. [72] that there may be at least two variants in *M. yongonense* strains, one of which was subjected to the *rpoB* LGT event from *M. parascrofulaceum*, including strains such as *M. yongonense* DSM 45126^{T} (*M. yongonense* Type I), and another, phylogenetically older variant, that was not subject to the *rpoB* LGT, including strains such as the INT-5 strains MOTT-36Y and MOTT-H4Y (*M. yongonense* Type II)..

Third, the IS*Myo2* targeting assay exhibited high sensitivity in detecting *M. yongonense* strains. Our genome analysis showed that more than 4 copies of IS*Myo2* are present in *M. yongonense* genomes (Figure 2-8, Table 2-6). Furthermore, the sequence conservation between the IS*Myo2* alleles found in the *M. yongonense* genomes facilitated the development of a common primer set capable of PCR amplification of all of these alleles (Figure 2-9). Thus, I successfully developed a real-time PCR assay capable of PCR amplification of all of the alleles via performing multiple sequence alignments of 15 IS*Myo2* alleles from the genomes of 3 *M. yongonense* strains. This real-time PCR assay can detect PCR amplicons at a DNA level of 100 fg in all 6 *M. yongonense* strains (Figure 2-12, Table 2-9) suggesting its feasibility as a diagnostic method for *M. yongonense* strains.

In conclusion, genome-based phylogenetic analysis indicated that the taxonomic status of the two INT-5 strains, MOTT-H4Y and MOTT-36Y previously identified as *M. intracellulare* should be revised to *M. yongonense*. Taken together, *M. yongonense* could be divided into 2 distinct genotypes depending on the presence of the LGT event of *rpoB* from *M. parascrofulaceum*: the Type I genotype with the *M. parascrofulaceum rpoB* gene and the Type II genotype with the *M. intracellulare rpoB* gene. Additionally, I developed a novel SNP-based phylogenetic analysis to enable the

taxonomic identification of *M. yongonense* clinical strains. Together, I identified a novel IS*Myo2* IS element belonging to the IS*21* family that is specific to *M. yongonense* strains via genome analysis and developed a real-time PCR method based on its sequences.

CHAPTER 3

Molecular taxonomic study for proving the lateral gene transfer event between *Mycobacterium yongonense* Type I and Type II genotype strains via genome-based analysis

INTRODUCTION

Recently, accumulation of whole bacterial genome sequences showed that the general importance of lateral gene transfer (LGT) in bacterial evolution. The LGT mechanism is the main driving force for bacterial genetic diversity, facilitating maintenance and enhancement of virulence or spread of drug resistance, particularly in pathogenic bacteria [86, 93, 94]. Although all genes could theoretically be transferred between bacteria by LGT, even distantly related species, it could be generally accepted that dispensable genes involved in virulence determinants or drug resistance may be more prone to LGT event than informative genes of the central cellular machinery such as DNA replication, transcription or translation related genes [95].

In the case of the genus of *Mycobacterium*, especially *M. tuberculosis* and *M. leprae*, genome sequences of those bacteria indicate that the LGT event has contributed only a minor role in their evolution [96, 97]. However, in some cases, the LGT events of informative genes within the genus *Mycobacterium* have been reported [40, 41].

In various bacterial strains, homeologous recombination is an important mechanism to gain the genomic diversity. In this procedure, to restrict recombination between moderately divergent (up to ~10%) DNA sequences at the DNA hybridization step, prokaryotes and eukaryotes have utilized a post-replicative mismatch repair (MMR) system [98, 99]. Interestingly, genome sequences of the genus *Mycobacterium* suggest

that mycobacteria lack the highly conserved DNA mismatch repair gene (*mutS*)-based MMR system [96, 97, 99].

A novel mycobacterial species, *M. yongonense*, was proved to have the *rpoB* sequence identical with *M. parascrofulaceum*, rather than that of *M. intracellulare*. This novel strain was genetically related with INT-5 strains, and I described the close relationship between the INT-5 strains and *M. yongonense* by genome-based analyses. So, I proposed the existence of two distinct genotypes of *M. yongonense* strains in Chapter 2. These *M. yongonense* strains, including INT-5 strains (*M. yongonense* Type II genotype) and *M. yongonense* DSM 45126^T (*M. yongonense* Type I genotype) could be divided into two genotypes. *M. yongonense* Type I strains have *rpoB* gene which is closely related with that of *M. parascrofulaceum*, however, Type II strains, such as INT-5 strains, are not. I focused on this fact and searched another putative LGT regions in the genome of *M. yongonense* strains.

The aim of this chapter is to find out genetic differences between *M. yongonense* Type I and Type II genotypes. For this purpose, I searched and compared another putative transferred regions (from *M. parascrofulaceum*) among a total of 9 strains including 3 strains of *M. intracellulare* [one strain of INT1 (*M. intracellulare* MOTT-64) and two strains of INT2 (*M. intracellulare* ATCC 13950^T and *M. intracellulare* MOTT-02)], 5 strains of *M. yongonense* [three strains of Type I (*M. yongonense* DSM 45126^T, *M. yongonense* MOTT-12, and MOTT-27) and two strains of Type II (*M. yongonense* MOTT-36Y and MOTT-H4Y)], *M. parascrofualceum* ATCC BAA-614^T and *M. avium* 104. And *M. yongonense* Type I strain specific ORFs were also found. Interestingly, among these ORFs, DNA mismatch repair gene (*mutS4*) was found in the genome of *M. yongonense* Type I strains. To prove the hypothesis that the DNA

mismatch repair genes in the genome of *M. yongonense* strains could facilitate the homologous recombination event under the harsh environment, I generated a recombinant *M. smegmatis* strain harboring DNA mismatch repair gene and cultured on 7H10 agar plate containing rifampin after introducing the partial *M. tuberculosis rpoB* sequences which are related to rifampin resistance. After that the colonies which could resistant to rifampin were selected and their *rpoB* sequences determined and examined the frequency of homologous recombination.

MATERIALS AND METHODS

1. Genome sequences used in this study

Ten mycobacterial genome sequences, from strains belonging to the *M. avium* complex (3 *M. intracellulare* strains: ATCC 13950^T, MOTT-02, and MOTT-64; 5 *M. yongonense* strains: DSM 45126^T, MOTT-12, MOTT-27, MOTT-36Y and MOTT-H4Y; one *M. avium* strain: *M. avium* 104; and one *M. parascrofulaceum* strain: *M. parascrofulaceum* ATCC BAA-614^T) were retrieved from the GenBank database (Table 2-1) and used for comparative genome analysis.

2. Identification of putative lateral gene transferred regions in *M. yongonense* Type I strains from *M. parascrofulaceum* or other genus strains

To identify putative lateral gene transferred regions of *M. yongonense*, all the ORFs were compared and analyzed by BLASTN and BLASTP programs. The ORFs which showed high sequence similarities with *M. parascrofulaceum* (compared length > 80% and sequence similarities > 80% in nucleotide sequence) were selected and analyzed for possibility of recombination event. Also, among the ORFs which showed high sequence similarities with other genus species in the genome of *M. yongonense* DMS 45126^T (*M. yongonense* Type I), were selected and compared with those of other two *M. yongonense* Type II strains (MOTT-36Y and MOTT-H4Y). And the not matched ORFs with the two *M. yongonense* Type II strains were identified.

3. Construction of the phylogenetic trees and SimPlot analysis

All the identified ORFs from the genome of *M. yongonense* DSM 45126^T were compared with other *M. intracellulare* (ATCC 13950^T, MOTT-02 and MOTT-64), *M. yongonense* Type I (MOTT-12 and MOTT-27), *M. yongonense* Type II (MOTT-36Y and MOTT-H4Y), *M. avium* and *M. parascrofulaceum* strains (Table 2-1). In the case of DNA mismatch repair genes in the *M. yongonense* Type I strains (DSM 45126^T, MOTT-12, and MOTT-27) were compared with DNA mismatch repair gene homologs from other bacteria [100]. Amino acid or nucleotide sequences were aligned by ClustalW method and the phylogenetic trees were constructed using the neighbor-joining method [55] in MEGA 4 software [58].

To visualization the putative recombination site in the genome of *M. yongonense* Type I strains, identified ORFs which showed high sequence similarities with *M. parascrofualceum* were aligned with other *M. intracellulare* (ATCC 13950^T, MOTT-02 and MOTT-64), *M. yongonense* Type II (MOTT-36Y and MOTT-H4Y), *M. avium* and *M. parascrofulaceum* strains by the MegAlign program in the DNASTAR package. Also, the possibility of the recombination event in the genome of *M. yongonense* Type I strains from *M. parascrofulaceum* was examined by the SimPlot program (http://sray.med.som.jhmi.edu/SCRoftware/) and boot scanning analysis [101]. The used parameters are as follows; Window: 200 bp, Step: 20 bp, GapStrip: on, Reps: 100, Kimura (2-parameter), T/t: 2.0, Neighbor-Joining.

4. Construction of the recombinant *M. smegmatis* harboring DNA mismatch repair genes from *M. yongonense* DSM 45126^T

To generate recombinant *M. smegmatis* harboring DNA mismatch repair genes from *M. yongonense* DSM 45126^T, about 3.8 kb of DNA fragment having the DNA mismatch repair gene (3,069 bp) and a putative promoter (770 bp) was amplified using a primer set as follows: forward primer; 5' – TTGCGGCCGCCGACCGAGTTGGCGTGG – 3' and reverse primer; 5' – GCTCTAGACCTTTAGACGGCAGTCAG – 3'. The underlined sequence of the forward and reverse primer indicated *Not*I and *Xba*I restriction enzyme sites, respectively. Genomic DNA of *M. yongonense* DSM 45126^T was used as a template, and the DNA repair mismatch gene was amplified with *i*-MAX^{TM} II DNA polymerase (iNtRON Biotechnology, Gyeonggi-do, Korea) and a primer set as described above by the following PCR amplification condition: 5 min at 95 ℃; 40 cycles of 30 sec at 95 ℃, 30 sec at 68 ℃, and 3 min at 72 ℃; 5 min at 72 ℃. The PCR amplicon was digested with *Not*I and *Xba*I restriction enzymes and ligated into pMV306 vector [102-104] that was also digested with the same enzyme.

The pMV306 vector comprising the DNA mismatch repair gene was electroporated into competent *M. smegmatis* mc^2 155 using the GenePulser II electroporation apparatus (Bio-Rad, Hercules, CA, USA) [105]. Transformants were cultured in Middlebrook 7H9 broth (Difco, Detroit, MI, USA) containing 10 % ADC (albumin-dextrose-catalase; Difco) for 3 hours and plated onto Middlebrook 7H10 agar plate (supplemented with OADC, oleic acid-albumin-dextrose-catalase; Difco) containing 100 μg/ ml of kanamycin.

5. Examine the frequency of homologous recombination in the recombinant *M. smegmatis* harboring DNA mismatch repair gene

Partial RNA polymerase β-subunit gene (*rpoB*) from *M. tuberculosis* was amplified by PCR using genomic DNA of *M. tuberculosis* which have mutations [at codon 522 TCG (Ser) → TTG (Leu), designated with '317' and at codon 526 CAC (His) → TAC (Tyr), designated with '309'] related with rifampin resistance [106]. Each DNA fragment of '317' and '309' was provided from the Korean Institute of Tuberculosis (KIT). The partial *rpoB* gene fragments (684 bp) were amplified with a primer set as follows: forward primer, 5' – CGGGATCCCGTCGGTCGCTATAAGGTCAACA – 3' and reverse primer, 5' – CCCAAGCTTCTCGTCGGCGGTCAGGTA – 3'. The underlined sequence of the forward and reverse primers indicated *Bam*HI and *Hind*III, respectively. The PCR amplification condition was as follows: 5 min at 95 °C; 30 cycles of 30 sec at 95 °C, 30 sec at 63 °C, and 45 sec at 72 °C; 5 min at 72 °C. The amplified fragment was then cloned into the *Bam*HI and *Hind*III sites of pSE100 [107] to construct pSE100-309 and -317.

Each of pSE100-309 and -317 vector was then electroporated into each of *M. smegmatis* strain transformed with the present DNA mismatch repair gene or an empty vector pMV306 as described above. The transformants were then plated onto 7H10 agar plate with 50 μg/ml of hygromycin and incubated for 72 hours at 37 °C. After that, colonies were picked and suspended in 7H9 broth with 50 μg/ml of hygromycin and cultured for 72 hours at 37 °C. The cultured bacteria was then adjusted to 0.2 OD (optical density at 600 nm) and plated onto the 7H10 agar plate with 100 μg/ml of rifampin. The number of colonies was counted after 3 days incubation and the *rpoB* gene was amplified by PCR using primers of 7940F (forward, 5' – TCAAGGAGAAGCGCTACGACC – 3') and MR (reverse, 5' –

TCGATCGGGCACATCCGG – 3') from the extracted genomic DNA of the colonies. PCR amplicons were then sequenced using the 7940F and MR primers.

RESULTS

Two putative lateral gene transferred regions in the *M. yongonense* Type I genome from *M. parascrofulaceum*

As described in Chapter 1 and 2, *M. yongonense* Type I strains (DSM 45126^T, MOTT-12 and MOTT-27) have an *rpoB* gene that might have been laterally transferred from the distantly-related scotochromogenic species *M. parascrofulaceum* (Figures 1-5C, 1-7, and 2-4A). And this difference in *rpoB* gene could be used for discrimination between the *M. yongonense* Type I (DSM 45126^T, MOTT-12 and MOTT-27) and Type II strains (MOTT-36Y and MOTT-H4Y). To find other ORFs that were laterally transferred from *M. parascrofulaceum*, and no hits in *M. yongonense* Type II strains, all the ORFs were compared and analyzed using the BLASTN and BLASTP programs. In the genomes of *M. yongonense* Type I strains, two locus which showed high sequence similarities with *M. parascrofulaceum*. The first region is *rpoBC* operon, which contains ABC transporter, *rpoB* and *rpoC* genes (corresponding to OEM_44170~44190 in *M. yongonense* DSM 45126^T) and the second region contains 57 ORFs comprised with dehydrogenase, virulence factor MCE family genes, fatty acid biosynthesis related genes, and so on (corresponding to OEM_08030~08590 in *M. yongonense* DSM 45126^T) (Table 3-1). Interestingly, *M. yongonense* Type I strains (MOTT-12 and MOTT-27) showed high sequence

similarities with *M. yongonens* DSM 45126^T (99~100 % in nucleotide sequence). Also, in these regions, *M. yongonense* Type I strains (DMS 45126^T, MOTT-12, and MOTT-27) were closely related with *M. parascrofulaceum* (in the *rpoBC* operon, 97~99 % sequence similarities and in the second region, 95~100 % sequence similarities) but not other *M. intracellulare* and *M. yongonense* Type II strains (in the *rpoBC* operon, 94~95 % sequence similarities and in the second region, 64~93 % sequence similarities) (Table 3-1).

Also, ORFs lied on the border within the putative lateral gene transferred regions were analyzed by construction phylogenetic trees. In the first region, *rpoBC* operon (OEM_44170~44190), ORFs corresponding to ABC transporter (OEM_44190) and *rpoC* (OEM_44170) (ORFs in the putative laterally transferred region from *M. parascrofulaceum*), and their neighboring ORFs (OEM_44200; sim14 and OEM_44160; endonuclease IV) from *M. yongonense* DSM 45126^T were compared with other strains. As expected, in the putative gene transferred region (from OEM_44170~44190), *M. yongonense* Type I strains (DSM 45126^T, MOTT-12 and MOTT-27) were clustered with *M. parascrofulaceum* (92 or 100% bootstrap values) (Figures 2-4A, 3-1B and C). However, in the neighboring ORFs (OEM_44200; sim14 and OEM_44160; endonuclease IV) based trees, *M. yongonense* Type I strains were closely grouped with other *M. intracellulare* (ATCC 13950^T, MOTT-02 and MOTT-64) or *M. yongonense* Type II (MOTT-36Y and MOTT-H4Y) strains (Figure 3-1A and D). Comparison of ORFs in the second putative lateral gene transferred region (OEM_08030~08590) were correlated with the result in the first transferred region (Figure 3-2).

Table 3-1. Sequence similarities between ORFs which consist of the putative transferred regions in *M. yongonense* DSM 45126^T and other *Mycobacterium* species

(A) *rpoBC* operon (OEM_44170~44190)

M. yongonense DSM 45126^T	Gene annotation	Sequence similarities (%)								
		M. intracellulare ATCC 13950^T	MOTT-02	MOTT-64	MOTT-36Y	MOTT-H4Y	MOTT-12	MOTT-27	*M. parascrofulaceum* ATCC BAA-614^T	*M. avium* 104
OEM_44170	DNA-directed RNA polymerase subunit beta prime	94	94	94	94	94	99	99	99	95
OEM_44180	DNA-directed RNA polymerase subunit beta	94	94	94	94	94	99	99	99	94
OEM_44190	sulfate/thiosulfate ABC superfamily ATP binding cassette transporter, ABC protein	95	95	94	94	94	100	100	97	93

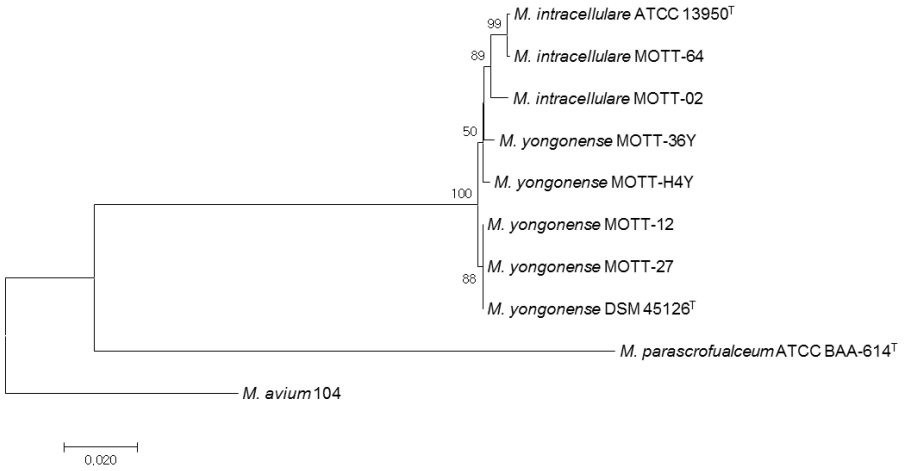
(B) OEM_08030~08590

M. yongonense DSM 45126^T	Gene annotation	Sequence similarities (%)								
		M. intracellulare ATCC 13950^T	MOTT-02	MOTT-64	MOTT-36Y	MOTT-H4Y	MOTT-12	MOTT-27	*M. parascrofulaceum* ATCC BAA-614^T	*M. avium* 104
OEM_08030	Rieske (2Fe-2S) domain protein	90	90	89	90	89	100	100	100	90
OEM_08040	amidohydrolase	91	91	91	90	90	100	100	100	85
OEM_08050	amidohydrolase	87	87	87	87	89	100	100	100	84
OEM_08060	possible 3-hydroxyisobutyrate dehydrogenase	81	81	81	82	81	100	100	100	81
OEM_08070	4-carboxymuconolactone decarboxylase	86	87	87	87	87	100	100	100	ND
OEM_08080	probable acyl-CoA dehydrogenase	87	87	87	86	86	100	100	100	86
OEM_08090	acyl-CoA dehydrogenase	85	85	85	85	90	100	100	100	87
OEM_08100	twin-arginine translocation pathway signal	85	84	85	85	83	100	100	100	87
OEM_08110	ISRSO5-transposase transposase	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08120	conserved hypothetical protein	85	85	85	85	86	100	100	100	ND
OEM_08130	conserved hypothetical protein	87	87	87	87	87	100	100	100	81
OEM_08140	cytochrome P450 family protein	85	85	85	85	85	100	100	100	85
OEM_08150	conserved hypothetical protein	89	89	89	87	88	100	100	100	90
OEM_08160	acetoacetyl-CoA reductase	85	85	85	86	85	100	100	100	87
OEM_08170	aldehyde dehydrogenase	88	88	88	88	86	100	100	99	87
OEM_08180	formyl-CoA transferase	88	88	88	88	87	100	100	100	88
OEM_08190	hypothetical protein	ND	ND	ND	ND	ND	100	100	ND	ND
OEM_08200	hypothetical protein	84	84	84	84	81	100	100	100	82
OEM_08210	ABC superfamily ATP binding cassette transporter, membrane protein	84	84	83	83	83	100	100	99	81
OEM_08220	virulence factor mce family protein	72	72	72	72	72	100	100	99	70
OEM_08230	virulence factor mce family protein	75	75	75	75	76	100	100	100	76
OEM_08240	virulence factor mce family protein	73	73	73	73	71	100	100	100	72
OEM_08250	virulence factor mce family protein	73	73	73	73	73	100	100	100	74

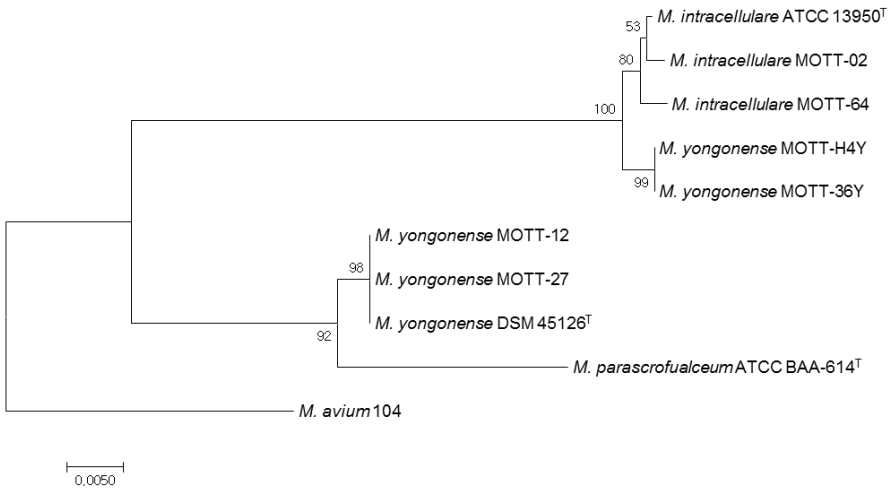
OEM_08260	virulence factor mce family protein	76	76	76	76	75	100	100	100	ND
OEM_08270	virulence factor mce family protein	73	73	73	72	73	100	100	99	74
OEM_08280	conserved hypothetical protein	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08290	twin-arginine translocation pathway signal	67	ND	ND	ND	68	100	100	100	ND
OEM_08300	conserved hypothetical protein	65	65	65	64	ND	100	100	98	67
OEM_08310	conserved hypothetical protein	73	73	73	74	ND	100	100	100	ND
OEM_08320	ISMsm2 transposase	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08330	TetR family transcriptional regulator	70	70	71	71	72	100	100	100	71
OEM_08340	conserved hypothetical protein	84	84	84	84	84	100	100	100	84
OEM_08350	IS111A/IS1328/IS1533 transposase	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08360	conserved hypothetical protein	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08370	conserved hypothetical protein	85	85	85	86	85	100	100	100	84
OEM_08380	cytochrome c oxidase subunit III family protein	84	84	84	84	83	99	99	99	85
OEM_08390	conserved hypothetical protein	76	76	76	77	74	100	100	100	79
OEM_08400	unspecific monooxygenase	84	84	84	84	85	100	100	100	85
OEM_08410	transcriptional regulator	83	83	83	81	82	100	100	100	81
OEM_08420	enoyl-CoA hydratase/isomerase	83	83	83	82	81	100	100	100	83
OEM_08430	conserved hypothetical protein	72	72	72	72	72	100	100	100	72
OEM_08440	amidohydrolase	92	92	92	92	93	100	100	100	93
OEM_08450	acyl-CoA synthetase	86	86	86	86	85	100	100	100	86
OEM_08460	conserved hypothetical protein	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08470	twin-arginine translocation pathway signal	ND	ND	ND	ND	ND	100	100	100	ND
OEM_08480	FadD family protein	81	81	81	82	82	100	100	99	82
OEM_08490	BFD family (2Fe-2S)-binding region	84	84	84	84	84	100	100	100	81
OEM_08500	indolepyruvate decarboxylase	86	86	86	86	86	100	100	99	87
OEM_08510	cyclase/dehydrase	88	88	88	89	88	100	100	100	88
OEM_08520	integrase catalytic subunit	ND	ND	ND	ND	ND	100	100	ND	ND
OEM_08530	transposase, IS4 family protein	ND	ND	ND	ND	ND	100	100	ND	ND
OEM_08540	hypothetical protein	ND	ND	ND	ND	ND	100	100	ND	ND
OEM_08550	fatty-acid-CoA racemase	86	86	86	86	86	100	100	100	85
OEM_08560	cyclase/dehydrase	87	88	87	87	87	100	100	100	87
OEM_08570	cyclase/dehydrase	89	89	89	88	88	100	100	100	89
OEM_08580	possible cysteine-S-conjugate beta-lyase	86	86	86	85	86	100	100	100	85
OEM_08590	fatty oxidation complex	93	92	93	93	93	100	100	95	93

ND, not detected.

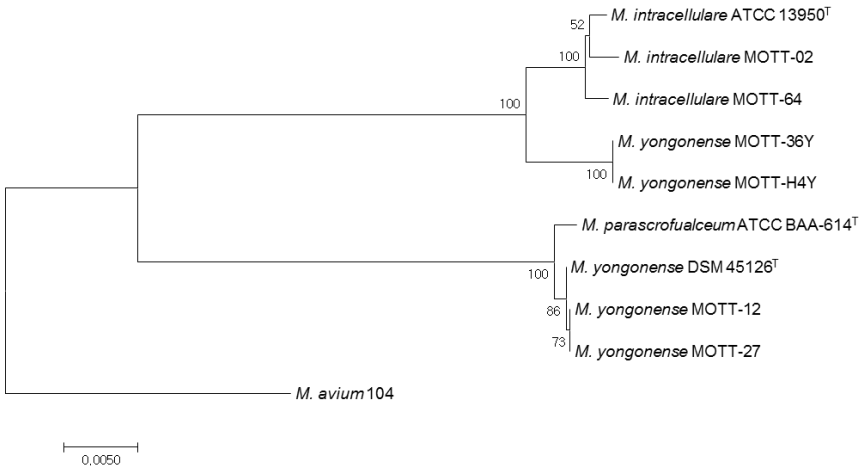
(A) OEM_44200 (sim14)



(B) OEM_44190 (ABC transporter)



(C) OEM_44170 (*rpoC*)



(D) OEM_44160 (endonuclease IV)

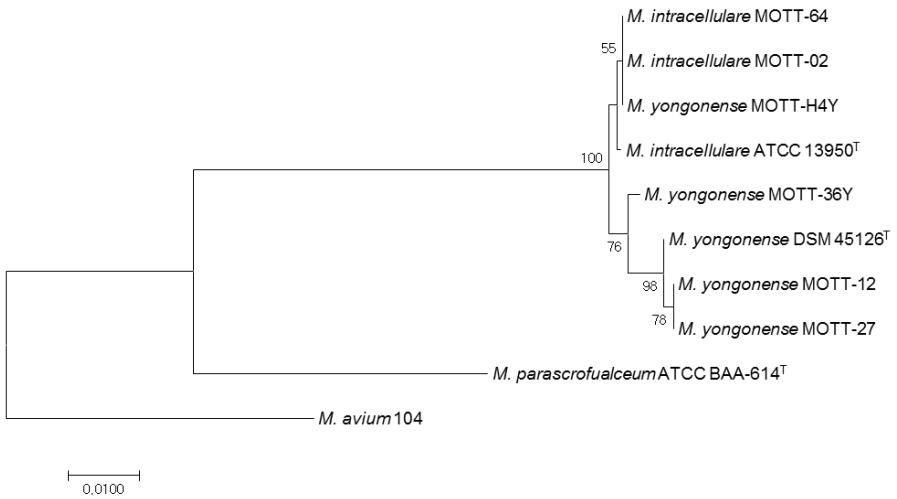
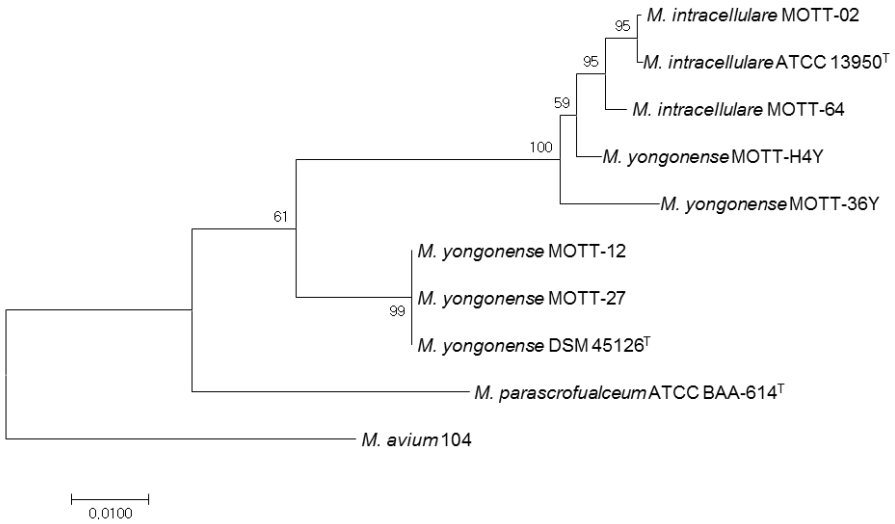


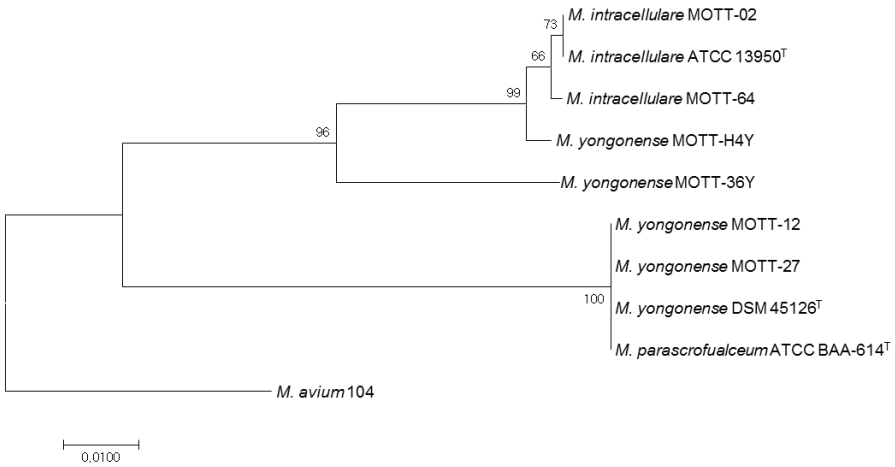
Figure 3-1. Phylogenetic analysis based on the ORFs lied on the border within the first putative lateral gene transferred region (OEM_44170~44190). (A and D) Trees based on the adjacent ORFs (OEM_44200; sim14 and OEM_44160; endonuclease IV) from the putative lateral gene transferred region. (B and C) Trees based on the putative lateral transferred ORFs (OEM_44190; ABC transporter and OEM_44170; *rpoC*) in the *M. yongonense* Type I strains from *M. parascrofulaceum*. The bootstrap values

were calculated from 1,000 replications and values <50% were not shown. The bar indicates the numbers of base substitutions per site. ORFs from *M. avium* 104 were used as an out-group.

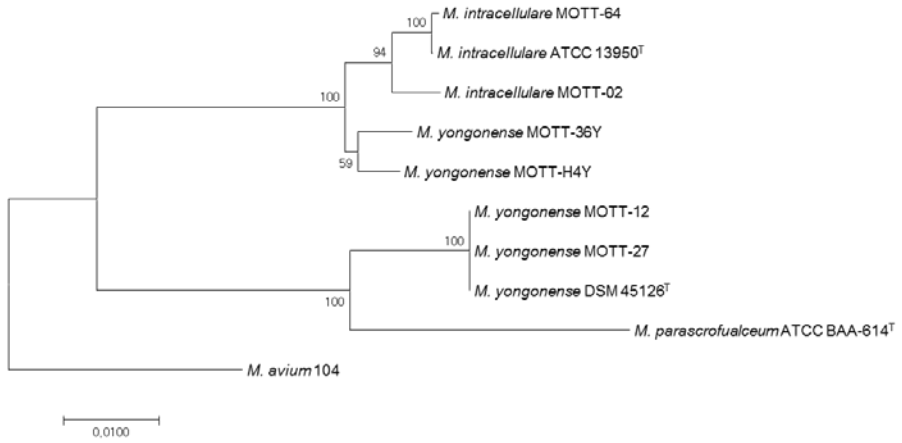
(A) OEM_08020 (hypothetical protein)



(B) OEM_08030 [Rieske (2Fe-2S) domain protein]



(C) OEM_08590 (fatty oxidation complex)



(D) OEM_08600 (putative acyl-CoA dehydrogenase)

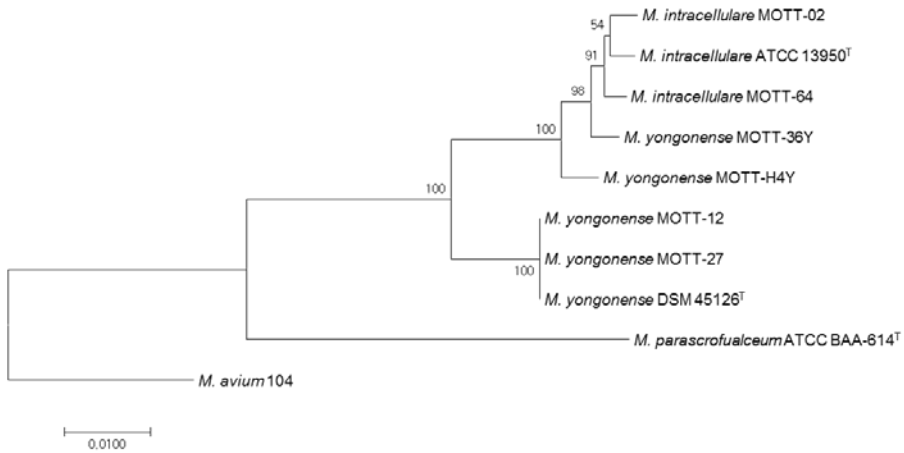


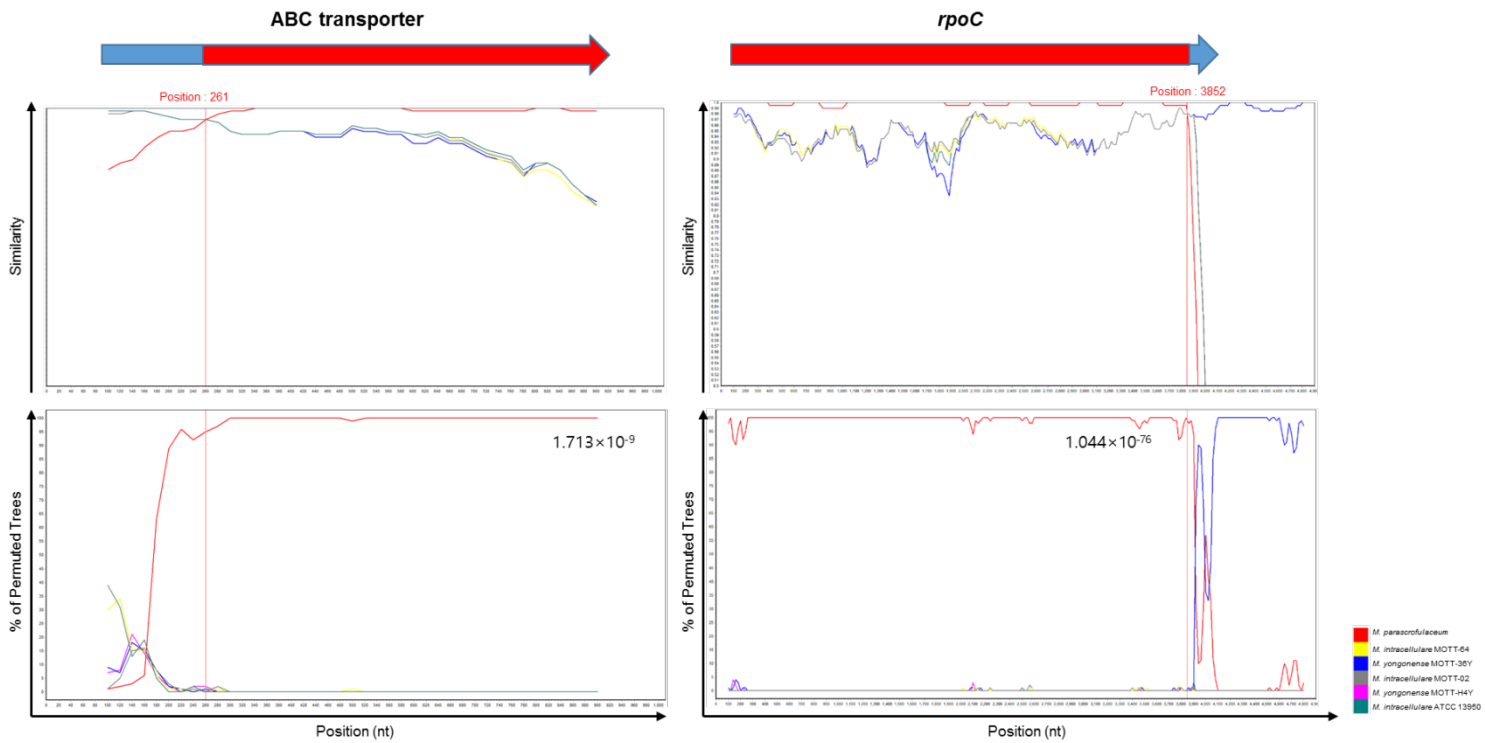
Figure 3-2. Phylogenetic analysis based on the ORFs lied on the border within the second putative lateral gene transferred region (OEM_08030~08590). (A and D) Trees based on the adjacent ORFs (OEM_08020; hypothetical protein and OEM_08600; putative acyl-CoA dehydrogenase) from the putative lateral gene transferred region. (B and C) Trees based on the putative lateral transferred ORFs [OEM_08030; Rieske (2Fe-2S) domain protein and OEM_08590; fatty oxidation complex) in the *M. yongonense* Type I strains from *M. parascrofulaceum.* The bootstrap values were

calculated from 1,000 replications and values <50% were not shown. The bar indicates the numbers of base substitutions per site. ORFs from *M. avium* 104 were used as an out-group.

Identification of homologous recombination sites in the *M. yongonense* Type I genome from *M. parascrofulaceum*

To detect evidence of lateral gene transfer (LGT) events and identify potential breakpoints for any LGT events, SimPlot analysis was conducted using the partial sequences in the regions of putative LGT events occurred described above (The first region of *rpoBC* operon and the second region containing ORFs from OEM_08030 to 08590). Correlated with the phylogenetic analysis, *M. yongonense* Type I strains showed high similarity to *M. parascrofulaceum* in the putative LGT regions (Figure 3-3). Also, in order to determine the breakpoints of possible recombination events, BootScan analysis was carried out for the *M. yongonense* Type I strains. In the first recombination region, the potential locations of the recombination breakpoints were determined at 261 nucleotide (nt) of OEM_44190 (ABC transporter) and at 3,852 nt of OEM_44170 (*rpoC*). In the case of the second region, the recombination breakpoints were detected at 355 nt of OEM_08030 and at 408 nt of OEM_08590 (Figure 3-3).

(A) First LGT region (*rpoBC* operon; OEM_44170~44190)



(B) Second LGT region (OEM_08030~08590)

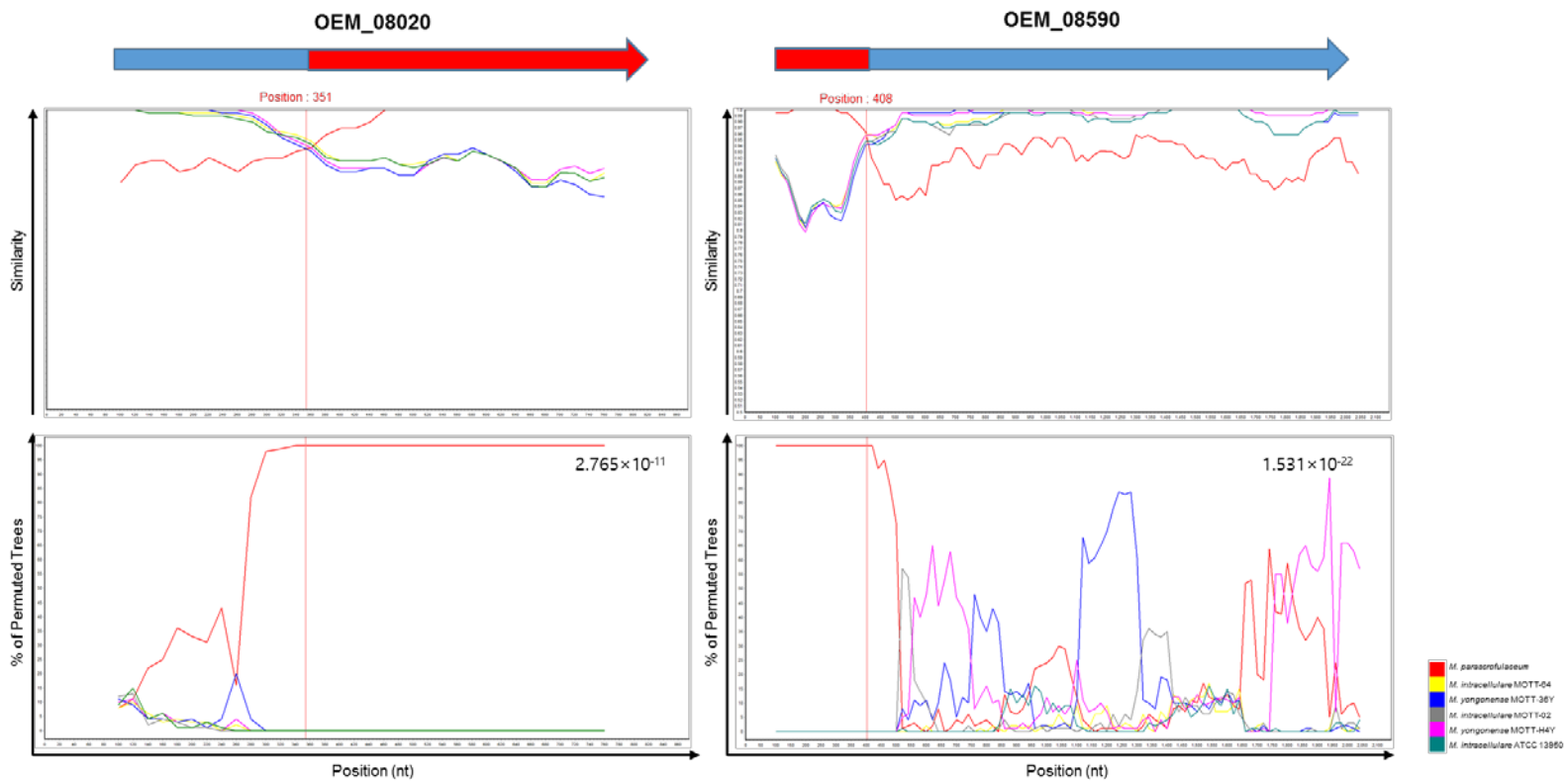
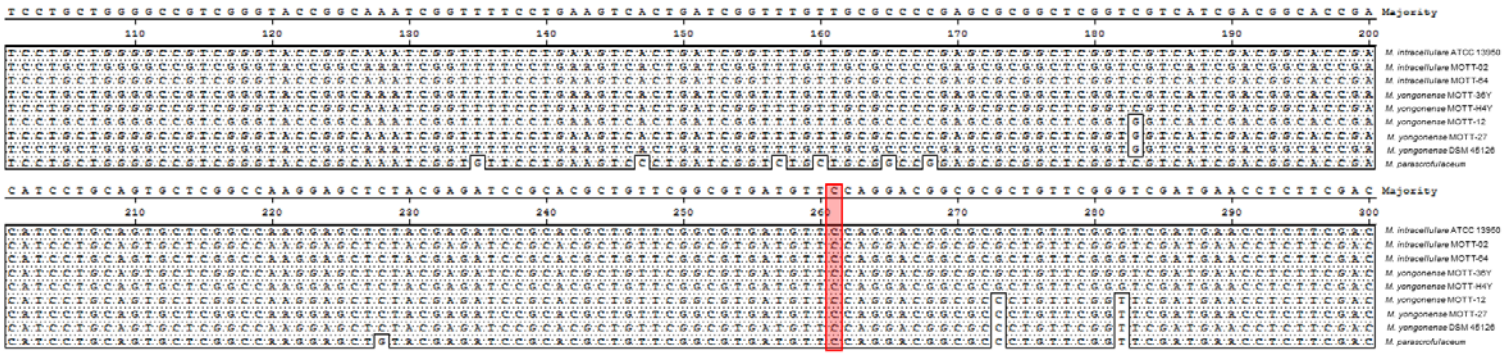
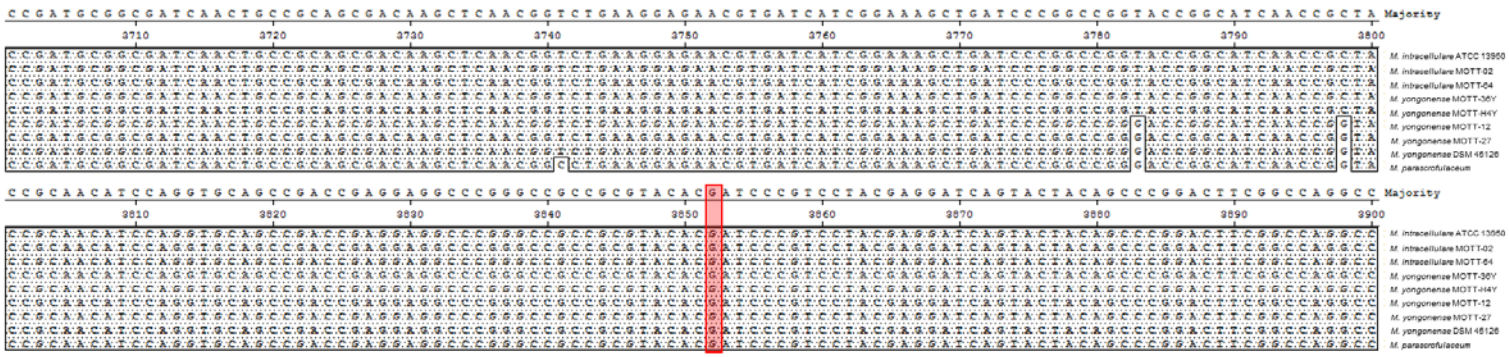


Figure 3-3. Plots of similarity of a set of sequences in the putative recombination site from *M. intracellulare* (ATCC 13950^{T}, MOTT-02 and MOTT-64) or *M. yongonense* Type II (MOTT-36Y and MOTT-H4Y) and *M. parascrofulaceum* strains to the *M. yongonense* Type I (DSM 45126^{T}, MOTT-12 and MOTT-27) strains. (A) The first recombination site (from OEM_44170 to 44190). (B) The second recombination site (from OEM_08030 to 08590). Upper panels in each figure indicates the SimPlot results of the *M. yongonense* Type I strains compared with *M. intracellulare*, *M. yongonense* Type II and *M. parascrofulaceum* strains. Lower panels in each figure indicates the BootScan results of the *M. yongonense* Type I strains compared with *M. intracellulare*, *M. yongonense* Type II and *M. parascrofulaceum* strains. Each point plotted is the percent identity within a sliding window 200 bp wide centered on the position plotted, with a step size between points of 20 bp. Detailed parameters used for the analysis are as follows; Window: 200 bp, Step: 20 bp, GapStrip: On, Reps: 100, Kimura (2-parameter), T/t: 2.0, Neighbor-joining.

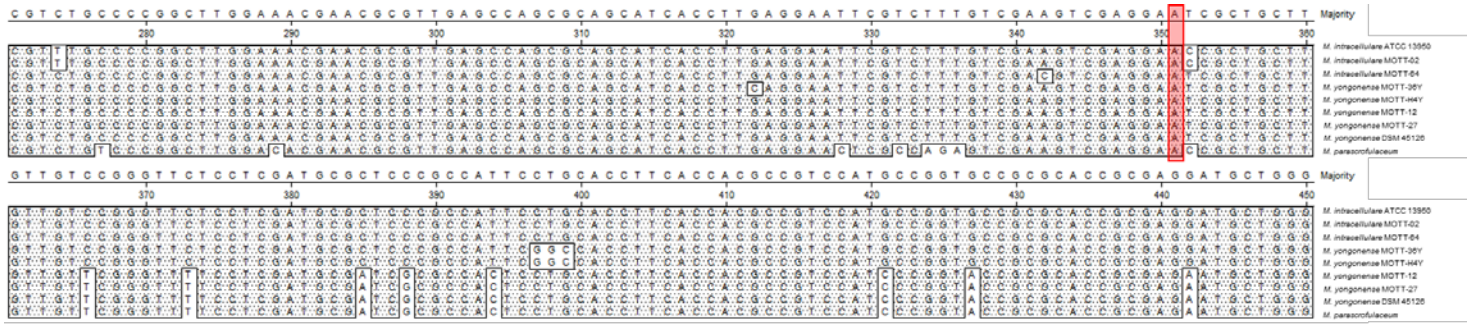
(A) ABC transporter



(B) *rpoC*



(C) OEM_08020



(D) OEM_08590

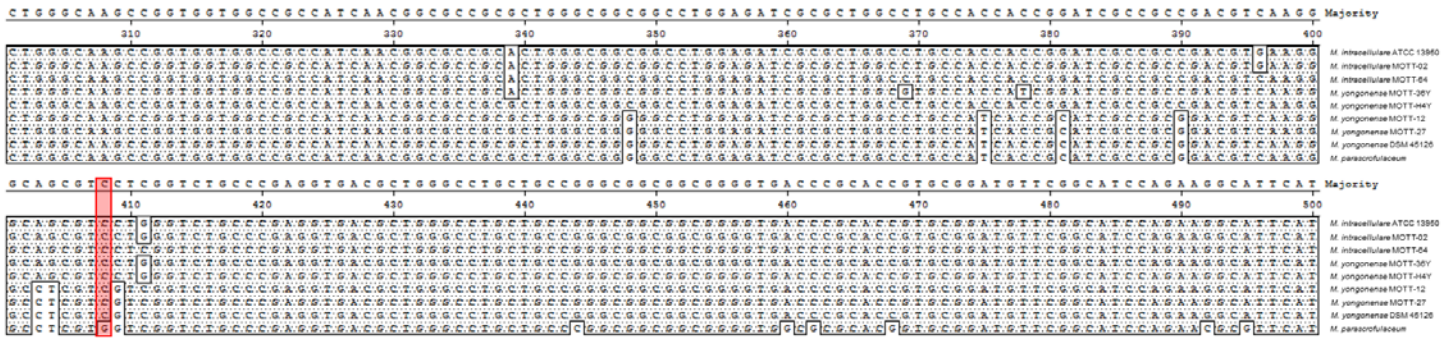


Figure 3-4. Multiple alignments of sequences corresponding to the two putative LGT regions of *M. intracellulare* (ATCC 13950^T, MOTT-02, and MOTT-64), *M. yongonense* Type I (DSM 45126^T, MOTT-12, and MOTT-27), *M. yongonense* Type II (MOTT-36Y and MOTT-H4Y) and *M. parascrofualceum* ATCC BAA-614^T strains. (A and B) The first LGT region. (C and D) The second LGT region. Red box indicates the breakpoints determined by BootStrap analysis in Figure 3-3.

DNA mismatch repair genes in the genome of *M. yongonense* Type I strains

To find out other ORFs which could differentiate *M. yongonense* strains into Type I and Type II genotype, non-mycobacterial genes were analyzed and compared among the *M. yongonense* strains. Among the non-mycobacterial genes, a locus containing 9 ORFs derived from other genus of bacteria was found in the genome of *M. yongonense* Type I strains (DSM 45126^T, MOTT-12 and MOTT-27), but not in the genome of *M. yongonense* Type II strains (MOTT-36Y and MOTT-H4Y) (Table 3-2 and Figure 3-5). In this locus, enolase (OEM_51290), NADH dehydrogenase complex (OEM_51300, 51310, 51320, 51330 and 51350), and hydrogenase subunit (OEM_51340) related genes were located (Figure 3-5). Interestingly, DNA mismatch repair genes (OEM_51400; MutS4A and 51410; MutS4B) were found in this locus. These genes have not been found in the genome of *Mycobacterium* species yet [100, 108] and grouped in MutS4 homologs (Figure 3-6).

Table 3-2. A locus containing non-mycobacterial genes in the *M. yongonense* Type I strains, but not in the Type II strains.

ORFs	Descriptions	Species	Sequence similarities (%)
OEM_51290	Enolase 2	*Nocardia nova*	70
OEM_51300	NADH/ubiquinone/plastoquinone (complex I)	*Nocardia nova*	67
OEM_51310	NADH ubiquinone oxidoreductase	*Nocardia nova*	72
OEM_51320	NADH-ubiquinone oxidoreductase	*Acidothermus cellulolyticus*	66
OEM_51330	NADH dehydrogenase	*Nocardia nova*	74
OEM_51340	hydrogenase subunit	*Nocardia nova*	74
OEM_51350	Respiratory-chain NADH dehydrogenase, subunit 1	*Nocardia nova*	72
OEM_51400	DNA mismatch repair protein MutS domain-containing protein	*Acidothermus cellulolyticus*	66
OEM_51410	DNA mismatch repair protein MutS domain protein	*Acidothermus cellulolyticus*	68

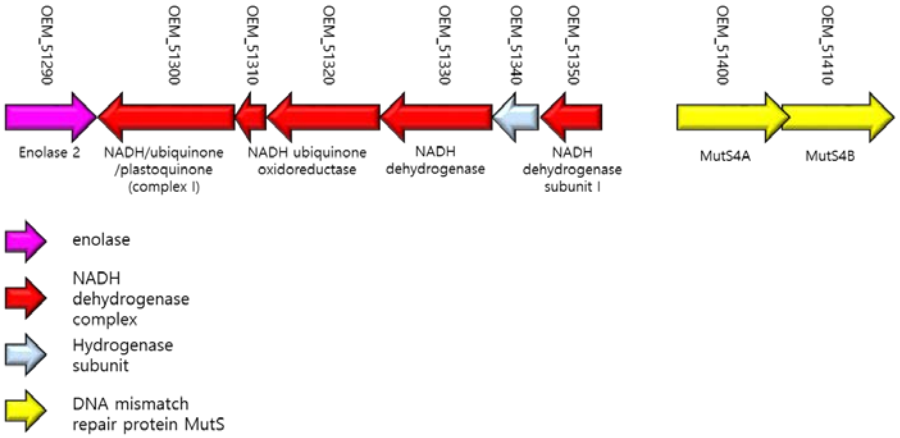


Figure 3-5. Schematic presentation of a locus containing non-mycobacterial genes in the *M. yongonense* Type I strains, but not in the Type II strains.

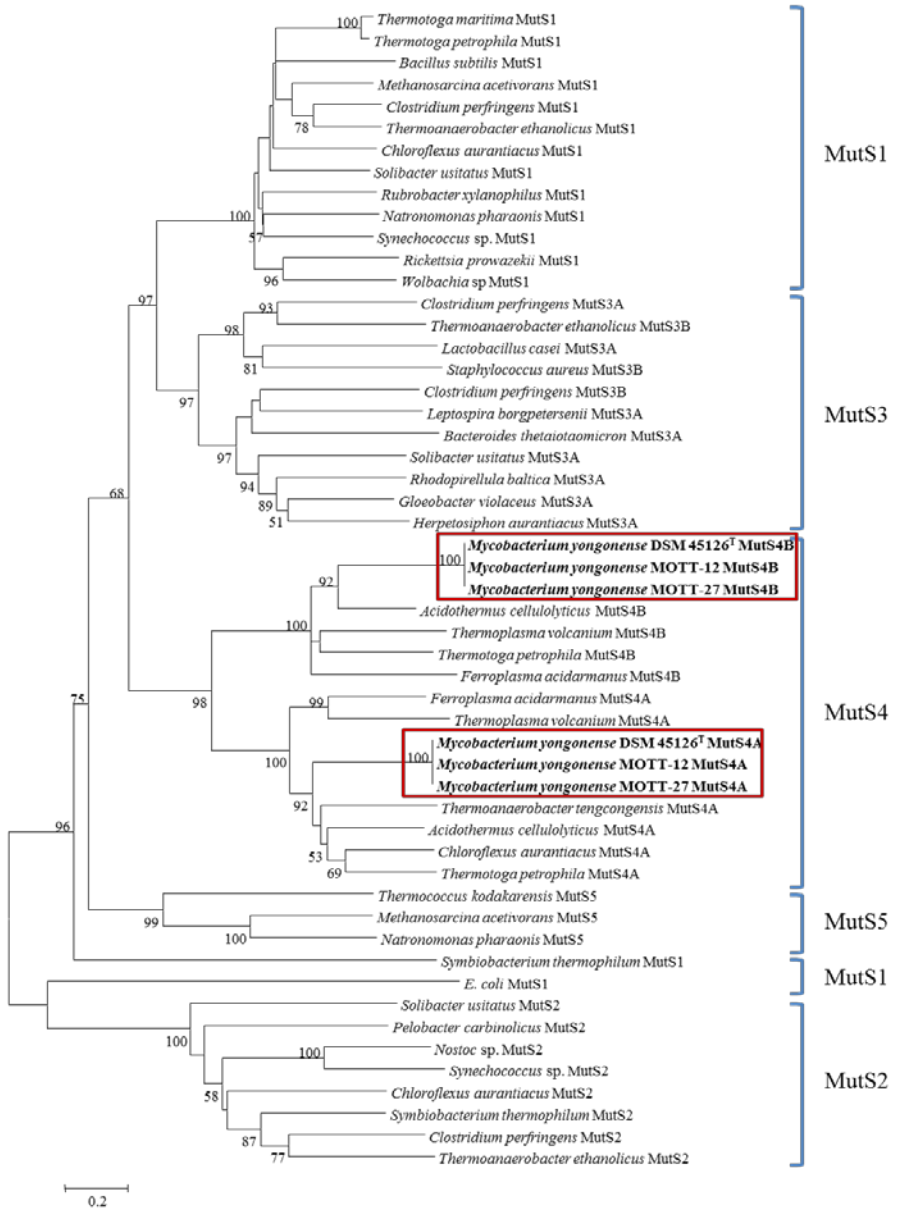


Figure 3-6. Neighbor-joining phylogenetic tree based on the DNA mismatch repair gene homologs. MutS4A and MutS4B homologs from *M. yongonense* Type I strains (DSM 45126^T, MOTT-12 and MOTT-27) were compared with DNA mismatch repair gene homologs derived from other bacterial species. The compared sequences were retrieved from GenBank database [100].

Frequency of homologous recombination in the recombinant *M. smegmatis* harboring DNA mismatch repair gene

To examine the role of DNA mismatch repair gene in the *M. yongonense* Type I strains, total DNA mismatch repair genes (OEM_51400 and 51410, 3,069 bp) and promoter region (770 bp) were amplified as described in Methods (Figure 3-7). This amplicon was cloned into pMV306 vector, and transformed into competent *M. smegmatis* to generate recombinant *M. smegmatis* harboring DNA mismatch repair gene. The recombinant *M. smegmatis* harboring DNA mismatch repair gene was confirmed by colony PCR and RT-PCR targeting the partial MutS4B sequence (from 275 bp to 597 bp in MutS4B gene, 323 bp) (Figure 3-8). Also, *M. tuberculosis* partial *rpoB* sequence (684 bp) containing mutations in codon 522 (TCG → TTG; 317) and 526 (CAC → TAC; 309) (which confers resistance to rifampin) was amplified and cloned into pSE100 vector. Each of pSE100-309 and -317 was then transformed into each recombinant *M. smegmatis* harboring DNA mismatch repair gene or an empty vector. After plating the transformed strains onto 7H10 agar medium with 100 μg/ml of rifampin, colony forming units were enumerated and from the randomly selected colonies, *rpoB* region was sequenced and analyzed.

The result showed that more colonies were formed from the *M. smegmatis* harboring the DNA mismatch repair gene than those from control strain. Also, in the multiple alignment of sequenced partial *rpoB* gene, homologous recombination occurred more frequent in the recombinant *M. smegmatis* harboring DNA mismatch repair gene. And the length of putative nucleotide sequences that underwent homologous recombination superior to those of control strain (Figures 3-9, 3-10 and Table 3-3).

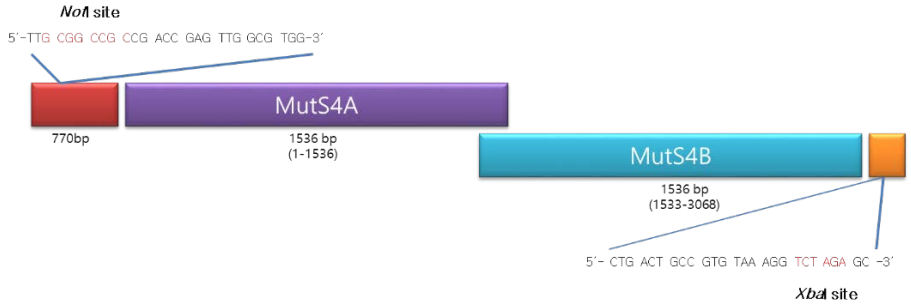
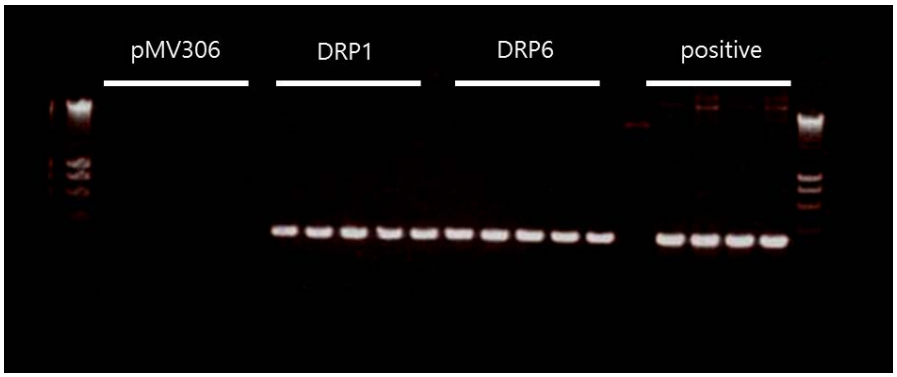


Figure 3-7. Schematic representation of DNA mismatch repair genes from *M. yongonense* DSM 45126^{T}. Primer sequences were annotated in the forward and reverse region.

(A)



(B)

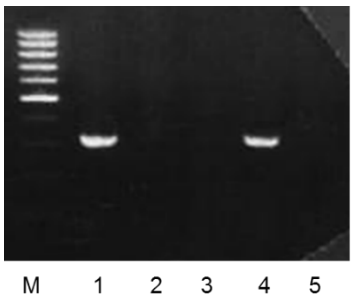


Figure 3-8. Confirmation of the recombinant *M. smegmatis* harboring DNA mismatch repair gene by colony PCR and RT-PCR. (A) Colony PCR results. pMV306, recombinant *M. smegmatis* harboring an empty vector, pMV306; DR1 and DR6, recombinant *M. smegmatis* harboring DNA mismatch repair gene; positive control, pMV306 vector cloned with DNA mismatch repair gene. (B) RT-PCR results. M, 100 bp DNA ladder; 1, *M. yongonense* DSM 45126^{T}; 2, *M. intracellulare* ATCC 13950^{T}; 3, recombinant *M. smegmatis* harboring an empty vector, pMV306; 4, recombinant *M. smegmatis* harboring DNA mismatch repair gene; 5, Negative control.

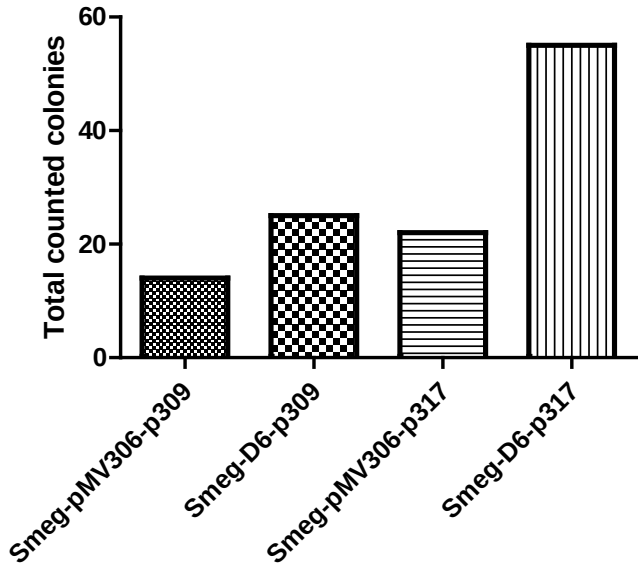


Figure 3-9. Total numbers of colonies from 7H10 agar plate containing 100 μg/ml of rifampin. Smeg-pMV306-p309, recombinant *M. smegmatis* harboring empty vector and transformed with pSE100-309 *rpoB*; Smeg-pMV306-p317, recombinant *M. smegmatis* harboring empty vector and transformed with pSE100-317 *rpoB*; Smeg-D6-p309, recombinant *M. smegmatis* harboring DNA mismatch repair gene and transformed with pSE100-309 *rpoB*; Smeg-D6-p317, recombinant *M. smegmatis* harboring DNA mismatch repair gene and transformed with pSE100-317 *rpoB*.

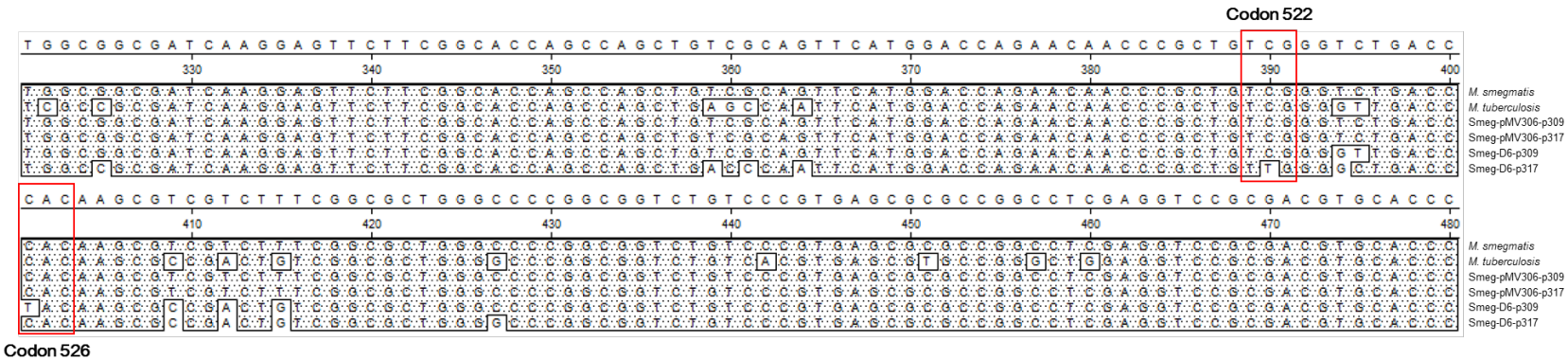


Figure 3-10. Representative multiple alignment of sequences in the putative homologous recombination sites from each transformed recombinant *M. smegmatis*. As a control, *rpoB* sequences of *M. smegmatis* and *M. tuberculosis* were also aligned. Red box indicates mutation codons which confer the resistance to rifampin in *M. tuberculosis*. Smeg-pMV306-p309, recombinant *M. smegmatis* harboring empty vector and transformed with pSE100-309 *rpoB*; Smeg-pMV306-p317, recombinant *M. smegmatis* harboring empty vector and transformed with pSE100-317 *rpoB*; Smeg-D6-p309, recombinant *M. smegmatis* harboring DNA mismatch repair gene and transformed with pSE100-309 *rpoB*; Smeg-D6-p317, recombinant *M. smegmatis* harboring DNA mismatch repair gene and transformed with pSE100-317 *rpo*

Table 3-3. Numbers and average length of putative homologous recombination sequences in the transformed recombinant *M. smegmatis* strains.

(A) Numbers of putative homologous recombination sequences

	Smeg-pMV306	Smeg-D6	*P* value [a]
522 mutation (p317)	5/16 (31.3 %)	27/36 (75 %)	0.020
526 mutation (p309)	5/12 (41.7 %)	16/24 (66.7 %)	0.151
Total	10/28 (33.3 %)	43/60 (71.7 %)	0.007

(B) Average length of putative homologous recombination sequences

	Smeg-pMV306	Smeg-D6	*P* value [a]
522 mutation (p317)	43	101.35	0.003
526 mutation (p309)	46	60.3	0.262
Total	44.5	86	0.004

[a] *P* value was calculated by Chi-Square and Student's t-tests.

DISCUSSION

Although, several papers were reported the possibilities of LGT events in mycobacterial strains, LGT events have not been clearly demonstrated in the genus of *Mycobacterium* [40, 41, 109-111]. Especially, the genome sequences of prominent pathogen, *M. tuberculosis* and *M. leprae*, indicate that endogenous gene rearrangement, duplication, or point mutation have played a major role in their evolution rather than LGT events [96, 97, 112].

An homologous recombination (HR)-associated sequence transfer mechanism has been shown to contribute to genome plasticity in various organisms [113]. However, genome-wide multi-locus analyses of *M. tuberculosis* and *M. bovis* environmenta strains have shown that the HR events in these mycobacteria may be extremely rare in natural settings [114, 115]. These results suggested that there may be tolerance mechanisms in the mycobacterial genome rejecting the recombination between moderately divergent DNA sequences. Ironically, sequence conservation of essential genes between different species could provide a better chance for an HR-associated LGT mechanism.

In this study, I found out two putative laterally transferred loci from donor bacteria, *M. parascrofualceum* into recipient bacteria, *M. yongonense* strains, especially in the genome of Type I strains. One is the *rpoBC* operon (from OEM_44170 to 44190) and another is a region including 57 genes (from OEM_08030 to 08590). The two putative HR sites observed in this study (Table 3-1, Figures 3-1 and 3-2) showed high sequence homology between *M. yongonense* and *M. parascrofulaceum*, also the breakpoint of

each putative HR sites was estimated by SimPlot and BootScan programs with statistical significance (Figure 3-3 and 3-4).

It is difficult to explain why their *rpoBC* operon or another region (including 57 genes) were replaced with those of *M. parascrofulaceum*. Because it is generally accepted that informative genes, particularly those involved in the central dogma, might be recalcitrant to LGT event during evolution. Both recipient and donor bacterial strains involved in LGT of informative genes, such as 16S rRNA or RNA polymerase genes, have not been disclosed, despite the existence of some clues based on phylogenetic analysis [108]. The most likely explanation for this question is that the transferred genes (*rpoBC* operon and another 57 genes) from *M. parascrofulaceum* could provide better survival in macrophages, enhancement the virulence, or metabolic advantages in *M. yongonense* Type I strains. In the case of *rpoBC* operon region, this region might be laterally transferred for the effective transcription of other LGT region (including 57 genes) from *M. parascrofulaceum*. Also, virulence factor MCE (mammalian cell entry) family genes (OEM_08200~08270) and seven genes related with lipid transport and metabolism (OEM_08060, 08080, 08090, 08160, 08420, 08450, and 08590) were found in the LGT locus containing 57 genes. These MCE family genes and Actually, *M. yongonense* Type I strains show significantly higher intracellulare survival than *M. intracellulare* strains in macrophage cell line, J774A.1 (data not shown). However, to address this issue, elaborate examinations should be conducted in the further study.

These two laterally transferred regions (*rpoBC* operon and another 57 genes) could discriminate the *M. yongonense* strains into Type I (DSM 45126^{T}, MOTT-12 and MOTT-27 strains) and Type II (MOTT-36Y and MOTT-H4Y strains) genotypes. Together with these regions, another region including 9 non-mycobacterial genes from

other genus of bacteria (Figure 3-5) could differentiate the *M. yongonense* Type I and Type II strains. Interestingly, in this locus, DNA mismatch repair gene (*mutS4*), which was known to lack in the genome of mycobacteria [96, 97, 99, 100, 108], was found (Figures 3-5 and 3-6). This gene is responsible to correct DNA replication errors and reject recombination between moderately divergent (up to ~10%) DNA sequences [98, 99, 108]. However, because of lack the highly conserved *mutS*-based DNA mismatch repair genes in mycobacterial strains, the function of the gene in the mycobacterial strains is not clear.

To evaluate the possible function of DNA mismatch repair gene in the genome of *M. yongonense* Type I strains, a recombinant *M. smegmatis* harboring DNA mismatch repair gene (Smeg-D6) was constructed (Figures 3-7 and 3-8). Interestingly, Smeg-D6 showed more frequent homologous recombination with partial *rpoB* sequence-associated rifampin resistant (684 bp of *M. tuberculosis rpoB* gene sequence) under rifampin pressure than control *M. smegmatis* (Smeg-pMV306). This result suggested that the DNA mismatch repair gene in the genome of *M. yongonense* Type I strains could facilitate homologous recombination event under harsh environment such as antibiotics exposure.

Furthermore, in the region including non-mycobacterial ORFs, enolase and NADH dehydrogenase subunit genes were also found. Cell surface enolase of group A streptococci, such as *Streptococcus pyogenes* and *S. pneumoniae,* has been focused due to its role in the virulence of these organisms [116]. NADH dehydrogenase subunit genes are known to transport electrons from NADH to the quinone pool (in the case of *Mycobacterium* species, menaquinone). In this process, protons are translocated across the membrane, and the result, proton motive force is generated.

These genes could be activated in the low pH environment [117]. With regard to this regult, additional electron transfer related genes (NADH dehydrogenase subunit genes) from other genus might give *M. yongonense* the characteristics of resistance in low pH environment (Table 1-4). Aside from this, one of the subunits of NADH dehydrogenase, *nuoG*, is critical for host macrophage apoptosis inhibition and virulence in mouse [118]. However, the relationship of these genes and virulence of *M. yongonense* should be verified in the further study.

Finally, taking into collective consideration of all the data in this study, previously *M. intracellulare* INT-5 strains proved to be closely related to *M. yongonense*, so the INT-5 strains should be taxonomically located on the *M. yongonense* (*M. yongonense* Type II genotype). Also, probably, LGT events of the *rpoBC* operon and another 57 genes from *M. parascrofulaceum* into a common ancestor of *M. yongonense* Type II strains may contribute to speciation into *M. yongonense* Type I strains. And this LGT event might be facilitated by DNA mismatch repair gene which was acquired by accident, under harsh environment. Also, enolase and NADH dehydrogenase subunit genes acquired from other genus may contribute to the pathogenicity of *M. yongonense* strains, however, it is necessary to examine the actual function of these genes (Figure 3-11).

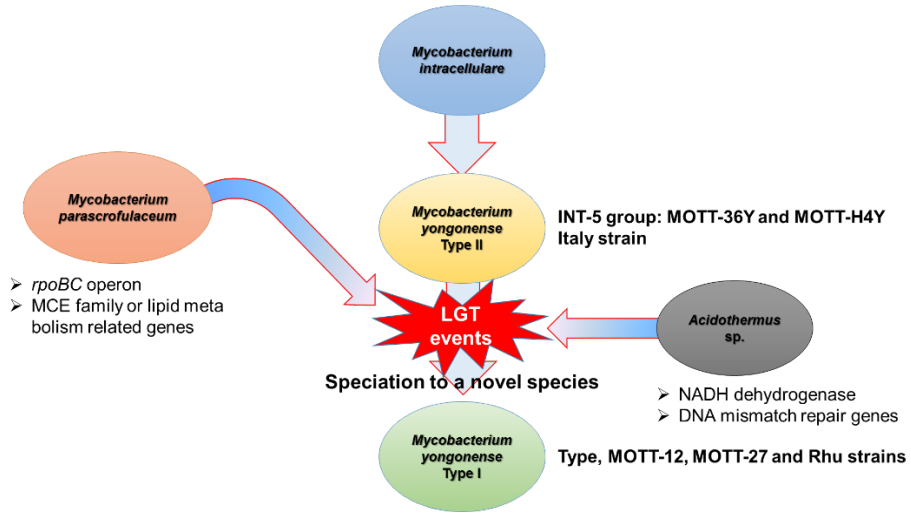


Figure 3-11. Schematic diagram for the process of speciation to a novel mycobacterial species, *M. yongonense*. Previously *M. intracellulare* INT-5 strains proved to be closely related to *M. yongonense*, so the INT-5 strains should be taxonomically located on the *M. yongonense* (*M. yongonense* Type II genotype). Probably, LGT events of the *rpoBC* operon and another 57 genes from *M. parascrofulaceum* into a common ancestor of *M. yongonense* Type II strains may contribute to speciation into *M. yongonense* Type I strains.

REFERENCES

1. Hoefsloot W, van Ingen J, Andrejak C, Angeby K, Bauriaud R, Bemer P, et al. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. Eur Respir J. 2013;42(6):1604-13.

2. Simons S, van Ingen J, Hsueh PR, Nguyen VH, Dekhuijzen PNR, Boeree MJ, et al. Nontuberculous Mycobacteria in Respiratory Tract Infections, Eastern Asia. Emerg Infect Dis. 2011;17(3):343-9.

3. Ryoo SW, Shin S, Shim MS, Park YS, Lew WJ, Park SN, et al. Spread of nontuberculous mycobacteria from 1993 to 2006 in Koreans. Journal of clinical laboratory analysis. 2008;22(6):415-20. J Clin Lab Anal. 2008;22(6):415-20.

4. Falkinham JO, 3rd. Epidemiology of infection by nontuberculous mycobacteria. Clinical microbiology reviews. 1996;9(2):177-215. Clin Microbiol Rev. 1996;9(2):177-215.

5. Inderlied CB, Kemper CA, Bermudez LE. The *Mycobacterium avium* complex. Clin Microbiol Rev. 1993;6(3):266-310.

6. Turenne CY, Wallace R, Jr., Behr MA. *Mycobacterium avium* in the postgenomic era. Clin Microbiol Rev. 2007;20(2):205-29.

7. Thorel MF, Krichevsky M, Levy-Frebault VV. 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. Int J Syst Bacteriol. 1990;40(3):254-60.

8. Mijs W, de Haas P, Rossau R, Van der Laan T, Rigouts L, Portaels F, et al. Molecular evidence to support a proposal to reserve the designation *Mycobacterium avium* subsp. *avium* for bird-type isolates and '*M. avium* subsp. *hominissuis*' for the human/porcine type of *M. avium*. Int J Syst Evol Microbiol. 2002;52(Pt 5):1505-18.

9. Murcia MI, Tortoli E, Menendez MC, Palenque E, Garcia MJ. *Mycobacterium colombiense* sp. nov., a novel member of the *Mycobacterium avium* complex and description of MAC-X as a new ITS genetic variant. Int J Syst Evol Microbiol. 2006;56(Pt 9):2049-54.

10. van Ingen J, Boeree MJ, Kosters K, Wieland A, Tortoli E, Dekhuijzen PN, et al. Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. Int J Syst Evol Microbiol. 2009;59(Pt 9):2277-82.

11. Bang D, Herlin T, Stegger M, Andersen AB, Torkko P, Tortoli E, et al. *Mycobacterium arosiense* sp. nov., a slowly growing, scotochromogenic species causing osteomyelitis in an immunocompromised child. Int J Syst Evol Microbiol. 2008;58(Pt 10):2398-402.

12. Ben Salah I, Cayrou C, Raoult D, Drancourt M. *Mycobacterium marseillense* sp. nov., *Mycobacterium timonense* sp. nov. and *Mycobacterium bouchedurhonense* sp. nov., members of the *Mycobacterium avium* complex. Int J Syst Evol Microbiol. 2009;59(Pt 11):2803-8.

13. Tortoli E, Rindi L, Garcia MJ, Chiaradonna P, Dei R, Garzelli C, et al. Proposal to elevate the genetic variant MAC-A, included in the *Mycobacterium avium* complex, to species rank as *Mycobacterium chimaera* sp. nov. Int J Syst Evol Microbiol. 2004;54(Pt 4):1277-85.

14. Park JH, Shim TS, Lee SA, Lee H, Lee IK, Kim K, et al. Molecular characterization of *Mycobacterium intracellulare*-related strains based on the sequence analysis of *hsp65*, internal transcribed spacer and 16S rRNA genes. J Med Microbiol. 2010;59(Pt 9):1037-43.

15. Iinuma Y, Ichiyama S, Hasegawa Y, Shimokata K, Kawahara S, Matsushima T. Large-restriction-fragment analysis of *Mycobacterium kansasii* genomic DNA and its application in molecular typing. J Clin Microbiol. 1997;35(3):596-9.

16. Mazurek GH, Hartman S, Zhang Y, Brown BA, Hector JS, Murphy D, et al. Large DNA restriction fragment polymorphism in the *Mycobacterium avium-M. intracellulare* complex: a potential epidemiologic tool. J Clin Microbiol. 1993;31(2):390-4.

17. Green EP, Tizard ML, Moss MT, Thompson J, Winterbourne DJ, McFadden JJ, et al. Sequence and characteristics of IS*900*, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. Nucleic Acids Res. 1989;17(22):9063-73.

18. Kunze ZM, Portaels F, McFadden JJ. Biologically distinct subtypes of *Mycobacterium avium* differ in possession of insertion sequence IS*901*. J Clin Microbiol. 1992;30(9):2366-72.

19. Moss MT, Malik ZP, Tizard MLV, Green EP, Sanderson JD, Hermontaylor J. IS*902*, an Insertion Element of the Chronic-Enteritis-Causing *Mycobacterium-Avium* Subsp *Silvaticum*. J Gen Microbiol. 1992;138:139-45.

20. Inagaki T, Nishimori K, Yagi T, Ichikawa K, Moriyama M, Nakagawa T, et al. Comparison of a variable-number tandem-repeat (VNTR) method for typing

Mycobacterium avium with mycobacterial interspersed repetitive-unit-VNTR and IS*1245* restriction fragment length polymorphism typing. J Clin Microbiol. 2009;47(7):2156-64.

21. Thibault VC, Grayon M, Boschiroli ML, Hubbans C, Overduin P, Stevenson K, et al. New variable-number tandem-repeat markers for typing *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* strains: comparison with IS*900* and IS*1245* restriction fragment length polymorphism typing. J Clin Microbiol. 2007;45(8):2404-10.

22. Dauchy FA, Degrange S, Charron A, Dupon M, Xin Y, Bebear C, et al. Variable-number tandem-repeat markers for typing *Mycobacterium intracellulare* strains isolated in humans. BMC Microbiol. 2010;10:93.

23. Linton CJ, Jalal H, Leeming JP, Millar MR. Rapid discrimination of *Mycobacterium tuberculosis* strains by random amplified polymorphic DNA analysis. J Clin Microbiol. 1994;32(9):2169-74.

24. Kim BJ, Lee KH, Park BN, Kim SJ, Bai GH, Kook YH. Differentiation of mycobacterial species by PCR-restriction analysis of DNA (342 base pairs) of the RNA polymerase gene (*rpoB*). J Clin Microbiol. 2001;39(6):2102-9.

25. Pfaller SL, Aronson TW, Holtzman AE, Covert TC. Amplified fragment length polymorphism analysis of *Mycobacterium avium* complex isolates recovered from southern California. J Med Microbiol. 2007;56(Pt 9):1152-60.

26. Wasem CF, McCarthy CM, Murray LW. Multilocus enzyme electrophoresis analysis of the *Mycobacterium avium* complex and other mycobacteria. J Clin Microbiol. 1991;29(2):264-71.

27. Yakrus MA, Reeves MW, Hunter SB. Characterization of isolates of *Mycobacterium avium* serotypes 4 and 8 from patients with AIDS by multilocus enzyme electrophoresis. J Clin Microbiol. 1992;30(6):1474-8.

28. Feizabadi MM, Robertson ID, Cousins DV, Dawson DJ, Hampson DJ. Use of multilocus enzyme electrophoresis to examine genetic relationships amongst isolates of *Mycobacterium intracellulare* and related species. Microbiology. 1997;143:1461-9.

29. Baker L, Brown T, Maiden MC, Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis. 2004;10(9):1568-77.

30. Kim BJ, Choi BS, Lim JS, Choi IY, Lee JH, Chun J, et al. Complete genome sequence of *Mycobacterium intracellulare* strain ATCC 13950(T). J Bacteriol. 2012;194(10):2750.

31. Kim BJ, Choi BS, Choi IY, Lee JH, Chun J, Hong SH, et al. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-36Y, belonging to the INT5 genotype. J Bacteriol. 2012;194(15):4141-2.

32. Kim BJ, Choi BS, Lim JS, Choi IY, Kook YH. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-64, belonging to the INT1 genotype. J Bacteriol. 2012;194(12):3268.

33. Kim BJ, Choi BS, Lim JS, Choi IY, Lee JH, Chun J, et al. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-02. J Bacteriol. 2012;194(10):2771.

34. Lee H, Kim BJ, Kim K, Hong SH, Kook YH. Whole-Genome Sequence of *Mycobacterium intracellulare* Clinical Strain MOTT-H4Y, Belonging to INT5 Genotype. Genome Announc. 2013;1(1).

35. Tortoli E, Kroppenstedt RM, Bartoloni A, Caroli G, Jan I, Pawlowski J, et al. *Mycobacterium tusciae* sp. nov. Int J Syst Bacteriol. 1999;49 Pt 4:1839-44.

36. Yajko DM, Chin DP, Gonzalez PC, Nassos PS, Hopewell PC, Reingold AL, et al. *Mycobacterium-Avium* Complex in Water, Food, and Soil Samples Collected from the Environment of Hiv-Infected Individuals. J Acquir Immune Defic Syndr Hum Retrovirol. 1995;9(2):176-82.

37. Saini V, Raghuvanshi S, Talwar GP, Ahmed N, Khurana JP, Hasnain SE, et al. Polyphasic taxonomic analysis establishes *Mycobacterium indicus pranii* as a distinct species. PLoS One. 2009;4(7):e6263.

38. Koh WJ, Kwon OJ, Jeon K, Kim TS, Lee KS, Park YK, et al. Clinical significance of nontuberculous mycobacteria isolated from respiratory specimens in Korea. Chest. 2006;129(2):341-8.

39. Koh WJ, Kwon OJ, Lee KS. Diagnosis and treatment of nontuberculous mycobacterial pulmonary diseases: a Korean perspective. J Korean Med Sci. 2005;20(6):913-25.

40. Macheras E, Roux AL, Bastian S, Leao SC, Palaci M, Sivadon-Tardy V, et al. Multilocus sequence analysis and *rpoB* sequencing of *Mycobacterium abscessus* (sensu lato) strains. J Clin Microbiol. 2011;49(2):491-9.

41. Simmon KE, Brown-Elliott BA, Ridge PG, Durtschi JD, Mann LB, Slechta ES, et al. *Mycobacterium chelonae-abscessus* complex associated with sinopulmonary disease, Northeastern USA. Emerg Infect Dis. 2011;17(9):1692-700.

42. Rogall T, Flohr T, Bottger EC. Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. J Gen Microbiol. 1990;136(9):1915-20.

43. Springer B, Stockman L, Teschner K, Roberts GD, Bottger EC. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. J Clin Microbiol. 1996;34(2):296-303.

44. McNabb A, Eisler D, Adie K, Amos M, Rodrigues M, Stephens G, et al. Assessment of partial sequencing of the 65-kilodalton heat shock protein gene (*hsp65*) for routine identification of *Mycobacterium* species isolated from clinical sources. J Clin Microbiol. 2004;42(7):3000-11.

45. Adekambi T, Colson P, Drancourt M. *rpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. J Clin Microbiol. 2003;41(12):5699-708.

46. Ben Salah I, Adekambi T, Raoult D, Drancourt M. *rpoB* sequence-based identification of *Mycobacterium avium* complex species. Microbiology. 2008;154(Pt 12):3715-23.

47. Kent PT, Kubica GP. Public Health Mycobacteriology: a Guide for the Level III Laboratory: Centers for Disease Control and Prevention; 1985.

48. Minnikin DE. Isolation and Purification of Mycobacterial Wall Lipids. Chichester: John Wiley & Sons.; 1988.

49. Butler WR, Thibert L, Kilburn JO. Identification of *Mycobacterium-Avium* Complex Strains and Some Similar Species by High-Performance Liquid-Chromatography. J Clin Microbiol. 1992;30(10):2698-704.

50. Perez E, Constant P, Lemassu A, Laval F, Daffe M, Guilhot C. Characterization of three glycosyltransferases involved in the biosynthesis of the

phenolic glycolipid antigens from the *Mycobacterium tuberculosis* complex. J Biol Chem. 2004;279(41):42574-83.

51. Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, Chae GT, et al. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). J Clin Microbiol. 1999;37(6):1714-20.

52. Adekambi T, Drancourt M. Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. Int J Syst Evol Microbiol. 2004;54:2095-105.

53. Kim H, Kim SH, Shim TS, Kim MN, Bai GH, Park YG, et al. Differentiation of *Mycobacterium* species by analysis of the heat-shock protein 65 gene (*hsp65*). Int J Syst Evol Microbiol. 2005;55(Pt 4):1649-56.

54. Frothingham R, Wilson KH. Sequence-based differentiation of strains in the *Mycobacterium avium* complex. J Bacteriol. 1993;175(10):2818-25.

55. Saitou N, Nei M. The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. Mol Biol Evol. 1987;4(4):406-25.

56. Fitch WM. Toward Defining Course of Evolution - Minimum Change for a Specific Tree Topology. Syst Zool. 1971;20(4):406-416.

57. Jukes TH, Cantor CR. In Mammalian Protein Metabolism. Munro HN, editor. New York: Academic Press; 1969.

58. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 2008;9(4):299-306.

59. Felsenstein J. Confidence-Limits on Phylogenies - an Approach Using the Bootstrap. Evolution. 1985;39(4):783-91.

60. Crump JA, van Ingen J, Morrissey AB, Boeree MJ, Mavura DR, Swai B, et al. Invasive disease caused by nontuberculous mycobacteria, Tanzania. Emerg Infect Dis. 2009;15(1):53-5.

61. Roth A, Fischer M, Hamid ME, Michalke S, Ludwig W, Mauch H. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. J Clin Microbiol. 1998;36(1):139-47.

62. Turenne CY, Cook VJ, Burdz TV, Pauls RJ, Thibert L, Wolfe JN, et al. *Mycobacterium parascrofulaceum* sp. nov., novel slowly growing, scotochromogenic clinical isolates related to *Mycobacterium simiae*. Int J Syst Evol Microbiol. 2004;54(Pt 5):1543-51.

63. Adekambi T, Shinnick TM, Raoult D, Drancourt M. Complete *rpoB* gene sequencing as a suitable supplement to DNA-DNA hybridization for bacterial species and genus delineation. Int J Syst Evol Microbiol. 2008;58:1807-14.

64. Mun HS, Kim HJ, Oh EJ, Kim H, Park YG, Bai GH, et al. Direct application of *Ava*II PCR restriction fragment length polymorphism analysis (*Ava*II PRA) targeting 644 bp heat shock protein 65 (*hsp65*) gene to sputum samples. Microbiol Immunol. 2007;51(1):105-10.

65. Kim HJ, Mun HS, Kim H, Oh EJ, Ha Y, Bai GH, et al. Differentiation of Mycobacterial species by *hsp65* duplex PCR followed by duplex-PCR-based restriction analysis and direct sequencing. J Clin Microbiol. 2006;44(11):3855-62.

66. Stackebrandt E, Goebel BM. A Place for DNA-DNA Reassociation and 16s Ribosomal-Rna Sequence-Analysis in the Present Species Definition in Bacteriology. Int J Syst Bacteriol. 1994;44(4):846-9.

67. Morita Y, Maruyama S, Kabeya H, Nagai A, Kozawa K, Kato M, et al. Genetic diversity of the *dnaJ* gene in the *Mycobacterium avium* complex. J Med Microbiol. 2004;53(Pt 8):813-7.

68. Fox GE, Wisotzkey JD, Jurtshuk P. How Close Is Close - 16s Ribosomal-Rna Sequence Identity May Not Be Sufficient to Guarantee Species Identity. Int J Syst Bacteriol. 1992;42(1):166-70.

69. Kim BJ, Hong SK, Lee KH, Yun YJ, Kim EC, Park YG, et al. Differential identification of *Mycobacterium tuberculosis* complex and nontuberculous mycobacteria by duplex PCR assay using the RNA polymerase gene (*rpoB*). J Clin Microbiol. 2004;42(3):1308-12.

70. Johnson MM, Odell JA. Nontuberculous mycobacterial pulmonary infections. J Thorac Dis. 2014;6(3):210-20.

71. Kim BJ, Math RK, Jeon CO, Yu HK, Park YG, Kook YH. *Mycobacterium yongonense* sp. nov., a slow-growing non-chromogenic species closely related to *Mycobacterium intracellulare*. Int J Syst Evol Microbiol. 2013;63(Pt 1):192-9.

72. Tortoli E, Mariottini A, Pierotti P, Simonetti TM, Rossolini GM. *Mycobacterium yongonense* in pulmonary disease, Italy. Emerg Infect Dis. 2013;19(11):1902-4.

73. Hong SK, Kim EC. Possible misidentification of *Mycobacterium yongonense*. Emerg Infect Dis. 2014;20(6):1089-90.

74. Bisercic M, Ochman H. The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium*. J Bacteriol. 1993;175(24):7863-8.

75. Dale JW. Mobile genetic elements in mycobacteria. Eur Respir J Suppl. 1995;20:633s-48s.

76. Wall S, Ghanekar K, McFadden J, Dale JW. Context-sensitive transposition of IS*6110* in mycobacteria. Microbiology. 1999;145 (Pt 11):3169-76.

77. Whittington R, Marsh I, Choy E, Cousins D. Polymorphisms in IS*1311*, an insertion sequence common to *Mycobacterium avium* and *M. avium* subsp. *paratuberculosis*, can be used to distinguish between and within these species. Mol Cell Probes. 1998;12(6):349-58.

78. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32-6.

79. Clewley JP, Arnold C. MEGALIGN. The multiple alignment module of LASERGENE. Methods Mol Biol. 1997;70:119-29.

80. Sharma D, Issac B, Raghava GPS, Ramaswamy R. Spectral Repeat Finder (SRF): identification of repetitive sequences using Fourier transformation. Bioinformatics. 2004;20(9):1405-12.

81. Macheras E, Konjek J, Roux AL, Thiberge JM, Bastian S, Leao SC, et al. Multilocus sequence typing scheme for the *Mycobacterium abscessus* complex. Res Microbiol. 2014;165(2):82-90.

82. Zelazny AM, Root JM, Shea YR, Colombo RE, Shamputa IC, Stock F, et al. Cohort study of molecular identification and typing of *Mycobacterium abscessus*, *Mycobacterium massiliense*, and *Mycobacterium bolletii*. J Clin Microbiol. 2009;47(7):1985-95.

83. Kim BJ, Hong SH, Kook YH, Kim BJ. Molecular Evidence of Lateral Gene Transfer in *rpoB* Gene of *Mycobacterium yongonense* Strains via Multilocus Sequence Analysis. PLoS One. 2013;8(1).

84. Berger B, Haas D. Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS*21* and related elements. Cell Mol Life Sci. 2001;58(3):403-19.

85. Mahillon J, Chandler M. Insertion sequences. Microbiol Mol Biol Rev. 1998;62(3):725-74.

86. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000;405(6784):299-304.

87. Syvanen M. Evolutionary implications of horizontal gene transfer. Annu Rev Genet. 2012;46:341-58.

88. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, et al. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 2005;1(1):e5.

89. Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat Rev Microbiol. 2006;4(9):670-81.

90. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, et al. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A. 1997;94(18):9869-74.

91. Rindi L, Garzelli C. Genetic diversity and phylogeny of *Mycobacterium avium*. Infect Genet Evol. 2014;21:375-83.

92. Lefevre P, Braibant M, Content J, Gilot P. Characterization of a *Mycobacterium bovis* BCG insertion sequence related to the IS*21* family. FEMS Microbiol Lett. 1999;178(2):211-7.

93. Morschhauser J, Kohler G, Ziebuhr W, Blum-Oehler G, Dobrindt U, Hacker J. Evolution of microbial pathogens. Philos Trans R Soc Lond B Biol Sci. 2000;355(1397):695-704.

94. Ochman H. Lateral and oblique gene transfer. Curr Opin Genet Dev. 2001;11(6):616-9.

95. Lawrence JG, Hendrickson H. Lateral gene transfer: when will adolescence end? Mol Microbiol. 2003;50(3):739-49.

96. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998;393(6685):537-44.

97. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001;409(6823):1007-11.

98. Modrich P, Lahue R. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem. 1996;65:101-33.

99. Vulic M, Dionisio F, Taddei F, Radman M. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc Natl Acad Sci U S A. 1997;94(18):9763-7.

100. Lin Z, Nei M, Ma H. The origins and early evolution of DNA mismatch repair genes--multiple horizontal gene transfers and co-evolution. Nucleic Acids Res. 2007;35(22):7591-603.

101. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999;73(1):152-60.

102. Blokpoel MC, Murphy HN, O'Toole R, Wiles S, Runn ES, Stewart GR, et al. Tetracycline-inducible gene regulation in mycobacteria. Nucleic Acids Res. 2005;33(2):e22.

103. Andreu N, Zelmer A, Fletcher T, Elkington PT, Ward TH, Ripoll J, et al. Optimisation of Bioluminescent Reporters for Use with Mycobacteria. PLoS One. 2010;5(5).

104. Murry J, Sassetti CM, Moreira J, Lane J, Rubin EJ. A new site-specific integration system for mycobacteria. Tuberculosis (Edinb). 2005;85(5-6):317-23.

105. Snapper SB, Melton RE, Mustafa S, Kieser T, Jacobs WR, Jr. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. Mol Microbiol. 1990;4(11):1911-9.

106. Mariam DH, Mengistu Y, Hoffner SE, Andersson DI. Effect of *rpoB* mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2004;48(4):1289-94.

107. Guo XV, Monteleone M, Klotzsche M, Kamionka A, Hillen W, Braunstein M, et al. Silencing essential protein secretion in *Mycobacterium smegmatis* by using tetracycline repressors. J Bacteriol. 2007;189(13):4614-23.

108. Springer B, Sander P, Sedlacek L, Hardt WD, Mizrahi V, Schar P, et al. Lack of mismatch correction facilitates genome evolution in mycobacteria. Mol Microbiol. 2004;53(6):1601-9.

109. Gamieldien J, Ptitsyn A, Hide W. Eukaryotic genes in *Mycobacterium tuberculosis* could have a role in pathogenesis and immunomodulation. Trends Genet. 2002;18(1):5-8.

110. Kinsella RJ, Fitzpatrick DA, Creevey CJ, McInerney JO. Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution, and gene duplication. Proc Natl Acad Sci U S A. 2003;100(18):10320-5.

111. Le Dantec C, Winter N, Gicquel B, Vincent V, Picardeau M. Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. J Bacteriol. 2001;183(7):2157-64.

112. Fang Z, Doig C, Kenna DT, Smittipat N, Palittapongarnpim P, Watt B, et al. IS*6110*-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. J Bacteriol. 1999;181(3):1014-20.

113. Humbert O, Prudhomme M, Hakenbeck R, Dowson CG, Claverys JP. Homeologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. Proc Natl Acad Sci U S A. 1995;92(20):9052-6.

114. Smith NH, Dale J, Inwald J, Palmer S, Gordon SV, Hewinson RG, et al. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. Proc Natl Acad Sci U S A. 2003;100(25):15271-5.

115. Supply P, Warren RM, Banuls AL, Lesjean S, van der Spuy GD, Lewis LA, et al. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. Mol Microbiol. 2003;47(2):529-38.

116. Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. Infect Immun. 2011;79(9):3476-91

117. Mehta M, Rajmani RS, Singh A. *Mycobacterium tuberculosis* WhiB3 Responds to Vacuolar pH-induced Changes in Mycothiol Redox Potential to Modulate Phagosomal Maturation and Virulence. J Biol Chem. 2016;291(6):2888-903.

118. Velmurugan K, Chen B, Miller JL, Azogue S, Gurses S, Hsu T, et al. *Mycobacterium tuberculosis nuoG* is a virulence gene that inhibits apoptosis of infected host cells. PLoS Pathog. 2007;3(7):972-80.

국문 초록

서론: 조류형 결핵균 군 (*Mycobacterium avium* complex)에 속하는 균종은 가장 높은 빈도로 분리되고 있는 비결핵 항산성균 (non-tuberculous mycobacteria)이다. 일반적으로 조류형 결핵균 군은 마이코박테리움 아비움 (*M. avium*)과 마이코박테리움 인트라셀룰라레 (*M. intracellulare*)로 구성되었다고 알려져 있다. 최근, 국내 환자에서 분리된 *M. intracellulare* 균주들을 대상으로, 세 종류의 독립적인 유전자 - *hsp65*, ITS1 과 16S rRNA - 를 분석한 결과, 유전적 다양성이 확인되어 5 종류의 유전자형 그룹 (INT-1~ -5)으로 나뉘어 진다고 알려져 있다.

방법: 국내 폐질환 환자에서 분리된 05-1390 균주는 *M. intracellulare,* 특히 INT-5 와 연관된 균주로, 생화학적 및 계통학적 분석을 통해 신종 균주로 제안되었다 (마이코박테리움 연건넨스, *M. yongonense* DSM 45126^{T}). 이 균주는 계통학적으로 *M. intracellulare* 와 유사하지만, RNA 중합효소 베타 서브유닛 유전자 (RNA polymerase β-subunit gene, *rpoB*)가 마이코박테리움 파라스크로풀라세움 (*M. parascrofulaceum*)과 높은 상동성을 보이는 특징을 지니고 있으며, 이는 잠재적 측방성 유전자 전달 기전을 통해 *M. parascrofulaceu* 의 *rpoB* 유전자를 획득할 가능성을 제시한다. *M. yongonense* 의 측방성 유전자 전달 기전에 대한 더 나은 이해를 위해 네 가지 시퀀싱 방법을 통해 *M. yongonense* 의 전체 유

전체 염기서열을 확보하였다. 또한 *M. yongonense* 와 다른 INT-5 유전자형 균주 (MOTT-36Y 와 MOTT-H4Y)간 정확한 분류학적 위치를 파악하기 위해 유전체 기반의 계통학적 분석과 삽입 서열 (IS-element) 동정을 수행하였다. 이를 위해, 두 INT-5 균주의 유전체 염기서열을 *M. intracellulare* ATCC 13950^{T}과 *M. yongonense* DSM 45126^{T}의 유전체와 비교하였다. 35 개의 타겟 유전자의 다중염기서열타이핑 (multilocus sequence typing, MLST)과 단일염기 다형성 (single nucleotide polymorphism, SNP) 분석을 통해 INT-5 균주와 *M. yongonense* DSM 45126^{T} 균주간의 관계를 비교, 분석하였다. 또한, *M. yongonense* 의 유전체 정보를 통해 신규 IS-element 를 동정하였고, 실시간 중합효소 연쇄반응 (real-time PCR)을 기반으로 한 새로운 진단법을 개발하였다. *M. parascrofulaceum* 으로 부터 *M. yongonense* 로 일어난 측방성 유전자 전달 기전을 이해하기 위해, 비교 유전체 분석을 진행하였고 SimPlot 과 BootScan 분석을 통해 잠재적 측방성 유전자 전달 지역을 조사하였다. 또한 *M. yongonense* 유전체에서 DNA 부조화 수리 유전자 (DNA mismatch repair genes, MutS4A and MutS4B)가 확인되었고, 이 유전자의 상동재조합 (homologous recombination) 능력을 조사하였다. 이를 위해 DNA mismatch repair 유전자를 지닌 재조합 마이코박테리움 스메그마티스 (*M. smegmatis*)를 제조하였고, 리팜핀 내성과 연관된 돌연변이를 지닌 *rpoB* 절편을 재조합 *M. smegmatis* 에 도입시켰다. 그 후, 재조합 균주의 리팜핀에 대한 저항능과 상동재조합 빈도를 조사하였다.

결과: 신종 균주 05-1390 은 생화학적 특성이 *M. intracellulare* 와 유사하였고, ITS1 과 *hsp65* 유전자형 또한 *M. intracellulare* 와 밀접한 연관이 있었지만, *rpoB* 유전자 염기서열이 *M. parascrofulaceum* ATCC BAA-614^{T} 균주와 높은 상동성을 보이는 특징을 지니고 있었다. 이러한 결과를 토대로, 05-1390 균주를 신종 균주인 *M. yongonense* DSM 45126^{T} 로 제안하였다. 이 균주는 5,521,023 염기쌍으로 구성된 원형의 염색체 (GenBank 등록번호: CP003347), 122,976 염기쌍으로 구성된 원형의 플라스미드 (pMyong1, GenBank 등록번호: JQ657805)와 18,089 염기쌍으로 구성된 선형의 플라스미드 (pMyong2, GenBank 등록번호: JQ657806)를 지니고 있었다. 이러한 유전체 염기서열 정보를 토대로 한 MLST 와 SNP 분석 결과는 INT-5 균주가 다른 *M. intracellulare* 균주 들보다 *M. yongonense* DSM 45126^{T} 균주와 밀접한 연관이 있다는 것을 시사하였다. 또한 *M. yongonense* DSM 45126^{T} 균주에서 동정된 신규 IS-element 인 IS*Myo2* 를 타겟으로 한 실시간 중합효소 연쇄반응법은 *M. yongonense* 균주와 INT-5 균주만을 검출할 수 있었다. *M. yongonense* DSM 45126^{T} 균주의 유전체에서 잠재적 측방성 유전자 전달이 일어난 두 지역을 확인할 수 있었다. 첫 번째 지역은 *rpoBC* 오페론 (OEM_44170~44190)에 해당하는 지역이었고, 두 번째 지역은 57 개의 개방형 해독틀 (open reading frame, ORF)을 포함하는 지역이었다 (OEM_08030~08590). SimPlot 과 BootStrap 분석 결과는 이 두 지역에서의 측방성 유전자 전달 가능성을 지지하였다. 또한 DNA mismatch repair 유전자를 지닌 재조합 *M. smegmatis* 균주에 리팜핀 내성과 연관된

유전자 절편을 도입시키면 리팜핀 존재하에 상동재조합이 일어나, 리팜핀 저항능을 지닐 수 있다는 것을 확인하였다.

결론: 본 논문에서, 계통학적으로 독특한 신종 마이코박테리아인 *M. yongonense* 균종을 생화학적 검사와 유전적 분석을 통해 유전체 기반의 계통학적 분석 결과를 바탕으로 기술하였다. 그리고 이 균주와 유전적으로 가까운 INT-5 균주들의정확한 분류학적 관계를 규명하고자 전체 유전체를 기반으로 한 분석을 수행하였고, 이를 토대로 *M. yongonense* 균주가 분류학적으로 INT-5 균주 (MOTT-36Y 와 MOTT-H4Y)와 밀접한 연관성이 있다는 것을 확인하였다. 결론적으로, 기존의 INT-5 균주들의 분류학적 위치가 *M. yongonense* 균주로 변경되어야 한다고 생각한다. *M. yongonense* 균주는 두 종류의 유전자형으로 나뉘어 질 수 있는데, 제 1 형 *M. yongonense* 는 *M. parascrofulaceum* 으로 부터 *rpoBC* 오페론 등을 측방 전달 받은 특징을 지니고 있고, 제 2 형 *M. yongonense* 는 측방성 유전자 전달 현상이 일어나지 않아 *M. intracellulare* 의 *rpoBC* 오페론을 유지하고 있는 특징을 지니고 있다. 측방성 유전자 전달 현상의 결과가 *M. yongonense* 균주들을 제 1 형 및 2 형의 유전자형으로 나눌 수 있는 기준이 되고, 측방성 유전자 전달 현상은 DNA mismatch repair 유전자의 도입으로 인해 더 쉽게 일어날 수 있을 거라 생각된다.

* 본 졸업 논문은 현재 Internationl Journal of Systematic and Evolutionary Microbiology (Kim BJ, Math RK, Jeon CO, Yu HK, Park YG, Kook YH, et al.

Mycobacterium yongonense sp. nov., a slow-growing non-chromogenic species closely related to *Mycobacterium intracellulare*. Int J Syst Evol Microbiol. 2013; 63(Pt 1):192-9), PLoS One (Kim BJ, Hong SH, Kook YH, and Kim BJ. Molecular evidence of lateral gene transfer in *rpoB* gene of *Mycobacterium yongonense* strains via multilocus sequence analysis. PLoS One. 2013;8(1):e51846.; Kim BJ, Kim BR, Lee SY, Kim GN, Kook YH, and Kim BJ. Molecular Taxonomic Evidence for Two Distinct Genotypes of *Mycobacterium yongonense* via Genome-Based Phylogenetic Analysis. PLoS One. 2016; 11(3):e0152703), Genome Announcements (Kim BJ, Kim BR, Lee SY, Seok SH, Kook YH, and Kim BJ. Whole-Genome Sequence of a Novel Species, *Mycobacterium yongonense* DSM 45126^{T}. Genome Announc. 2013; 1(4)),와 BMC Genomics (Kim BJ, Kim K, Kim BR, Kook YH, and Kim BJ. Identification of IS*Myo2*, a novel insertion sequence element of IS*21* family and its diagnostic potential for detection of *Mycobacterium yongonense*. BMC Genomics. 2015; 16(1):794.) 저널에 출판 완료된 내용을 포함하고 있습니다.

주요어 : 마이코박테리움 연건넨스, 마이코박테리움 인트라셀룰라레 INT-5 유전자형 그룹, 유전체 염기서열, 비교 유전체 분석, 다중염기서열 타이핑, 단일염기 다형성, RNA 중합효소 베타 서브유닛, 측방성 유전자 전달 현상, DNA 부조화 수리 유전자

학 번 : 2010-30606