JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2005, p. 3811–3817 Vol. 43, No. 8
0095-1137/05/$08.00+0 doi:10.1128/JCM.43.8.3811–3817.2005

Computational Approach Involving Use of the Internal Transcribed Spacer 1 Region for Identification of *Mycobacterium* Species

Amr M. Mohamed,[1]† Dan J. Kuyper,[2] Peter C. Iwen,[1]* Hesham H. Ali,[2] Dhundy R. Bastola,[1] and Steven H. Hinrichs[1]

Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska 68198-6495,[1] *and Department of Computer Science, University of Nebraska at Omaha, Omaha, Nebraska 68182-0116*[2]

Received 9 November 2004/Returned for modification 20 December 2004/Accepted 17 April 2005

The rapid and reliable identification of clinically significant *Mycobacterium* species is a challenge for diagnostic laboratories. This study evaluates a unique sequence-dependent identification algorithm called *MycoAlign* for the differential identification of *Mycobacterium* species. The *MycoAlign* system uses pan-*Mycobacterium*-specific primer amplification in combination with a customized database and algorithm. The results of testing were compared with conventional phenotypic assays and GenBank sequence comparisons using the 16S rRNA target. Discrepant results were retested and evaluated using a third independent database. The custom database was generated using the hypervariable sequences of the internal transcribed spacer 1 (ITS-1) region of the rRNA gene complex from characterized *Mycobacterium* species. An automated sequence-validation process was used to control quality and specificity of evaluated sequence. A total of 181 *Mycobacterium* strains (22 reference strains and 159 phenotypically identified clinical isolates) and seven nonmycobacterial clinical isolates were evaluated in a comparative study to validate the accuracy of the *MycoAlign* algorithm. *MycoAlign* correctly identified all referenced strains and matched species in 94% of the phenotypically identified *Mycobacterium* clinical isolates. The ITS-1 sequence target showed a higher degree of specificity in terms of *Mycobacterium* identification than the 16S rRNA sequence by use of GenBank BLAST. This study showed the *MycoAlign* algorithm to be a reliable and rapid approach for the identification of *Mycobacterium* species and confirmed the superiority of the ITS-1 region sequence over the 16S rRNA gene sequence as a target for sequence-based species identification.

The genus *Mycobacterium* is comprised of more than 90 different species of aerobic, acid-fast bacilli, with about 30 species associated with a wide variety of human and animal diseases (30, 37, 40). Until the onset of human immunodeficiency virus infection, tuberculosis caused by *Mycobacterium tuberculosis* was considered the most common infectious disease in the world (31). Other mycobacterial species are now known to cause a wide variety of nontuberculous *Mycobacterium* infectious diseases involving a range of different body sites (30, 43). This increase in both tuberculous and nontuberculous *Mycobacterium* infections has intensified the need for alternatives to conventional culture-based identification methods, which are time consuming and labor intensive and lack discriminatory power to differentiate between closely related species (17, 20, 38). To overcome this limitation, recent efforts have focused on the development of molecular approaches for identification (33, 35, 39).

DNA probe assays for identification of a select group of *Mycobacterium* species from culture have been widely accepted (19). These probes have been shown to be sensitive and specific for the identification of the most common *Mycobacterium* species (19). However, they are limited in application and are not capable of identifying all mycobacterial species that may be encountered in the clinical laboratory (32).

An alternative approach to the use of DNA probes is sequence analysis of specific genetic elements (39). Three major genomic targets in the *Mycobacterium* genus have been investigated and include the 16S rRNA gene, the heat shock protein 65 gene (*hsp65*), and the recombinase A gene (*recA*) (2, 4, 33, 34, 38, 39). The 16S rRNA gene has been the target most widely used; however, the presence of identical or highly similar 16S rRNA sequences limits the use of this target for differentiation (4, 8, 18, 21, 27, 35, 36, 42). More recently, the internal transcribed spacer 1 (ITS-1) region sequence located between the 16S rRNA and the 23S rRNA genes has been proposed as an alternative target to the 16S rRNA gene due to a high level of expression and to the greater sequence variability among species and strains (12). Several ITS-1 sequence-based assays have been successfully developed as alternative approaches for the identification of *Mycobacterium* species (10–12).

Utilization of the ITS-1 sequence for species identification requires the availability of a reliable database for comparison studies and a computational approach to sequence alignment analysis (3, 5). The most popular approach currently used for sequence comparison analysis employs a BLAST search of the GenBank database (National Center for Biotechnology Information [NCBI], Washington, D.C.) (http://www.ncbi.nlm.nih.gov/BLAST/). Although this approach is valuable, the absence of a specific validation system to limit the number of low-quality sequences due to sequencing errors and the presence of improperly or ambiguously named sequences contributes to

* Corresponding author. Mailing address: Department of Pathology and Microbiology, University of Nebraska Medical Center, 986495 Nebraska Medical Center, Omaha, NE 68198-6495. Phone: (402) 559-7774. Fax: (402) 559-4077. E-mail: piwen@unmc.edu.

† Present address: Department of Animal Medicine, Faculty of Veterinary Medicine, Assiut University, Assiut, Egypt.

TABLE 1. Utilization of referenced *Mycobacterium* isolates to compare sequence analysis identification results using the 16S rRNA and the internal transcribed spacer-1 (ITS-1) regions as targets

Referenced isolate species	ATCC no.[a]	Sequence analysis species identification[b]		ITS-1 amplicon size (bp)[c]
		16S rRNA	ITS-1	
M. abscessus	19977	*M. abscessus*/*M. chelonae*	*M. abscessus*	300
M. asiaticum	25276	*M. asiaticum*	*M. asiaticum*	265
M. avium[d]	35766	*M. avium*	*M. avium*	263
M. bovis	35741	*M. tuberculosis* complex	*M. tuberculosis* complex	265
M. chelonae	35751	*M. abscessus*/*M. chelonae*	*M. chelonae*	300
M. farcinogenes	35753	*M. farcinogenes*	*M. farcinogenes*	341
M. flavescens	23033	*M. flavescens*	*M. flavescens*	331
M. fortuitum	49404	*M. fortuitum*	*M. fortuitum*	302
M. gastri	15754	*M. gastri*/*M. kansasii*	*M. gastri*	264
M. gordonae	14470	*M. gordonae*	*M. gordonae*	255
M. intracellulare	13950	*M. intracellulare*	*M. intracelluare*	263
M. kansasii	35775	*M. gastri*/*M. kansasii*	*M. kansasii*	*264*
M. lentiflavum	51985	*M. lentiflavum*	*M. lentiflavum*	268
M. marinum	927	*M. marinum*/*M. ulcerans*	*M. marinum*/*M. ulcerans*	264
M. mucogenicum	49650	*M. mucogenicum*	*M. mucogenicum*	267
M. scrofulaceum	19981	*M. scrofulaceum*	*M. scrofulaceum*	264
M. simiae	25275	*M. simiae*	*M. simiae*	*269*
M. szulgai	35799	*M. szulgai*	*M. szulgai*	264
M. terrae	15755	*M. terrae*	*M. terrae*	360
M. tuberculosis	25177	*M. tuberculosis* complex	*M. tuberculosis* complex	265
M. ulcerans	19423	*M. marinum*/*M. ulcerans*	*M. marinum*/*M. ulcerans*	264
M. xenopi	19250	*M. xenopi*	*M. xenopi*	247

[a] ATCC, American Type Culture Collection.
[b] Strain/sequevar variations not listed.
[c] Amplicon size following amplification testing using the *Mycobacterium*-specific primers.
[d] Listed in the ATCC as *M. intracellulare*.

unreliability of the database. In addition, the large size of the GenBank database makes certain operations such as running optimal alignment algorithms impractical due to time constraints (1, 3).

To overcome these challenges, an algorithm-based method to rapidly and reliably identify *Mycobacterium* species was developed using the ITS-1 region as a molecular target. In addition, the algorithm was developed for evaluation of new sequences. The system, called *MycoAlign*, was investigated for functionality under clinical circumstances. The evaluation included three components: a comparison with the results generated using the GenBank BLAST with the 16S rRNA region as a target, comparison with the results of conventional culture testing, and analysis of discrepant isolates by use of an independent database.

MATERIALS AND METHODS

Bacterial isolates. A total of 22 referenced American Type Culture Collection (ATCC) *Mycobacterium* species were used in the current study as controls for evaluation of the sequence-dependent identification system. The ATCC reference strains included in the study are shown in Table 1. In addition, seven different nonmycobacterial strains (*Escherichia coli*, *Nocardia asteroides*, *Corynebacterium urealyticum*, *Pseudomonas aeruginosa*, *Enterococcus faecalis*, *Staphylococcus aureus*, and *Streptococcus pneumonia*) were obtained from the clinical laboratory at the Nebraska Medical Center and used to verify the specificity of the ITS-1 and the 16S rRNA sequences as targets for *Mycobacterium* identification. A total of 159 phenotypically identified clinical *Mycobacterium* isolates collected over a 4-year period (2000 to 2003) were obtained from the Michigan Public Health Laboratory and the Nebraska Medical Center. All isolates were cultured with Löwenstein Jensen (L-J) agar before testing.

Conventional identification. All clinical isolates were identified using conventional phenotypic assays (22). The recognition of an acid-fast positive isolate (by the Ziehl-Neelsen method) growing on L-J and/or Middlebrook7H11 medium was indicative of the presence of a *Mycobacterium* species. Phenotypic studies were performed to determine growth rate, pigment production, and colony morphology. Slow-growing nonchromogenic isolates were tested using the AccuProbe assay (GenProbe, Inc., San Diego, CA) for *M. tuberculosis* complex and the *M. avium* complex according to the manufacturer's directions. Probe assays targeting *M. gordonae* and *M. kansasii* were performed on the chromogenic isolates. Isolates that could not be identified with the nucleic acid probes were identified through a battery of biochemical tests that included niacin accumulation, nitrate reductase, arylsulfatase on day 3 and 14, semiquantitative catalase (45 mm), 68°C catalase production, Tween 80 hydrolysis, urease activity, pyrazinamidase production, tolerance to 5% sodium chloride, and growth on MacConkey agar without crystal violet (22).

DNA extraction, target amplification, and sequencing. An inoculating loopful of mature culture on L-J medium was removed and subjected to DNA extraction. Genomic DNA was extracted from the isolates by the glass bead agitation method as previously described (28). The crude DNA extract was purified using a QIAmp blood kit (QIAGEN Inc., Valencia, Calif.) according to protocols provided by the manufacturer. Both of the 16S rRNA and ITS-1 region targets were amplified from all isolates for sequence-dependent identification. The hypervariable segment of the 16S rRNA (approximate 500 bp) was amplified using the previously described primer set 5′-TGG AGA GTT TGA TCC TGG CTC AG-3′ and 5′-TAC CGC GGC TGC TGG CAC-3′ (13). The hypervariable region of the ITS-1 region (approximately 250 to 350 bp) utilized by the *MycoAlign* custom database was amplified using a newly specified pan-*Mycobacterium* primer set. The forward primer ITS-A1 (5′-GAA GTC GTA ACA AGG TAG CCG-3′) amplified from the 3′ end of the 16S rRNA, while the reverse primer ITS-A6 (5′-G ATG CTC GCA ACC ACT ATC CA-3′) amplified from within the ITS-1 target. The PCR assay for each assay was performed using 5 μl of template DNA (10 ng/μl) in a total reaction volume of 50 μl to include PCR buffer (20 mM Tris-HCl [pH 8.4] and 50 mM KCl); 0.1 mM (each) dATP, dGTP, dTTP, and dCTP; 1.5 mM $MgCl_2$; 0.3 μM (each) primer; and 1.5 U of Platinum *Taq* High-Fidelity DNA polymerase (Gibco BRL, Life Technologies, Gaithersburg, Md.). Amplification was performed on a Stratagene Robocycler model 96 thermocycler (Stratagene, La Jolla, CA), starting with an initial denaturation step at 95°C for 10 m, followed by 35 cycles, each cycle consisting of a denaturation step at 95°C for 1 min, an annealing step at 64°C for 1 min, and an extension step at 72°C for 1 min. An additional extension step at 72°C for 7 min was performed after the last cycle. Ten microliters of amplicon was loaded onto a 2% agarose gel and subjected to electrophoresis to evaluate the size of the PCR

products. PCR products were purified before being sent for sequencing using a QIAquick PCR purification kit (QIAGEN Inc., Valencia, Calif.). Purified PCR products of either 16S rRNA or ITS-1 targets were sequenced at the Eppley Molecular Biology Core Laboratory (University of Nebraska Medical Center, Omaha, NE) using the same forward and reverse PCR amplification primers for both targets.

Sequence source. The National Center for Biotechnology Information GenBank database (NCBI, Washington, D.C.) (http://www.ncbi.nlm.nih.gov/) and the Ribosomal Differentiation of Medical Microorganisms database (RIDOM, Würzburg, Germany) (http://www.ridom-rdna.de) were examined for the availability of *Mycobacterium* species ITS-1 sequences. When identified, the ITS-1 sequences were analyzed by the validation program as described in the custom database design section before incorporation into the *MycoAlign* database. A total of 74 *Mycobacterium* species ITS-1 region sequences were obtained from GenBank, 23 were obtained from the RIDOM database, and 1 sequence (*M. nebraskense*) was generated as a result of this study (23). The GenBank sequences include those of the following species (accession no.): *M. abscessus* (AJ314870), "*M. acapulcensis*" (AF191094), *M. asiaticum* (AB026703), *M. africanum* (AB026699), *M. avium* subsp. *avium* seq. I/Mav-A (AB026690), *M. avium* subsp. *avium* seq. I/Mav-D (L07858), *M. bohemicum* (AJ277282), *M. botniense* (AJ012756), *M. bovis* (L26328), *M. celatum* (AF375990), *M. chelonae* seq. I/Mche-A (AJ291582), *M. chelonae* seq. I/Mche-B (AJ291583), *M. chelonae* seq. II/Mche-C (AJ291584), *M. chimaera* (AJ548480), *M. conspicuum* (X92668), *M. diernhoferi* (AJ314877), *M. farcinogenes* (Y10384), *M. flavescens* seq. I/Mfla-A (AJ291586), *M. fortuitum* subsp. *fortuitum* seq. I/Mfo-A (AJ291587), *M. fortuitum* subsp. *acetamidolyticum* seq. I/Mfo-B (AJ291588), *M. fortuitum* seq. II/Mfo-C (AJ291589), *M. fortuitum* seq. III/Mfo-D (AJ291590), *M. fortuitum* seq. III/Mfo-E (AJ291591), *M. fortuitum* seq. IV/Mfo-F (AJ291592), *M. fortuitum* seq. IV/Mfo-G (AJ291593), *M. gastri* (AB026697), *M. genavense* (Y14183), *M. gilvum* (AJ314876), *M. gordonae* seq. I/Mgo-A (L42258), *M. gordonae* seq. II/Mgo-B (L42259), *M. gordonae* seq. III/Mgo-C (L42260), *M. gordonae* seq. IV/Mgo-D (L42261), "*M. habana*" (X74056), *M. holsaticum* (AJ310470), *M. intracellulare* seq. Min-A (AB026691), *M. intracellulare* seq. Min-B (Z46423), *M. intracellulare* seq. Min-C (Z46424), *M. intracellulare* seq. Min-D (Z46425), *M. kansasii* seq. I/Mka-A (AB026695), *M. kansasii* seq. II/Mka-B (L42263), *M. kansasii* seq. III/Mka-C (L42264), *M. lentiflavum* (AF317658), *M. leprae* (AL583920), *M. malmoense* (AB026696), "*M. manitobense*" (AY082001), *M. marinum* (AJ315572), *M. microti* (L26329), *M. montefiorense* (AF330038), *M. palustre* (AJ308603), *M. parascrofulaceum* (AY337279), *M. peregrinum* seq. Mpe-A (AY291594), *M. peregrinum* seq. Mpe-B (AJ291595), *M. phlei* (AJ291596), *M. porcinum* (AJ291598), *M. saskatchewanense* (AY208857), "*M. savoniae*" (AJ48836), *M. scrofulaceum* seq. Mscro-A (AB026702), *M. senegalense* (Y10385), *M. shimoidei* (AJ005005), *M. simiae* seq. I/Msi-A (AB026694), *M. simiae* seq. I/Msi-C (Y14187), *M. simiae* seq. I/Msi-D (Y14188), *M. smegmatis* seq. Msm-A (AJ291599), *M. smegmatis* seq. Msm-B (U07955), *M. szulgai* (X99220), *M. tokaiense* (AY642533), *M. triviale* (X99221), *M. tuberculosis* (L15623), *M. ulcerans* (X99217), *M. vaccae* (AJ291600), *M. vanbaalenii* (X84977), *M. xenopi* seq. I/Mxe-A (X14190), *M. xenopi* seq. II/Mxe-B (A14191), and *M. xenopi* seq. III/Mxe-C (X14192). The RIDOM sequences included the following: *M. avium* subsp. *silvaticum* seq. I/Mav-A, *M. avium* subsp. *paratuberculosis* seq. I/Mav-A, *M. flavescens* seq. I/Mfla-B, *M. haemophilum*, *M. intracellulare* seq. V/Mac-A, *M. intracellulare* seq. III/Mac-D, *M. intracellulare* seq. II/Mac-I, *M. intracellulare* seq. III/Mac-J, *M. intracellulare* seq. III/Mac-K, *M. intracellulare* seq. IV/Mac-L, *M. kansasii* seq. IV/Mka-D, *M. kansasii* seq. V/Mka-E, *M. kansasii* seq. VI/Mka-F, *M. mucogenicum*, *M. novocastrense*, *M. peregrinum* seq. Mpe-C, *M. scrofulaceum* seq. Mscro-B, *M. septicum*, *M. simiae* seq. II/Msi-E, *M. terrae* seq. I, *M. terrae* seq. II, *M. terrae* seq. III, and *M. triplex*.

Custom database design and validation. The *MycoAlign* system was implemented in a PostGreSOL relational database using JAVA programming language on a Linux platform. The sequence analysis program used an optimal algorithmic approach as described by Pevzner (26). The algorithm was modified to incorporate prioritized filtering criteria that determined the acceptability of a sequence for inclusion in the database. The filtering criteria for input sequences included not having more than two ambiguous bases (N) in the entire sequence, having no more than six continuous stretches of any of the four nucleotide bases (C, A, T, G), and the presence of the 5′-end and 3′-end recognition sequences. Input sequences were first analyzed for the end recognition patterns. These recognition patterns were 5′-CACCTCCTTTCT-3′ as start sequence and 5′-G GGGTGTGG-3′ as end sequence for the ITS-1 target analysis. If these recognition patterns were identified, then the two other validation parameters for limiting ambiguous bases and long stretches of nucleotides were evaluated. Sequences that passed the validation conditions were subsequently included in the customized database.

Unknown sequence evaluation. Unknown sequences submitted for analysis by the MycoAlign system were first validated by fulfillment of the validation parameters. Once the parameters were fulfilled, the sequence underwent alignment analysis against the sequences within the database. Results were given in terms of relative similarity for percentage of identity (RI%) to any of the known mycobacterial sequences in the customized database. The organism identification with the highest RI% value was used for comparison with other methods.

Comparison with 16S rRNA sequence identification. The taxonomic identification of a *Mycobacterium* species by the *MycoAlign* method was compared in tabular form with identification generated from 16S rRNA sequences. The submitted 16S rRNA sequence was evaluated using a nucleotide-nucleotide BLAST analysis against NCBI GenBank database sequences. The sequence was filtered for low complexity.

Discrepancy resolution. In instances when phenotypic culture identification of the clinical isolate did not match the *MycoAlign* sequence-based identification, the isolates were reevaluated by phenotypic tests and the sequences of both molecular targets were retested using a second independent database analysis tool from the University of Würzburg, Würzburg, Germany (Ribosomal Differentiation of Medical Microorganisms [RIDOM]) (available at www.RIDOM.de) (14–16).

RESULTS

***Mycobacterium* target specificity.** The panmycobacterial ITS-1 primer set generated an approximate 250-bp product from *M. avium* and *M. tuberculosis*, while no products were obtained from the seven different nonmycobacterial species. The universal 16S rRNA primer set generated products from both the *Mycobacterium* species as well as nonmycobacterial species.

Conventional versus sequence-dependent identification of referenced strains. A total of 18 of 22 (81.8%) referenced strains were correctly identified to the species level using the ITS-1 sequence as the target for sequence analysis (Table 1). The ITS-1 target was unable to differentiate between *M. tuberculosis* and *M. bovis* and between *M. marinum* and *M. ulcerans*. In comparison, only 14 of 22 (63.6%) referenced strains were correctly identified to the species level when the 16S rRNA sequence was used as a target. Sequence analysis using the 16S rRNA target was unable to differentiate between *M. tuberculosis* and *M. bovis*, between *M. chelonae* and *M. abscessus*, between *M. kansasii* and *M. gastri*, and between *M. marinum* and *M. ulcerans*.

Conventional versus sequence-dependent identification of clinical isolates. Of 159 phenotypically identified *Mycobacterium* species, the *MycoAlign* sequence-dependent identification tool using the ITS-1 target matched the conventional identification of 149 species (Table 2). All 10 discrepant results (comparing the ITS-1 to the conventional approach) were also discrepant using the GenBank 16S rRNA gene target. In seven cases, the identification between ITS-1 and 16S rRNA target matched, while in the other three cases, the identification did not match. In one other case, the 16S rRNA identified a phenotypically identified *M. terrae* as *M. nonchromogenicum*. The ITS-1 identification (*M. terrae*) was compatible with the phenotypic result.

Discordant identification resolution. The 11 isolates with discordant results between conventional and sequence-based identification methods were retested by culture analysis and by the RIDOM database tool (Table 3). Seven of the 11 isolates retested with supplemental culture methods had a change in their identification that matched the *MycoAlign* identification. Six of these included the change of *M. avium* complex to *M.*

TABLE 2. Comparison of the sequence-based *MycoAlign* algorithm using the ITS-1 targets with the GenBank BLAST using the 16S rRNA gene target and conventional culture methods to identify *Mycobacterium* species

Species detected by conventional method[a]	No. of isolates detected by conventional method[a]	16S rRNA sequence analysis result			ITS-1 sequence analysis result		
		Species[c]	No. of isolates	RI %[b]	Species[c]	No. of isolates	RI %[b]
M. tuberculosis	5	*M. tuberculosis* complex	5	100	*M. tuberculosis* complex	5	100
M. bovis	5	*M. tuberculosis* complex	5	100	*M. tuberculosis* complex	5	100
M. avium complex	28	*M. avium*	17	100	*M. avium* subsp. *avium*	12	100
					M. avium subsp. *silvaticum*	5	100
		M. intracellulare seq. V	2	100	*M. intracellulare* seq. Min-A	2	100
		M. intracellulare seq. III	6	100	*M. intracellulare* seq. III/Mac-D	4	100
					M. intracellulare seq. III/Mac-K	2	100
		M. intracellulare seq. I	2	100	*M. intracellulare* seq. Min-A	2	100
		M. kansasii/*M. gastri*	1	100	*M. kansasii* seq. I/Mka-A	1	100
M. scrofulaceum	10	*M. scrofulaceum*	9	100	*M. scrofulaceum* seq. Mscro-A	6	100
					M. scrofulaceum seq. Mscro-B	3	100
		M. simiae	1	99	*M. simiae* seq. II/Msi-E	1	98
M. kansaii	18	*M. kansasii*/*M. gastri*	18	100	*M. kansasii* seq. I/Mka-A	9	100
					M. kansasii seq. III/Mka-C	3	100
					M. kansasii seq. V/Mka-E	6	100
M. gastri	3	*M. kansasii*/*gastri*	3	100	*M. gastri*	3	100
M. chelonae complex	22	*M. chelonae*/*M. abscessus*	22	100	*M. chelonae* seq. I/Mche-A	11	100
					M. chelonae seq. I/Mche-B	2	100
					M. chelonae seq. II/Mche-C	4	100
					M. abscessus	5	100
M. fortuitum	18	*M. fortuitum* subsp. *fortuitum*/*M. fortuitum* subsp. *acetamidolyticum*	16	100	*M. fortuitum* subsp. *fortuitum*	11	100
					M. fortuitum subsp. *acetamidolyticum*	5	100
		M. septicum	1	100	*M. septicum*	1	100
		Nocardia flavorosea	1	100	*M. septicum*	1	100
M. flavescens	3	*M. flavescens* seq. II/*M. novocastrense*	2	100	*M. flavescens* seq. I/Mfla-B	2	100
		"*M. acapulcensis*"	1	98	"*M. acapulcensis*"	1	78
M. gordonae	17	*M. gordonae* seq. I	6	100	*M. gordonae* seq. I/Mgo-A	6	100
		M. gordonae seq. II	4	100	*M. gordonae* seq. II/Mgo-B	4	100
		M. gordonae seq. IV	4	100	*M. gordonae* seq. IV/Mgo-D	4	100
		M. gordonae seq. V	2	100	*M. gordonae* seq. IV/Mgo-D	2	100
		M. szulgai	1	99	*M. szulgai*	1	98
M. lentiflavum	1	*M. lentiflavum*	1	100	*M. lentiflavum*	1	100
M. marinum	5	*M. marinum*/*M. ulcerans*	5	100	*M. marinum*/*M. ulcerans* Mul-A	4	100
					M. ulcerans seq. Mul-B	1	100
M. ulcerans	1	*M. marinum*/*M. ulcerans*	1	100	*M. marinum*/*M. ulcerans* Mul-A	1	100
M. xenopi	9	*M. xenopi* seq. I	6	100	*M. xenopi* seq. I/Mxe-A	6	100
		M. xenopi seq. II	1	100	*M. xenopi* seq. II/Mxe-B	1	100
		M. celatum	1	100	*M. celatum*	1	99
		M. montefiorense	1	98	*M. scrofulaceum*	1	93
M. szulgai	2	*M. szulgai*	1	100	*M. szulgai*	1	100
		M. gordonae	1	100	*M. gordonae* seq. I/Mgo-A	1	100
M. terrae	8	*M. terrae* seq. I	2	100	*M. terrae* seq. I	3	100
		M. terrae seq. II	1	100			
		M. terrae seq. III	4	100	*M. terrae* seq. III	4	98
		M. nonchromogenicum	1	99	*M. terrae* seq. II	1	97
M. mucogenicum	2	*M. mucogenicum*	1	100	*M. mucogenicum*	1	100
		M. farcinogenes	1	99	*M. farcinogenes*	1	98
M. asiaticum	1	*M. asiaticum*	1	100	*M. asiaticum*	1	100
M. simiae	1	*M. simiae*	1	100	*M. simiae* seq. I/Msi-A	1	100

[a] Isolates were identified from clinical materials.
[b] RI% is the relative identity percentage between the database sequence and the clinical isolate sequence.
[c] seq., sequevar.

kansasii, *M. scrofulaceum* to *M. simiae*, *M. gordonae* to *M. szulgai*, *M. xenopi* to *M. celatum*, *M. szulgai* to *M. gordonae*, and *M. fortuitum* to *M. septicum*. The seventh isolate, which was originally identified as *M. fortuitum* by the conventional methods and as *M. septicum* by *MycoAlign* ITS-1 sequence analysis and *Nocardia flavorosea* by GenBank 16S rRNA sequence analysis, showed two different colonies upon subculture that were identified as *M. septicum* (by ITS-1 sequence analysis) and *Nocardia flavorosea* (by 16S rRNA sequence analysis) (Table 3). Sequence analysis using the RIDOM tool confirmed the *MycoAlign* ITS-1 sequence identification for all seven isolates

In the first two of the four remaining cases discordant between *MycoAlign* and conventional culture, the RIDOM system confirmed the identification by *MycoAlign* and GenBank BLAST ("*M. acapulcensis*" and *M. farcinogenes*). Retesting by supplemental conventional culture methods did not change the original identification of *M. flavescens* and *M. mucogenicum*. In the third case identified as *M. xenopi*, the RIDOM system also confirmed the *MycoAlign* identification (*M. scrofulaceum*) but

TABLE 3. Reevaluation of discrepant *Mycobacterium* identification results between conventional and sequence-based test analysis using repeat conventional tests and sequence analysis with the RIDOM database tool[a]

Discrepant identification result					Reevaluation result					
Conventional ID	Sequence analysis ID				Conventional Mycobacterium ID	Sequence analysis ID				
	16S rRNA	RI%[b]	ITS	RI%		16S rRNA	RI%	ITS	RI%	RIDOM
M. avium complex	*M. kansasii/gastri*	100	*M. kansasii*	100	*M. kansasii*	*M. kansasii/gastri*	99	*M. kansasii*	98	*M. kansasii*
M. scrofulaceum	*M. simiae*	99	*M. simiae*	98	*M. simiae*	*M. simiae*	99	*M. simiae*	98	*M. simiae*
M. flavescens	"*M. acapulcensis*"	98	"*M. acapulcensis*"	78	*M. flavescens*	"*M. acapulcensis*"	98	"*M. acapulcensis*"	64	"*M. acapulcensis*"
M. xenopi	*M. celatum*	100	*M. celatum*	99	*M. celatum*	*M. celatum*	99	*M. celatum*	99	*M. celatum*
M. xenopi	*M. montefiorense*	98	*M. scrofulaceium*	93	*M. xenopi*	*M. heidelbergense*	98	*M. scrofulaceum*	87	*M. scrofulaccum*
M. gordonae	*M. szulgai*	99	*M. szulgai*	98	*M. szulgai*	*M. szulgai*	99	*M. szulgai*	99	*M. szulgai*
M. terrae	*M. nonchromo-genicum*	99	*M. terrae*	97	*M. terrae* complex	*M. nonchromo-genicum*	98	*M. terrae*	98	*M. nonchromo-genicum*
M. mucogenicum	*M. farcinogenes*	99	*M. farcinogenes*	98	*M. mucogenicum*	*M. farcinogenes*	99	*M. farcinogenes*	98	*M. farcinogenes*
M. szulgai	*M. gordonae*	100	*M. gordonae*	99	*M. gordonae*	*M. gordonae*	99	*M. gordonae*	98	*M. gordonae*
M. fortuitum	*M. septicum*	100	*M. septicum*	100	*M. septicum*	*M. septicum*	100	*M. septicum*	100	*M. septicum*
M. fortuitum	*Nocardia flavorosea*	100	*M. septicum*	100	*M. septicum/ N. flavorosea*[c]	*N. flavorosea*	100	*M. septicum*	100	*M. septicum*

[a] Only species have been included for simplification of results. ID, identification.
[b] RI%, relative identity percentage between the database sequence and the unknown sequence results.
[c] Both isolates were recognized in culture.

not the identification based on the 16S rRNA results (*M. heidelburgense*). In the final case where there was a discrepancy between conventional culture and the GenBank 16S rRNA result (*M. nonchromogenicum*) but not with the *MycoAlign* result (*M. terrae*), the RIDOM database result was identical to the GenBank 16S rRNA result.

DISCUSSION

New approaches to the identification of bacterial species on the basis of comparative DNA sequence analysis hold great promise for improvement in microbial diagnostics (39). Sequence-dependent identification has been shown to be an especially effective molecular tool for the rapid differentiation of *Mycobacterium* species, although a number of challenges remain before this approach can be routinely applied in the clinical laboratory (34).

One major challenge is the development and characterization of a validated database for sequence comparison analysis. Limitations of currently available public databases have been widely discussed and include the presence of sequencing errors and ambiguous bases within the database sequences as well as incorrectly identified sequences (3, 6, 40). These errors may not be evident to most users but can affect the accuracy of the search process; consequently, discrepancies and inconclusive results have been reported (25, 40). To overcome the limitations within public databases, several curated databases have been created, including the MicroSeq 500 database (Applied Biosystems, Foster City, Calif.) and the RIDOM database from the University of Würzburg, Würzburg, Germany (25, 36). These custom database identification systems have been shown to be useful; however, they do lack ITS-1 sequence information for the more unusual *Mycobacterium* species as well as for the recently described species (5, 40).

In the creation of the *MycoAlign* system, a series of steps was incorporated to overcome limitations of sequence validation in addition to incorporating an automated updating tool in the *MycoAlign* software program. This allowed for periodic automated searching of the GenBank database for the recently deposited sequences of newly established *Mycobacterium* species. New sequences identified for inclusion into the *MycoAlign* custom database were subjected to restricted validation conditions to ensure quality.

The *MycoAlign* assay was shown to be reliable for the identification of *Mycobacterium* species and illustrated the greater utility of choosing the ITS-1 hypervariable sequence target over the 16S rRNA sequence as a target for sequence-based differential identification. This was not a surprise, since earlier studies by Park and others had shown that mycobacterial strains could be identified only to the group or complex level using 16S rRNA sequence as a target but that they could be further differentiated to the species level using the ITS-1 target (12, 24, 36, 41). In the present study, the *MycoAlign* system was able to differentiate between *M. abscessus* and *M. chelonae* and between *M. kansasii* and *M. gastri* when the ITS-1 target was used. These isolates were not differentiated when using GenBank analysis and the 16S rRNA gene target. In addition, it was also possible to identify several mycobacterial isolates at the subspecies or strain level.

Although the *MycoAlign* system showed greater capability than other analysis systems based on 16S rRNA sequences alone, the algorithm was still unable to differentiate between some clinically relevant mycobacterial species. Most notable was the inability to differentiate *M. tuberculosis* from *M. bovis* and *M. marinum* from *M. ulcerans*. This finding was consistent with studies that showed these latter species or subspecies to be highly related, with identical sequences for both the 16S rRNA and ITS-1 region targets (10, 12, 29).

In the evaluation of identifications of clinical isolates by conventional culture methods versus the *MycoAlign* approach using the ITS-1 target, 10 discrepancies were noted whereas 11 discrepancies were noted using the 16S rRNA target. Retesting using expanded biochemicals and the RIDOM sequence analysis tool showed the *MycoAlign* identification was correct in seven cases. In two of the three remaining ITS-1 discrepant cases, the GenBank, RIDOM, and *MycoAlign* approaches were

in agreement and conflicted with the culture identification. In the case of one isolate identified as *M. xenopi* by conventional culture methods, MycoAlign and RIDOM were in agreement but differed from the 16S rRNA sequence identification. Since conventional assays were designed primarily to detect and identify *M. tuberculosis* and other clinically important nontuberculous mycobacteria, it was not unexpected that this method failed to identify all the species correctly (17, 20). The increasing number of newly defined mycobacterial species and the "difficult-to-identify" variants of known species represent a significant challenge for conventional approaches (9).

During evaluation of the discrepant *Mycobacterium* species identification results between the conventional and the sequence-based test analysis, a difference in RI% of more than five units was noted when using the *MycoAlign* program for two samples (Table 3). In each case, the sequence identification using the 16S rRNA gene sequence and GenBank BLAST analysis, the ITS-1 region sequence and *MycoAlign*, and RIDOM system analysis gave the same identification for the isolates. In one case identified as "*M. acapulcensis*" by ITS-1 sequence analysis, the RI% value decreased from 78% in the original test to 64% in the repeat test, while in another case identified as *M. scrofulaceum*, the RI% decreased from 93% to 87%. Upon further evaluation of the *MycoAlign* algorithm, it was noted that minor changes within the sequence (base pair change or the addition of a no-read base) had a significant effect on the value. This "rigidity" of the database had originally been allowed for greater accuracy in species and strain identification. It appears that an RI% value $< 98\%$ suggests that the organism may be new to the database. In both of the discrepant cases, neither the "*M. acapulcensis*" nor the *M. scrofulaceum* isolate was identified with certainty using the ITS-1 region target, suggesting that these isolates may be unrecognized species with sequences not available in the *MycoAlign* database. Additional testing of phenotypically similar strains will help to resolve this issue. Additional testing of isolates has now shown the ability of the MycoAlign system to identify new species (23).

Contamination of clinical specimens by nonmycobacterial species is a known problem that not only reduces the accuracy of the diagnostic process but also extends the time required to separate to purity prior to retesting (7). This problem was addressed in the *MycoAlign* system through the use of *Mycobacterium*-specific primers in the amplification reaction. Whereas universal 16S rRNA primers generated products from mixed samples or samples containing only nonmycobacterial species, a product was generated with the ITS-1 primers only in those samples containing mycobacterial DNA. The value of this capability was clearly demonstrated by the identification of a mixed clinical isolate as *Nocardia flavorosea* by use of 16S rRNA sequence, whereas the ITS-1 region sequence analysis identified *M. septicum* as the *Mycobacterium* species present in the sample.

This study showed the *MycoAlign* identification system to be a reliable alternative to conventional phenotypic methods for the identification of *Mycobacterium* species. It also confirmed the superiority of the ITS-1 sequence over the 16S rRNA sequence as a target for sequence-based species identification. Additional evaluation of the computational software with the addition of new ITS-1 sequences as they become available and undergo validation will allow for increased discriminatory power of the *MycoAlign* system in the future.

Nucleotide sequence accession numbers. The following sequences have been deposited in the GenBank database (accession no.) as a result of this study: *Mycobacterium nebraskense* ATCC BAA-837 complete 16S rRNA gene (AY368456) and 16S–23S rRNA intergenic spacer region (AY368458).

ACKNOWLEDGMENTS

We thank Steve Church, Michigan Department of Community Health Laboratories, and the technologists in the Nebraska Medical Center Mycobacteriology Laboratory for providing selected mycobacterial isolates for testing.

A.M.M. was supported by a research fellowship from the Egyptian government.

In the interest of full disclosure, the authors acknowledge that the commercial application of this approach is being explored.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Blackwood, K. S., C. He, J. Gunton, C. Y. Turenne, J. Wolfe, and A. M. Kabani.** 2000. Evaluation of *recA* sequences for identification of *Mycobacterium* species. J. Clin. Microbiol. **38:**2846–2852.
3. **Bork, P., and A. Bairoch.** 1996. Go hunting in sequence databases but watch out for the traps. Trends Genet. **12:**425–427.
4. **Clarridge, J. E., III.** 2004. Impact of 16S rRNA gene sequence analysis of identification of bacteria on clinical microbiology and infectious diseases. Clin. Microbiol. Rev. **17:**840–862.
5. **Cloud, J. L., H. Neal, R. Rosenberry, C. Y. Turenne, M. JAMA, D. R. Hillyard, and K. C. Carroll.** 2002. Identification of *Mycobacterium* spp. by using a commercial 16S ribosomal DNA sequencing kit and additional sequencing libraries. J. Clin. Microbiol. **40:**400–406.
6. **Cook, V. J., C. Y. Turenne, J. Wolfe, R. Pauls, and A. Kabani.** 2003. Conventional methods versus 16S ribosomal DNA sequencing for identification of nontuberculous mycobacteria: cost analysis. J. Clin. Microbiol. **41:**1010–1015.
7. **Cornfield, D. B., K. G. Beavis, J. A. Greene, M. Bojak, and J. Bondi.** 1997. Mycobacterial growth and bacterial contamination in the mycobacteria growth indicator tube and BACTEC 460 culture systems. J. Clin. Microbiol. **35:**2068–2071.
8. **Dobner, P., K. Feldmann, M. Rifai, T. Loscher, and H. Rinder.** 1996. Rapid identification of mycobacterial species by PCR amplification of hypervariable 16S rRNA gene promoter region. J. Clin. Microbiol. **34:**866–869.
9. **Floyd, M. M., W. M. Gross, D. A. Bonato, V. A. Silcox, R. W. Smithwick, B. Metchock, J. T. Crawford, and W. R. Butler.** 2000. *Mycobacterium kubicae* sp. nov., a slowly growing, scotochromogenic *Mycobacterium*. Int. J. Syst. Evol. Microbiol. **50**(Pt. 5)**:**1811–1816.
10. **Frothingham, R., H. G. Hills, and K. H. Wilson.** 1994. Extensive DNA sequence conservation throughout the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **32:**1639–1643.
11. **Frothingham, R., and K. H. Wilson.** 1994. Molecular phylogeny of the *Mycobacterium avium* complex demonstrates clinically meaningful divisions. J. Infect. Dis. **169:**305–312.
12. **Frothingham, R., and K. H. Wilson.** 1993. Sequence-based differentiation of strains in the *Mycobacterium avium* complex. J. Bacteriol. **175:**2818–2825.
13. **Hall, L., K. A. Doerr, S. L. Wohlfiel, and G. D. Roberts.** 2003. Evaluation of the MicroSeq system for identification of mycobacteria by 16S ribosomal DNA sequencing and its integration into a routine clinical mycobacteriology laboratory. J. Clin. Microbiol. **41:**1447–1453.
14. **Harmsen, D., S. Dostal, A. Roth, S. Niemann, J. Rothganger, M. Sammeth, J. Albert, M. Frosch, and E. Richter.** 2003. RIDOM: comprehensive and public sequence database for identification of *Mycobacterium* species. BMC Infect. Dis. **13:**26–35.
15. **Harmsen, D., J. Rothganger, M. Frosch, and J. Albert.** 2002. RIDOM: ribosomal differentiation of medical microorganisms database. Nucleic Acids Res. **30:**416–417.
16. **Harmsen, D., J. Rothganger, C. Singer, J. Albert, and M. Frosch.** 1999. Intuitive hypertext-based molecular identification of microorganisms. Lancet **353:**291.
17. **Kent, P. T., G. P. Kubica.** 1985. Public health mycobacteriology. A guide for the level III laboratory. U.S. Department of Health and Human Services publication. Centers for Disease Control and Prevention, Atlanta, GA.
18. **Kim, B.-J., S.-H. Lee, M.-A. Lyu, S.-J. Kim, G.-H. Bai, G.-T. Chae, E.-C. Kim, C.-Y. Cha, and Y.-H. Kook.** 1999. Identification of mycobacterial spe-

cies by comparative sequence analysis of the RNA polymerase gene (*rpoB*). J. Clin. Microbiol. **37:**1714–1720.
19. **Kirschner, P., J. Rosenau, B. Springer, K. Teschner, K. Feldmann, and E. C. Bottger.** 1996. Diagnosis of mycobacterial infections by nucleic acid amplification: 18-month prospective study. J. Clin. Microbiol. **34:**304–312.
20. **Kirschner, P., B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F. C. Bange, and E. C. Bottger.** 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.
21. **Lappayawichit, P., S. Rienthong, D. Rienthong, C. Chuchottaworn, A. Chaiprasert, W. Panbangred, H. Saringcarinkul, and P. Palittapongarnpim.** 1996. Differentiation of *Mycobacterium* species by restriction enzyme analysis of amplified 16S–23S ribosomal DNA spacer sequences. Tuber. Lung. Dis. **77:**257–263.
22. **Metchock, B. G., F. S. Nolte, and R. J. Wallace, Jr.** 1999. *Mycobacterium*, pp. 399–437. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of clinical microbiology, 7th ed. ASM Press, Washington, D.C.
23. **Mohamed, A. M., S. Tarantolo, P. C. Iwen, and S. H. Hinrichs.** 2004. *Mycobacterium nebraskense* sp. nov., a novel slowly growing scotochromogenic species. Int. J. Syst. Evol. Microbiol. **54:**2056–2060.
24. **Park, H., H. Jang, C. Kim, B. Chung, C. L. Chang, S. K. Park, and S. Song.** 2000. Detection and identification of mycobacteria by amplification of the internal transcribed spacer regions with genus- and species-specific PCR primers. J. Clin. Microbiol. **38:**4080–4085.
25. **Patel, J. B., D. G. Leonard, X. Pan, J. M. Musser, R. E. Berman, and I. Nachamkin.** 2000. Sequence-based identification of *Mycobacterium* species using the MicroSeq 500 16S rDNA bacterial identification system. J. Clin. Microbiol. **38:**246–251.
26. **Pevzner, P. A.** 2000. Computational molecular biology: an algorithmic approach. MIT Press, London, United Kingdom.
27. **Pitulle, C., M. Dorsch, J. Kazda, J. Wolters, and E. Stackebrandt.** 1992. Phylogeny of rapidly growing members of the genus *Mycobacterium*. Int. J. Syst. Bacteriol. **42:**337–343.
28. **Plikaytis, B. B., R. H. Gelber, and T. M. Shinnick.** 1990. Rapid and sensitive detection of *Mycobacterium leprae* using a nested-primer gene amplification assay. J. Clin. Microbiol. **28:**1913–1917.
29. **Portaels, F., J. Agular, K. Fissette, P. A. Fonteyne, H. De Beenhouwer, P. de Rijk, A. Guedenon, R. Lemans, C. Steunou, C. Zinsou, J. M. Dumonceau, and W. M. Meyers.** 1997. Direct detection and identification of *Mycobacterium ulcerans* in clinical specimens by PCR and oligonucleotide-specific capture plate hybridization. J. Clin. Microbiol. **35:**1097–1100.
30. **Primm, T. P., C. A. Lucero, and J. O. Falkinham III.** 2004. Health impacts of environmental mycobacteria. Clin. Microbiol. Rev. **17:**98–106.
31. **Raviglione, M. C., D. E. Snider, Jr., and A. Kochi.** 1995. Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. JAMA **273:**220–226.
32. **Reisner, B. S., A. M. Gatson, and G. L. Woods.** 1994. Use of Gen-Probe AccuProbes to identify *Mycobacterium avium complex*, *Mycobacterium tuberculosis* complex, *Mycobacterium kansasii*, and *Mycobacterium gordonae* directly from BACTEC TB broth cultures. J. Clin. Microbiol. **32:**2995–2998.
33. **Ringuet, H., C. Akoua-Koffi, S. Honore, A. Varnerot, V. Vincent, P. Berche, J. L. Gaillard, and C. Pierre-Audigier.** 1999. *hsp65* sequencing for identification of rapidly growing mycobacteria. J. Clin. Microbiol. **37:**852–857.
34. **Rogall, T., J. Wolters, T. Flohr, and E. C. Bottger.** 1990. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. Int. J. Syst. Bacteriol. **40:**323–330.
35. **Roth, A., M. Fischer, M. E. Hamid, S. Michalke, W. Ludwig, and H. Mauch.** 1998. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S–23S rRNA gene internal transcribed spacer sequences. J. Clin. Microbiol. **36:**139–147.
36. **Roth, A., U. Reischl, A. Streubel, L. Naumann, R. M. Kroppenstedt, M. Habicht, M. Fischer, and H. Mauch.** 2000. 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:**1094–1104.
37. **Shinnick, T. M., and R. C. Good.** 1994. Mycobacterial taxonomy. Eur. J. Clin. Microbiol. Infect. Dis. **13:**884–901.
38. **Springer, B., L. Stockman, K. Teschner, G. D. Roberts, and E. C. Bottger.** 1996. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. J. Clin. Microbiol. **34:**296–303.
39. **Tortoli, E.** 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin. Microbiol. Rev. **16:**319–354.
40. **Turenne, C. Y., L. Tschetter, J. Wolfe, and A. Kabani.** 2001. Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous *Mycobacterium* species. J. Clin. Microbiol. **39:**3637–3648.
41. **van der Giessen, J. W., R. M. Haring, and B. A. van der Zeijst.** 1994. Comparison of the 23S ribosomal RNA genes and the spacer region between the 16S and 23S rRNA genes of the closely related *Mycobacterium avium* and *Mycobacterium paratuberculosis* and the fast-growing *Mycobacterium phlei*. Microbiology **140:**1103–1108.
42. **van der Vliet, G. M., R. A. Schukkink, B. van Gemen, P. Schepers, and P. R. Klatser.** 1993. Nucleic acid sequence-based amplification (NASBA) for the identification of mycobacteria. J. Gen. Microbiol. **139:**2423–2429.
43. **Wayne, L. G., and H. A. Sramek.** 1992. Agents of newly recognized or infrequently encountered mycobacterial diseases. Clin. Microbiol. Rev. **5:**1–25.