

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

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

### Correspondence

Bum-Joon Kim

kbumjoon@snu.ac.kr

<sup>1</sup>Department of Microbiology, College of Medicine, Seoul National University, 28 Yongon-dong, Chongno-gu, Seoul 110-799, Korea

<sup>2,3</sup>Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine<sup>2</sup> and Department of Laboratory Medicine<sup>3</sup>, University of Ulsan College of Medicine, Asan Medical Center, Seoul 138-600, Korea

<sup>4</sup>The Korean Institute of Tuberculosis, The Korean National Tuberculosis Association, Seoul 137-140, Korea

<sup>5</sup>Department of Microbiology, College of Medicine, Konkuk University, Chungju 380-230, Korea

<sup>6</sup>Institute 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.*,

Published online ahead of print on 27 May 2005 as DOI 10.1099/ijs.0.63553-0.

**Abbreviations:** NTM, non-tuberculous mycobacterium; PRA, PCR-restriction fragment analysis.

The GenBank/EMBL/DDBJ accession numbers for the partial *hsp65* sequences determined in this study are AF057449–AF057493.

Details of reference strains, including GenBank accession numbers, trees comparing the *hsp65* and 16S rRNA genes of the same sample of reference strains and details of a protocol to distinguish *M. tuberculosis* complex and NTM isolates are available as supplementary material in IJSEM Online.

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 kansasii* and *Mycobacterium gastri*, 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 *M. avium* and *Mycobacterium paratuberculosis* (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 *XhoI* fragment of the *hsp65* gene (Devallois *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 epidemic studies were also limited to the sequence databases of only some specific species, like the *M. 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 paurometabola* KCTC 9821<sup>T</sup> 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 the amplification of DNA encoding an 18-kDa protein (Williams *et al.*,

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

Species	Sequence similarity (%)	Strains (n)	<i>XhoI</i> PRA pattern
<i>M. tuberculosis</i>	100	25	A
<i>M. leprae</i> *	100	16	B
<i>M. avium</i>	99.2–100	6	C
<i>M. intracellulare</i>	99.2–100	10	C
<i>M. scrofulaceum</i>	100	3	B
<i>M. szulgai</i>	98.7–100	8	C
<i>M. kansasii</i> †	99.2–100	18	B
<i>M. fortuitum</i>	98.2–100	5	D
<i>M. peregrinum</i>	100	3	D
<i>M. abscessus</i>	98.5–100	11	D
Total	98.2–100	105	

\*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

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

\*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*.

1990) were provided or performed by the Institute of Hansen's Disease, Catholic University Medical College, Seoul, Korea. Two heat-killed reference strains of *M. kansasii* (subspecies II and III) were kindly provided by Veronique Vincent (TB and Mycobacteria Lab, Institute of Pasteur, Paris, France). The type strain of *T. paurometabola* (KCTC 9821<sup>T</sup>) was provided by the Korean Collection for Type Cultures (KCTC). The 70 non-tuberculous mycobacterium (NTM) clinical isolates for the blinded *hsp65* direct sequencing analysis were identified by conventional biochemical testing and provided by the Department of Internal Medicine at the Asan Medical Center.

**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.

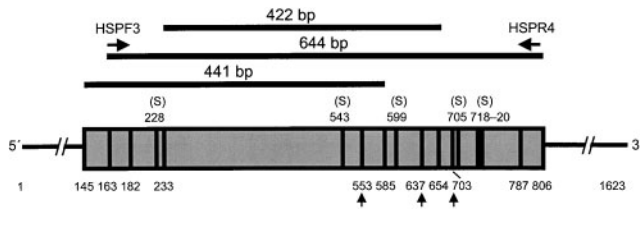
A set of primers [forward primer HSPF3 (5'-ATCGCCAA-GGAGATCGAGCT-3'; positions 163–182 in the *hsp65* gene of *Mycobacterium tuberculosis*, GenBank accession no. M15467), reverse primer HSPR4 (5'-AAGGTGCCGCGGATCTTGTT-3'; positions 806–787)] were used for the amplification of 644 bp PCR products from *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), 40 mM KCl, 1.5 mM MgCl<sub>2</sub> 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<sup>-1</sup>) 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 9821<sup>T</sup> 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 9821<sup>T</sup> as an outgroup. Evolutionary-distance matrices were



**Fig. 1.** Locations of primers for PCR targeting sites for direct sequencing and restriction sites for *XhoI* 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 *XhoI*.

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 PHYLIP 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 sequence 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 *XhoI* 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 (*XhoI*) 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 *XhoI* were also calculated using the MapDraw program. *XhoI* 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

*XhoI* recognition sequence ( $703\text{C}\downarrow\text{TCGAC}$ ). PCR was performed using the set of primers described above (HSPF3 and HSPR4) and using the same PCR conditions. Restriction enzyme *XhoI* (Takara) was then added to the 644 bp PCR product of the 56 reference strains. Briefly, 10  $\mu\text{l}$  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  $\mu\text{l}$  per reaction. Digestion was performed for 2 h at 37  $^{\circ}\text{C}$ . 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. paurometabola* KCTC 9821<sup>T</sup> 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 *Mycobacterium phlei* ATCC 11758<sup>T</sup>) was observed among the *hsp65* DNA sequences of mycobacterial strains (interspecies divergence > 16.9%). In a variety of *M. tuberculosis* complex strains, namely *Mycobacterium africanum* ATCC 25420<sup>T</sup>, *Mycobacterium bovis* ATCC 19210<sup>T</sup>, *Mycobacterium bovis* BCG, *Mycobacterium microti* ATCC 19422<sup>T</sup> and *M. tuberculosis* ATCC 27294<sup>T</sup>, the nucleotide sequences of the amplified *hsp65* genes were identical. The *hsp65* nucleotide sequences of *Mycobacterium simiae* ATCC 25275<sup>T</sup> and *Mycobacterium genavense* ATCC 51233<sup>T</sup>, 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

Differences between the two groups are highlighted in bold.

Group	Codon (nucleotide positions)				
	76 (226–228)	181 (541–543)	200 (597–599)	235 (703–705)	240 (718–720)
<i>M. tuberculosis</i> complex	GTA (V)	TTT (F)	GGG (G)	CTC (L)	<b>GGA (G)</b>
NTM	GTC (V), GTT (V)	TTC (F)	GGC (G), GGT (G)	CTG (L)	CAG (Q)

*hsp65* DNA sequence similarity, except for *M. genavense* ATCC 51233<sup>T</sup> and *M. simiae* ATCC 25275<sup>T</sup>, 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% (intraspecies 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. kansasii* showed similarity in the range 99.2–100% with the *M. kansasii* type I reference strain (ATCC 12478<sup>T</sup>), with the other strain showed a similarity of 99.7% with *M. kansasii* 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 partially supported by the result of biochemical testing (Goodfellow & Magee, 1998; Wayne & Kubica, 1986) that *Mycobacterium triviale* 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 29548<sup>T</sup>, 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**

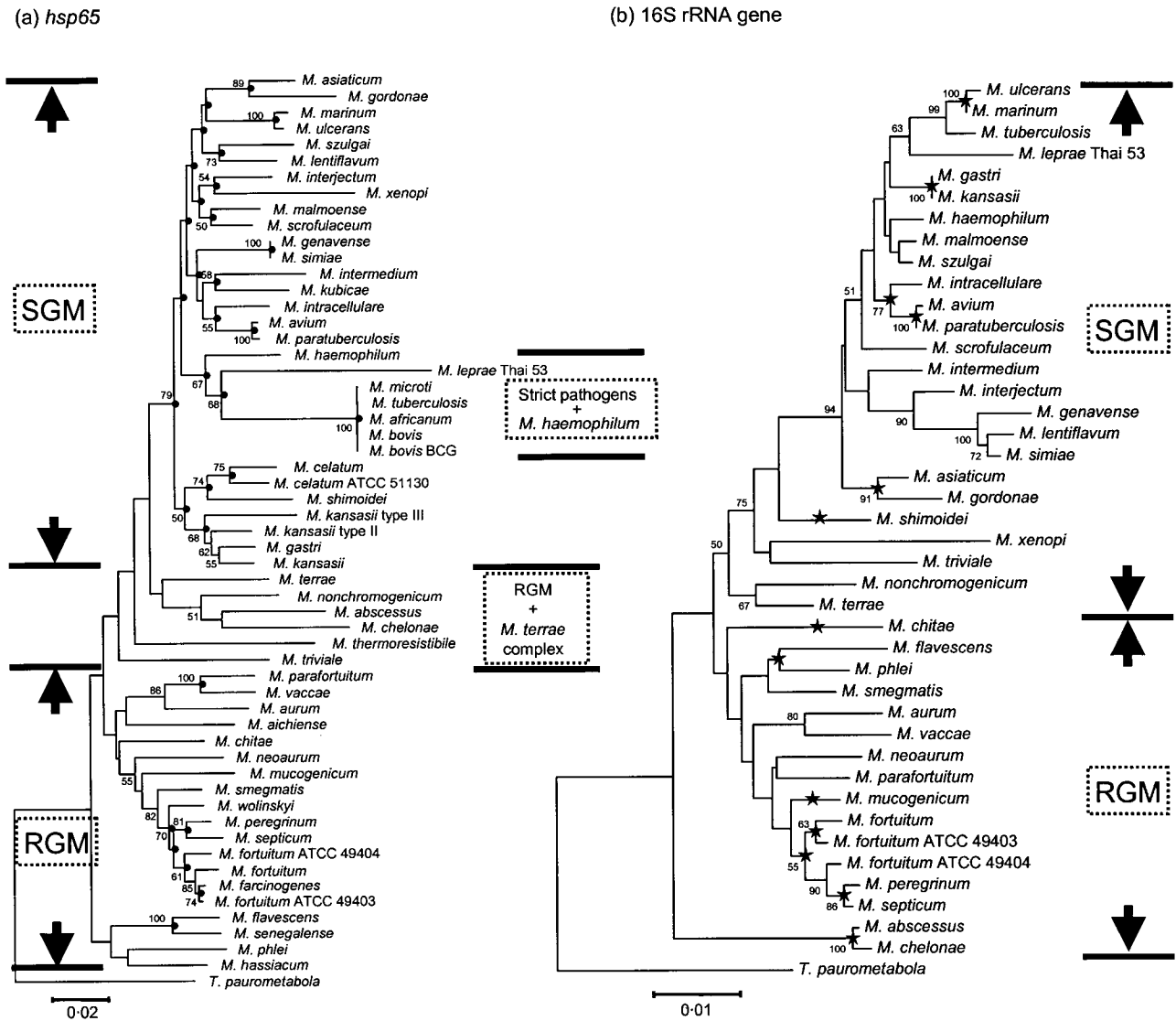
Signature nucleotides specific to *M. tuberculosis* complex strains were determined by analysing *hsp65* sequence alignments, and seven signature nucleotides were detected at five codons (Table 3). Since the signature nucleotides in four codons (positions 76, 181, 200 and 235) were located only at synonymous 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, 240<sup>Q(CAG)</sup>; *M. tuberculosis* complex, 240<sup>G(GGA)</sup>] (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 *Xho*I 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. kansasii* 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. kansasii* and *M. gastri* is poorly achieved by 16S rRNA gene analysis due to the small amount of sequence variation (Kirschner



**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 9821<sup>T</sup> 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.

*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 1S 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

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 for the differentiation 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**

- Chang, Y. H., Shangkuan, Y. H., Lin, H. C. & Liu, H. W. (2003). PCR assay of the *groEL* gene for detection and differentiation of *Bacillus cereus* group cells. *Appl Environ Microbiol* **69**, 4502–4510.
- Clayton, R. A., Sutton, G., Hinkle, P. S., Jr, Bult, C. & Fields, C. (1995). Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. *Int J Syst Bacteriol* **45**, 595–599.
- Devallois, A., Goh, K. S. & Rastogi, N. (1997). Rapid identification of mycobacteria to species level by PCR-restriction fragment length polymorphism analysis of the *hsp65* gene and proposition of an algorithm to differentiate 34 mycobacterial species. *J Clin Microbiol* **35**, 2969–2973.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Felsenstein, J. (1993). PHYLIP – Phylogeny Inference Package, version 3.5. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.
- Fitch, W. M. (1972). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Zool* **20**, 406–416.
- Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.
- Goodfellow, M. & Magee, J. G. (1998). Taxonomy of mycobacteria. In *Mycobacteria*, vol. 2, *Basic Aspects*, pp. 1–71. Edited by P. R. J. Gangadharam & P. A. Jenkins. New York & London: Chapman & Hall.
- Hafner, B., Haag, H., Geiss, H. K. & Nolte, O. (2004). Different molecular methods for the identification of rarely isolated non-tuberculous mycobacteria and description of new *hsp65* restriction fragment length polymorphism patterns. *Mol Cell Probes* **18**, 59–65.
- Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.
- Kim, B. J., Lee, S. H., Lyu, M. A., Kim, S. J., Bai, G. H., Chae, G. T., Kim, E. C., Cha, C. Y. & Kook, Y. H. (1999). Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). *J Clin Microbiol* **37**, 1714–1720.
- Kim, B. J., Lee, K. H., Park, B. N., Kim, S. J., Bai, G. H., Kim, S. J. & Kook, Y. H. (2001). Differentiation of mycobacterial species by PCR-restriction analysis of DNA (342 base pairs) of the RNA polymerase gene (*rpoB*). *J Clin Microbiol* **39**, 2102–2109.
- Kirschner, P., Springer, B., Vogel, U., Meier, A., Wrede, A., Kiekenbeck, M., Bange, F. C. & Bottger, E. C. (1993). Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory. *J Clin Microbiol* **31**, 2882–2889.
- Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, AZ, USA.
- Kwok, A. Y., Su, S. C., Reynolds, R. P., Bay, S. J., Av-Gay, Y., Dovichi, N. J. & Chow, A. W. (1999). Species identification and phylogenetic relationships based on partial HSP60 gene sequences within the genus *Staphylococcus*. *Int J Syst Bacteriol* **49**, 1181–1192.
- Lee, J. H., Park, H. S., Jang, W. J. & 9 other authors (2003). Differentiation of rickettsiae by *groEL* gene analysis. *J Clin Microbiol* **41**, 2952–2960.
- Ninet, B., Monod, M., Emler, S., Pawlowski, J., Metral, C., Rohner, P., Auckenthaler, R. & Hirschel, B. (1996). Two different 16S rRNA genes in a mycobacterial strain. *J Clin Microbiol* **34**, 2531–2536.

- Palys, T., Nakamura, L. K. & Cohan, F. M. (1997). Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* **47**, 1145–1156.
- Pitulle, C., Dorsch, M., Kazda, J., Wolters, J. & Stackebrandt, E. (1992). Phylogeny of rapidly growing members of the genus *Mycobacterium*. *Int J Syst Bacteriol* **42**, 337–343.
- Reischl, U., Feldmann, K., Naumann, L., Gaugler, B. J. M., Ninet, B., Hirschel, B. & Emler, S. (1998). 16S rRNA sequence diversity in *Mycobacterium celatum* strains caused by presence of two different copies of 16S rRNA gene. *J Clin Microbiol* **36**, 1761–1764.
- Ringuet, H., Akoua-Koffi, C., Honore, S., Varnerot, A., Vincent, V., Berche, P., Gaillard, J. L. & Pierre-Audigier, C. (1999). *hsp65* sequencing for identification of rapidly growing mycobacteria. *J Clin Microbiol* **37**, 852–857.
- Rogall, T., Wolters, J., Flohr, T. & Böttger, E. C. (1990). Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. *Int J Syst Bacteriol* **40**, 323–330.
- Roth, A., Fischer, M., Hamid, M. E., Michalke, S., Ludwig, W. & Mauch, H. (1998). Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. *J Clin Microbiol* **36**, 139–147.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Segal, G. & Ron, E. Z. (1996). Regulation and organization of the *groE* and *dnaK* operons in *Eubacteria*. *FEMS Microbiol Lett* **138**, 1–10.
- Smole, S. C., McAleese, F., Ngampasutadol, J., Von Reyn, C. F. & Arbeit, R. D. (2002). Clinical and epidemiological correlates of genotypes within the *Mycobacterium avium* complex defined by restriction and sequence analysis of *hsp65*. *J Clin Microbiol* **40**, 3374–3380.
- Stahl, D. A. & Urbance, J. W. (1990). The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *J Bacteriol* **172**, 116–124.
- Stone, B. B., Nietupski, R. M., Breton, G. L. & Weisburg, W. G. (1995). Comparison of *Mycobacterium* 23S rRNA sequences by high-temperature reverse transcription and PCR. *Int J Syst Bacteriol* **45**, 811–819.
- Swanson, D. S., Pan, X. & Musser, J. M. (1996). Identification and subspecific differentiation of *Mycobacterium scrofulaceum* by automated sequencing of a region of the gene (*hsp65*) encoding a 65-kilodalton heat shock protein. *J Clin Microbiol* **34**, 3151–3159.
- Takewaki, S. I., Okuzumi, K., Manabe, I., Tanimura, M., Miyamura, K., Nakahara, K. I., Yazaki, Y., Ohkubo, A. & Nagai, R. (1994). Nucleotide sequence comparison of the mycobacterial *dnaJ* gene and PCR-restriction fragment length polymorphism analysis for identification of mycobacterial species. *Int J Syst Bacteriol* **44**, 159–166.
- Telenti, A., Marchesi, F., Balz, M., Bally, F., Böttger, E. C. & Bodmer, T. (1993). Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *J Clin Microbiol* **31**, 175–178.
- Turenne, C. Y., Tschetter, L., Wolfe, J. & Kabani, A. (2001). Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous *Mycobacterium* species. *J Clin Microbiol* **39**, 3637–3648.
- Viale, A. M., Arakaki, A. K., Soncini, F. C. & Ferreyra, R. G. (1994). Evolutionary relationships among eubacterial groups as inferred from GroEL (chaperonin) sequence comparisons. *Int J Syst Bacteriol* **44**, 527–533.
- Wayne, L. G. & Kubica, G. P. (1986). The mycobacteria. In *Bergey's Manual of Determinative Bacteriology*, 9th edn, pp. 1435–1457. Edited by R. E. Buchanan & N. E. Gibbons. Baltimore: Williams & Wilkins.
- Williams, D. L., Gillis, T. P., Booth, R. J., Looker, D. & Watson, J. D. (1990). The use of a specific DNA probe and polymerase chain reaction for the detection of *Mycobacterium leprae*. *J Infect Dis* **162**, 193–200.