Available online at www.sciencedirect.com
ScienceDirect
Systematic and Applied Microbiology 30 (2007) 453–462
SYSTEMATIC AND APPLIED MICROBIOLOGY
www.elsevier.de/syapm

Mycobacterium species identification – A new approach via *dnaJ* gene sequencing ☆

Makiko Yamada-Noda[a,b], Kiyofumi Ohkusu[a], Hiroyuki Hata[c], Mohammad Monir Shah[a], Pham Hong Nhung[a], Xiao Song Sun[a], Masahiro Hayashi[a], Takayuki Ezaki[a,*]

[a]*Department of Microbiology, Regeneration and Advanced Medical Science, Gifu University Graduate School of Medicine, 1-1 Yanagido, Gifu 501-1194, Japan*
[b]*Gifu Prefectural Institute of Health and Environmental Sciences, 1-1 Naka-fudogaoka, Kakamigahara 504-0838, Japan*
[c]*Department of Research & Development, Kyokuto Pharmaceutical Industrial Company Limited, 3333-26 Aza-Asayama, Kamitezuna, Takahagi-shi, Ibaraki 318-0004, Japan*

Abstract

The availability of the *dnaJ1* gene for identifying *Mycobacterium* species was examined by analyzing the complete *dnaJ1* sequences (approximately 1200 bp) of 56 species (54 of them were type strains) and comparing sequence homologies with those of the 16S rRNA gene and other housekeeping genes (*rpoB*, *hsp65*). Among the 56 *Mycobacterium* species, the mean sequence similarity of the *dnaJ1* gene (80.4%) was significantly less than that of the 16S rRNA, *rpoB* and *hsp65* genes (96.6%, 91.3% and 91.1%, respectively), indicating a high discriminatory power of the *dnaJ1* gene. Seventy-one clinical isolates were correctly clustered to the corresponding type strains, showing isolates belonging to the same species. In order to propose a method for strain identification, we identified an area with a high degree of polymorphism, bordered by conserved sequences, that can be used as universal primers for PCR amplification and sequencing. The sequence of this fragment (approximately 350 bp) allows accurate species identification and may be used as a new tool for the identification of *Mycobacterium* species.

Keywords: *Mycobacterium*; *dnaJ*; Taxonomic study; Identification

Introduction

The genus *Mycobacterium* comprises a wide range of organisms, including obligate parasites causing serious human and animal diseases, opportunistic pathogens, and saprophytic species found in nature. At present, the genus *Mycobacterium* contains more than 110 species, the vast majority of which have been isolated from human or animal samples. Traditionally, the definitive diagnosis of mycobacterial infections has been dependent on the isolation and identification of causative agents and requires a series of specialized physiological and biochemical tests. The procedures for these tests are complex, laborious, and usually impeded by the slow growth of mycobacteria in clinical laboratories.

The use of molecular genetic methods such as 16S rRNA gene sequence analysis has facilitated a much

☆ Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession numbers described in Table 1.

*Corresponding author. Tel.: +81 58 230 6487; fax: +81 58 230 6489.
E-mail address: tezaki@gifu-u.ac.jp (T. Ezaki).

0723-2020/$ - see front matter © 2007 Published by Elsevier GmbH.
doi:10.1016/j.syapm.2007.06.003

tighter circumscription of the genus, and the availability of comparative 16S rRNA gene sequence data with improved phenotypic data has resulted in much improved and more reliable species identification [5,10,27]. Although the sequence of the 16S rRNA gene is the most widely used molecular marker to determine the phylogenetic relationships of bacteria, low intragenus polymorphism limits its usefulness for taxonomic analysis or identification to the species level. As an example, the species *Mycobacterium kansasii* and *Mycobacterium gastri* and the species *Mycobacterium mucogenicum* and *Mycobacterium phocaicum* have 100% 16S rDNA identity values.

DnaJ, one of the members of the Hsp40 (heat-shock protein) family, co-regulates the activity of heat-shock sigma factor 32, and it has been used to study several unrelated genera, including *Legionella* spp. [13], *Streptococcus* spp. [7], *Staphylococcus* spp. [26], *Vibrio* spp. [17], *Aeromonas* spp. [19] and the *Enterobacteriaceae* family [18]. In the genus *Mycobacterium*, two genes encoding Hsp40 homologues are annotated as *dnaJ1* and *dnaJ2*, but the gene and protein similarities are low (56.6% and 44.9%, respectively). Overexpression studies suggest that DnaJ1 has an important role in the upregulation of other genes encoding heat-shock proteins, such as DnaK and GrpE, but that DnaJ2 has no detectable effect on the transcription of other genes [28]. Recently, the *dnaJ* gene was used as an alternative tool to identify *Mycobacterium* species [16,30]. However, only a limited number of reference species (17 slowly growing and 2 rapidly growing) and a limited region of *dnaJ* sequences in the genus *Mycobacterium* were used.

In the study described here, we investigated the complete *dnaJ1* sequences of 56 *Mycobacterium* species, compared with 16S rRNA and other conserved gene (*hsp65*, *rpoB*) sequences available in public databases. The complete *dnaJ1* sequences of 88 clinical strains were also obtained for investigation of strain diversity. In addition, we designed universal primers for amplification of a small but discriminatory sequence for use in the routine identification of *Mycobacterium* species.

Materials and methods

Bacterial strains and genomic material

The 57 strains belonging to 56 *Mycobacterium* species (54 of them were type strains) and *Nocardia farcinica* used in this study are listed in Table 1. Clinical isolates of *Mycobacterium tuberculosis* (32 strains), *M. genavense* (1 strain), *M. ulcerans* (1 strain), *M. avium* complex (15 strains), *M. scroflaceum* (1 strain), *M. kansasii* (17 strains), *M. gastri* (2 strains), *M. parafortuitum* (1 strain) and *M. mucogenicum* (1 strain) were identified by conventional methods and 16S rRNA gene sequences, and were used for *dnaJ1* sequence analysis. All strains were cultured on 2% Ogawa medium, Middlebrook 7H9 broth (Kyokuto Pharmaceutical Industrial Co., Ltd., Tokyo, Japan) or Middlebrook 7H11 medium (BD Co., Franklin Lakes, NJ, USA) and incubated at 37 °C. Bacterial DNAs were extracted from a heavy suspension of bacteria by using the MORA-EXTRACTION kit (AMR Co., Ltd., Gifu, Japan), according to the protocol of the supplier.

Primer design for determination of complete *dnaJ1* sequences

The *dnaJ1* sequences of *M. tuberculosis*, *M. avium* subsp. *paratuberculosis*, *M. leprae*, (GenBank accession numbers X58406, NC_002944, and NC_002677, respectively) were aligned in order to generate primers to be used for PCR and sequencing. Additional primers were selected from ongoing base sequence determinations. All primers used in this study are summarized in Table 2.

PCR and sequencing

Amplification was performed with a 20-μl reaction mixture containing 2 μl of 10 × PCR buffer, 0.5 U of *Taq* polymerase (Takara shuzo Co., Ltd., Otsu, Japan), 0.2 mM dNTPs, 2.5 μl of each primer (0.5 μM), 1 μl of dimethylsulfoxide (DMSO), 8.3 μl of distilled water and 2 μl of DNA (10 ng). DMSO was used at a concentration of 5% (v/v) in all PCR mixtures so as to reduce the formation of secondary structures in the single DNA strands. The thermal profile involved initial denaturation for 5 min at 95 °C, 35 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 58 °C, and extension for 1 min at 72 °C followed by 1 cycle at 72 °C for 7 min. PCR products were purified for sequencing by use of a Wizard SV Gel and PCR Clean-Up System (Promega Co., Ltd., Madison, WI, USA), according to the instructions of the supplier. Sequencing reactions were carried out with the reagents of the ABI Prism 3100 DNA sequencer (Applied Biosystems, Hitachi, Japan) by a standard automated sequencer protocol.

The nucleotide sequences of the *dnaJ1* gene fragments obtained were processed into sequence data with DNASIS Pro Software (Hitachi Software Engineering Co., Ltd., Tokyo, Japan) and partial sequences were combined into a single consensus sequence. The sequences of *M. avium* subsp. *paratuberculosis* strain K-10 and *Nocardia farcinica* strain IFM 10152 were obtained from the NCBI database. DNA sequences from 16S rRNA, *rpoB* and *hsp65* genes were obtained mostly from public databases. Additional sequences were obtained using published primers [14]. All GenBank accession numbers of sequences are listed in Table 1.

Table 1. *Mycobacterium* strains used in this study

Species	Strain no.	Original no.	Genbank accession no.	Size (bp) of *dnaJ1* sequence determined		Genbank accession no.		
			dnaJ1	Complete	Partial	16S rRNA	*rpoB*	*hsp65*
Slow growers								
M. africanum	GTC 605^T	ATCC 25420^T	AB292525	1188	386	AF480605	AY544880	AF547803
M. asiaticum	GTC 606^T	ATCC 25276^T	AB292526	1173	371	AF480595	AY544884	AF547806
M. arupense	GTC 2730	CST 7052	AB292527	1185	380	DQ157760	AB239921	DQ168662
M. avium								
subsp. *avium*	GTC 603^T	ATCC 25291^T	AB292528	1179	377	AJ536037	AY544887	AF547808
subsp. *silvaticum*	GTC 1695^T		AB292529	1179	377	AB292582	AY544889	AF547810
subsp. *paratuberculosis*		Strain K-10	NC_002944 [a]	1179	377	X52934	AY544888	AF547809
M. bovis	GTC 602^T	ATCC 19210^T	AB292530	1188	386	AB292583	AY544894	AF547813
M. branderi	GTC 811^T	ATCC 51789^T	AB292531	1185	380	AF480574	AY544895	AF547815
M. caprae		ATCC BAA-824^T	AB292532	1188	386	AJ131120	AY544972	AF547884
M. celatum	GTC 816^T	ATCC 51131^T	AB292533	1197	392	L08169	AY544897	AF547817
M. chimaera	GTC 3154^T	CIP 107892^T	AB292534	1179	377	AJ548480	AY943187	AY943198
M. cookie	GTC 856^T		AB292535	1185	380	AF480598	AY544904	AF547824
M. gastri	GTC 610^T	ATCC 15754^T	AB292536	1188	386	AF480602	AY544916	AF547836
M. genavense	GTC 1694^T		AB292537	1179	377	X60070	AF057467 [b]	AB292585
M. gordonae	GTC 612^T	ATCC 14470^T	AB292538	1170	368	X52923	AY544919	AF547840
M. haemophilum	GTC 3151^T	ATCC 29548^T	AB292539	1176	374	L24800	AY544920	AF547841
M. hiberniae	GTC 2687^T	DSM 44241^T	AB292540	1200	395	X67096	AY544924	AY438083
M. interjectum	GTC 679^T	DSM 44064^T	AB292541	1179	377	X70961	AY544928	AF547846
M. intermedium	GTC 689^T	DSM 44049^T	AB292542	1167	365	X67847	AY544929	AF547847
M. intracellulare	GTC 613^T	ATCC 13950^T	AB292543	1179	377	AJ536036	AY544930	AF547848
M. kansasii	GTC 614^T	ATCC 12478^T	AB292544	1188	386	AJ536035	AY544931	AF547849
M. kumamotonense	GTC 2729^T	CST 7247^T	AB292545	1194	389	AB239925	AB239919	AB239920
M. lentiflavum		ATCC 51985^T	AB292546	1179	377	AF480583	AY544939	AF547851
M. leprae	DN 1780[c]	Clinical sample	AB292547	1167	371	X53999	Z14314	AY772718
M. malmoense	GTC 615^T	ATCC 29571^T	AB292548	1176	374	X52930	AY544942	AF547854
M. marinum	GTC 616^T	ATCC 927^T	AB292549	1191	389	AJ536032	AY544943	AF547855
M. microti	GTC 604^T	NCTC 8710^T	AB292550	1188	386	AF480584	AY544944	AF547856
M. nonchromogenicum	GTC 617^T	ATCC 19530^T	AB292551	1194	392	X52928	AY544949	AF547861
M. scrofulaceum	GTC 619^T	ATCC 19981^T	AB292552	1179	377	AF480604	AY544959	AF547871
M. shimoidei	GTC 627^T	E4796^T	AB292552	1188	386	X82459	AY544962	AF547874
M. simiae	GTC 620^T	ATCC 25275^T	AB292554	1188	386	X52931	AY544963	AF547875
M. szulgai	GTC 622^T	NCTC 10831^T	AB292555	1173	371	X52926	AY544966	AF547878
M. terrae	GTC 623^T	ATCC 15755^T	AB292556	1182	377	X52925	AY544967	AF547879
M. triviale	GTC 624^T	ATCC 23292^T	AB292557	1185	377	X88924	AY544971	AF547883
M. tuberculosis	GTC 601^T	ATCC 27294^T	AB292558	1188	386	X58890	AY544974	AF547885
M. ulcerans	GTC 859^T		AB292559	1191	389	X88926 [b]	AY544975	AF547888
M. xenopi	GTC 626^T	NCTC 10042^T	AB292560	1179	377	AJ536033	AY544979	AF547891

Table 1. (*continued*)

Species	Strain no.	Original no.	Genbank accession no.	Size (bp) of *dnaJ1* sequence determined		Genbank accession no.		
			dnaJ1	Complete	Partial	16S rRNA	*rpoB*	*hsp65*
Rapid growers								
M. abscessus	GTC 857^T		AB292561	1182	380	M29559	AY544879	AY458075
M. alvei	GTC 1997^T	DSM 44176^T	AB292562	1167	365	AF023664	AY544883	AF547805
M. aubagnense	GTC 3158^T	CCUG 50186^T	AB292563	1191	380	AY859683	AY859694 [a]	AY859677
M. bolletii	GTC 3156^T	CCUG 50184^T	AB292564	1182	380	AY859681	AY859692 [a]	AY859675
M. chelonae	GTC 858^T		AB292565	1185	383	AF480594	AY544898	AY458074
M. chitae	GTC 607^T	ATCC 19627^T	AB292566	1179	374	X67874	AY544899	AF547819
M. flavescens	GTC 608^T	ATCC 14474^T	AB292567	1182	380	X52932	AY544911	AF547831
M. fortuitum	GTC 609^T	ATCC 6841^T	AB292568	1167	365	AJ536039	AY544913	AY458072
M. gilvum	GTC 2689^T	JCM 6395^T	AB292569	1179	377	X55599	AY544917	AF547838
M. houstonense	GTC 1070^T	ATCC 49403^T	AB292570	1167	365	AY012579	AY147173 [a]	AY458077
M. mucogenicum	GTC 3155^T	CCUG 47451^T	AB292571	1191	380	AY457074	AY544946	AF547858
M. neworleansense	GTC 1069^T	ATCC 49404^T	AB292572	1167	365	AY012575	AY943188	AY943199
M. parafortuitum	GTC 2688^T	JCM 6367^T	AB292573	1185	380	X93183	AY544952	AF547864
M. peregrinum	GTC 1725^T		AB292574	1167	365	AF130308	AY544953	AY458069
M. phlei	GTC 618^T	ATCC 11758^T	AB292575	1173	371	AF480603	AY544954	AF547866
M. phocaicum	GTC 3157^T	CCUG 50185^T	AB292576	1185	374	AY859682	AY859693	AY859676
M. porcinum	GTC 1797^T	JCM 6378^T	AB292577	1167	365	AF480588	AY544955	AF547867
M. senegalense	GTC 1622^T		AB292578	1167	365	AF480596	AY544960	AY458067
M. septicum	GTC 1996^T	JCM 44393^T	AB292579	1167	365	AF111809	AY544961	AF547873
M. smegmatis	GTC 621^T	ATCC 19420^T	AB292580	1173	371	AJ131761	AY544964	AY458065
M. vaccae	GTC 625^T	ATCC 15483^T	AB292581	1182	377	X55601	AY544976	AF547889
Outgroups								
Nocardia farcinica		IFM 10152	AP006618 [a]	1158	359	AP006618 [a]	AP006618 [a]	AP006618 [a]

[a]Sequences were trimmed to start and finish at the corresponding regions.
[b]Not a type strain.
[c]Patient derived DNA.

Table 2. Primers used for amplification and sequencing of the entire *dnaJ1* gene in this study

Primer name	Sequence (5′→3′)	Target gene	Position[a]
MYCO-DNK-F1	TGGCTCGTGCGGTCGGTATC	*dnaK*	−2615
MYCO-grpE-F1	CTGAAGTCGGTCGCCGACAAG	*grpE*	−399
MYCO-grpE-F2	CAGGCSGTGCAGCACGAGG	*grpE*	−268
MYCO-DN-F1	GAATGGGTCGAAAAAGACTTC	*dnaJ1*	13
MYCO-DN-F1-2	GAGTGGGTTGAGAAGGACTTC	*dnaJ1*	13
MYCO-DN-F14	TACAAGGAGCTCGGCGTCTCC	*dnaJ1*	34
MYCO-DN-R18-1	GTCTCGTCGTACTCCTTGCG	*dnaJ1*	202
MYCO-DN-R18-2	GTTTCGTCGTATTCTTTGCG	*dnaJ1*	202
MYCO-DN-R18muc	GGTTTCGTCGTATTCGGCAC	*dnaJ1*	202
MYCO-DN-F4	GATGGCGCCGAGTTCAAYC	*dnaJ1*	310
MYCO-DN-R2	CACCGAACAAGTCACCGATG	*dnaJ1*	372
MYCO-DN-F12	CGGTTCGATCATCGAGCACC	*dnaJ1*	666
MYCO-DN-F19	TCGATCATCGAGCACCCGTG	*dnaJ1*	680
MYCO-DN-R11	ATCCGCACGTTGATGGTTCG	*dnaJ1*	727
MYCO-DN-R17	ACCCCGACCTTGCCGTCCAG	*dnaJ1*	946
MYCO-DN-R7	TCASCGATTACCTGCCCATCC	*dnaJ1*	1168
DN4138R	TCRATRATSCGCTTGATGC	*hspR*	1408
MYCO-hspR-R1	AGGTTGACCCCCTCGTCCTG	*hspR*	1383
MYCO-hspR-R2	CCGGCSAGCTCGGCSGCCA	*hspR*	1154
MYCO-hspR-R3	TGCGCTGCGGGSTSACCAGG	*hspR*	1212
J10F	CGIGARTGGGTYGARAARG	*dnaJ1*	10
J335R	ARICCICCGAAIARRTCICC	*dnaJ1*	376

[a]Position relative to the *M. tuberculosis dnaJ1* sequence.

Phylogenetic analysis

For phylogenetic analysis of the complete *dnaJ1* gene, sequences were trimmed to start at the start codon (ATG or GTG) and finish at the stop codon (TGA or TGG) for all strains and clinical isolates studied (1167–1200 bp). Multiple-sequence alignments and percent similarities of these DNA sequences between the different species were obtained with the CLUSTAL W software, version 1.83 [32]. Phylogenetic trees were obtained from DNA sequences by the neighbor-joining method. Bootstrap replicates (1000 replicates) were made in order to estimate the reliabilities of the nodes of the phylogenetic trees obtained. Maximum-likelihood trees were also generated by the DNAML program in the PHYLIP software package and they were drawn with TreeView [21].

Determination of discriminatory partial sequences in *dnaJ1* genes

Based on complete *dnaJ1* sequences, several *dnaJ1* universal primers were designed for identification of *Mycobacterium* species. After trying several primers, the primer pair J10F and J335R (Table 2) was finally selected, with amplicon sizes from 365 to 395 bp, depending on the species. The sequences from 326 to 356 bp in length (excluding 39 nucleotides at the two ends corresponding to primer binding sites) were used for sequence comparison. The thermal profile was initial denaturation for 5 min at 95 °C, 35 cycles of denaturation for 30 s at 95 °C, annealing for 30 s at 55 °C, and extension for 1 min at 72 °C followed by 1 cycle at 72 °C for 7 min.

Results and discussion

Determination of *dnaJ1* sequences in *Mycobacterium* species

Complete *dnaJ1* gene sequences were determined for all 56 *Mycobacterium* species and *N. farcinica*. The *dnaJ1* gene of *Mycobacterium* species was 1167 to 1200 bp in length (Table 1), and the G+C content of the *dnaJ1* gene ranged from 59.6 to 72.3 mol%, which is representative of the G+C content of the DNA in *Mycobacterium* species (62–70 mol%) [34]. The similarity between different species ranged from 69.6% to 100%. In the case of the assumption that the *M. tuberculosis* complex (*M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti* and *M. caprae*) was regarded as a single species, the similarity between different species ranged from 69.6% to 98.8% (Table 3).

For those mycobacterial taxa that could not be discriminated on the basis of their 16S rRNA gene sequence, other house-keeping genes such as *gyrB* [8,20], *rpoB* [9], *hsp65* [4,15,23], *recA* [3], *sodA* [2] and the 16S-23S

Table 3. Comparison of similarities for 16S rRNA, *dnaJ1*, *rpoB* and *hsp65* gene sequences

Species or subspecies compared	% Similarity				
	16S rDNA	*dnaJ1*		*rpoB*	*hsp65*
		Complete	Partial		
56 *Mycobacterium* species					
Mean[a]	96.6	80.4	75.7	91.3	91.1
Range[a]	93.4–100	69.6–98.8	63.0–98.9	85.7–99.5	82.1–100
Closely related species[b]					
M. tuberculosis complex	99.9–100	99.9–100	99.7–100	100	99.7–100
M. kansasii – *M. gastri*	100	95.5	94.5	94.4	97.7
M. marinum – *M. ulcerans*	99.9	98.8	98.9	99.3	99.4
M. intracellulare – *M. chimaera*	99.9	98.2	97.6	100	99.1
M. szulgai – *M. malmoense*	99.7	86.4	79.1	91.0	93.7
M. mucogenicum – *M. phocaicum*	100	95.8	95.5	98.7	100
M. abscessus – *M. bolletii*	99.9	97.5	97.7	96.7	98.3
M. porcinum – *M. neworleansense*	99.9	96.0	96.6	97.0	96.9
Subspecies					
M. avium subsp. *avium* – *M. avium* subsp. *silvaticum*	100	99.9	100	100	100

[a] *M. tuberculosis* complex (*M. tuberculosis*, *M. bovis*, *M. africanum*, *M. microti* and *M. caprae*) was regarded as a single species.
[b] Determined on the basis of 99.7% similarity of the 16S rRNA gene sequence.

rRNA gene internal transcribed spacer (ITS) [24,25] were targeted for gene sequence-based identification. As compared with 16S rRNA and other conserved gene (*hsp65*, *rpoB*) sequences that could be obtained from public databases or sequenced by using published primers, the mycobacterial *dnaJ1* gene showed statistically higher discrimination (mean similarity 80.4%) than the 16S rDNA (96.6%), *rpoB* (91.3%) and *hsp65* (91.1%) genes ($P = 0$; Student's t test). This higher degree of polymorphism was particularly evident for species not well differentiated by 16S rDNA sequence analysis, since, among 7 pairs of species with 16S rRNA gene similarities ranging from 99.7% to 100%, the similarities of the *dnaJ1* gene ranged from 86.4% to 98.8%. The mean values of the similarities between the 16S rRNA gene and *dnaJ1* gene sequences among these 7 pairs were statistically significant. Also, the complete *dnaJ1* gene showed more discriminatory power than the *rpoB* and *hsp65* genes. Among these 7 pairs, 6 pairs were more polymorphic compared with *rpoB* and *hsp65* genes. Thus, the complete *dnaJ1* gene will clearly facilitate the differentiation of any two bacterial species. The similarity between the two *M. avium* reference subspecies was 99.9% and, thus, it was 1.1% above the highest level of similarity (except for the *M. tuberculosis* complex) between two species.

Phylogenetic analysis

On the basis of *dnaJ1* gene sequence analysis, phylogenetic trees were constructed by the neighbor-joining (Fig. 1; with the alignment in Supplementary Fig. S1) and maximum-likelihood methods (See Supplementary Fig. S2). Whichever algorithm was used, most phylogenetic relationships of *Mycobacterium* species were stably maintained, except for some minor differences at the deepest branching points. The tree proved to be reliable for deeply branching nodes, which was supported by high bootstrap values. The bootstrap values at the nodes were in all cases much higher than those observed by 16S rRNA ($P<0.05$; χ^2-test), *rpoB* ($P<0.00001$) and *hsp65* ($P<0.001$) gene sequencing. Among their nodes, bootstrap values $>80\%$ were observed for 24 of 56 nodes for the 16S rRNA gene, 13 of 55 nodes for *rpoB* and 18 of 56 nodes for *hsp65*, respectively, whereas values $>80\%$ were observed for 35 of 56 nodes for the *dnaJ1* gene.

The phylogenetic relationships of some medically important groups by complete *dnaJ1* sequences corresponded well with those of the 16S rRNA gene sequence (Fig. 1). For example, the *M. avium* complex (*M. avium*, *M. intracellulare* and *M. chimaera*), the *M. chelonae*–*M. abscessus* group and the *M. terrae* complex (*M. terrae, M. nonchromogenicum*, *M. hiberniae*, *M. arupense* and *M. kumamotonense*) are known as groups of closely related species by phenotypic characteristics or the DNA–DNA hybridization method [11,12,14,33]. These respective groups formed clusters in the 16S rDNA and *dnaJ1* trees, although *M. avium* and *M. intracellulare* were separated in the *rpoB* tree, whereas, in contrast, the *M. chelonae*–*M. abscessus* group and species of the *M. terrae* complex were separated in the *hsp65* tree.

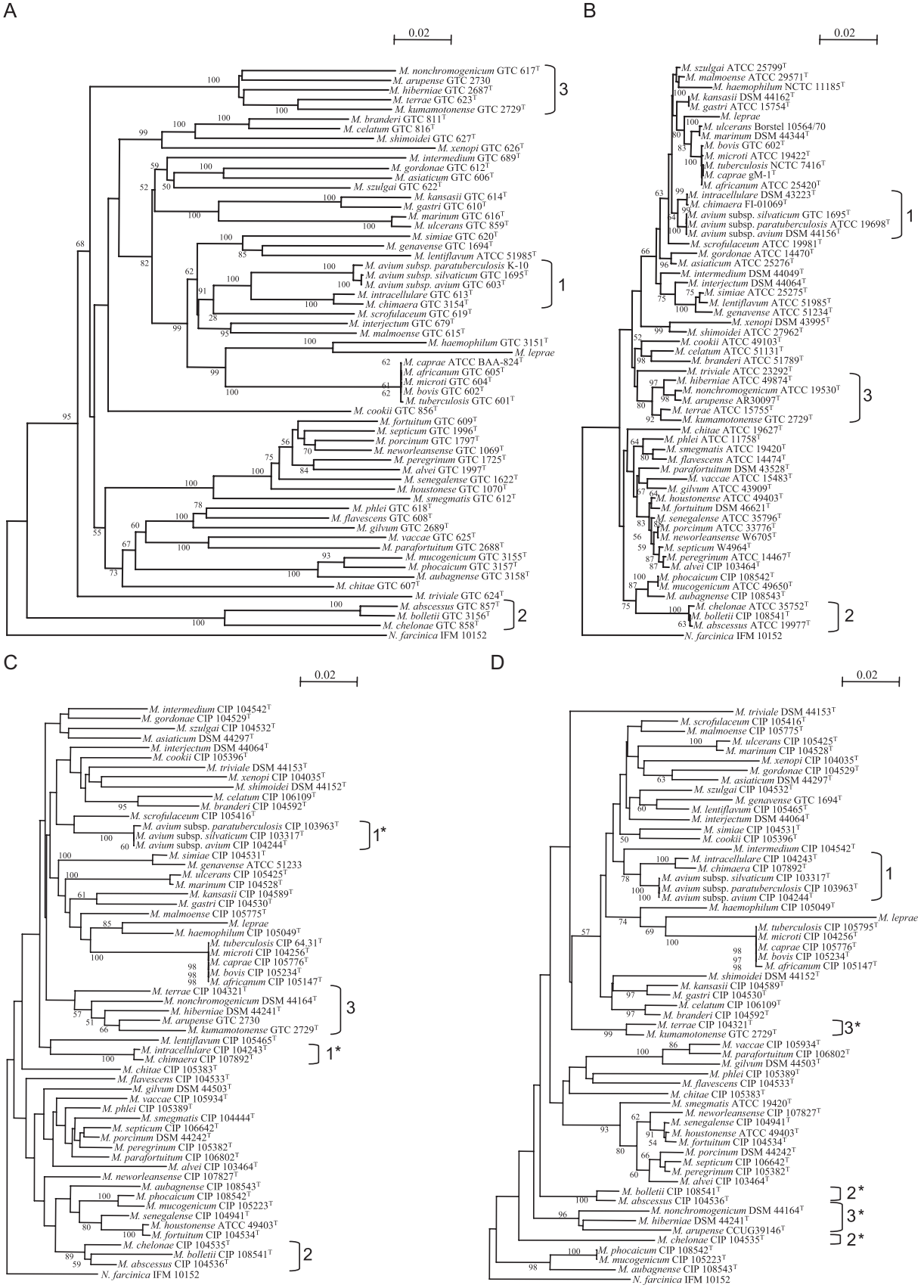

Fig. 1. Neighbor-joining phylogenetic trees based on the *dnaJ1* complete sequence (1188 bp), 16S rDNA (1536 bp), *rpoB* (395 bp), and *hsp65* (424 bp) using 56 *Mycobacterium* species. Bootstrap values (expressed as a percentage of 1000 replicates) greater than 50% are shown at the tree nodes. The bar represents 2% sequence divergence. Three medically important groups are shown as groups 1–3. Groups: 1, *M. avium* complex; 2, *M. chelonae–M. abscessus* group; 3, *M. terrae* complex. Separated clusters are denoted by asterisks. (A) *dnaJ*—Complete. (B) 16S rDNA. (C) *rpoB*. (D) *hsp65*.

The correlation coefficient between *dnaJ1* and 16S rDNA sequence similarities ($r = 0.696$) was statistically higher than that between *rpoB* and 16S rDNA ($r = 0.566$, $P < 0.001$) or that between *hsp65* and 16S rDNA ($r = 0.599$, $P < 0.001$) sequence similarities, suggesting that the relationship between a pair of taxa in *dnaJ1* is more coherent with that of 16S rDNA (Fig. 2).

For accurate sequence-based identification, it is important for the phylogenetic positions in the housekeeping gene used to be in agreement with DNA reassociation studies. Although a relationship between the DNA–DNA hybridization similarity value and each housekeeping gene similarity has not been revealed in the genus *Mycobacterium*, a stronger relationship between *dnaJ* sequence similarities and DNA–DNA hybridization values was observed in the genus *Aeromonas* [19]. Thus, a *dnaJ1* gene sequence-based assay may be a more effective alternative to the currently used methods for identification and taxonomical analysis of *Mycobacterium* species.

Phylogenetic analysis in clinical strains

Nucleotide sequence variations of complete *dnaJ1* among the clinical isolates were observed and the range of variation between most strains in each species was narrow (intraspecies divergence, <1%). In general, 1–10 nucleotide variations were observed among the clinical isolates (99.2–100% similarity) (Table 4). Variants of *M. tuberculosis* were not found among 32 strains. The clinical isolates of the *M. avium* complex were divided into three groups: *M. avium*, *M. intracellulare* and *M. chimaera*. Interestingly, the *dnaJ1* sequence of one *M. intracellulare* strain, KPM 3108, exhibited a low level of similarity (96.5%) to that of the reference strain (GTC 613). The *M. avium* complex heterogeneity had been recently revealed by molecular genotyping studies of the *hsp65* sequence [29], ITS sequence [6], and DT1/DT6 sequence [31].

Table 4. Sequence similarities of complete *dnaJ1* among the 71 clinical isolates

Species	Range of % similarity[a]	No. of strains
M. tuberculosis	100	32
M. genavense	100	1
M. ulcerans	99.8	1
MAC complex		
M. avium	99.9–100	9
M. intracellulare	99.2–100	3
	94.8[b]	1
M. chimaera	99.3–99.8	2
M. scrofulaceum	100	1
M. kansasii	99.8–100	16
	96.5[c]	1
M. gastri	100	2
M. parafortuitum	99.8	1
M. mucogenicum	100	1

[a]% similarities between the clinical strain and type strain were calculated.
[b]KPM 3108.
[c]KPM 1002.

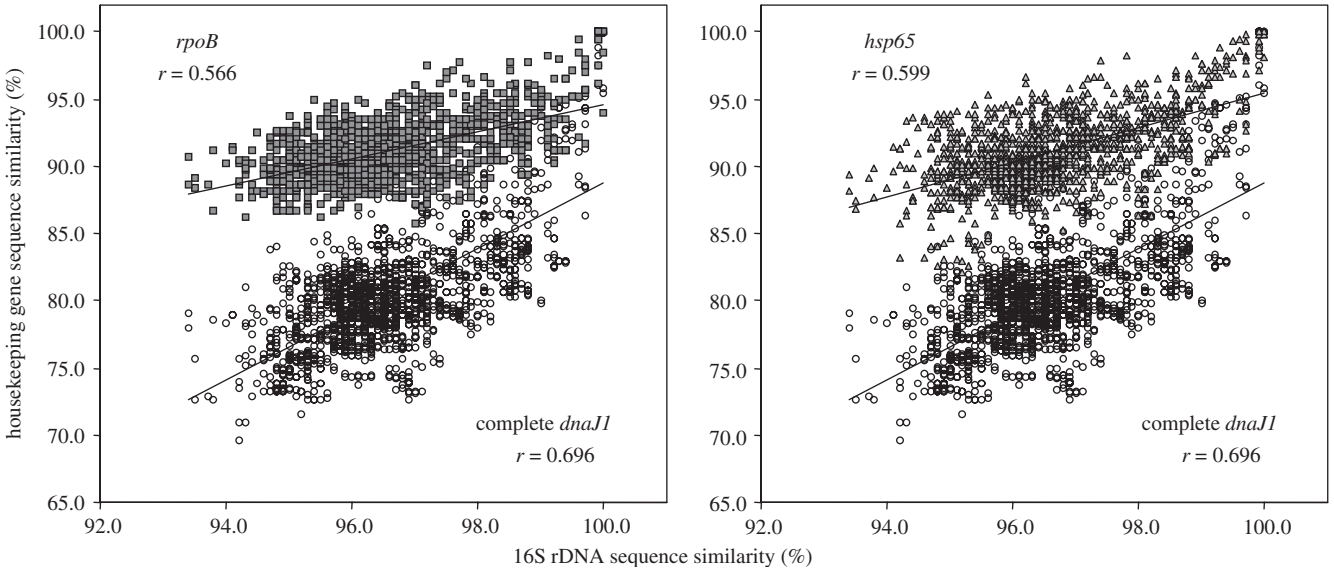

Fig. 2. Scatter plots of *dnaJ1* (open circles) and *rpoB* (squares) sequence similarities vs 16S rDNA sequence similarities in the left graph, and *dnaJ1* (open circles) and *hsp65* (triangles) sequence similarities vs 16S rDNA sequence similarities in the right graph for 56 *Mycobacterium* strains. Each dot represents a pair of taxa, plotted according to their sequence similarities. The solid lines represent the regression lines. The value of *r* represents the correlation coefficient between each housekeeping gene vs 16S rDNA.

On the other hand, the *dnaJ1* sequence of one *M. kansasii* strain, KPM 1002, showed a low level of similarity (96.5%) to that of the reference strain (GTC 614), while 16 other *M. kansasii* strains were almost identical (99.8–100%). One possible and reasonable explanation is the *M. kansasii* heterogeneity that had been revealed by molecular genotyping studies of PCR-restriction fragment length pattern analysis (PRA) of the *hsp65* gene [22,35] and ITS sequence [1].

By referring to the phylogenetic tree and by using reference strains, we applied this procedure to the clinical isolates and thus could identify various *Mycobacterium* species that had previously been confirmed in the conventional way. The applicability of *dnaJ1* sequences for identification was confirmed by assigning all clinical isolates to the correct clusters containing the corresponding type strains and by showing that isolates belonging to the same species shared at least 96.5% *dnaJ1* sequence similarity. The *dnaJ1* gene will therefore be a promising tool for the classification and identification of such bacteria.

dnaJ1 partial sequences for strain identification

In order to propose a method for strain identification that does not require sequencing of the complete *dnaJ1* sequence (approximately 1200 bp), an area was identified with a high degree of polymorphism, bordered by conserved sequences, that can be used as universal primers for PCR amplification and sequencing. Among the *Mycobacterium* species, a consensus primer pair (J10F and J335R) that allowed the successful amplification of all *Mycobacterium* type strains, except for 5 species belonged to the *M. tuberculosis* complex and *M. intermedium*. The amplified fragment sizes were from 365 to 391 bp, depending on the species (Table 1). The same as for complete *dnaJ1* sequences, the similarities observed in the partial *dnaJ1* sequence (63.0–98.9%, mean 75.7%) were also significantly less than those observed in the 16S rRNA, *rpoB* and *hsp65* genes (Table 3). The similarity between the two *M. avium* subspecies was 100% and was thus 1.1% greater than the highest degree of similarity (except for the *M. tuberculosis* complex) between two species.

Our data, based on the *dnaJ1* sequences of these bacteria, confirm that this gene is significantly more polymorphic than the 16S rRNA gene, and we propose that it be used to replace or complement the 16S rRNA gene for phylogenetic studies of *Mycobacterium*.

Appendix A. Supplementary Materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.syapm.2007.06.003.

References

[1] Y. Abed, C. Bollet, P. de Micco, Demonstration of *Mycobacterium kansasii* species heterogeneity by the amplification of the 16S-23S spacer region, J. Med. Microbiol. 43 (1995) 156–158.
[2] T. Adekambi, M. Drancourt, 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. 54 (2004) 2095–2105.
[3] K.S. Blackwood, C. He, J. Gunton, C.Y. Turenne, J. Wolfe, A.M. Kabani, Evaluation of *recA* sequences for identification of *Mycobacterium* species, J. Clin. Microbiol. 38 (2000) 2846–2852.
[4] F. Brunello, M. Ligozzi, E. Cristelli, S. Bonora, E. Tortoli, R. Fontana, Identification of 54 mycobacterial species by PCR-restriction fragment length polymorphism analysis of the *hsp65* gene, J. Clin. Microbiol. 39 (2001) 2799–2806.
[5] J.L. Cloud, H. Neal, R. Rosenberry, C.Y. Turenne, M. Jama, D.R. Hllyard, K.C. Carroll, Identification of *Mycobacterium* spp. by using a commercial 16S ribosomal DNA sequencing kit and additional sequencing libraries, J. Clin. Microbiol. 40 (2002) 400–406.
[6] R. Frothingham, K.H. Wilson, Molecular phylogeny of the *Mycobacterium avium* complex, J. Infect. Dis. 169 (1994) 305–312.
[7] Y. Itoh, Y. Kawamura, H. Kasai, M.M. Shah, P.H. Nhung, M. Yamada, X.S. Sun, T. Koyama, M. Hayashi, K. Ohkusu, T. Ezaki, *dnaJ* and *gyrB* gene sequence relationship among species and strains of the genus *Streptococcus*, Syst. Appl. Microbiol. 29 (2006) 368–374.
[8] H. Kasai, T. Ezaki, S. Harayama, Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences, J. Clin. Microbiol. 38 (2000) 301–308.
[9] B.J. Kim, S.H. Lee, M.A. Lyu, S.J. Kim, G.H. Bai, S.J. Kim, G.T. Chae, E.C. Kim, C.Y. Cha, Y.H. Kook, Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*), J. Clin. Microbiol. 37 (1999) 1714–1720.
[10] P. Kirschner, B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F.-C. Bange, E.C. Bottger, Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory, J. Clin. Microbiol. 31 (1993) 2882–2889.
[11] S. Kusunoki, T. Ezaki, M. Tamesada, Y. Hatanaka, K. Asano, Y. Hashimoto, E. Yabuuchi, Application of colorimetric microdilution plate hybridization for rapid genetic identification of 22 *Mycobacteium* species, J. Clin. Microbiol. 29 (1991) 1596–1603.
[12] S. Kusunoki, T. Ezaki, Proposal of *Mycobacterium peregrinum* sp. nov., nom. rev., and elevation of *Mycobacterium chelonae* subsp. *abscessus* (Kubica et al.) to species status: *Mycobacterium abscessus* comb. nov, Int. J. Syst. Bacteriol. 42 (1992) 240–245.
[13] H. Liu, Y. Li, X. Huang, Y. Kawamura, T. Ezaki, Use of the *dnaJ* gene for the detection and identification of all *Legionella pneumophila* serogroups and description of the primers used to detect 16S rRNA gene sequences of major members of the genus *Legionella*, Microbiol. Immunol. 47 (2003) 859–869.

[14] T. Masaki, K. Ohkusu, H. Hata, N. Fujiwara, H. Iihara, M. Yamada-Noda, P.H. Nhung, M. Hayashi, Y. Asano, Y. Kawamura, T. Ezaki, *Mycobacterium kumamotonense* sp. nov. recovered from clinical specimen and the first isolation report of *Mycobacterium arupense* in Japan: novel slowly growing, nonchromogenic clinical isolates related to *Mycobacterium terrae* complex, Microbiol. Immunol. 50 (2006) 889–897.

[15] A. McNabb, D. Eisler, K. Adie, M. Amos, M. Rodrigues, G. Stephens, W.A. Black, J. Isaac-Renton, Assessment of partial sequencing of the 65-kilodalton heal shock protein gene (*hsp65*) for routine identification of *Mycobacterium* species isolated from clinical sources, J. Clin. Microbiol. 42 (2004) 3000–3011.

[16] Y. Morita, S. Maruyama, H. Kabeya, A. Nagai, K. Kozawa, M. Kato, T. Nakajima, T. Mikami, Y. Katsube, H. Kimura, Genetic diversity of the *dnaJ* gene in the *Mycobacterium avium* complex, J. Med. Microbiol. 53 (2004) 813–817.

[17] P.H. Nhung, M.M. Shah, K. Ohkusu, M. Noda, H. Hata, X.S. Sun, H. Iihara, K. Goto, T. Masaki, J. Miyasaka, T. Ezaki, The *dnaJ* gene as a novel phylogenetic marker for identification of *Vibrio species*, Syst. Appl. Microbiol. 30 (2007) 309–315.

[18] P.H. Nhung, K. Ohkusu, N. Mishima, M. Noda, M.M. Shah, X.S. Sun, M. Hayashi, T. Ezaki, Phylogeny and species identification of the family *Enterobacteriaceae* based on *dnaJ* sequences, Diagn. Microbiol. Infect. Dis. 58 (2007) 153–161.

[19] P.H. Nhung, H. Hata, K. Ohkusu, M. Noda, M.M. Shah, K. Goto, T. Ezaki, Novel phylogenetic marker *dnaJ* and DNA–DNA hybridization for clarifying interrelationships among the genus *Aeromonas*, Int. J. Syst. Evol. Microbiol. 57 (2007) 1232–1237.

[20] S. Niemann, D. Harmsen, S. Rusch-Gerdes, E. Richter, Differentiation of clinical *Mycrobacterium tuberculosis* complex isolates by *gyrB* DNA sequence polymorphism analysis, J. Clin. Microbiol. 38 (2000) 3231–32347.

[21] R.D.M. Page, TREEVIEW: an application to display-phylogenetic trees on personal computers, Comp. Appl. Biosci. 12 (1996) 357–358.

[22] M. Picardeau, G. Prod'hom, L. Raskine, M.P. LePennec, V. Vincent, Genotypic characterization of five subspecies of *Mycobacterium kansasii*, J. Clin. Microbiol. 35 (1997) 25–32.

[23] H. Ringuet, C. Akoua-Koffi, S. Honore, A. Varnerot, V. Vincent, P. Berche, J.L. Gaillard, C. Pierre-Audigier, *hsp65* sequencing for identification of rapidly growing mycobacteria, J. Clin. Microbiol. 37 (1999) 852–857.

[24] A. Roth, M. Fischer, M.E. Hamid, S. Michalke, W. Ludwig, H. Mauch, Differentiation of phylogenetically related slowly growing mycrobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences, J. Clin. Microbiol. 36 (1998) 139–147.

[25] A. Roth, U. Reischl, A. Streubel, L. Naumann, R.M. Kroppenstedt, M. Habicht, M. Fischer, H. Mauch, Novel diagnostic algorithm for identification of mycobacteria using genus-specific amplification of the 16S-23S rRNA gene spacer and restriction endonucleases, J. Clin. Microbiol. 38 (2000) 1094–1104.

[26] M.M. Shah, H. Iihara, M. Noda, X.S. Sun, P.H. Nhung, K. Ohkusu, Y. Kawamura, T. Ezaki, *dnaJ* gene sequence-based assay for species identification and phylogenetic grouping in the genus *Staphylococcus*, Int. J. Syst. Evol. Microbiol. 57 (2007) 25–30.

[27] B. Springer, L. Stockman, K. Teschner, G.D. Roberts, E.C. Bottger, Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods, J. Clin. Microbiol. 34 (1996) 296–303.

[28] G.R. Stewart, B.D. Robertson, D.B. Young, Analysis of the function of mycobacterial DnaJ proteins by over-expression and microarray profiling, Tuberculosis 84 (2004) 180–187.

[29] D.S. Swanson, V. Kapur, K. Dtockbauer, X. Pan, R. Fronthingham, J.M. Musser, Subspecific differentiation of *Mycobacterium avium* complex strains by automated sequencing of a region of the gene (*hsp65*) encoding a 65-Kilodalton heat shock protein, Int. J. Syst. Evol. Microbiol. 47 (1997) 414–419.

[30] S. Takewaki, K. Okuzumi, H. Ishiko, K. Nakahara, S. Ohkubo, R. Nagai, Genus-specific polymerase chain reaction for the mycobacterial *dnaJ* gene and species-specific oligonucleotide probes, J. Clin. Microbiol. 31 (1993) 446–450.

[31] D. Thierry, V. Vincent, F. Clement, J.L. Guesdon, Isolation of specific DNA fragments of *Mycobacterium avium* and their possible use in diagnosis, J. Clin. Microbiol. 31 (1993) 1048–1054.

[32] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTALW, Nucleic Acids Res 22 (1994) 4673–4680.

[33] E. Tortoli, L. Rindi, M.J. Garcia, P. Chiaradonna, R. Dei, C. Garzelli, R.M. Kroppenstedt, N. Lari, R. Mattei, A. Mariottini, G. Mazzarelli, M.I. Murcia, A. Nanetti, P. Piccoli, C. Sca, 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. 54 (2004) 1277–1285.

[34] L.G. Wayne, G.P. Kubica, The Mycobacteriaceae, in: P.H.A. Sneath, N.S. Mair, M.E. Sharpe, J.G. Holt (Eds.), Bergey's Manual of Systematic Bacteriology, vol. 2, Williams and Wilkins, Baltimore, 1986, pp. 1436–1457.

[35] M. Yang, B.C. Ross, B. Dwyer, Isolation of a DNA probe for identification of *Mycobacterium kansasii*, including the genetic subgroup, J. Clin. Microbiol. 31 (1993) 2769–2772.