

Research Article

Evaluation of the Ribosomal Protein S1 Gene (*rpsA*) as a Novel Biomarker for *Mycobacterium* Species Identification

Hongfei Duan, Guan Liu, Xiaobo Wang, Yuhong Fu, Qian Liang, Yuanyuan Shang, Naihui Chu, and Hairong Huang

National Clinical Laboratory on Tuberculosis, Beijing Key Laboratory on Drug-Resistant Tuberculosis, Beijing Chest Hospital, Capital Medical University, Beijing Tuberculosis & Thoracic Tumor Research Institute, 97 Beimeichang, Tongzhou Qu, Beijing 101149, China

Correspondence should be addressed to Naihui Chu; dongchu1994@sina.com and Hairong Huang; hairong.huangcn@gmail.com

Received 26 July 2014; Accepted 7 November 2014

Academic Editor: Om Parkash

Copyright © 2015 Hongfei Duan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objectives. To evaluate the resolution and reliability of the *rpsA* gene, encoding ribosomal protein S1, as a novel biomarker for mycobacteria species identification. **Methods.** A segment of the *rpsA* gene (565 bp) was amplified by PCR from 42 mycobacterial reference strains, 172 nontuberculosis mycobacteria clinical isolates, and 16 *M. tuberculosis* complex clinical isolates. The PCR products were sequenced and aligned by using the multiple alignment algorithm in the MegAlign package (DNASTAR) and the MEGA program. A phylogenetic tree was constructed by the neighbor-joining method. **Results.** Comparative sequence analysis of the *rpsA* gene provided the basis for species differentiation within the genus *Mycobacterium*. Slow- and rapid-growing groups of mycobacteria were clearly separated, and each mycobacterial species was differentiated as a distinct entity in the phylogenetic tree. The sequences discrepancy was obvious between *M. kansasii* and *M. gastri*, *M. chelonae* and *M. abscessus*, *M. avium* and *M. intracellulare*, and *M. szulgai* and *M. malmoense*, which cannot be achieved by 16S ribosomal DNA (rDNA) homologue genes comparison. 183 of the 188 (97.3%) clinical isolates, consisting of 8 mycobacterial species, were identified correctly by *rpsA* gene blast. **Conclusions.** Our study indicates that *rpsA* sequencing can be used effectively for mycobacteria species identification as a supplement to 16S rDNA sequence analysis.

1. 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. Both slow-growing mycobacteria and rapid-growing mycobacteria can cause human infections. Traditionally, taxonomy based on biochemical characteristics has been used for species determination of mycobacteria, but this approach is limited due to the overlapping biochemical and phenotypic patterns among the different mycobacterial species. Another approach using analysis of cell-wall fatty acid and mycolic acid composition is also limited by profile similarity among some emerging nontuberculosis mycobacteria (NTM) [1]. 16S rDNA homologue gene sequence comparison has been used as an important method for the mycobacterial species identification; however, ambiguous results have been obtained either due to the presence of more than one copy of the 16S

rDNA gene within the genome, for example, in *M. celatum* and *M. terrae* complex [2, 3], or due to sequence homology between species [4]. Therefore, alternative phylogenetic markers which are capable of complementing 16S rRNA gene would be useful for the phylogenetic study and species identification of the genus *Mycobacterium*.

Ribosomal protein S1 (*rpsA*), which is in the 30S ribosome subunit, contains the S1 domain that has been found in a large number of RNA-associated proteins. *RpsA* is a vital protein involved in protein translation and the ribosome-sparing process of translation. In addition, it has been reported that *RpsA* is the target of pyrazinoic acid, the active form of the antituberculosis drug pyrazinamide [5]. *Mycobacterium tuberculosis* has a single copy of the *rpsA* gene in the genome [6], while the sequence homology among different mycobacteria species in which the gene has already been sequenced is between 86.7 and 100%. This suggests that *rpsA* may be

suitable for phylogenetic study of the genus *Mycobacterium*. In this paper, we report an evaluation of the *rpsA* gene as a novel biomarker for mycobacteria species identification.

2. Materials and Methods

2.1. Mycobacterial Reference Strains and Clinical Isolates. 42 type and reference strains of the genus *Mycobacterium* (Table 1), 172 clinical NTM isolates, and 16 *M. tuberculosis* complex (MTC) isolates were investigated. All the type and reference strains were purchased from the American Type Culture Collection (ATCC) and all the clinical isolates used in this study were obtained from the Clinical Database and Sample Bank of tuberculosis of Beijing, National Clinical Lab on Tuberculosis, Beijing Chest Hospital. All the clinical NTM isolates were identified to the species level by sequence alignment of at least two of the following: 16S rDNA, 16-23S rRNA gene internal transcribed spacer (ITS), and *rpoB* and *hsp65* genes as described before [2, 7–9]. Following sequencing, the 172 NTM clinical strains were found to include 10 strains of *M. avium*, 75 strains of *M. intracellulare*, 23 strains of *M. kansasii*, 39 strains of *M. abscessus*, 17 strains of *M. fortuitum* complex, 7 strains of *M. goodii*, and 1 strain of *M. neoaurum*.

2.2. *rpsA* Gene Amplification and Sequencing. DNA was released from cultured mycobacteria by boiling the cultured mycobacterial suspension in TE buffer for 10 min. After centrifugation, the supernatant was used for PCR amplification [10]. The primers used were forward primer, 5'-CCCTAC-ATCGGCAAGGAG-3', position 487–504 in the *rpsA* gene of *Mycobacterium tuberculosis*, GenBank accession number NC.000962.2, and reverse primer, 5'-TGTCGATGACCTTGACCATC-3', position 1032–1051 in the *rpsA* gene of *Mycobacterium tuberculosis*, GenBank accession number NC.000962.2. The amplified product was 565 bp. PCR products were purified and sequenced by a commercial company (YINGJUN Biotech Company, Beijing, China) using ABI 3730 DNA Analyzer (Applied Biosystems, California, USA).

2.3. Sequence Analysis and the Phylogenetic Tree Constructions. In addition to the 42 reference strains, the *rpsA* sequences of 5 other mycobacterial species were obtained directly from GenBank, including *M. avium* subspecies *paratuberculosis* (GenBank accession number NC_002944.2), *M. ulcerans* (GenBank accession number NC_008611.1), *M. vanbaalenii* (GenBank accession number NC_008726.1), *M. abscessus* subspecies *massiliense* (GenBank accession number NC_018150.1), and *M. canettii* (GenBank accession number NC_015848.1). All sequences were aligned and homology was calculated by using the multiple alignment algorithm in the MegAlign package (Windows version 7.1.0; DNASTAR, Madison, Wis). 527 bp of sequence (excluding 38 nucleotides at each end of the amplicon, corresponding to the primer binding sites) was analyzed for the phylogenetic tree construction by the neighbor-joining method using the MEGA 5.05 package (<http://www.megasoftware.net/mega.php>). A bootstrap analysis (1000 repeats) using *Rhodococcus equi*

(GenBank accession number NC.006361.1) as the outgroup was performed to evaluate the topology of the phylogenetic tree.

2.4. Species Identification of the Clinical Isolates. The 188 clinical isolates were analyzed blindly. Sequences, minus the known PCR primer sequences, were assembled and edited by using SeqMan software (version 7.1.0; DNASTAR, Madison, WI). Isolates were identified by comparing sequences by using a FASTA BLASTn search with MegAlign (version 7.1.0; DNASTAR, Madison, WI) to an in-house database of sequences consisting of type and reference strains from external culture collections.

3. Results

3.1. *rpsA* Sequence Alignment of the Reference Strains. Between 85.4% and 100% sequence homology (interspecies divergence, 0% to 14.6%) was observed among the 42 tested reference strains and the 5 additional mycobacterial species whose sequences were obtained from the GenBank database (Table 2). All the *rpsA* gene sequences of the analyzed mycobacteria strains were distinct from the outgroup strain *R. equi*. Among the 19 reference strains of the slow-growing *Mycobacterium* genus, 11 strains were greater than 97% homology, including 5 *M. tuberculosis* complex strains. Among the 28 reference strains of the rapid-growing *Mycobacterium* genus, 14 strains were greater than 97% homology (Table 1). The pathogenic *M. kansasii* was easily differentiated from the nonpathogenic *M. gastri* (95.8% homology), while those two species were not distinguishable by the 16S rDNA sequence alignment. The sequence homologies between various species were 91.8% between *M. chelonae* and *M. abscessus*, 95.6% between *M. avium* and *M. intracellulare*, and 93.9% between *M. szulgai* and *M. malmoense*. However, the sequence homology between other species was higher; for example, it was 99.6% between *M. ulcerans* and *M. marinum* and 98.7% between *M. abscessus* subspecies *massiliense* and *M. abscessus*. All members of the *M. tuberculosis* complex had identical sequences as did *M. senegalense* and *M. thermoresistibile*, *M. parafortuitum* and *M. trivialis*, *M. diernhoferi* and *M. duvalii*, and *M. austroafricanum* and *M. terrae*.

3.2. Phylogenetic Tree Construction. A phylogenetic tree, which provided the basis for species differentiation in the genus *Mycobacterium*, was constructed (Figure 1). The absolute majority tested species showed good separation. The rapid-growing species were well defined from the slow-growing species in the tree. *M. chelonae*, *M. abscessus* subspecies *massiliense*, and *M. abscessus*, which are categorized in the pathogenic taxonomic group of rapid-growing mycobacteria, formed a distinctive cluster which was much closer to the slow-growing species compared with the nonpathogenic group of rapid-growing mycobacteria. The reliability of the phylogenetic tree was verified by the bootstrap method, using *R. equi* as the outgroup.

TABLE I: Percentage *rpsA* sequence match among type strain.

Species	Type strain	First match (%) ^a	Second match (%) ^a
<i>M. thermoresistibile</i>	ATCC19527	<i>M. senegalense</i> (100%)	<i>M. porcinum</i> (98.7%)
<i>M. senegalense</i>	ATCC35796	<i>M. thermoresistibile</i> (100%)	<i>M. porcinum</i> (98.7%)
<i>M. porcinum</i>	ATCC33776	<i>M. senegalense</i> / <i>M. thermoresistibile</i> (98.7%)	<i>M. fortuitum</i> (98.3%)
<i>M. fortuitum</i>	ATCC6481	<i>M. porcinum</i> (98.3%)	<i>M. senegalense</i> / <i>M. thermoresistibile</i> (98.1%)
<i>M. phlei</i>	ATCC11758	<i>M. porcinum</i> (96.5%)	<i>M. fortuitum</i> (96.1%)
<i>M. gadium</i>	ATCC27726	<i>M. porcinum</i> / <i>M. fortuitum</i> (95.7%)	<i>M. smegmatis</i> (95.3%)
<i>M. flavescens</i>	ATCC14474	<i>M. smegmatis</i> (96.6%)	<i>M. vaccae</i> (96.0%)
<i>M. smegmatis</i>	ATCC19420	<i>M. vaccae</i> / <i>M. parafortuitum</i> / <i>M. triviale</i> (96.6%)	<i>M. senegalense</i> / <i>M. thermoresistibile</i> (96.3%)
<i>M. vaccae</i>	ATCC15483	<i>M. diernhoferi</i> / <i>M. duvalii</i> (96.8%)	<i>M. parafortuitum</i> / <i>M. triviale</i> (96.6%)
<i>M. parafortuitum</i>	ATCC19686	<i>M. triviale</i> (100%)	<i>M. terrae</i> / <i>M. austroafricanum</i> (96.6%)
<i>M. triviale</i>	ATCC23292	<i>M. parafortuitum</i> (100%)	<i>M. terrae</i> / <i>M. austroafricanum</i> (96.6%)
<i>M. neoaurum</i>	ATCC25795	<i>M. diernhoferi</i> / <i>M. duvalii</i> (96.4%)	<i>M. porcinum</i> (95.5%)
<i>M. diernhoferi</i>	ATCC19340	<i>M. duvalii</i> (100%)	<i>M. vaccae</i> (96.8%)
<i>M. duvalii</i>	ATCC43910	<i>M. diernhoferi</i> (100%)	<i>M. vaccae</i> (96.8%)
<i>M. aurum</i>	ATCC23366	<i>M. gilvum</i> (93.9%)	<i>M. chubuense</i> / <i>M. vanbaalenii</i> (93.0%)
<i>M. chubuense</i>	ATCC27278	<i>M. gilvum</i> (96.6%)	<i>M. vanbaalenii</i> (96.0%)
<i>M. gilvum</i>	ATCC43909	<i>M. vanbaalenii</i> (97.2%)	<i>M. terrae</i> / <i>M. austroafricanum</i> (97.0%)
<i>M. vanbaalenii</i>	PYR-1	<i>M. terrae</i> / <i>M. austroafricanum</i> (99.8%)	<i>M. gilvum</i> (97.2%)
<i>M. terrae</i>	ATCC15755	<i>M. austroafricanum</i> (100%)	<i>M. vanbaalenii</i> (99.8%)
<i>M. austroafricanum</i>	ATCC33464	<i>M. terrae</i> (100%)	<i>M. vanbaalenii</i> (99.8%)
<i>M. rhodesiae</i>	ATCC27024	<i>M. terrae</i> / <i>M. austroafricanum</i> (93.9%)	<i>M. obuense</i> (93.7%)
<i>M. obuense</i>	ATCC27023	<i>M. porcinum</i> / <i>M. fortuitum</i> (96.1%)	<i>M. senegalense</i> / <i>M. thermoresistibile</i> (96.5%)
<i>M. aichiense</i>	ATCC27280	<i>M. sphagni</i> (95.8%)	<i>M. vaccae</i> (94.7%)
<i>M. sphagni</i>	ATCC33027	<i>M. aichiense</i> (95.8%)	<i>M. porcinum</i> (95.5%)
<i>M. tokaiense</i>	ATCC27282	<i>M. porcinum</i> (95.3%)	<i>M. smegmatis</i> (93.7%)
<i>M. chelonae</i>	ATCC35752	<i>M. obuense</i> (93.9%)	<i>M. phlei</i> (93.5%)
<i>M. abscessus</i> subspecies <i>massiliense</i>	str. GO 06	<i>M. abscessus</i> (98.7%)	<i>M. sphagni</i> (91.4%)
<i>M. abscessus</i>	ATCC19977	<i>M. abscessus</i> subspecies <i>massiliense</i> (98.7%)	<i>M. chubuense</i> (92.0%)
<i>M. tuberculosis</i> H37Rv	ATCC27294	<i>M. tuberculosis</i> H37Ra/ <i>M. africanum</i> / <i>M. bovis</i> BCG/ <i>M. bovis</i> (100%)	<i>M. canettii</i> (99.8%)
<i>M. tuberculosis</i> H37Ra	ATCC25177	<i>M. tuberculosis</i> H37Rv/ <i>M. africanum</i> / <i>M. bovis</i> BCG/ <i>M. bovis</i> (100%)	<i>M. canettii</i> (99.8%)
<i>M. africanum</i>	ATCC25420	<i>M. tuberculosis</i> H37Ra/ <i>M. tuberculosis</i> Rv/ <i>M. bovis</i> BCG/ <i>M. bovis</i> (100%)	<i>M. canettii</i> (99.8%)
<i>M. bovis</i> BCG	ATCC35735	<i>M. tuberculosis</i> H37Ra/ <i>M. tuberculosis</i> H37Rv/ <i>M. africanum</i> / <i>M. bovis</i> (100%)	<i>M. canettii</i> (99.8%)
<i>M. bovis</i>	ATCC19210	<i>M. tuberculosis</i> H37Ra/ <i>M. tuberculosis</i> H37Rv/ <i>M. africanum</i> / <i>M. bovis</i> BCG (100%)	<i>M. canettii</i> (99.8%)
<i>M. canettii</i>	CIPT140010059	<i>M. tuberculosis</i> / <i>M. africanum</i> / <i>M. bovis</i> BCG/ <i>M. bovis</i> (99.8%)	<i>M. asiaticum</i> (93.2%)
<i>M. xenopi</i>	ATCC19250	<i>M. asiaticum</i> (93.0%)	<i>M. avium</i> subspecies <i>paratuberculosis</i> (92.4%)
<i>M. avium</i>	ATCC25291	<i>M. avium</i> subspecies <i>paratuberculosis</i> (99.6%)	<i>M. asiaticum</i> (95.8%)

TABLE I: Continued.

Species	Type strain	First match (%) ^a	Second match (%) ^a
<i>M. avium</i> subspecies <i>paratuberculosis</i>	K-10	<i>M. avium</i> (99.4%)	<i>M. asiaticum</i> (96%)
<i>M. malmoense</i>	ATCC29571	<i>M. intracellulare</i> (94.3%)	<i>M. szulgai</i> (93.9%)
<i>M. intracellulare</i>	ATCC13950	<i>M. avium</i> (95.6%)	<i>M. avium</i> subspecies <i>paratuberculosis</i> (95.4%)
<i>M. scrofulaceum</i>	ATCC19981	<i>M. asiaticum</i> (95.4%)	<i>M. szulgai</i> (95.1%)
<i>M. marinum</i>	ATCC927	<i>M. ulcerans</i> (99.6%)	<i>M. szulgai</i> (96.0%)
<i>M. ulcerans</i>	Agy99	<i>M. marinum</i> (99.6%)	<i>M. szulgai</i> (96.0%)
<i>M. szulgai</i>	ATCC25799	<i>M. marinum</i> / <i>M. ulcerans</i> (96.0%)	<i>M. asiaticum</i> (95.6%)
<i>M. gordonae</i>	ATCC14470	<i>M. asiaticum</i> (95.3%)	<i>M. kansasii</i> (95.1%)
<i>M. asiaticum</i>	ATCC25276	<i>M. simiae</i> (97.7%)	<i>M. kansasii</i> (96.4%)
<i>M. kansasii</i>	ATCC12478	<i>M. asiaticum</i> (96.3%)	<i>M. gastri</i> (95.8%)
<i>M. gastri</i>	ATCC15754	<i>M. kansasii</i> (95.8%)	<i>M. asiaticum</i> (94.5%)
<i>Rhodococcus equi</i>	103S	<i>M. smegmatis</i> (90.3%)	<i>M. intracellulare</i> (90.1%)

^aBeside the queried type strain itself, the first matched or the second matched type strain.

3.3. *Species Identification Outcomes of the Clinical Isolates.* *rpsA* sequence and alignment efficiently identified clinical isolates representing 8 mycobacterial species. We found that a criterion of first distinct *rpsA* sequence match >97% confirmed the identification of 183 of 188 (97.3%) clinical isolates. 20 *M. abscessus* strains, 2 *M. fortuitum* strains, and 10 *M. avium* strains were identified to the subspecies level. Identification discrepancies between *rpsA* and other sequences were only encountered with 4 *M. gordonae* strains and 1 *M. kansasii* strain (see Table 3). No discrepancies were found in the species submitted as *M. intracellulare*, *M. avium*, *M. abscessus*, and *M. tuberculosis* complex.

The sequence divergence among MTC members was 0.4%. The intraspecies divergence among the 188 clinical strains ranged from 0.6% to 5.3% (Table 3). The sequence diversity among the *M. abscessus* was 0.6%, while that among the *M. gordonae* clinical isolates was 5.3%. No *M. gordonae* clinical strain had an identical sequence with that of the *M. gordonae* reference strain (ATCC14470). Interestingly, the *rpsA* sequence of one strain among the 23 *M. kansasii* isolates exhibited a relatively low level of sequence similarity (95.6%) to that of the reference strain (ATCC 12478), while those of all the other *M. kansasii* isolates were identical. Since the strain was confirmed as *M. kansasii* by *rpoB* and *hsp65* gene alignment, it might suggest a distant variant or a new subtype.

4. Discussion

16S rDNA gene sequence alignment has been used as the reference method for mycobacterial species identification. However, it has been reported that, by using the 16S rDNA gene alone for species identification of clinical NTM, 37% of such isolates remained unclassified which illustrates the need for additional molecular tools for proper phylogenetic assignment and accurate NTM identification [11]. The common assumption that bacterial isolates belong to the same species if they have fewer than 5–15 bp differences within

the 16S rDNA gene sequence [12] or if they have more than 97% 16S rDNA gene sequence identities [13] may not be applicable to genus of *Mycobacterium*, whose members are much more closely related to each other. Furthermore, some mycobacterial strains shared uniform sequence of 16S rDNA gene, such as *M. intracellulare* and *M. avium*, *M. kansasii* and *M. gastri*, *M. abscessus* and *M. chelonae*, and *M. szulgai* and *M. malmoense*.

In this study, we found that the *rpsA* gene is useful as a complementary method to the 16S rDNA for mycobacterial species identification. Among the 47 reference strains and totally 1081 paired comparisons {1081 = [$na_1 + n(n-1)d/2$]; $a_1 = 1, n = 46, d = 1$ }, the *rpsA* gene alone can differentiate 97.2% (1051 out of 1081) of them if using 97% sequence homology as cutoff value. When referring to the phylogenetic tree, the *rpsA* resolution might increase. For the most commonly seen pathogenic NTM species such as *M. intracellulare*, *M. avium*, *M. abscessus*, *M. chelonae*, and *M. xenopi*, *rpsA* gene sequence could differentiate these species easily. Furthermore, 183 of the 188 (97.3%) clinical isolates, representing 8 mycobacterial species, were identified correctly by *rpsA* gene blast. Discrepancies of identification were only encountered with 4 *M. gordonae* strains and 1 *M. kansasii* strain. Both these species have reportedly more intraspecies sequence divergence [9, 14–16]. Three out of the 4 *M. gordonae* strains with discrepant identification still had the first distinction as *M. gordonae* by *rpsA*, but the sequence identity was lower than 97%, which means the identification might be correct when a more complete in-house database is being developed.

rpsA alone cannot differentiate the following: between members of the *M. tuberculosis* complex, between *M. senegalense* and *M. thermoresistibile*, between *M. parafortuitum* and *M. trivialis*, between *M. diernhoferi* and *M. duvalii*, and between *M. austroafricanum* and *M. terrae*; however, these species are also difficult to be separated by other markers alone such as 16S rDNA, *ITS*, *rpoB*, and *hsp65*. Several

TABLE 2: Sequence pair distances of 47 type strains of mycobacteria and 1 out-group strain determined by using the multiple alignment algorithm in the MegAlign package (version 7.1.0; DNASTAR, Madison, WI).

Table with 48 columns representing pairwise distances between 47 strains. The strains are listed in the first column, and the remaining 47 columns represent the distance values between each pair of strains. The diagonal elements are all 1.0, representing the distance of a strain to itself. The table is symmetric across the diagonal.

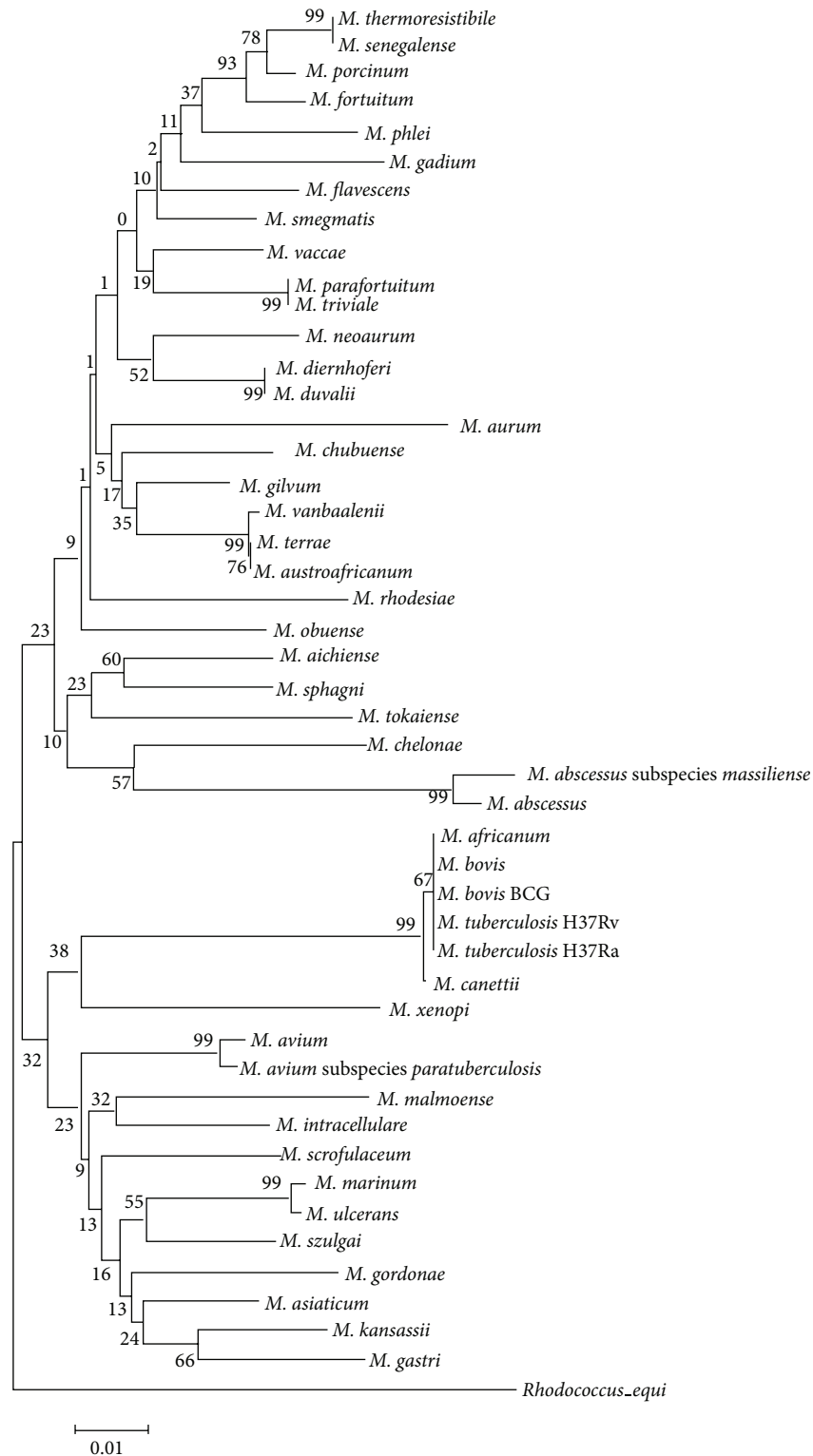


FIGURE 1: Phylogenetic tree based on *rpsA* gene sequences shows the relationship of the 47 type strains of mycobacteria and 1 outgroup strain. This tree was constructed by the neighbor-joining method. Topology was also evaluated by bootstrap analysis (MEGA program, 1000 repeats, with *R. equi* as the outgroup). The numerical values in the tree represent bootstrap results. The distance between two strains is the sum of the branch lengths between them.

TABLE 3: Comparison of mycobacteria identified by 16S rDNA/ITS/hsp65/rpoB gene sequencing and by rpsA sequencing among the 188 clinical isolates^a.

Primary identification ^b (n)	Primary identification ^b (n)			
	First choice	%match	Second choice	%match
<i>M. intracellulare</i> (75)	Identification (n)		Identification (n)	
	<i>M. intracellulare</i> (4)	100%	<i>M. avium</i> (4)	95.6%
	<i>M. intracellulare</i> (1)	99.8%	<i>M. avium</i> (1)	95.3%
	<i>M. intracellulare</i> (53)	98.9%	<i>M. avium</i> (53)	95.1%
	<i>M. intracellulare</i> (10)	98.7%	<i>M. avium</i> (10)	94.9%
	<i>M. intracellulare</i> (1)	98.5%	<i>M. avium</i> (1)	95.3%
	<i>M. intracellulare</i> (3)	98.1%	<i>M. avium</i> (3)	94.7%
	<i>M. intracellulare</i> (2)	97.5%	<i>M. avium</i> (2)	93.2%
<i>M. abscessus</i> (39)	<i>M. abscessus</i> (9)	100%	<i>M. abscessus</i> subspecies <i>massiliense</i> (9)	98.7%
	<i>M. abscessus</i> (10)	99.8%	<i>M. abscessus</i> subspecies <i>massiliense</i> (10)	98.5%
	<i>M. abscessus</i> subspecies <i>massiliense</i> (20)	99.4%	<i>M. abscessus</i> (20)	99.1%
<i>M. kansasii</i> (23)	<i>M. kansasii</i> (22)	100%	<i>M. asiaticum</i> (22)	96.3%
	<i>M. szulgai</i> (1)	96.6%	<i>M. marinum</i> / <i>M. ulcerans</i> (1)	96.0%
<i>M. avium</i> (10)	<i>M. avium</i> subspecies <i>paratuberculosis</i> (5)	99.6%	<i>M. avium</i> (5)	99.4%
	<i>M. avium</i> subspecies <i>paratuberculosis</i> (4)	99.8%	<i>M. avium</i> (4)	99.2%
	<i>M. avium</i> subspecies <i>paratuberculosis</i> (1)	97.5%	<i>M. avium</i> (1)	97.0%
<i>M. fortuitum</i> complex (17)	<i>M. fortuitum</i> (2)	100%	<i>M. senegalense</i> / <i>M. thermoresistibile</i> (2)	98.1%
	<i>M. fortuitum</i> (11)	99.8%	<i>M. porcinum</i> (11)	98.5%
	<i>M. fortuitum</i> (1)	99.6%	<i>M. porcinum</i> (1)	98.3%
	<i>M. fortuitum</i> / <i>M. porcinum</i> (1)	98.9%	<i>M. senegalense</i> / <i>M. thermoresistibile</i> (1)	98.3%
	<i>M. thermoresistibile</i> / <i>M. senegalense</i> (2)	99.6%	<i>M. fortuitum</i> (2)	98.1%
<i>M. gordonae</i> (7)	<i>M. gordonae</i> (1)	99.2%	<i>M. asiaticum</i> (1)	95.5%
	<i>M. gordonae</i> (2)	98.5%	<i>M. asiaticum</i> (2)	96.3%
	<i>M. gordonae</i> (3)	95.9%	<i>M. asiaticum</i> (3)	95.5%
	<i>M. szulgai</i> (1)	95.9%	<i>M. gordonae</i> (1)	94.7%
<i>M. tuberculosis</i> complex (16)	<i>M. tuberculosis</i> (2)	100%	<i>M. canetti</i> (2)	99.8%
	<i>M. tuberculosis</i> (12)	99.8%	<i>M. canetti</i> (12)	99.6%
	<i>M. tuberculosis</i> (2)	99.6%	<i>M. canetti</i> (2)	99.4%
<i>M. neoaurum</i> (1)	<i>M. neoaurum</i> (1)	98.7%	<i>M. smegmatis</i> (1)	95.4%

^aNumbers in parentheses (n) represent the numbers of isolates identified as a particular species.

^bIdentification based on sequencing of at least two of the following: 16S rDNA, 16-23S rRNA gene internal transcribed spacer (ITS), and rpoB and hsp65 genes.

reference strains had similar homology with two or more species which suggests inadequate taxonomy within the currently described species. Based on nucleotide sequences of *rpoB*, *hsp65*, and *sodA* in clinical isolates of the *Mycobacterium abscessus* group, one-fourth of isolates had discordant identification [16]. Multilocus sequence typing and sequence analysis of several genes facilitate the identification of closely related species or subspecies [17].

In our study, we have demonstrated that the *rpsA* homology gene is a promising marker for mycobacterial species identification for both fast- and slow-growing mycobacteria. To be a good marker for species differentiation, the target gene should be stable and sequence variations should occur randomly; additionally, an extremely conserved or highly variable gene may not be adequate. As a single-copy house-keeping gene, *rpsA* gene could work well as a target gene for

Mycobacterium species discrimination without ambiguous identification. The 16S rDNA gene has higher homology within the mycobacteria compared to the *rpsA* gene: 94.3% to 100% compared to 85.4% to 100%. According to our work, when used as sole marker, *rpsA* had better resolution than 16S rDNA but had similar resolution as ITS, *rpoB*, and *hsp65*.

Due to the high resolution power of *rpsA* for species identification, we presume it has potential of clinical use. From our own experience, we recommend that when performing species identification by homologue DNA sequence comparison, one should start with 16S RNA gene, plus at least one of the other markers such as ITS, *rpoB*, *hsp65*, and *rpsA*. When conflicting outcome or dubious outcomes yield, more markers should be added further. Even though 16S RNA gene is inferior to the other markers considering the capacity, it should be chosen firstly since 16S RNA gene has

the most robust sequence database, which can help to avoid big identification errors due to unreliable database of other markers [18].

The main deterrents for the primary use of *rpsA* sequencing as a routine means of identifying mycobacteria reside in the need for a comprehensive database. Therefore, we constructed our own *rpsA* sequence database by including as many type strains as possible and integrating *rpsA* sequences deposited in GenBank such as *M. avium* subspecies *paratuberculosis*, *M. ulcerans*, *M. vanbaalenii*, *M. abscessus* subspecies *massiliense*, and *M. canettii*. Besides type strains, *rpsA* sequences of confirmed clinical strains were also included. We will constantly upgrade our database and the capability to identify the most recent described species.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contribution

Hongfei Duan and Guan Liu contributed equally to this study.

Acknowledgments

The work was supported by the research funding from Infectious Diseases Special Project, Ministry of Health of China (2012ZX10003002-09), and Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special Funding (ZYLX201304). All the type strains and clinical isolates used in this study were obtained from the "Beijing Bio-Bank of Clinical Resources on Tuberculosis (D09050704640000)," Beijing Chest Hospital.

References

- [1] T. Adékambi, P. Berger, D. Raoult, and M. Drancourt, "rpoB gene sequence-based characterization of emerging nontuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov.," *International Journal of Systematic and Evolutionary Microbiology*, vol. 56, no. 1, Article ID 63969, pp. 133–143, 2006.
- [2] B. Ninet, M. Monod, S. Emler et al., "Two different 16S rRNA genes in a mycobacterial strain," *Journal of Clinical Microbiology*, vol. 34, no. 10, pp. 2531–2536, 1996.
- [3] U. Reischl, K. Feldmann, L. Naumann et al., "16S rRNA sequence diversity in *Mycobacterium celatum* strains caused by presence of two different copies of 16S rRNA gene," *Journal of Clinical Microbiology*, vol. 36, no. 6, pp. 1761–1764, 1998.
- [4] D. E. Griffith, T. Aksamit, B. A. Brown-Elliott et al., "An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases," *American Journal of Respiratory and Critical Care Medicine*, vol. 175, no. 4, pp. 367–416, 2007.
- [5] W. Shi, X. Zhang, X. Jiang et al., "Pyrazinamide inhibits translation in *Mycobacterium tuberculosis*," *Science*, vol. 333, no. 6049, pp. 1630–1632, 2011.
- [6] S. T. Cole, R. Brosch, J. Parkhill et al., "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence," *Nature*, vol. 393, no. 6685, pp. 537–544, 1998.
- [7] A. Roth, M. Fischer, M. E. Hamid, S. Michalke, W. Ludwig, and H. Mauch, "Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences," *Journal of Clinical Microbiology*, vol. 36, no. 1, pp. 139–147, 1998.
- [8] H. Ringuet, A. Honore, A. Varnerot et al., "hsp65 sequencing for identification of rapidly growing mycobacteria," *Journal of Clinical Microbiology*, vol. 37, no. 3, pp. 852–857, 1999.
- [9] B.-J. Kim, S.-H. Lee, M.-A. Lyu et al., "Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB)," *Journal of Clinical Microbiology*, vol. 37, no. 6, pp. 1714–1720, 1999.
- [10] J. Amita, T. Vandana, R. S. Guleria, and R. K. Verma, "Qualitative evaluation of mycobacterial DNA extraction protocols for polymerase chain reaction," *Molecular Biology Today*, vol. 3, no. 2, pp. 43–49, 2002.
- [11] R. J. Pauls, C. Y. Turenne, J. N. Wolfe, and A. Kabani, "A high proportion of novel mycobacteria species identified by 16S rDNA analysis among slowly growing AccuProbe-negative strains in a clinical setting," *American Journal of Clinical Pathology*, vol. 120, no. 4, pp. 560–566, 2003.
- [12] G. E. Fox, J. D. Wisotzkey, and P. Jurtschuk Jr., "How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity," *International Journal of Systematic Bacteriology*, vol. 42, no. 1, pp. 166–170, 1992.
- [13] M. Drancourt, P. Berger, and D. Raoult, "Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans," *Journal of Clinical Microbiology*, vol. 42, no. 5, pp. 2197–2202, 2004.
- [14] P. Kirschner, E. C. Bottger, D. T. Walton, and M. Valesco, "Microheterogeneity within rRNA of *Mycobacterium goodii*," *Journal of Clinical Microbiology*, vol. 30, no. 4, pp. 1049–1050, 1992.
- [15] C. Taillard, G. Greub, R. Weber et al., "Clinical implications of *Mycobacterium kansasii* species heterogeneity: swiss national survey," *Journal of Clinical Microbiology*, vol. 41, no. 3, pp. 1240–1244, 2003.
- [16] E. Macheras, A.-L. Roux, F. Ripoll et al., "Inaccuracy of single-target sequencing for discriminating species of the *Mycobacterium abscessus* group," *Journal of Clinical Microbiology*, vol. 47, no. 8, pp. 2596–2600, 2009.
- [17] L. A. Almeida and R. Araujo, "Highlights on molecular identification of closely related species," *Infection, Genetics and Evolution*, vol. 13, no. 1, pp. 67–75, 2013.
- [18] J. Dai, Y. Chen, and M. Lauzardo, "Web-accessible database of hsp65 sequences from *Mycobacterium* reference strains," *Journal of Clinical Microbiology*, vol. 49, no. 6, pp. 2296–2303, 2011.