

Mycobacterium Species Identification and Rifampin Resistance Testing with High-Density DNA Probe Arrays

A. TROESCH,^{1*} H. NGUYEN,² C. G. MIYADA,² S. DESVARENNE,¹ T. R. GINGERAS,²
P. M. KAPLAN,² P. CROS,¹ AND C. MABILAT¹

bioMérieux, 69280 Marcy-L'Etoile, France,¹ and Affymetrix, Santa Clara, California²

Received 27 July 1998/Returned for modification 3 September 1998/Accepted 15 October 1998

Species identification within the genus *Mycobacterium* and subsequent antibiotic susceptibility testing still rely on time-consuming, culture-based methods. Despite the recent development of DNA probes, which greatly reduce assay time, there is a need for a single platform assay capable of answering the multitude of diagnostic questions associated with this genus. We describe the use of a DNA probe array based on two sequence databases: one for the species identification of mycobacteria (82 unique 16S rRNA sequences corresponding to 54 phenotypical species) and the other for detecting *Mycobacterium tuberculosis* rifampin resistance (*rpoB* alleles). Species identification or rifampin resistance was determined by hybridizing fluorescently labeled, amplified genetic material generated from bacterial colonies to the array. Seventy mycobacterial isolates from 27 different species and 15 rifampin-resistant *M. tuberculosis* strains were tested. A total of 26 of 27 species were correctly identified as well as all of the *rpoB* mutants. This parallel testing format opens new perspectives in terms of patient management for bacterial diseases by allowing a number of genetic tests to be simultaneously run.

Tuberculosis (TB), caused by members of the *Mycobacterium tuberculosis* complex, is one of the most common human infectious diseases, causing three million deaths a year worldwide (20). While the disease is associated with impoverished economic conditions, TB is on the rise in many industrialized nations. The spread in TB is due to immigration, the emergence of drug-resistant strains, and the AIDS epidemic. In advanced stages of AIDS, where TB infections are commonly found, mycobacterial infections due to members of the *M. avium-intracellulare* complex (MAC) are also on the increase (10). Reduced and compromised immune function as found in newborns, infants, and immunosuppressed individuals allows opportunistic infections caused by mycobacteria other than *M. tuberculosis* (MOTT). Such species infect many sites within the body but primarily cause pulmonary disease, cervical lymphadenopathy, and localized skin and soft tissue lesions. Mycobacterial species associated with MOTT disease states include *M. avium-intracellulare*, *M. chelonae*, *M. fortuitum*, *M. kansasii*, *M. xenopi*, *M. marinum*, *M. scrofulaceum*, and *M. szulgai*.

The increasing number of mycobacterial infections has made it clinically important to quickly identify mycobacteria at the species level. The diagnosis of a pathogenic versus a nonpathogenic species not only has epidemiological implications but is also relevant to the demands of patient management. Individuals with highly contagious infections may be isolated to prevent the spread of the disease. Antibiotic treatments may vary according to the species encountered.

The emergence of drug-resistant *M. tuberculosis* has created additional concern in the event of TB diagnosis. These strains are resistant to the primary antituberculosis drugs and require a specialized antibiotic treatment. Recently, a number of genetic changes leading to rifampin, isoniazid, pyrazinamide, and ethambutol resistance have been characterized (4). For example, rifampin resistance has been shown to be conferred by missense mutations within a short motif of the beta subunit of

DNA-dependent RNA polymerase encoded by the *rpoB* gene (26) in over 90% of rifampin-resistant isolates (12). The availability of sequence changes resulting in a resistant phenotype allows for the development of probe-based assays for antibiotic resistance.

To date, the design of molecular tests has sped up the diagnostics of this fastidious genus but still suffers some drawbacks. For species identification, a highly polymorphic region of the 16S rRNA gene has been shown to contain species-specific polymorphisms (2, 14). This region is currently used in several commercially available assays, applicable either directly on specimen (Accu-Probe; Gen-Probe, San Diego, Calif.) or after enzymatic amplification of the target for improved sensitivity (AMTD; Gen-Probe; Amplicor MTB, Roche Diagnostics Systems, Somerville, N.J.). However, these kits are mostly designed for the diagnosis of TB and do not take full advantage of the information-rich content of this region. The vast majority of mycobacterial species can be discriminated by this region, as originally reported by Springer et al. (23) and demonstrated directly on clinical specimens by Kirschner et al. (15).

The GeneChip technology, recently developed by Affymetrix (Santa Clara, Calif.), is a promising new method for assessing genetic diversity on a larger scale. This method relies upon the hybridization of the nucleic acid target to large sets of oligonucleotides synthesized at precise locations on a miniaturized glass substrate (7). This technology has been already successfully applied for monitoring gene expression and screening of mutations and polymorphisms in several human and viral genes (3, 6, 17, 18). We have investigated this probe array strategy for bacteriology testing, focusing on mycobacterial diseases. An assay that is able to interrogate the sequence of regions from the 16S rRNA and *rpoB* loci has been developed. Unique hybridization patterns allow for the identification of *Mycobacterium* species and the rifampin-resistant alleles.

MATERIALS AND METHODS

Bacterial strains, phenotypic identification, and susceptibility testing. The strains originated from reference collections (American Type Culture Collection and Deutsche Sammlung von Mikroorganismen) or were clinical isolates obtained through several laboratories. All isolates were grown on either Lowenstein-Jensen or Coletsos medium and were examined for growth rate, gross and

* Corresponding author. Mailing address: Affymetrix, 3380 Central Expressway, Santa Clara, CA 95051. Phone: (408) 731 5572. Fax: (408) 481 0435. E-mail: alain_troesch@affymetrix.com.

microscopic colony morphology, and pigmentation. They were subjected to biochemical tests for niacin, nitrate reduction, catalase (drop method), arylsulfatase, pyrazinamidase, urease, and lipase (Tween 80 hydrolysis) (21) and also to complementary tests for certain isolates. Rifampin resistance testing was performed by the proportion method with medium containing 1 mg of rifampin per ml (13). The isolate was considered resistant if there was more than 1% growth on the antibiotic-containing medium compared with the growth on the drug-free medium.

Bacterial *rpoB* clones. Clones were generated from *rpoB* PCR products that were ligated into pGEM-T Easy plasmids (Promega, Madison, Wis.). The *rpoB* PCR products were generated from clinical isolates by using the following primer pair: MTX 2281 TB (CCC AGG ACG TGG AGG CGA TCA CAC CGC A) and MTX 2985 TB (ACG TCG CCG CGT CGA TCG CCG) (8). MTX 2281 TB is located at coordinates 2281 to 2308 of GenBank accession no. L27989 sequence. The resulting amplicon was approximately 700 bp in length. The insert sequence was characterized by automated DNA sequencing.

Probe array design and tiling strategy. An array tiling strategy similar to that described by Cronin et al. (6), Kozal et al. (17), and Chee et al. (3) was used to identify the sequences differentiating the *Mycobacterium* species and the *rpoB* mutations. For every base within a given reference sequence, four probes of equal length are synthesized on the array. The interrogated base is centrally located within the probes, which also have common 3' and 5' termini. One probe is an exact complement to the reference sequence, while the three other probes represent the possible single base mismatches to the interrogated base. Base calls are determined by comparing the hybridization intensity of a labeled target to the four probes (see Fig. 2). For the 16S rRNA sequences, probe redundancy was eliminated by synthesizing probes shared by two or more references only once on the array. In addition to the wild-type probes for *rpoB*, each rifampin resistance mutation is represented by its own set of probes. The array also contains 2.2 kb of the *M. tuberculosis katG* gene sequence, which encodes catalase peroxidase. The probe array is divided into four distinct zones corresponding to 16S rRNA, *rpoB* antisense, *rpoB* sense, and *katG* antisense sequences, as shown in Fig. 1. The array is divided into specific 50-by-50- μ m units or cells over a 1.28- by 1.28-cm area, making a total number of 65,000 different synthesis sites. A database of 82 unique 16S rRNA sequences was utilized to design the array, which enables the discrimination of 54 phenotypically distinct species. Certain species or taxonomic complexes are represented by more than one reference sequence due to the sequence heterogeneity observed in the region of the 16S rRNA tiled on the array (i.e., the MAC). Also, a database of *rpoB* sequences from 61 rifampin-resistant isolates is represented on the array. Those sequences contain 51 unique *rpoB* mutations.

Target preparation. For isolates, one or two freshly grown colonies of bacteria (3- to 5-mm diameter; ca. 10^8 bacteria) were scraped on the end of a spatula and resuspended into 250 μ l of sterile water in a 1.5-ml Eppendorf tube. Total nucleic acids were released from culture material by vortexing the bacterial suspension in the presence of glass beads. A 5- μ l aliquot of the lysate was added directly to the PCR. Alternatively, 20 ng of plasmid DNA was added to the PCR. The 16S hypervariable region was PCR amplified with *Mycobacterium* genus primers (positions 213 to 236 and 394 to 415 on the *M. tuberculosis* reference sequence M20940 [GenBank]; *M. tuberculosis* amplicon size is 202 bp). The primers also contained either a bacteriophage T3 or T7 promoter sequence at their 5' ends (17): T3-M1, 5'-AATTAACCTCACTAAAGGGAACACGTGGG TGATCTGCCTGCA, and T7-M2, 5'-GTAATACGACTCACTATAGGGCTGT GGCCGGACACCCTCTCA (promoter sequences in boldface and mycobacterial sequence in lightface). The *M. tuberculosis rpoB* Rif^r locus was amplified with MTX 2281 TB and MTX 2985 TB primers (8), described above, and tailed with T3 or T7 promoter sequences. Amplification of the *M. tuberculosis katG* region analyzed on the array was performed with T3- or T7-tailed primers 33f (5'-TC ACAGCCCGATAACACCAAC) and 2288r (5'-GGCCGATCAACCCGAATC AGC) (positions 1942 to 1962 and 4197 to 4217 on X68081, respectively; *M. tuberculosis* amplicon size is 2,275 bp).

PCR was carried out in a 100- μ l reaction volume containing 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM MgCl₂, 0.001% (wt/vol) gelatin, 5% (vol/vol) dimethyl sulfoxide, 0.5 μ M (each) primer, 200 μ M (each) deoxynucleotide triphosphates, and 1.5 U of *Taq* polymerase (AmpliTaq; Perkin-Elmer, Norwalk, Conn.). PCR was performed in a Perkin-Elmer 2400 thermal cycler with an initial denaturation step at 94°C for 5 min and cycling conditions of 94°C for 45 s, 60°C for 30 s, and 72°C for 30 s (2 min for *katG* target) for 35 cycles and 72°C for 10 min for the last cycle. PCR products were analyzed by agarose gel electrophoresis.

The promoter-tagged PCR amplicons were used for generating labeled single-stranded RNA targets by *in vitro* transcription. Each 20- μ l reaction mixture contained approximately 50 ng of PCR product; 20 U of T3 or T7 RNA polymerase (Promega); 40 mM Tris acetate (pH 8.1); 100 mM Mg(acetate)₂; 10 mM dithiothreitol; 1.25 mM (each) ATP, CTP, and GTP; 0.5 mM UTP; and 0.25 mM fluorescein-UTP. The reaction was carried out at 37°C for 1 h. *In vitro*-transcribed RNA was fragmented either by adjustment of the concentration of MgCl₂ to 30 mM and heating at 94°C for 30 min (17) or by incubation with 30 mM MnCl₂ and 30 mM imidazole at 65°C for 30 min. The efficiency of fragmentation was analyzed by denaturing polyacrylamide gel electrophoresis.

Probe array hybridization and analysis. Hybridizations of the probe arrays were performed with the GeneChip Fluidics Station (Affymetrix). One to five microliters of the fragmented labeled RNA target was diluted in 500 μ l of

hybridization buffer. Hybridization buffer 1 consisted of 4.5 \times SSPE (0.675 M NaCl, 45 mM NaH₂PO₄, 4.5 mM EDTA, pH 7.4) and 0.05% (vol/vol) Triton X-100. Hybridization buffer 2 consisted of 6 \times SSPE (0.9 M NaCl, 60 mM NaH₂PO₄, 6 mM EDTA, pH 7.4), 0.05% (vol/vol) Triton X-100, and a proprietary mix of detergents and denaturing agents. The probe array was incubated in the presence of these solutions for 30 min at 50°C and then washed twice in 3 \times SSPE (0.45 M NaCl, 30 mM NaH₂PO₄, 3 mM EDTA, pH 7.4)–0.005% (vol/vol) Triton X-100 at 30°C. Fluorescent signal emitted by target bound to the array was detected at a pixel resolution of 6 μ m by using the GeneArray scanner (Hewlett-Packard, Palo Alto, Calif.). Probe array cell intensities, nucleotide base call, sequence determination, and reports were generated by functions available on GeneChip software (Affymetrix). A candidate selection index was determined by the percentage of homology between the experimentally derived sequence and all of the reference sequences tiled on the array (Table 1).

RESULTS

Optimization of the DNA-probe array hybridization. Experimental fragmentation and hybridization conditions for the probe array assay were optimized by using wild-type reference targets derived from *M. tuberculosis* 16S rRNA and from two genes involved in *M. tuberculosis* drug resistance, i.e., *rpoB* and *katG*. Hybridization of these targets to the 20-mer nucleotide probes on the array combined with the MnCl₂-imidazole fragmentation procedure and hybridization buffer 2 produced sequence base call accuracies ranging from 98.9 to 100% (Fig. 1). Our assay conditions also produced a high degree of specificity during the hybridization. We observed little cross-hybridization to probes on different regions of the array, despite the use of one set of experimental conditions for several targets hybridizing to thousands of different array sequences (Fig. 1B).

Identification of *Mycobacterium* species. We tested 70 mycobacterial isolates from 27 different species characterized by conventional phenotypic methods. For each isolate tested, results are expressed as percent homology to reference sequences tiled on the array as produced by the GeneChip software. Results for the best two scores are presented in Table 1. The highest score was taken as the identification result (range, 96.4 to 100%; mean, 98.97%). Over the approximately 169-nucleotide region analyzed for the 70 isolates, the highest homology scores had less than two discordant base calls on average. Discordant calls occurred all in a conserved region, so there was no effect on identification. Based on our results, all isolates from the following species were unambiguously identified by the probe array (i.e., the highest identification score being the one expected): *M. asiaticum*, *M. chelonae*, *M. celatum*, *M. conspicuum*, *M. cookii*, *M. fortuitum*, *M. flavescens*, *M. gordonae*, *M. interjectum*, *M. intermedium*, *M. malmoense*, *M. scrofulaceum*, *M. shimodei*, *M. simiae*, *M. smegmatis*, *M. sphagni*, *M. tuberculosis*, and *M. xenopi*. Also, 100% of MAC isolates were identified by the probe array as species belonging to the *M. avium* complex. However, we obtained minor discrepancies between the phenotypic and the genotypic identifications for two of the eight isolates tested. One of the four *M. avium* isolates was identified as *M. intracellulare* (score, 99.4%), and one of the four *M. intracellulare* isolates was identified as *M. avium/M. paratuberculosis* by the array (score, 97%).

Six species shared the highest reference selection with another species. *M. avium* and *M. paratuberculosis*, *M. kansasii* and *M. gastri*, *M. marinum* and *M. ulcerans*, *M. bovis* and *M. tuberculosis*, and *M. chelonae* and *M. abscessus* (observed for two of the four *M. chelonae* isolates) were not individually identified in the study. The identification of both strains was predictable since their respective 16S rRNA sequences are identical in the region tiled on the array. This problem can be solved by expanding the region interrogated by the array for the pair *M. avium* and *M. paratuberculosis* (data not shown). This solution, however, will not work for the other species.

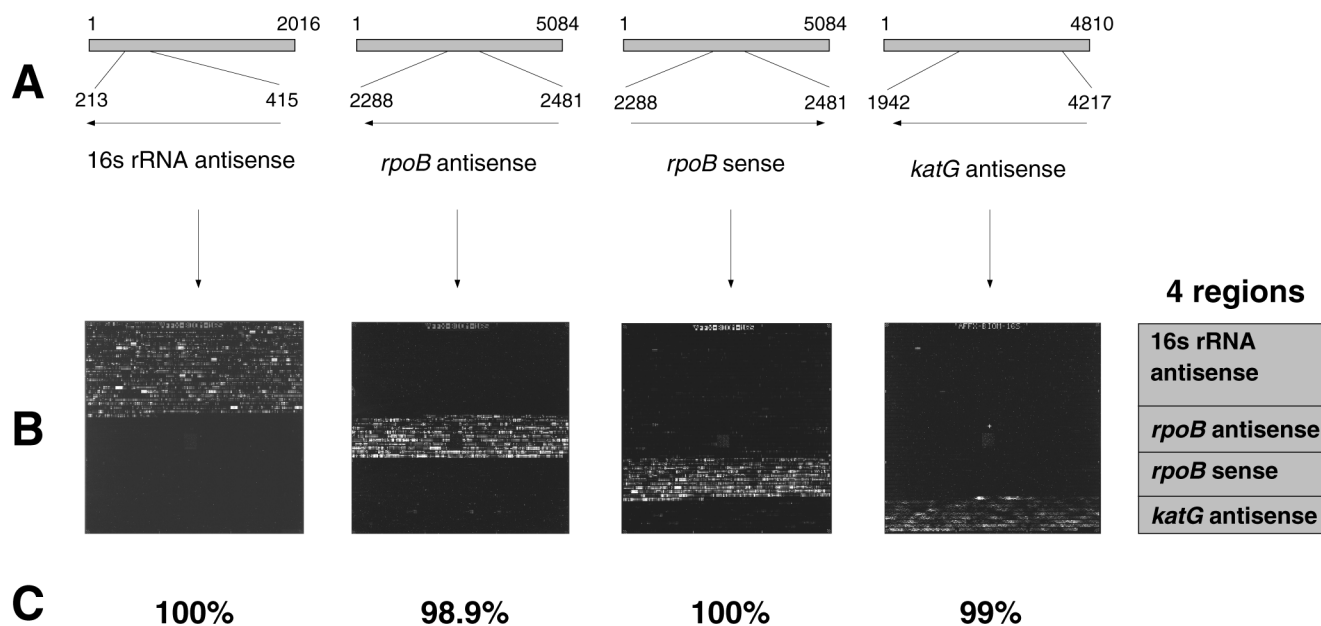


FIG. 1. The *Mycobacterium* probe array. The *Mycobacterium* probe array can be divided into four separate regions, with specific probes for the analysis of either 16S rRNA antisense, *rpoB* antisense, *rpoB* sense, or *katG* antisense targets. (A) Diagram of regions analyzed on the array. First line shows the *M. tuberculosis* 16S rRNA, *rpoB*, and *katG* reference sequences (GenBank accession no. M20940, L27989, and X68081, respectively) with the nucleotide numbers (1 to 2016 for 16S rRNA, 1 to 5084 for *rpoB*, and 1 to 4810 for *katG*). Second line shows the region of each reference sequence that is analyzed on the array (213 to 415 for 16S rRNA, 2288 to 2481 for *rpoB*, and 1942 to 4217 for *katG*). Arrows indicate whether antisense (←) or sense (→) RNA single-stranded targets are being used. (B) Hybridization experiments with the *M. tuberculosis* 16S rRNA antisense, *rpoB* antisense, *rpoB* sense, and *katG* antisense targets, described for panel A. The fluorescence images were obtained following hybridization of fluorescein-labeled fragmented RNA, generated by in vitro transcription from PCR amplicons, as described in Materials and Methods (use of hybridization buffer 2). (C) Results (percentages of correct base calling) are given for each experiment and represent performances routinely achieved with $MnCl_2$ -imidazole fragmentation protocol and hybridization buffer 2.

Two isolates did not have their respective reference sequences tiled on the array. *M. branderi* was identified as *Mycobacterium* sp. strain sp.6 and *M. lentiflavum* was identified as *Mycobacterium* sp. strain MCRO 8. The description of *M. branderi* as a new species includes the *Mycobacterium* sp. strain sp.6 isolate sequence as a reference (16). *M. lentiflavum* isolates have previously been described as being phylogenetically related to *M. simiae* (27) as well as the *Mycobacterium* sp. strain MCRO 8 isolate sequence (23). The identifications produced by probe array analysis are consistent with the available literature for the two isolates.

Table 1 also includes a blinded study. Twenty PCR amplicons were kindly provided by E. Böttger, and all were accurately identified with the probe array assay. This panel included a mix of pathogenic species (*M. tuberculosis*, MAC, *M. marinum*, *M. kansasii*, and *M. xenopi*) and nonpathogenic species.

Assuming proper identification for *M. branderi* and *M. lentiflavum*, the sequences of 67 of 70 isolates representing 27 species were identified correctly. Only one species, *M. szulgai*, was not correctly identified. All three *M. szulgai* isolates were diagnosed as *M. malmoense*. The reference sequences used to discriminate these two species are identical except at two locations, where a single base difference exists. At one location, an error in the *M. szulgai* reference sequence created probes that were identical to *M. malmoense* sequence (data not shown). This error resulted in a base call that incorrectly lowered the *M. szulgai* score while raising the *M. malmoense* score. The unique remaining polymorphism differentiating *M. malmoense* from *M. szulgai* was efficiently discriminated in all of the samples. The error in the reference sequence has been corrected in a second version of the probe array.

Detection of *M. tuberculosis rpoB* mutants. We tested 16 *rpoB* sequences, generated from 15 rifampin-resistant isolates, and one sensitive *M. tuberculosis* isolate. To study the influence of the strand polarity on base-calling accuracy, we hybridized *rpoB* sense and antisense transcripts on separate probe arrays. Mutations included substitutions (single-base and double-base) and deletions (three-base and six-base) (Table 2). All mutations were detected on both strands. Detection of the CAC→TAC point mutation responsible for the His526Tyr substitution is presented in Fig. 2. This mutation accounts for 30% of *M. tuberculosis* rifampin-resistant strains (12). Redundancy in the probes used in the sequence interrogation provides robustness for the test. The C→T change could be detected by using either the probes specific for the mutation (Fig. 2B) or the probes specific for the wild-type sequence (Fig. 2A). At all positions interrogated by mutant-specific probes, the mutant target hybridized, producing intensities that are roughly the same. However, intensities of the wild-type-specific probes are different for the same target. At the point mutation position, a single probe is perfectly complementary to the target and produces a relatively strong hybridization signal. At the neighboring positions, hybridization to the probes is reduced since there is a one-base mismatch with respect to the target sequence. The opposite experiment, hybridization of the *M. tuberculosis rpoB* wild-type target to the same probe tiles, is shown in Fig. 2C and D. Hybridization signals to the wild-type C probes are clearly observed.

The ability to detect a new point mutation not included in the array design was demonstrated for one isolate (G047). The wild-type probes detected the TTC→TTG mutation responsible for the Phe505Leu substitution (data not shown).

TABLE 1. *Mycobacterium* species identification

Isolate reference no. (ATCC-DSM) ^d	Phenotypic identification	Blind-testing status	Genetic identification (GeneChip results) ^a			
			Highest reference selection	Score	Second reference selection	Score
1	<i>M. tuberculosis</i>		<i>M. tuberculosis</i> (X52917)	98.2	<i>M. asiaticum</i>	94.6
2	<i>M. tuberculosis</i>	+	<i>M. tuberculosis</i> (X52917)	98.2	<i>M. asiaticum</i>	94.0
3	<i>M. tuberculosis</i>		<i>M. tuberculosis</i> (X52917)	98.8	<i>M. asiaticum</i>	94.6
4	<i>M. tuberculosis</i>	+	<i>M. tuberculosis</i> (X52917)	99.4	<i>M. asiaticum</i>	94.6
5 (27294) A	<i>M. tuberculosis</i>		<i>M. tuberculosis</i> (X52917)	98.2	<i>M. asiaticum</i>	95.8
6 (clinic)	<i>M. tuberculosis</i>		<i>M. tuberculosis</i> (X52917)	99.4	<i>M. asiaticum</i>	94.6
7 (clinic)	<i>M. tuberculosis</i>		<i>M. tuberculosis</i> (X52917)	100	<i>M. asiaticum</i>	94
1	<i>M. bovis</i>		<i>M. tuberculosis</i> (X52917)/ <i>M. bovis</i>	98.8 ^{b,c}	<i>M. asiaticum</i>	94.6
1	<i>M. avium</i>		<i>M. avium</i> (X52918)/ <i>M. paratuberculosis</i>	100 ^f	<i>M. intracellulare</i> 4	94.7
2	<i>M. avium</i>	+	<i>M. avium</i> (X52918)/ <i>M. paratuberculosis</i>	100 ^f	<i>M. intracellulare</i> 4	94.1
3 (43216) D	<i>M. avium</i>		<i>M. avium</i> (X52918)/ <i>M. paratuberculosis</i>	100 ^f	<i>M. intracellulare</i> 4	94.1
4 (29555) A	<i>M. avium</i>		<i>M. intracellulare</i> 4 (M61683)	99.4	<i>M. intracellulare</i> 1	98.2
1	<i>M. intracellulare</i>		<i>M. intracellulare</i> 1 (X52927)	99.4	Strain sp1	98.2
2	<i>M. intracellulare</i>	+	<i>M. intracellulare</i> 1 (X52927)	99.4	Strain sp1	98.2
3 (43223) D	<i>M. intracellulare</i>		<i>M. intracellulare</i> 1 (X52927)	100	Strain sp1	98.8
1 (35767) A	<i>M. intracellulare</i>		<i>M. avium</i> (X52918)/ <i>M. paratuberculosis</i>	97 ^c	<i>M. leprae</i>	94.2
1 (25276) A	<i>M. asiaticum</i>		<i>M. asiaticum</i> 2 (X55604)	98.2	<i>M. ulcerans</i> 1	97
2 (25276) A	<i>M. asiaticum</i>		<i>M. asiaticum</i> 2 (X55604)	100	<i>M. celatum</i> 2	97
1	<i>M. chelonae</i>		<i>M. chelonae</i> 1- <i>fortuitum</i> (X52921)	98.8	<i>M. fortuitum-chelonae</i>	97.6
2 (35752) A	<i>M. chelonae</i>		<i>M. chelonae</i> 7 (X82236)/ <i>M. abscessus</i>	99.4 ^c	<i>M. fortuitum</i> 1	94.1
3 (clinic)	<i>M. chelonae</i>		<i>M. chelonae</i> 1- <i>fortuitum</i> (X52921)	97	<i>M. fortuitum-chelonae</i>	94.7
4 (14472) A	<i>M. chelonae</i>		<i>M. chelonae</i> 7 (X82236)/ <i>M. abscessus</i>	97 ^c	<i>M. chelonae</i> 4	92.3
1	<i>M. fortuitum</i>		<i>M. fortuitum-chelonae</i> (X52933)	98.2	<i>M. chelonae</i> 4	97.6
2 (49403) A	<i>M. fortuitum</i>		<i>M. fortuitum</i> 1 (X65528)/ <i>M. senegalense</i>	98.2 ^c	<i>M. chelonae</i> 4	97
1	<i>M. gordonae</i>	+	<i>M. gordonae</i> (X52923)	97	<i>M. shimodei</i>	91.1
2	<i>M. gordonae</i>		<i>M. gordonae</i> (X52923)	97.7	Strain MCRO 6	92.3
3	<i>M. gordonae</i>		<i>M. gordonae</i> (X52923)	98.8	Strain sp. 6	92.4
4 (14470) A	<i>M. gordonae</i>		<i>M. gordonae</i> (X52923)	98.8	<i>M. asiaticum</i>	94
5 (clinic)	<i>M. gordonae</i>		<i>M. gordonae</i> (X52923)	98.8	<i>M. shimodei</i>	94.1
6 (clinic)	<i>M. gordonae</i>	+	<i>M. gordonae</i> (X52923)	97.6	Strain sp1	92.3
1 (43224) D	<i>M. kansasii</i>		<i>M. kansasii</i> (X15916)/ <i>M. gastri</i>	100 ^f	<i>M. simiae</i>	98.8
2 (12478) A	<i>M. kansasii</i>		<i>M. kansasii</i> (X15916)/ <i>M. gastri</i>	99.4 ^c	<i>M. haemophilum</i>	96.5
1	<i>M. smegmatis</i>		<i>M. smegmatis</i> (X52922)	97.7	Strain MCRO 17	95.3
1	<i>M. malmoense</i>	+	<i>M. malmoense</i> (X52930)	99.4	<i>M. simiae</i>	98.8
2	<i>M. malmoense</i>		<i>M. malmoense</i> (X52930)	100	<i>M. szulgai</i> 2	97.6
3	<i>M. malmoense</i>		<i>M. malmoense</i> (X52930)	100	<i>M. szulgai</i> 2	97.6
4 (29571) A	<i>M. malmoense</i>		<i>M. malmoense</i> (X52930)	100	<i>M. interjectum</i>	97
5 (29571) A	<i>M. malmoense</i>		<i>M. malmoense</i> (X52930)	100	<i>M. szulgai</i> 2	98.8
1	<i>M. shimodei</i>		<i>M. shimodei</i> (X82459)	98.8	Strain sp.6	97.7
2 (27962) A	<i>M. shimodei</i>		<i>M. shimodei</i> (X82459)	98.2	Strain sp.6	97.1
3 (clinic)	<i>M. shimodei</i>		<i>M. shimodei</i> (X82459)	100	Strain sp.6	97.1
1	<i>M. flavescens</i>		<i>M. flavescens</i> (X52932)	97.7	<i>M. smegmatis</i>	96.5
1	<i>M. simiae</i>		<i>M. simiae</i> (X52931)	100	<i>M. kansasii</i> /M. <i>gastri</i>	98.2 ^c
2 (25275) A	<i>M. simiae</i>		<i>M. simiae</i> (X52931)	99.4	<i>M. kansasii</i> /M. <i>gastri</i>	98.2 ^c
3 (25275) A	<i>M. simiae</i>		<i>M. simiae</i> (X52931)	99.4	<i>M. kansasii</i> /M. <i>gastri</i>	98.2 ^c
1	<i>M. szulgai</i>		<i>M. malmoense</i> (X52930)	99.4	<i>M. szulgai</i>	98.8
2 (clinic)	<i>M. szulgai</i>		<i>M. malmoense</i> (X52930)	99.4	<i>M. szulgai</i>	98.8
3 (35799) A	<i>M. szulgai</i>		<i>M. malmoense</i> (X52930)	98.8	<i>M. szulgai</i>	98.8
1	<i>M. cookii</i>		<i>M. cookii</i> (X53896)	96.4	<i>M. farcinogenes</i>	93.5
1	<i>M. intermedium</i>		<i>M. intermedium</i> (X67847)	100	<i>M. kansasii</i> /M. <i>gastri</i>	94.1 ^c
2	<i>M. intermedium</i>	+	<i>M. intermedium</i> (X67847)	100	<i>M. kansasii</i> /M. <i>gastri</i>	94.1 ^c
1	<i>M. marinum</i>		<i>M. marinum</i> (X52920)/ <i>M. ulcerans</i>	100 ^f	<i>M. asiaticum</i>	97
2	<i>M. marinum</i>	+	<i>M. marinum</i> (X52920)/ <i>M. ulcerans</i>	100 ^f	<i>M. asiaticum</i>	97.6
1	<i>M. ulcerans</i>	+	<i>M. ulcerans</i> (X58954)/ <i>M. marinum</i>	99.4 ^c	<i>M. asiaticum</i>	97.6
2	<i>M. ulcerans</i>	+	<i>M. ulcerans</i> (X58954)/ <i>M. marinum</i>	99.4 ^c	<i>M. asiaticum</i>	98.2
3	<i>M. ulcerans</i>		<i>M. ulcerans</i> (X58954)/ <i>M. marinum</i>	99.4 ^c	<i>M. asiaticum</i>	96.4
1	<i>M. celatum</i>		<i>M. celatum</i> 2 (L08169)	98.2	<i>M. asiaticum</i>	95.8
2	<i>M. celatum</i>	+	<i>M. celatum</i> 2 (L08169)	98.2	<i>M. asiaticum</i>	95.8
3 (51130) A	<i>M. celatum</i>		<i>M. celatum</i> 1 (L08170)	99.4	<i>M. celatum</i> 2	95.3
4 (51131) A	<i>M. celatum</i>		<i>M. celatum</i> 2 (L08169)	99.4	<i>M. celatum</i> 1	95.9
1	<i>M. conspicuum</i>	+	<i>M. conspicuum</i> (X88922)	100	<i>M. simiae</i>	97
1	<i>M. interjectum</i>	+	<i>M. interjectum</i> (X70961)	100	<i>M. haemophilum</i> 1	97.1
1	<i>M. branderi</i>	+	<i>Mycobacterium</i> sp. strain sp.6 (X82234)	98.8	<i>M. shimodei</i>	91.1
1	<i>M. scrofulaceum</i>	+	<i>M. scrofulaceum</i> (X52924)	99.4	<i>M. simiae</i>	98.8
2	<i>M. scrofulaceum</i>	+	<i>M. scrofulaceum</i> (X52924)	99.4	<i>M. simiae</i>	98.2
1	<i>M. xenopi</i>	+	<i>M. xenopi</i> (X52929)	97.1	<i>M. shimodei</i>	88.8
2 (19250) A	<i>M. xenopi</i>		<i>M. xenopi</i> (X52929)	98.3	<i>M. shimodei</i>	89.9
3 (clinic)	<i>M. xenopi</i>		<i>M. xenopi</i> (X52929)	99.4	<i>M. paratuberculosis</i>	89.3
1	<i>M. sphagni</i>	+	<i>M. sphagni</i> (X55590)	98.2	<i>M. chitae</i> 2	95.3
1	<i>M. lentiflavum</i>	+	<i>Mycobacterium</i> sp. strain MCRO 8 (X93034)	98.2	Strain MCRO 18	97.6

^a Targets have been generated from the various *Mycobacterium* isolates by in vitro transcription from PCR amplicons and fragmented by the MgCl₂ protocol. Hybridization experiments were performed with hybridization buffer 1 as described in Materials and Methods. Sequence giving the highest reference score is referenced both in terms of origin (EMBL accession number) and biological significance (species identification).

^b *M. bovis* isolates have the same 16S rRNA sequence as *M. tuberculosis* isolates.

^c Same base calling (percent) obtained for two reference sequences (see explanation in Results).

^d ATCC, American Type Culture Collection; DSM, Deutsche Sammlung von Mikroorganismen.

TABLE 2. *M. tuberculosis rpoB* mutant isolates analyzed with the *Mycobacterium* probe array

Clinical isolate	Codon	Nucleotide change	Amino acid substitution
G049	526	CAC→GAC	His→Asp
G055	526	CAC→CGC	His→Arg
G051	526	CAC→TAC	His→Tyr
J3919	531	TCG→TTG	Ser→Leu
J37	513	CAA→CTA	Gln→Leu
	515	ATG→GTG	Met→Val
J499	513	CAA→GAA	Gln→Glu
J535	533	CTG→CCG	Leu→Pro
G054	531	TCG→TGG	Ser→Trp
J80	516-517	Del ^a GACCAG	Del Asp-Gln
J978	526	CAC→CTC	His→Leu
G052	511	CTG→CCG	Leu→Pro
G048	513	CAA→CTA	Gln→Leu
G050	516	GAC→TAC	Asp→Tyr
G047	505	TTC→TTG	Phe→Leu
	511	CTG→CCG	Leu→Pro
	531	TCG→TGT	Ser→Cys
G053	518	Del AAC	Del Asn

^a Del, deletion.

DISCUSSION

The work described in this study represents one of the first applications of high-density DNA probe arrays for bacteriology, focusing on fastidious bacterium diagnostics. The technology allowed us to design an array that contained all of the 16S rRNA polymorphisms over a 200-bp region present in a my-

cobacterial database. The array also contained 51 *M. tuberculosis* rifampin resistance-causing *rpoB* mutations in a 200-bp region and 2.2 kb of the *M. tuberculosis* wild-type *katG* gene, this fragment spanning codons where mutations were previously shown to confer resistance to isoniazid (9, 31). An experimental protocol was optimized to perform hybridization on the probe array, scanning, and data analysis in 1 h. The total process from culture, including sample preparation and amplification, takes less than 4 h manually. An additional objective of this study was to evaluate base-calling accuracy by using a single strand of target. Previous studies have combined the information from both strands (3, 6, 17). We found that base-calling accuracies with one target strand ranged routinely from 96.4 to 100%, even though the 16S, *rpoB*, and *katG* targets have a high G+C content (>60%) and contain potentially highly stable secondary structures. Moreover, cotesting of independently generated 16S rRNA and *rpoB* amplicons on the same array provided the same base-calling accuracy as did each target alone (data not shown).

The sequencing of the 16S rRNA gene has been utilized by numerous investigators as a means of discrimination among mycobacterial taxa (27). The hybridization-based probe array assay used here showed a specificity matching the sequence resolution (polymorphisms) of the 16S rRNA marker. In this study, the sequences for 26 of 27 species were correctly identified, with the only discrepancy being an artifact due to an error in the reference sequence. Closely related sequences specific for species of very different clinical importance were clearly differentiated (i.e., *M. tuberculosis* complex species from *M. celatum*, *M. marinum*, *M. asiaticum*, and *M. terrae*). Newly described species for which sequences were not tiled on the

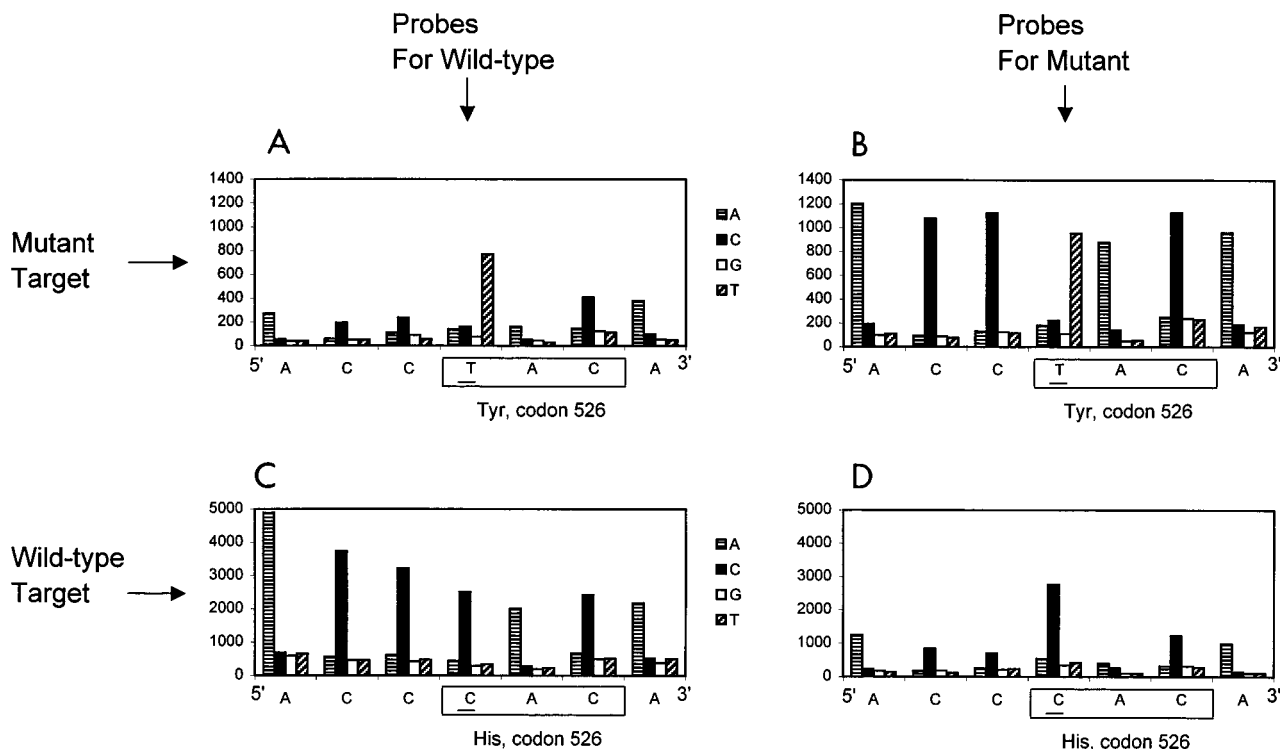


FIG. 2. *M. tuberculosis rpoB* mutation detection with the *Mycobacterium* probe array. Labeled *rpoB* targets (sense strand) originating from either a His526Tyr (CAC→TAC) *M. tuberculosis rpoB* mutant isolate (A and B) or a wild-type *M. tuberculosis rpoB* sequence (C and D) have been hybridized on the *Mycobacterium* probe array. Results (signal intensities at each interrogated position for the four different probes) have been obtained by using either the probes specific for the His526Tyr mutant sequence (B and D) or the probes specific for the wild-type sequence (A and C). The signal intensities are shown on the y axis of each panel.

array (*M. lentiflavum* and *M. branderi*) were identified as subspecies variants or closely related species. Recent genetic studies have demonstrated a higher degree of intraspecies sequence diversity in the 16S rRNA locus than was previously believed (8, 23). Given that the assay produces sequence information, a phylogenetic module could be added to the analysis to improve the accuracy of identifying polymorphic variants and strains whose sequences are not exactly represented on the array.

The ability of accurately detecting sequence variation is especially important when the clinical interpretation (i.e., in vitro resistance or sensitive status) depends on the discrimination of a single point mutation as seen in *rpoB*-mediated rifampin resistance in *M. tuberculosis*. All tested mutation types were detected by the *Mycobacteria* probe array, and these include single-base substitutions, double-base substitutions, deletions (three-base and six-base), and insertions (data not shown). The detection of these mutations was also independent of the strand used for hybridization. The ability to place many mutant sequences on the array may have its advantages. A commercially available rifampin resistance test (Inno-LiPA Rif.TB; Innogenetics, Zwijndrecht, Belgium) was reportedly unable to detect a three-base insertion mutation, presumably because the sequence was not present on the strip (5). Interestingly, we have also demonstrated the ability of the probe array to detect new point mutations, neither previously described nor tiled on the array, by simply using the probes defined for the wild-type sequence. Thus, the efficiency of our probe array strategy should not be diminished by the occurrence of a new point mutation or polymorphisms not on the array. This strategy is being advanced by examining *katG* sequences in isoniazid-resistant *M. tuberculosis* isolates to identify new mutations (data not shown). In the perspective of testing cultures or direct clinical samples with this molecular approach, we need to position it with growth-based interpretation with more extended studies so that the inference of in vivo resistance status remains clinically pertinent.

Simple, ready-to-use probe assays are commercially available for *Mycobacterium* identification (Accu-Probe and AMTD, Gen-Probe; Amplicor MTB; Roche Diagnostics Systems; SHARP Signal; Digene Diagnostics Inc., Silver Spring, Md.). However, they are applicable to isolates of the most common disease-causing mycobacterial species (*M. tuberculosis* and MAC strains). A mycobacterial probe array can expand the existing platform for identification of mycobacterial species and be used to perform complete strain-specific typing of clinical isolates. For example, the probe array could be used, in conjunction with analysis of clinical data, for determining the true incidence of infections due to environmental mycobacteria.

Today, the number of MOTT infections is difficult to assess because there is no system for notification as exists for *M. tuberculosis*. The current report frequency of these species is likely to be underestimated due to the lack of additional testing in cases of minimal disease or misidentification as *M. tuberculosis* (11). Despite the likely underestimation of MOTT disease, a growing number of MOTT isolates are submitted to laboratories for identification. This may reflect an increase in the prevalence of opportunistic mycobacterial disease (notably in the AIDS context), or it may also reflect an increase in the number and nature of specimens submitted for culture brought about by a greater awareness of tuberculosis. The increased use of endoscopy for diagnostic purposes and changes in environmental factors affecting the nature, number, and distribution of MOTT species in the environment, including piped water supplies (30), certainly have also contributed to this

increase. A single assay that quickly identifies *M. tuberculosis*, MAC, and other mycobacterial species will clearly aid in documenting their apparent rise in disease states and possibly in new environments.

The potential of the GeneChip probe array strategy for parallel testing of different targets has been demonstrated in this study, as the same hybridization conditions could be used for the three genes tiled on the mycobacterial array (16S rRNA, *rpoB*, and *katG*). The platform described here can be expanded to other *M. tuberculosis* drug resistance determinants since they are being gradually understood. For example, probe array-based mutation analysis of the catalase-peroxidase gene (*katG*) (31), the promoter region of the *inhA* and *ahpC* genes (1, 29), the recently described *kasA* protein (19) for isoniazid resistance, pyrazinamidase-nicotinamidase gene (*pncA*) for pyrazinamide resistance (22, 24), and the *emb* operon for ethambutol resistance (25) could all be done to monitor drug resistance simultaneously. Recently, a probe array interrogating most of these genes has been designed and is currently being used (8). Epidemiological markers could also be added to the array for tracing epidemic or sporadic dissemination of strains. Finally, the potential to perform direct testing on samples providing species identification results (15) and drug resistance genotyping (28) represents the next step in the application of this technology to clinical diagnostics.

ACKNOWLEDGMENTS

We thank E. Böttger for the generous gifts of *Mycobacterium* samples and for thoughtful discussions; S. Cole for the gift of *rpoB* clones; M. Mittmann for his assistance with lithographic mask design; D. Wu, E. Wang, and B. Lacroix for software design and development; J. Drenkow for preparation of the *rpoB* mutant targets; D. Do for technical assistance; A. Lau for graphics preparation; and C. Rogers for his support for this project.

REFERENCES

- Banerjee, A., E. Dubnau, A. Quemard, V. Balasubramanian, S. K. Um, T. Wilson, D. Collins, G. De Lisle, and W. R. Jacobs. 1994. *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* **263**:227-230.
- Bödinghaus, B., T. Rogall, T. Flohr, H. Blöcker, and E. C. Böttger. 1990. Detection and identification of mycobacteria by amplification of rRNA. *J. Clin. Microbiol.* **28**:1751-1759.
- Chee, M., R. Yang, E. Hubbell, A. Berno, X. C. Huang, D. Stern, J. Winkler, D. J. Lockhart, M. S. Morris, and S. P. A. Fodor. 1996. Accessing genetic information with high-density DNA arrays. *Science* **274**:610-614.
- Cole, S. T. 1994. *Mycobacterium tuberculosis*: drug-resistance mechanisms. *Trends Microbiol.* **2**:411-415.
- Cooksey, R. C., G. P. Morlock, S. Glickman, and J. T. Crawford. 1997. Evaluation of a Line Probe assay kit for characterization of *rpoB* mutations in rifampin-resistant *Mycobacterium tuberculosis* isolates from New York City. *J. Clin. Microbiol.* **35**:1281-1283.
- Cronin, M. T., R. V. Fucini, S. M. Kim, R. S. Masino, R. M. Wespi, and C. G. Miyada. 1996. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum. Mutat.* **7**:244-255.
- Fodor, S. P. A., R. P. Rava, X. C. Huang, A. C. Pease, C. P. Holmes, and C. L. Adams. 1993. Multiplexed biochemical assays with biological chips. *Nature* **364**:555-556.
- Gingeras, T. R., G. Ghandour, E. Wang, A. Berno, P. M. Small, F. Drobniowski, D. Alland, E. Desmond, M. Holodniy, and J. Drenkow. 1998. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res.* **8**:435-448.
- Heym, B., P. M. Alazari, N. Honoré, and S. T. Cole. 1995. Missense mutations in catalase-peroxidase gene, *katG*, are associated with isoniazid-resistance in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **15**:235-245.
- Horsburgh, C. R. 1991. *Mycobacterium avium* complex infection in the acquired immunodeficiency syndrome. *N. Engl. J. Med.* **324**:1332-1338.
- Jenkins, P. A. 1991. Mycobacteria in the environment. *J. Appl. Bacteriol. Symp.* **70**(Suppl.):137S-141S.
- Kapur, V., L. L. Li, S. Iordanescu, M. R. Hamrik, A. Wanger, B. N. Kreiswirth, and J. M. Musser. 1994. Characterization by automated DNA sequencing of mutations in the gene (*rpoB*) encoding the RNA polymerase

- β subunit in rifampin-resistant *Mycobacterium tuberculosis* strains from New York City and Texas. J. Clin. Microbiol. **32**:1095–1098.
13. **Kent, P. T., and G. P. Kubica.** 1985. Public health mycobacteriology: a guide for the level III laboratory. U.S. Department of Health and Human Services publication no 86-8230. U.S. Department of Health and Human Services, Washington, D.C.
 14. **Kirschner, P., B. Springer, U. Vogel, A. Meier, A. Wrede, M. Kiekenbeck, F. C. Bange, and E. C. Böttger.** 1993. Genotypic identification of mycobacteria by nucleic acids sequence determination: report of a 2-year experience in a clinical laboratory. J. Clin. Microbiol. **31**:2882–2889.
 15. **Kirschner, P., J. Rosenau, B. Springer, K. Teschner, K. Feldmann, and E. C. Böttger.** 1996. Diagnosis of mycobacterial infections by nucleic acid amplification: 18-month prospective study. J. Clin. Microbiol. **34**:304–312.
 16. **Koukila-Kahkola, P., B. Springer, E. C. Böttger, L. Paulin, E. Jantzen, and M. L. Katila.** 1995. *Mycobacterium branderi* sp. nov., a new potential human pathogen. Int. J. Syst. Bacteriol. **45**:549–553.
 17. **Kozal, M. J., N. Shah, N. Shen, R. Yang, R. Fucini, T. C. Merigan, D. D. Richman, D. Morris, E. Hubbell, M. Chee, and T. R. Gingeras.** 1996. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat. Med. **2**:753–759.
 18. **Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittman, C. Wang, M. Kobayashi, H. Horton, and E. L. Brown.** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. **14**:1675–1680.
 19. **Mdluli, K., R. A. Slayden, Y. Zhu, S. Ramaswamy, X. Pan, D. Mead, D. D. Crane, J. M. Musser, and C. E. Barry III.** 1998. Inhibition of a *Mycobacterium tuberculosis* β -ketoacyl ACP synthase by isoniazid. Science **280**:1607–1610.
 20. **Raviglione, M. C., D. E. Snider, and A. Kochi.** 1995. Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. JAMA **273**: 220–226.
 21. **Roberts, G. D., E. W. Koneman, and Y. K. Kim.** 1991. *Mycobacterium*, p. 304–339. In A. Balows, W. J. Hausler, Jr., K. L. Herrmann, H. D. Isenberg, and H. J. Shadomy (ed.), Manual of clinical microbiology, 5th ed. American Society for Microbiology, Washington, D.C.
 22. **Scorpio, A., and Y. Zhang.** 1996. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculosis drug pyrazinamide in tubercle bacillus. Nat. Med. **2**:662–667.
 23. **Springer, B., L. Stockman, K. Teschner, G. D. Roberts, and E. C. Böttger.** 1996. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. J. Clin. Microbiol. **34**:296–303.
 24. **Sreevatsan, S., X. Pan, Y. Zhang, B. N. Kreiswirth, and J. M. Musser.** 1997. Mutations associated with pyrazinamide resistance in *pncA* of *Mycobacterium tuberculosis* complex organisms. Antimicrob. Agents Chemother. **41**: 636–640.
 25. **Sreevatsan, S., K. E. Stockbauer, X. Pan, B. N. Kreiswirth, S. L. Moghazeh, W. R. Jacobs, Jr., A. Telenti, and J. M. Musser.** 1997. Ethambutol resistance in *Mycobacterium tuberculosis*: critical role of *embB* mutations. Antimicrob. Agents Chemother. **41**:1677–1681.
 26. **Telenti, A., P. Imboden, F. Marchesi, D. Lowrie, S. T. Cole, M. J. Colston, L. Matter, K. Schopfer, and T. Bodmer.** 1993. Detection of rifampin-resistance mutations in *Mycobacterium tuberculosis*. Lancet **341**:647–650.
 27. **Tortoli, E., C. Piersimoni, P. Kirschner, A. Bartoloni, C. Burrini, C. Laccini, A. Mantella, G. Muzzi, C. Passerini Tosi, V. Penati, C. Scarparo, M. Tullia Simonetti, and E. C. Böttger.** 1997. Characterization of mycobacterial isolates phylogenetically related to, but different from, *Mycobacterium simiae*. J. Clin. Microbiol. **35**:697–702.
 28. **Whelen, A. C., T. A. Felmlee, J. M. Hunt, D. L. Williams, G. D. Roberts, L. Stockman, and D. H. Persing.** 1995. Direct genotypic detection of *Mycobacterium tuberculosis* rifampin resistance in clinical specimens by using single-tube heminested PCR. J. Clin. Microbiol. **33**:556–561.
 29. **Wilson, T. M., and D. M. Collins.** 1996. *ahpC*, a gene involved in isoniazid resistance of the *Mycobacterium tuberculosis* complex. Mol. Microbiol. **19**: 1025–1034.
 30. **Yates, M. D., A. Pozniak, A. H. C. Uttley, R. Clarke, and J. M. Grange.** 1997. Isolation of environmental mycobacteria from clinical specimens in south-east England: 1973–1993. Int. J. Tuberc. Lung Dis. **1**:75–80.
 31. **Zhang, Y., B. Heym, B. Allen, D. Young, and S. T. Cole.** 1992. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature **358**:501–503.