Differentiation of *Mycobacterium* species by analysis of the heat-shock protein 65 gene (*hsp65*)

Hong Kim,¹ Sun-Hyun Kim,¹ Tae-Sun Shim,² Mi-na Kim,³ Gill-Han Bai,⁴ Young-Gil Park,⁴ Sueng-Hyun Lee,⁵ Gue-Tae Chae,⁶ Chang-Yong Cha,¹ Yoon-Hoh Kook¹ and Bum-Joon Kim¹

Correspondence
Bum-Joon Kim
kbumjoon@snu.ac.kr

¹Department of Microbiology, College of Medicine, Seoul National University, 28 Yongon-dong, Chongno-gu, Seoul 110-799, Korea
²,³Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine² and Department of Laboratory Medicine³, University of Ulsan College of Medicine, Asan Medical Center, Seoul 138-600, Korea
⁴The Korean Institute of Tuberculosis, The Korean National Tuberculosis Association, Seoul 137-140, Korea
⁵Department of Microbiology, College of Medicine, Konkuk University, Chungju 380-230, Korea
⁶Institute of Hansen’s Disease, The Catholic University Medical College, Seoul 137-7014, Korea

The nucleotide sequences (604 bp) of partial heat-shock protein genes (*hsp65*) from 161 *Mycobacterium* strains containing 56 reference *Mycobacterium* species and 105 clinical isolates were determined and compared. *hsp65* sequence analysis showed a higher degree of divergence between *Mycobacterium* species than did 16S rRNA gene analysis. Generally, the topology of the phylogenetic tree based on the *hsp65* DNA sequences was similar to that of the 16S rRNA gene, thus revealing natural relationships among *Mycobacterium* species. When a direct sequencing protocol targeting 422 bp sequences was applied to 70 non-tuberculous mycobacterium (NTM) clinical isolates, all NTMs were clearly identified. In addition, an *Xho*I PCR restriction fragment length polymorphism analysis method for the differentiation of *Mycobacterium tuberculosis* complex from NTM strains was developed during this study. The results obtained suggest that 604 bp *hsp65* sequences are useful for the phylogenetic analysis and species identification of mycobacteria.

**INTRODUCTION**

Members of the genus *Mycobacterium* are widespread in nature and range from harmless saprophytic species to strict pathogens that cause serious human and animal diseases, such as tuberculosis and leprosy. Traditionally, numerical taxonomy based on biochemical characteristics presents criteria for species delimitation and the identification of mycobacteria, but is limited in terms of its ability to differentiate species such as *Mycobacterium avium* and *Mycobacterium intracellulare* (Goodfellow & Magee, 1998; Wayne & Kubica, 1986). Moreover, this culture-based taxonomic approach is laborious, complex and usually impeded by the slow growth of *Mycobacterium* species in clinical laboratories.

Several macromolecular sequence comparisons have been conducted to delimit more clearly species boundaries within the genus *Mycobacterium* and to determine phylogenetic relationships between species within this genus (Kim *et al.*, 1999; Pitulle *et al.*, 1992; Rogall *et al.*, 1990; Roth *et al.*, 1998; Stahl *et al.*, 1990; Stone *et al.*, 1995; Takewaki *et al.*, 1994). In particular, phylogenetic approaches targeting the 16S rRNA gene have been widely used and have demonstrated natural relationships among species within this genus (Pitulle *et al.*,...
1992; Rogall et al., 1990; Stahl & Urbance, 1990). However, this phylogenetic approach also has its limitations, for example, due to the existence, albeit rare, of two-copy genes with different sequences in a single organism (Ninet et al., 1996; Reischl et al., 1998), or with respect to the sequence identities of Mycobacterium kansasi and Mycobacterium gastrici, which are obviously different species (Goodfellow & Magee, 1998; Rogall et al., 1990; Stahl & Urbance, 1990). Another example of 16S rRNA sequence identity between two species is between Mycobacterium and Mycobacterium paradueterculosis (Goodfellow & Magee, 1998). Other examples of the limitations of relying on a single target molecule to determine phylogenetic relationships and for species identification among species in a genus have also been reported (Clayton et al., 1995; Fox et al., 1992; Palys et al., 1997; Turenne et al., 2001). Therefore, alternative phylogenetic and identification methods capable of complementing 16S rRNA gene analysis should be developed for the phylogenetic study and species identification of the genus Mycobacterium.

In recent years, the 60 kDa heat-shock protein family (HSP60 or HSP65), because of their highly conserved primary structures and ubiquity, have been considered as useful phylogenetic markers in several eubacterial genera (Chang et al., 2003; Kwok et al., 1999; Lee et al., 2003; Ringuet et al., 1999; Viale et al., 1994). It is also known that almost all bacterial species have a single copy of the hsp60 or hsp65 gene in their genome (Segal & Ron, 1996), which means that this gene is not easily transferred from one bacterium to another and that a phylogenetic approach, targeting this gene, is probably suitable for phylogenetic studies of closely related species or strains (Kwok et al., 1999). The hsp65 gene has also been used successfully for species identification of cultured clinical isolates of the genus Mycobacterium, although there are some drawbacks in the application of methods targeting the hsp65 gene, a single-copy gene, for the direct detection of mycobacterial organisms in clinical specimens due to limited sensitivity. However, identifications based on this gene are usually based on PCR-restriction fragment length polymorphism analysis (PRA) targeting the 441 bp Telent fragment of the hsp65 gene (Devalois et al., 1997; Hafner et al., 2004; Telenti et al., 1993). Although phylogenetic and molecular epidemiological studies based on the hsp65 sequence have also been performed, the phylogenetic relationships of only a limited number of reference strains, i.e. ‘fast growers’, were analysed (Ringuet et al., 1999). Molecular epidemiologic studies were also limited to the sequence databases of only some specific species, like the Mycobacterium avium complex (Smole et al., 2002) and Mycobacterium scrofulaceum (Swanson et al., 1996). Therefore, we undertook this study to perform a phylogenetic analysis and to develop species identification methods of mycobacteria, based on 604 bp hsp65 sequences of 56 reference and 101 clinical strains, in order to evaluate the usefulness of this gene as an alternative target molecule for identification and phylogenetic analysis of mycobacteria.

**METHODS**

**Mycobacterial reference strains and clinical isolates.** Fifty-seven reference strains composed of 56 mycobacterial strains and Tsukamurella paurometabolica KCTC 9821T and 175 clinical isolates composed of 105 strains for intraspecies similarity analysis (Table 1) and 70 strains for blinded hsp65 direct sequencing analysis (Table 2) were used in this study. The 53 Mycobacterium reference strains and 89 clinical isolates that had already been identified by conventional biochemical testing and rpoB PRA analysis (Kim et al., 2001) were provided by the Korean Institute of Tuberculosis (KIT). The single strain of Mycobacterium leprae (Thai 53 strain) and 16 punch-biopsy specimens obtained from active lesions of patients diagnosed on the basis of histological findings, acid-fast bacterium staining and amplification of DNA encoding an 18-kDa protein (Williams et al.,

---

**Table 1. Intraspecies similarity of hsp65 sequences observed among clinical isolates and their Xhol PRA patterns**

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence similarity (%)</th>
<th>Strains (n)</th>
<th>Xhol PRA pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. tuberculosis</td>
<td>100</td>
<td>25</td>
<td>A</td>
</tr>
<tr>
<td>M. leprae*</td>
<td>100</td>
<td>16</td>
<td>B</td>
</tr>
<tr>
<td>M. avium</td>
<td>99-2-100</td>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>M. intracellular</td>
<td>99-2-100</td>
<td>10</td>
<td>C</td>
</tr>
<tr>
<td>M. scrofulaceum</td>
<td>100</td>
<td>3</td>
<td>B</td>
</tr>
<tr>
<td>M. szulgai</td>
<td>98-7-100</td>
<td>8</td>
<td>C</td>
</tr>
<tr>
<td>M. kansasii†</td>
<td>99-2-100</td>
<td>18</td>
<td>B</td>
</tr>
<tr>
<td>M. fortuitum</td>
<td>98-2-100</td>
<td>5</td>
<td>D</td>
</tr>
<tr>
<td>M. peregrinun</td>
<td>100</td>
<td>3</td>
<td>D</td>
</tr>
<tr>
<td>M. abscessus</td>
<td>98-5-100</td>
<td>11</td>
<td>D</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>98-2-100</strong></td>
<td><strong>105</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Sixteen punch-biopsy specimens obtained from active lesions of patients diagnosed on the basis of histological findings, acid-fast bacterium staining and amplification of DNA encoding an 18-kDa protein.
†One strain among 18 clinical isolates of M. kansasii showed 99-7% sequence similarity with M. kansasii type II.
Table 2. Comparison of the identification results obtained by conventional biochemical testing and by hsp65 direct sequencing analysis of NTM clinical isolates

<table>
<thead>
<tr>
<th>Conventional biochemical tests</th>
<th>Isolates (n)</th>
<th>hsp65 direct sequencing*</th>
<th>Isolates (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. avium complex</td>
<td>35†</td>
<td>M. avium</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. intracellulare</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. celatum ATCC 51130</td>
<td>1</td>
</tr>
<tr>
<td>M. gordonae</td>
<td>5</td>
<td>M. gordonae</td>
<td>5</td>
</tr>
<tr>
<td>M. kansasii</td>
<td>7</td>
<td>M. kansasii type I</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. kansasii type II</td>
<td>1</td>
</tr>
<tr>
<td>M. terrae complex</td>
<td>11</td>
<td>M. terrae</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. nonchromogenicum</td>
<td>3</td>
</tr>
<tr>
<td>M. fortuitum complex</td>
<td>6</td>
<td>M. fortuitum ATCC 6841</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. peregrinum</td>
<td>3</td>
</tr>
<tr>
<td>M. chelonae complex</td>
<td>6</td>
<td>M. abscessus</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M. chelonae</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>Total</td>
<td>70</td>
</tr>
</tbody>
</table>

*Results obtained by hsp65 direct sequencing were confirmed by rpoB sequencing.
†One strain identified as M. avium complex by conventional testing was confirmed as M. celatum.

DNA extraction. Chromosomal DNA was extracted by the bead beater-phenol extraction method. To disrupt the Mycobacterium cell wall, a bacterial mixture containing phenol and glass beads was oscillated on a Mini-Bead beater (Kim et al., 1999). The aqueous phase was transferred to another clean tube and the DNA pellet was precipitated by adding isopropanol and then solubilized with 60 μl TE buffer (10 mM Tris/HCl, 1 mM EDTA, pH 8-0). Two micro-litres of purified DNA was used as a template in the PCRs.

PCR amplification of a 644 bp fragment of the hsp65 gene. A set of primers [forward primer HSPR3 (5’-ATCGCCCAA-GGAGATCGAGCT-3’; positions 163–182 in the hsp65 gene of Mycobacterium tuberculosis, GenBank accession no. M15467), reverse primer HSPR4 (5’-AAGGTCGCCCCATCTTGGT-3’; positions 806–787)] were used for the amplification of 644 bp PCR products from Mycobacterium strains (Fig. 1). Template DNA (50 ng) and 20 pmol of each primer were added to a PCR mixture tube (AccuPower PCR PreMix; Bioneer) containing 1 U Taq DNA polymerase, 250 μM of each dNTP, 50 mM Tris/HCl (pH 8.3), 400 mM KCl, 1–5 mM MgCl2, and gel-loading dye and the volume was then adjusted with distilled water to 20 μl. The reaction mixture was subjected to 30 cycles of amplification (60 s at 95 °C, 45 s at 62 °C and 90 s at 72 °C) followed by a 5 min extension at 72 °C (model 9600 Thermocycler; Perkin-Elmer Cetus). The PCR products were electrophoresed in a 3% agarose gel and visualized with ethidium bromide under UV light.

Cloning and sequencing of the PCR products. The expected PCR products (644 bp) were purified using a QIAEX II gel extraction kit (Qiagen), as recommended by the manufacturer. The purified PCR products were then ligated to PCR II-TOPO using a TOPO TA Cloning kit (Invitrogen) and transformed into chemically competent Escherichia coli, as recommended by the manufacturer. Ampicillin (100 μg ml⁻¹) and blue–white selection were used to select the required transformants and the plasmid DNAs were extracted using the alkaline lysis method. To verify the presence of the correct insert of the partial hsp65 gene, the PCR protocol, described above, was adopted using approximately 2 ng plasmid DNA as a template. In view of the fidelity of Taq polymerase, more than two colonies per strain were selected and sequenced. The 604 bp sequence, excluding those of the primers, was determined in both directions by using an Applied Biosystems model 373A automatic sequencer and a BigDye Terminator cycle sequencing kit (Perkin-Elmer Applied Biosystems). For the sequencing reaction, 300 ng plasmid DNA, 3-2 pmol of either M13 Forward (~20) primer (5’-GTAAAACGACGGCCAG-3’) or M13 Reverse primer (5’-CAGGAAACAGCTATGAC-3’) and 8 μl BigDye Terminator RR mix (Perkin-Elmer Applied Biosystems) were mixed and adjusted to a final volume of 20 μl with distilled water. The reaction was run with 5% (v/v) DMSO for 30 cycles of 15 s at 95 °C, 10 s at 50 °C and 4 min at 60 °C.

Analysis of the sequence data and phylogenetic tree construction. The partial hsp65 DNA sequences (604 bp), excluding both primers, from 56 mycobacterial reference strains, T. paurometabola KCTC 9821T and 105 clinical strains were aligned and their similarities were calculated using the multiple-alignment algorithm in the MEGALIGN software package (DNASTAR, Windows version 3.12e) to analyse the similarity levels among different species and within clinical isolates of the same species of the genus Mycobacterium and to select signature nucleotides specific to M. tuberculosis complex strains.

Two hsp65 phylogenetic trees were constructed. One used 41 mycobacterial reference strains for comparison with the 16S rRNA gene tree conducted from the same strains and the other used all 56 reference strains for analysis of phylogenetic relationships of overall strains. The hsp65 tree was inferred by using the neighbour-joining (Saitou & Nei, 1987), maximum-likelihood (Felsenstein, 1993) and maximum-parsimony (Fitch, 1972) methods using T. paurometabola KCTC 9821T as an outgroup. Evolutionary-distance matrices were
Fig. 1. Locations of primers for PCR targeting sites for direct sequencing and restriction sites for Xhol PRA. The 441 bp hsp65 sequences were used for the differentiation of mycobacteria in previous studies (Devallois et al., 1997; Telenti et al., 1993). Numbers indicate nucleotide positions of the hsp65 gene of \( M. \) tuberculosis. Positions of signature nucleotides (see Table 3) are indicated by (S). Upward-pointing arrows indicate sites cut by Xhol.

generated according to the model of Jukes & Cantor (1969). The neighbour-joining and maximum-parsimony methods were carried out using MEGA version 2.1 (Kumar et al., 2001) and the maximum-likelihood method (DNAML) was carried out using PHYML version 3.5 (Felsenstein, 1993). The resulting trees and topology were evaluated by bootstrap analyses (Felsenstein, 1985) based on 1000 resamplings. The 16S rRNA gene sequences (1396–1427 bp, positions 30–1443 in the 16S rRNA gene numbering of \( M. \) tuberculosis) of the same 41 species were aligned and compared. The 16S rRNA gene tree was inferred by using the neighbour-joining method and compared with the hsp65 tree.

hsp65 direct sequencing analysis. Purified PCR amplicons from 70 NTM clinical isolates were directly sequenced. In direct sequencing analysis, we used 644 bp PCR products as templates and used HSPR4, the reverse PCR primer, as a direct sequencing primer. Although approximately 500 bp sequences were obtained by direct sequencing, for reasons of simplicity only 422 bp sequences (positions 233–654 in the hsp65 gene of \( M. \) tuberculosis) were used for species differentiation (Fig. 1).

PRA with Xhol digestion. Based on the determined hsp65 sequences of the 56 mycobacterial reference strains used in this study (GenBank accession numbers AF057449–AF057493), a single endonuclease (Xhol) was selected using MapDraw (version 3.14; DNASTAR) for the differential identification of the \( M. \) tuberculosis complex from NTM strains, and for the grouping of NTM strains. The recognition sites and fragment sizes of Xhol were also calculated using the MapDraw program. Xhol recognizes and cuts at three sites (positions 553, 637 and 703) on the mycobacterial hsp65 gene (Fig. 1). Differential identification between NTM and \( M. \) tuberculosis strains is based on the presence of a signature nucleotide specific to \( M. \) tuberculosis complex strains (position 705) (Table 3) in the third Xhol recognition sequence (\( 700 \mathrm{C} \downarrow \mathrm{TGGAC} \)). PCR was performed using the set of primers described above (HSPF3 and HSPR4) and using the same PCR conditions. Restriction enzyme Xhol (Takara) was then added to the 644 bp PCR product of the 56 reference strains. Briefly, 10 \( \mu \)l of the PCR products, 2 U enzyme and restriction buffer were transferred to a fresh microcentrifuge tube and water was added to a final volume of 20 \( \mu \)l per reaction. Digestion was performed for 2 h at 37 °C. Following digestion, the mixtures were electrophoresed in a 3 % agarose gel.

Nucleotide sequence accession numbers. The hsp65 gene sequences of 56 mycobacterial reference strains and \( T. \) paumetabolobola KCTC 9821T determined in this study have been deposited in GenBank and their accession numbers are listed in Supplementary Table S1 available in IJSEM Online. To construct a tree based on 16S rRNA gene sequences and to compare sequence similarity levels with hsp65 gene analysis, the 16S rRNA sequences of 41 strains with available sequences more than 1400 bp were used. These sequence accession numbers are also listed in Supplementary Table S1.

RESULTS AND DISCUSSION

hsp65 DNA sequences of mycobacterial reference strains

The G+C content of the 604 bp hsp65 DNA sequences, except that of \( M. \) leprae Thai 53 strain, which showed a low G+C content of 58 mol%, was between 64 and 67 mol%. No insertions or deletions were observed in the sequence alignments of the hsp65 genes. More than 83–1 % similarity (between \( M. \) leprae Thai 53 strain and \( M. \) tuberculosis \( \mathrm{ATCC} \) 11758\(^T \)) was observed among the hsp65 DNA sequences of mycobacterial strains (interspecies divergence \( >16-9 \%) \). In a variety of \( M. \) tuberculosis complex strains, namely \( \mathrm{Mycobacterium} \) africanum ATCC 25420\(^T \), \( \mathrm{Mycobacterium} \) bovis ATCC 19210\(^T \), \( \mathrm{Mycobacterium} \) bovis BCG, \( \mathrm{Mycobacterium} \) microti ATCC 19422\(^T \) and \( M. \) tuberculosis ATCC 27294\(^T \), the nucleotide sequences of the amplified hsp65 genes were identical. The hsp65 nucleotide sequences of \( \mathrm{Mycobacterium} \) simiae ATCC 25275\(^T \) and \( \mathrm{Mycobacterium} \) genavense ATCC 51233\(^T \), two phylogenetically closely related strains, were also identical.

Comparison of sequence similarity of hsp65 and the 16S rRNA gene

When the nucleotide sequences of hsp65 and the 16S rRNA gene from the same 41 mycobacterial reference strains were compared, the 16S rRNA gene sequence similarity between any given pair was always higher than the corresponding

Table 3. Signature nucleotides and amino acids specific to \( M. \) tuberculosis complex strains as distinct from NTM strains

<table>
<thead>
<tr>
<th>Group</th>
<th>Codon (nucleotide positions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M. ) tuberculosis complex</td>
<td>GTA (V)</td>
</tr>
<tr>
<td>NTM</td>
<td>GTC (V), GTT (V)</td>
</tr>
</tbody>
</table>

Differences between the two groups are highlighted in bold.
hsp65 DNA sequence similarity, except for *M. genavense* ATCC 51233T and *M. simiae* ATCC 25275T, which had identical hsp65 sequences.

**hsp65 DNA sequences of clinical mycobacteria isolates**

The sequence similarities of hsp65 among the clinical isolates of the same species were investigated. The level of sequence similarity among strains in each species was more than 98.2% (intrasppecies divergence <1.8%). In general, fewer than 11 nucleotide variations were observed among the clinical isolates. For clinical isolates of *M. tuberculosis*, *M. leprae*, *M. scrofulaceum* and *Mycobacterium peregrinum*, a sequence similarity of 100% was observed, and 17 of 18 clinical isolates of *M. kansasi* showed similarity in the range 99–2–100% with the *M. kansasi* type I reference strain (ATCC 12478T), with the other strain showed a similarity of 99.7% with *M. kansasi* type II (Table 1).

**Phylogenetic analysis based on the hsp65 sequences**

On comparing the 16S rRNA tree based on the sequences of the reference strains with the phylogenetic trees based on hsp65 sequences, they were generally found to be topologically similar, in spite of some differences (Fig. 2a, b). This hsp65 tree confirms that the traditional division between rapidly and slowly growing mycobacteria is a natural division. The branching point of the rapidly growing members is close to that of total *Mycobacterium* strains, which shows that these are phylogenetically older than their slowly growing relatives.

Although the general topology was similar, some differences were observed between the hsp65 and 16S rRNA gene trees. First, abnormal taxonomic positions of *Mycobacterium terrae* complex strains were commonly conserved in the hsp65 trees inferred by three different methods. Abnormal taxonomic positions of *M. terrae* complex strains in the hsp65 tree are also in good agreement with the result from RpoB amino acid analysis (Kim et al., 1999) that all three species of the *M. terrae* complex have rapid-grower-specific leucine (L) instead of methionine (M) at codon 468 of RpoB. This phenomenon is also supported by the result of biochemical testing (Goodfellow & Magee, 1998; Wayne & Kubica, 1986) that *Mycobacterium terrae* is the only slow-grower known to grow in the presence of 5% NaCl, a trait it shares with rapid-growers. Secondly, the grouping of the strict pathogens *M. leprae* Thai 53 strain and *M. tuberculosis* complex strains and *Mycobacterium haemophilum* ATCC 29548T, the so-called ‘blood-loving mycobacterium’ distinct from other mycobacteria in its requirement for ferric ions in the form of haemin or ferric ammonium citrate, was observed in the hsp65 tree. This grouping reflecting pathogenic potential seems to be reliable and it was also supported by relatively high bootstrap values and conservation of different trees. Thirdly, an unusual branching of *M. leprae* Thai 53 strain, reflecting the rapid evolution of the hsp65 gene of this species, was also observed in the hsp65 tree. It may be associated with the extraordinarily low G+C content of the hsp65 gene of this species. These phylogenetic results emphasize that molecular markers other than the 16S rRNA gene must be applied to obtain a precise phylogenetic and evolutionary scheme.

**Signature nucleotides and amino acids specific to *M. tuberculosis* complex strains**

Signature nucleotides specific to *M. tuberculosis* complex strains were determined by analysing hsp65 sequence alignments, and seven signature nucleotides were detected at five codons (Table 3). Since the signature nucleotides in four codons (positions 76, 181, 200 and 235) were located only at synonymous sites of each codon, they did not affect the protein sequence, but the signature nucleotides in codon 240 were located at all three nucleotides and thus cause amino acid sequence variations [NTM, 240G+CAG; *M. tuberculosis* complex, 240G(GAG)] (Table 3).

The seven *M. tuberculosis* signature nucleotides, particularly the three nucleotides (GGA) concentrated at codon 240, might be used effectively for the development of *M. tuberculosis* detection methods. Actually, a specific polymorphism that distinguished *M. tuberculosis* from NTM strains in the XhoI PRA method developed in this study was also based on the recognition of a signature nucleotide at position 703.

**Identification of clinical isolates by direct sequencing analysis of hsp65**

We applied a direct sequencing protocol targeting 422 bp of hsp65, which enabled species differentiation of mycobacteria in a single sequencing reaction without a cloning procedure, to 70 NTM clinical isolates. No ambiguous results due to two-copy genes with different sequences were obtained using this direct sequencing protocol (data not shown). By referring to the phylogenetic tree constructed by using 56 reference strains, we were able to identify all 70 NTM strains to the species level. The result was concordant with that obtained by conventional biochemical testing and rpoB sequencing analysis (Kim et al., 1999). Furthermore, one *Mycobacterium celatum* strain misidentified as *M. avium* complex by conventional tests, seven *M. kansasi* strains and three *Mycobacterium fortuitum* isolates could be differentiated to the strain level (Table 2).

Compared with the approach targeting the 16S rRNA gene, hsp65 sequence-based analysis has several advantages. Firstly, hsp65 analysis provided a resolving power higher than 16S rRNA gene analysis for differentiation among species. Furthermore, separation between closely related species such as between *Mycobacterium abscessus* and *Mycobacterium chelonae*, between *Mycobacterium szulgai* and *Mycobacterium malmoense* and between *M. kansasi* and *M. gastri* is poorly achieved by 16S rRNA gene analysis due to the small amount of sequence variation (Kirschner...
et al., 1993), but these species can be clearly separated by hsp65 analysis. Secondly, in the multiple alignment of hsp65 sequences of 56 mycobacterial reference strains and 105 clinical isolates, no gaps or additions were found. This means that all the sequence information can be considered for phylogenetic analysis without deletion of gap sequences. This is not the case with the 16S rRNA gene. Thirdly, although rare, some mycobacterial strains (M. celatum and M. terrae complex) have two-copy 16S rRNA genes with different sequences in an organism (Ninet et al., 1996; Reischl et al., 1998). In this case, the application of direct sequencing requires some care in the interpretation of the results due to sequence polymorphisms. Analysis targeting the hsp65 gene, known as a single-copy gene, can resolve

Fig. 2. Comparison of trees based on the hsp65 gene (a) and 16S rRNA gene (b) sequences. The hsp65 sequence tree was constructed from all 56 mycobacterial reference strains but the 16S rRNA tree was from the 41 accessible mycobacterial reference strains. Both trees were constructed using the neighbour-joining method. Percentages indicated at nodes represent bootstrap levels supported by 1000 resampled datasets; values < 50% are not shown. Solid circles on the hsp65 tree indicate that the corresponding nodes (groupings) were also recovered in the maximum-likelihood and maximum-parsimony trees. Stars on the 16S rRNA gene tree indicate that the corresponding nodes (groupings) were conserved in the two trees. T. paurometabola KCTC 9821T was used as an outgroup in both trees. Bars, sequence differences of 0–01% (a) and 0–02% (b). SGM, Slowly growing mycobacteria; RGM, rapidly growing mycobacteria. Strain details and accession numbers are given in Supplementary Table S1 available in IJSEM Online; unless the strain name is given, sequences were obtained from type strains. Trees based on the hsp65 gene and 16S rRNA gene sequences from the same 41 reference strains are available as Supplementary Fig. S1 in IJSEM Online.
this problem. In fact, when direct sequencing protocols targeting 422 bp hsp65 fragments were applied to 70 clinical isolates, no ambiguous results were observed (data not shown).

**XhoI PRA and *M. tuberculosis* complex-specific PCR**

The XhoI PRA algorithm which could differentiate *M. tuberculosis* strains from NTM strains and mycobacterial strains into five groups [A (391, 150, 103 bp), B (391, 253 bp), C (391, 169, 84 bp), D (644 bp) and E (475, 169 bp)] was developed on the basis of the determined hsp65 DNA sequences of 56 mycobacterial reference strains (see Supplementary Fig. S2 available in IJSEM Online). When this PRA method was applied to 105 clinical isolates, all 25 *M. tuberculosis* strains could be clearly differentiated from the other 90 NTM strains, showing a distinct polymorphism (A type) on the 3% agarose gel (Table 1).

Compared with the previous PRA targeting a 441 bp fragment of hsp65 (Devallois et al., 1997; Telenti et al., 1993), which required two restriction enzymes, our XhoI PRA method using only one cheap enzyme offers an easy, rapid and inexpensive procedure for the differential identification of *M. tuberculosis* and NTM strains. However, since it has limitations in the differentiation of all mycobacterial strains, novel PRA algorithms using several restriction enzymes and inexpensive procedure for the differential identification of all the mycobacteria strains should be developed in future.

In the present study, the usefulness of hsp65 sequencing analysis in species identification and phylogenetic study of mycobacteria was evaluated with only 56 reference strains, about 50% of those officially recognized at present, and clinical isolates of frequently encountered species. Therefore, hsp65 sequencing analysis of the other type strains and application of this method to the identification of infrequently encountered strains remain to be studied.

In conclusion, this study shows that a 604 bp fragment of hsp65 is a good alternative molecular marker not only for the phylogenetic analysis of mycobacteria but also for species identification of mycobacterial clinical isolates. The sequence divergence shown by this 604 bp hsp65 fragment is sufficient for it to be applied to the development of molecular methods for the detection of mycobacteria strains with clinical significance, such as *M. tuberculosis*.

**ACKNOWLEDGEMENTS**

This study was supported by a grant from the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (01-PJ1-PG3-20200-0025) and in part by the BK21 project for Medicine, Dentistry, and Pharmacy.

**REFERENCES**


Felsenstein, J. (1993). *PHYLIP* – Phylogeny Inference Package, version 3.5. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, USA.


