CLINICAL AND DIAGNOSTIC LABORATORY IMMUNOLOGY, Nov. 1995, p. 770–775 Vol. 2, No. 6
1071-412X/95/$04.00+0
Copyright © 1995, American Society for Microbiology

Comparison of PCR-Generated Fragments of the *mce* Gene from *Mycobacterium tuberculosis*, *M. avium*, *M. intracellulare*, and *M. scrofulaceum*

STEPHEN L. PARKER, YU-LI TSAI, AND CAROL J. PALMER*

Environmental Sciences Laboratory, County Sanitation Districts of Orange County, Fountain Valley, California 92728-8127

Received 10 February 1995/Returned for modification 29 March 1995/Accepted 16 August 1995

Nineteen strains representing 13 species of mycobacteria were tested for the ability to serve as PCR templates for the production of a 293-bp fragment of the mycobacterial *mce* gene. The *mce* gene is a virulence factor recently sequenced from *Mycobacterium tuberculosis*. PCR products were obtained for only the species of the *Mycobacterium tuberculosis* complex (MTC) and the *Mycobacterium avium-M. intracellulare-M. scrofulaceum* complex. The fragment was sequenced from *M. tuberculosis* (one strain), *M. avium* (three strains), *M. intracellulare* (two strains), and *M. scrofulaceum* (two strains). Sequence comparisons suggest that the fragments from each of the species are regions that code for a similar product. One of the *M. scrofulaceum* strains yielded a sequence whose most probable reading frame was truncated by an amber stop codon caused by a single nucleic acid difference from the other sequences. The amino acid sequences from the non-MTC sequences cluster together, displaying fewer differences from each other than from the *M. tuberculosis* sequence and the truncated *M. scrofulaceum* sequence. Principal component analysis of the distance matrix displays the clustering of the *M. avium* and *M. intracellulare* sequences into single-species clusters. It is concluded that at least one open reading frame of the *mce* gene is found, although it is discernibly different, in pathogenic mycobacteria other than the MTC.

The history of mycobacterial disease can be traced back over a thousand years (16), and currently one third of the world's population is suspected of harboring *Mycobacterium tuberculosis* (2). The *Mycobacterium tuberculosis* complex (MTC) appears to be the most aggressive public health threat, but many infections are caused by members of the *Mycobacterium avium-M. intracellulare-M. scrofulaceum* complex. *M. avium-M. intracellulare-M. scrofulaceum* complex infections are particularly devastating in immunocompromised individuals and have contributed significantly to the high morbidity and mortality rates in the human immunodeficiency virus-infected population.

The transmission of disease by the MTC is primarily considered to occur from person to person, although the MTC has been isolated from environmental sources (13). The *M. avium-M. intracellulare-M. scrofulaceum* complex, however, is considered to be transmitted by various environmental sources, especially fresh- or ocean waters (5, 8, 9, 18). Risk assessments of potential environmental sources of mycobacterial infections have been hampered by the extremely slow growth rates of the *M. avium-M. intracellulare-M. scrofulaceum* complex and MTC. In order to address public health concerns over environmental sources of pathogenic mycobacteria, it is necessary to develop a more rapid method of assessing the prevalence of these organisms in environmental samples.

In recent years, there have been significant advances in the use of molecular techniques to rapidly detect mycobacterial pathogens in clinical samples. Several of these tests have displayed cross-reactivity between certain mycobacterial species (4, 6). It has also been observed that the presence of a large proportion of nontarget organisms in a sample can significantly decrease sensitivity (11). It is important, therefore, to continue examining mycobacterial genomes for genetic elements which may provide more specific and sensitive targets for molecular probes.

Genetic elements specific to virulence of the target species would constitute attractive probe targets. The presence of virulence factors can help establish the disease potential of environmental and clinical samples much more effectively than the presence of specific rRNA sequences. Investigations designed on this principle have enjoyed considerable success (14, 15).

The genetic elements responsible for the virulence of mycobacteria are not yet well established. Many virulence genes of various levels of complexity are probably found in the MTC. One virulence gene, the *mce* gene, has been identified as a 1,535-bp piece of *M. tuberculosis* genomic DNA able to confer the ability to invade HeLa cells and macrophages to *Escherichia coli* cells transformed with plasmids containing it (1). The *mce* gene sequence contains two open reading frames. The first (ORF1) has been shown to contain the element responsible for attachment and entry into mammalian cells; the second was shown to confer the ability to survive inside cells. Although the in vivo importance of the *mce* gene needs to be confirmed through animal experiments, it holds great promise as a template for probe development.

In this study, a portion of the *mce* gene ORF1 is identified for the first time in pathogenic mycobacteria outside of the MTC. The identification of this genetic element in other mycobacteria, especially the *M. avium-M. intracellulare-M. scrofulaceum* complex, may help pave the way for the design of genetic probes for use with both clinical and environmental samples.

The mycobacterial strains used in this study are listed in Table 1, along with their original sources as identified by the

* Corresponding author. Mailing address: Environmental Sciences Laboratory, County Sanitation Districts of Orange County, P.O. Box 8127, 10844 Ellis Ave., Fountain Valley, CA 92728-8127. Phone: (714) 962-2411. Fax: (714) 962-2591.

TABLE 1. Mycobacterial strains

Strain	Original source	ATCC no.	PCR result[a]
M. avium J. P. Tigpen 6	Opossum	ATCC 29555	SS
M. avium MAIS 2	Chicken	ATCC 35712	SS
M. avium MAIS 1	Human sputum	ATCC 35717	SS
M. bovis BCG	Vaccine strain	ATCC 27289	U
Fast-growing mycobacterium isolate	Municipal wastewater effluent		N
M. gordonae P-15	Gastric lavage	ATCC 14470	N
M. intracellulare type strain	Laboratory	ATCC 13950	SS
M. intracellulare MAIS 5	Avian	ATCC 35768	SS
M. kansasii G133	Fatal case	ATCC 12478	N
M. lepraemurium	Animal model of leprosy	ATCC 35779	N
M. marinum	Fish	ATCC 927	N
M. microti LV 285	Laboratory	ATCC 19422	U
M. microti OV 183	Vole	ATCC 11152	U
M. nonchromogenicum	Soil	ATCC 19530	N
M. scrofulaceum type strain	Cervical lymph node	ATCC 19981	SS
M. scrofulaceum MAIS 27	Human sputum and lung	ATCC 35785	SS
M. terrae	Laboratory	ATCC 15755	N
M. tuberculosis H37Ra	Laboratory	ATCC 25177	US
M. tuberculosis R1Rv	Human lung (attenuated)	ATCC 35818	U

[a] N, no PCR product observed; U, PCR product observed but insufficient yield to sequence from a single PCR; SS, PCR product sequenced from a single PCR; US, low-yield PCR product combined from several reactions and sequenced.

American Type Culture Collection (ATCC), Rockville, Md. Mycobacterial strains were cultivated on slants of Lowenstein-Jensen medium (Difco, Detroit, Mich.). The slants were incubated at 36°C except for those inoculated with *M. marinum*, which were incubated at room temperature. Cultures were incubated until visible growth was obtained.

DNA was extracted by a modification of the method of Bose et al. (3). A rice-grain-sized loopful of cell mass was suspended in 600 μl of lysis inactivation buffer (5 M guanidine thiocyanate, 0.5% Sarkosyl, 0.5% sodium dodecyl sulfate, 0.85% benzalkonium chloride, 100 mM EDTA) and heated to 99°C for 1 h. Insoluble debris was removed by centrifugation (16,000 × *g* for 3 min), and the supernatant (approximately 600 μl) was mixed with an equal volume of 100% isopropyl alcohol. The DNA was precipitated during a 30-min room temperature incubation and then collected by centrifugation (16,000 × *g* for 15 min). The pellet was washed in 75% isopropyl alcohol, dried under a vacuum, and resuspended in 20 μl of autoclaved high-pressure liquid chromatography water (Baker). Two microliters of this solution reliably yields more than 1 ng of template-quality DNA as determined spectrophotometrically.

The oligonucleotide primers used in these studies were developed from an *M. intracellulare* sequence, using Oligo primer development software version 4.1 (National Biosciences Inc., Plymouth, Minn.). The primers are MMSPU (5′-GCCGAG AAGGTGGAT-3′) and MMSPD (5′-GCCGCCGACAACA AC-3′), which correspond to bp 343 to 357 and 621 to 635, respectively, on the *mce* gene sequence for *M. tuberculosis* H37Ra (L. W. Riley; GenBank accession number, X70901). Oligonucleotide primers were synthesized on an automated DNA synthesizer (model 392; Applied Biosystems Inc., Foster City, Calif.) according to the manufacturer's instructions.

PCR was performed by using a model 9600 thermocycler (Perkin-Elmer, Norwalk, Conn.). The PCR mix consisted of final concentrations of 200 μM for each nucleotide, 1.5 mM for $MgCl_2$, 0.3 μM for each primer, 10 mM Tris-HCl (pH 8.3), and 50 mM KCl. Two microliters of template DNA produced as described above was used for 100 μl of reaction mix; this reliably results in less than 1 μg of template per reaction. AmpliTaq (5 U; Perkin-Elmer) was used to catalyze each 100-μl reaction mixture. The PCR temperature cycle for amplification was as follows: denaturation for 2 min at 94°C; two cycles at 98°C for 20 s, 62°C for 30 s, and 72°C for 45 s; 40 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min; the mixture was then held at 72°C for an additional 7 min and cooled to 4°C until analysis. The PCR product was electrophoretically separated in a 3% Bio-Gel low-melting-point agarose gel (Bio 101, San Diego, Calif.) containing 0.9 μg of ethidium bromide per ml. Fragments with sizes of approximately 293 bp were visualized under UV light (302 nm) and excised from the gel with a razor blade. The fragments were then purified from the gel pieces by using a Mermaid oligomer purification kit (Bio 101) according to the manufacturer's instructions.

DNA sequencing was performed in both directions on 50 nmol of purified PCR product, using MMSPU and MMSPD as the primers. A fluorescent-dye-linked dideoxy terminator single-tube mix (PRISM; Applied Biosystems Inc.) was used for the reaction mixture. Cycle sequencing was performed according to the manufacturer's instructions in a PCR thermal cycler (model 9600; Perkin-Elmer). Dye terminators were removed from the sequencing reaction mixture by two phenol-chloroform extractions. The sequencing reaction mixtures were analyzed on an automated DNA sequencer (model 373; Applied Biosystems Inc.).

The partial and inversely complemented sequences were aligned, and the automated base calling was checked by using the SeqEd analysis program (Applied Biosystems Inc.). Bases were assigned to positions only after three separate sequencing reactions produced the same base call. Sequences were exported as ASCII text files to the PC-Gene analysis suite (IntelliGenetics Inc., Mountain View, Calif.), where they were aligned by using the CLUSTAL algorithm (10). PC-Gene was also used to translate the sequences to amino acids, which were also aligned by CLUSTAL. The CLUSTAL alignment uses the unweighted pair-group method, which can also generate a dendrogram. A dendrogram was assembled for the amino acid sequences. The aligned nucleic acid sequences were converted to a coded matrix, using a word processor program. The code replaced A with 1, C with 2, G with 3, and T with 4. This matrix was imported into the NTSYS-pc numerical taxonomy and multivariate analysis system (Exeter Software, Setauket, N.Y.). Using NTSYS-pc, a distance matrix was calculated by the

M. tuberculosis	CCG	GTC	AAG	CTG	AAC	CTG	ACC	CTG	AGC	GCG	GCC	GCG	GAG	GCG	TTG	ACC	GGG	CTG		54
25177	P	V	K	L	N	L	T	L	S	A	A	A	E	A	L	T	G	L		18
M. avium	...	...	...	...	..T	T..	..G	T..	...	...	...	...	C..	T..	...	T.G	...	...		54
35717	.	.	.	.	.	.	.	.	.	.	.	.	Q	S	.	S	.	.		18
M. avium	...	...	...	...	..T	...	...	T..	...	...	..G	...	C..	T..	...	T.G	...	...		54
35712	.	.	.	.	.	.	.	.	.	.	.	.	Q	S	.	S	.	.		18
M. avium	...	...	...	...	..T	...	..G	T..	..T	...	..G	...	C..	T..	C..	T.G	...	...		54
29555	.	.	.	.	.	.	.	.	.	.	.	.	Q	S	.	S	.	.		18
M. scrofulaceum	...	...	...	...	..T	T..	..G	T..	...	...	..G	...	C..	T..	...	..G	...	T..		54
19981	.	.	.	.	.	.	.	.	.	.	.	.	Q	S	.	.	.	.		18
M. intracellulare	...	...	...	T..	..T	...	..G	T..	...	...	..G	..A	C..	T..	...	G..	...	T..		54
13950	.	.	.	.	.	.	.	.	.	.	.	.	Q	S	.	A	.	.		18
M. intracellulare	...	...	...	T..	..T	...	..G	T..	...	...	..G	...	C..	T..	...	G..	...	T..		54
35768	.	.	.	.	.	.	.	.	.	.	.	.	Q	S	.	A	.	.		18
M. scrofulaceum	...	...	...	...	...	...	..G	...	...	...	...	..C	C..	...	C..	...	...	T..		54
35785	.	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	.	.		18

M. tuberculosis	GGC	GAT	AAG	TTC	GGC	GAG	TCG	ATC	GTC	AAC	GCC	AAC	ACC	GTT	CTG	GAT	GAC	CTC		109
25177	G	D	K	F	G	E	S	I	V	N	A	N	T	V	L	D	D	L		36
M. avium	..T	..G	...	...	...	C..	...	G.G	...	...	..G	..T	G.T	T.G	..C	...	...	G.G		109
35717	.	E	.	.	.	Q	.	V	.	.	.	.	A	L	.	.	.	V		36
M. avium	...	..G	...	...	...	C..	...	G.G	...	...	..G	..T	G.G	T.G	..C	...	...	G.G		109
35712	.	E	.	.	.	Q	.	V	.	.	.	.	A	L	.	.	.	V		36
M. avium	..G	..G	...	...	...	C..	...	G.G	...	...	...	...	.T.	T.G	..C	...	...	G.G		109
29555	.	E	.	.	.	Q	.	V	.	.	.	.	I	L	.	.	.	V		36
M. scrofulaceum	..G	..G	...	...	..T	C..	...	G.G	AC.	...	..T	...	G..	T.G	..T	...	...	G.G		109
19981	.	E	.	.	.	Q	.	V	T	.	.	.	A	L	.	.	.	V		36
M. intracellulare	..G	..C	.G.	...	..A	C..	...	G.G	..G	..T	..G	...	G.G	...	T..	...	..G	G.G		109
13950	.	.	R	.	.	Q	.	V	.	.	.	.	A	.	.	.	E	V		36
M. intracellulare	..G	..C	.G.	...	..G	C..	...	G.G	..G	..T	A.G	...	G.G	...	T..	...	..G	G.G		109
35768	.	.	R	.	.	Q	.	V	.	.	T	.	A	.	.	.	E	V		36
M. scrofulaceum	..T	..C	...	..G	...	C..	...	G.G	...	...	...	...	G.G	A.C	..C	..C	...	G..		109
35785	.	.	.	L	.	Q	.	V	.	.	.	.	A	I	.	.	.	V		36

M. tuberculosis	AAT	TCG	CGG	ATG	CCG	CAG	TCG	CGC	CAC	GAC	ATT	CAG	CAA	TTG	GCG	GCT	CTG	GGC		164
25177	N	S	R	M	P	Q	S	R	H	D	I	Q	Q	L	A	A	L	G		54
M. avium	..C	C.A	..A	...	...	...	G..	...	A.G	..T	...	...	GGG	...	...	..G	...	..G		164
35717	.	P	Q	.	.	.	A	.	K	.	.	.	G	.	.	.	.	.		54
M. avium	..C	C.A	..A	...	...	...	G..	...	A.G	...	...	...	GGG	...	...	..G	..T	...		164
35712	.	P	Q	.	.	.	A	.	K	.	.	.	G	.	.	.	.	.		54
M. avium	..C	C.C	...	...	...	...	G..	...	A.G	..T	...	...	..G	...	...	..G	T..	..T		164
29555	.	P	.	.	.	.	A	.	K	.	.	.	.	.	.	.	.	.		54
M. scrofulaceum	...	C..	.A.	C..	...	...	G..	...	A.G	..T	...	GC.	.GG	...	...	..G	...	..G		164
19981	.	P	Q	L	.	.	A	.	K	.	.	A	R	.	.	.	.	.		54
M. intracellulare	...	C..	...	...	...	...	G..	...	A.G	...	...	...	.GG	...	...	..G	...	.C.		164
13950	.	P	.	.	.	.	A	.	K	.	.	.	R	.	.	.	.	A		54
M. intracellulare	...	C..	...	...	...	...	G..	...	A.G	...	...	...	.GG	...	...	..G	...	...		164
35768	.	P	.	.	.	.	A	.	K	.	.	.	R	.	.	.	.	.		54
M. scrofulaceum	..C	C..	.A.	...	..C	T..	G..	...	...	...	...	...	..G	C..	..C	.GG	...	...		164
35785	.	P	Q	.	.	amb	A	.	.	.	.	.	.	.	.	G	.	.		53

M. tuberculosis	GAC	GTC	TAC	GCC	GAC	GCG	GCG	CCG	GAC	CTG	TTC	GAC	TTT	CTC	GAC	AGT	TCG	GTG	219
25177	D	V	Y	A	D	A	A	P	D	L	F	D	F	L	D	S	S	V	72
M. avium	...	AC.	..T	..G	...	...	C	T.A	...	...	...	...	...	...	..C	...	A..	.A.	G..	...	219
35717	.	T	.	.	.	.	S	.	.	.	.	.	.	.	.	N	N	A	.		72
M. avium	...	AC.	..T	..G	...	...	T..	...	...	...	...	...	...	..C	...	A..	.A.	G..	...		219
35712	.	T	.	.	.	.	S	.	.	.	.	.	.	.	.	N	N	A	.		72
M. avium	..T	ACG	..T	..G	A..	...	...	...	..T	...	..G	..T	..C	T.G	A..	.A.	G..	...			219
29555	.	T	.	.	N	.	.	.	.	.	L	.	.	.	N	N	A	.			72
M. scrofulaceum	...	ACG	..T	..G	..T	...	T..	...	..T	...	...	...	..C	T.G	A..	.AC	G..	...			219
19981	.	T	.	.	.	.	S	.	.	.	.	.	.	.	N	N	A	.			72
M. intracellulare	...	ACG	...	..G	...	..C	T..	...	..T	...	...	...	...	T.G	A..	.AC	G..	...			219
13950	.	T	.	.	.	.	S	.	.	.	.	.	.	.	N	N	A	.			72
M. intracellulare	...	ACG	...	..G	...	..C	T..	...	..T	...	...	...	...	T.G	A..	.AC	...	...			219
35768	.	T	.	.	.	.	S	.	.	.	.	.	.	.	N	N	.	.			72
M. scrofulaceum	..T	AC.	...	...	A..	..C	..A	..C	...	...	...	..T	..C	..G	A..	.AC	G..	...			219
35785	.	T	.	.	N	.	.	.	.	.	.	.	.	.	N	N	A	.			71

M. tuberculosis	ACC	ACC	GCC	CGC	ACC	ATC	AAT	GCC	CAG	CAA	GCG	GAA	CTG	GAT	TCG	GC	263
25177	T	T	A	R	T	I	N	A	Q	Q	A	E	L	D	S		87
M. avium	.T.	...	T.T	...	..G	...	..C	...	...	..G	AA.	..T	...	..C	CA.	..	263
35717	I	.	S	.	.	.	.	.	.	.	K	D	.	.	Q		87
M. avium	.T.	...	T.G	...	..G	...	..C	.T.	...	..G	AA.	..T	...	..C	CA.	..	263
35712	I	.	S	.	.	.	.	V	.	.	K	D	.	.	Q		87
M. avium	.T.	...	T.G	...	...	...	..C	...	...	..G	AA.	..T	...	...	CG.	..	263
29555	I	.	S	.	.	.	.	.	.	.	K	D	.	.	Q		87
M. scrofulaceum	.T.	...	T.G	...	...	...	...	..G	...	..G	AA.	..T	T..	...	CA.	..	263
19981	I	.	S	.	.	.	.	.	.	.	K	D	.	.	Q		87
M. intracellulare	GTG	...	T.G	...	..G	...	...	..G	...	...	AA.	..T	T..	...	CA.	..	263
13950	V	.	S	.	.	.	.	.	.	.	K	D	.	.	Q		87
M. intracellulare	GTG	...	T.G	...	..G	...	...	..G	...	...	AA.	..T	T..	...	CA.	..	263
35768	V	.	S	.	.	.	.	.	.	.	K	D	.	.	Q		87
M. scrofulaceum	.T.	...	T.G	...	...	...	..C	.T.	...	..G	AAT	..T	...	..C	CA.	..	263
35785	I	.	S	.	.	.	.	V	.	.	N	D	.	.	Q		86

FIG. 1. Alignment of the 263-bp sequence from the 293-bp PCR fragment of the mycobacterial *mce* gene ORF1. Mycobacterial strains are distinguished by ATCC reference number. The sequence of *M. tuberculosis* H37Ra between nucleotides 358 and 620 was used as a reference. Nucleotides and amino acids that are different from the reference sequence are indicated by their single-letter code; dots indicate identity. The amber stop codon at amino acid position 42 in the sequence for *M. scrofulaceum* 35785 is indicated (amb). There are no insertions or deletions to be indicated.

method of Jukes and Cantor (12). Principal coordinate analysis was then performed on the distance matrix to create a three-dimensional model.

Of the 19 mycobacterial strains that were examined, 12 were observed to yield the target 293-bp fragment. Only 7 of these 12 fragment-yielding strains had high enough yields to sequence from a single PCR amplification reaction. These results are summarized in Table 1. All of the high-yield strains were *M. avium-M. intracellulare-M. scrofulaceum* complex. The low-yield strains were typified by the production of many different-sized products. These yield differences may be due to the fact that the primers were designed from an *M. intracellulare* template and are not homologous to the *M. tuberculosis* sequence. A suboptimal annealing temperature was chosen in order to yield a product from the MTC. The nonyielding strains may lack the ability to serve as targets of the primer pair, because the primers may not be completely homologous with their corresponding *mce* ORF1-like target sequences.

The species that did not yield the 293-bp fragment were *M. nonchromogenicum*, *M. terrae*, *M. marinum*, *M. kansasii*, *M. gordonae*, *M. lepraemurium*, and the fast-growing mycobacterial isolate. Several of these strains produced PCR products of a different size from that of the target fragment. The species which yielded less than 50 ng of product from the agarose gel were *M. tuberculosis*, *M. bovis*, and *M. microti.*

To confirm the published sequence (GenBank accession number, X70901) for the portion of the *mce* gene analyzed in this study, the fragment obtained from *M. tuberculosis* H37Ra (ATCC 25177) was purified from several PCR mixtures and combined for sequencing. The sequences from *M. tuberculosis* H37Ra allowed only a 232-contiguous-bp consensus sequence to be unambiguously called. All of the unambiguous

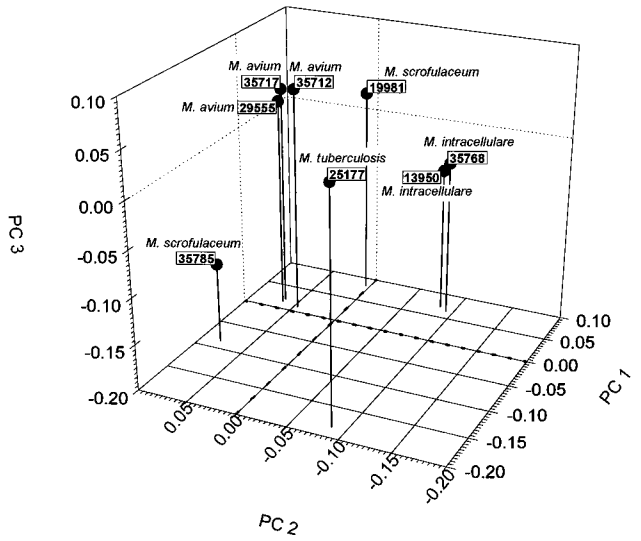

FIG. 2. Plot of the first three principal components (PC 1 to 3) of the genetic distance matrix created by the method of Jukes and Cantor (12). Mycobacterial strains are distinguished by ATCC reference number. Coordinate axes are represented by dotted lines. The strains of *M. avium* and *M. intracellulare* can be seen to form tight single-species clusters. The point for *M. scrofulaceum* 19981 falls in the same octant as that of the *M. avium* cluster, but the other *M. scrofulaceum* strain is distant from all the other strains. The *M. tuberculosis* strain's point falls with the greatest degree of distance from all other points.

bases from the resultant sequence were completely homologous to the published sequence (GenBank accession number, X70901). The fragment was identified as occurring between bp 343 and 635 as numbered from the published sequence. The *mce* gene ORF1 starts at base 208 and extends to base 807, so the PCR fragment's sequence represents most of the 5′ half of ORF1 of the *mce* gene.

The seven *M. avium-M. intracellulare-M. scrofulaceum* complex sequences were aligned with the sequence from *M. tuberculosis* H37Ra. Both the alignment of the nucleic acid sequence translations and that of the nucleic acid sequences are given in Fig. 1. As a result of the multiple DNA sequence alignment, the reading frame that corresponds to the ORF1 of the published *mce* gene sequence was identified. Both of the other reading frames were observed to yield several stop codons. The reading frames appear, therefore, to coincide.

Plotting the three principal components of the distance matrix, created by the method of Jukes and Cantor (12), yields the graph in Fig. 2. From the plot in Fig. 2, it appears as if the *M. avium-M. intracellulare-M. scrofulaceum* complex forms two cohesive groups delineated by the principal coordinate octants. The *M. intracellulare* and *M. avium* strains each group closely together.

The two *M. scrofulaceum* sequences do not appear as closely related to each other as do those from the strains of the other two *M. avium-M. intracellulare-M. scrofulaceum* complex species. One of the *M. scrofulaceum* sequences falls in the *M. avium* octant, while the other appears in an octant unpopulated by other sequences. The sequence from *M. scrofulaceum* 35785, which is an apparent outlier of the *M. avium-M. intracellulare-M. scrofulaceum* complex group, yields the only amino acid sequence that lacks a continuous open reading frame.

Despite the poor quality of sequence electrophoretograms derived from direct sequencing of PCR products, we elected to use this sequencing method, as it is less sensitive to the infidelity of AmpliTaq than subcloning of the PCR product. Subcloning would have allowed us to generate high-quality sequence electrophoretograms from even the low-yield strains, but the sequence data would represent that of only a single randomly selected strand of product.

All the sequences align with no insertions or deletions and were found to have the same most probable reading frame as the corresponding MTC *mce* ORF1. These features strongly

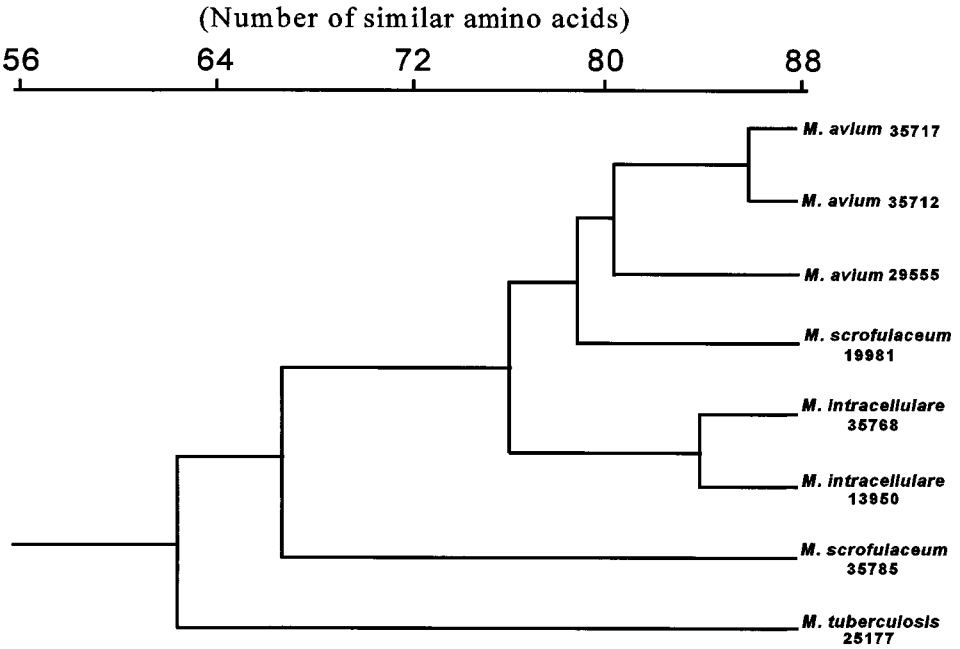

FIG. 3. Dendrogram derived from amino acid similarities between the sequence translations given in Fig. 1. Mycobacterial strains are distinguished by ATCC reference number. The number of similar amino acid positions from a possible total of 87 is given in a scale at the top. All of the *M. avium-M. intracellulare-M. scrofulaceum* complex species with the exception of *M. scrofulaceum* 35785 can be seen to cluster close together.

suggest that the source of the fragment is located in a coding region whose product is closely related to the MTC *mce* product. Coupled with infectivity experiments, the sequence data may help illuminate the relationship between the *mce* gene and mycobacterial virulence. However, the presence of target fragments in an avirulent strain (ATCC 25177), an attenuated strain (ATCC 35818), and a vaccine strain (ATCC 27289) of the MTC tested in this study suggests that this fragment of the *mce* ORF1 is not sufficient to confer virulence in vivo.

The single *M. avium-M. intracellulare-M. scrofulaceum* complex *mce* ORF1 sequence fragment that did not cluster with the other *M. avium-M. intracellulare-M. scrofulaceum* complex strains was also the only sequence whose amino acid translation would create a truncated product. This truncation is due to an amber stop codon at amino acid position 42. The amber codon is due to a single nucleic acid deviation at nucleic acid position 134 (Fig. 1). The difference between the *M. scrofulaceum* 35785 amino acid sequence and the other *M. avium-M. intracellulare-M. scrofulaceum* complex amino acid sequences is not limited to position 42. A dendrogram of the amino acid sequences (Fig. 3) assists in illuminating the magnitude of sequence deviation between *M. scrofulaceum* 35785 and the other *M. avium-M. intracellulare-M. scrofulaceum* complex species. If the stop codon results in a nonfunctional protein, the sequence is vestigial. There is little selective pressure to clear a population of individuals carrying mutations in a vestigial sequence. Therefore, the distance between a vestigial sequence and a related active sequence should increase with time at a rate far greater than that seen between two related active sequences. The *M. scrofulaceum* 35785 DNA sequence exhibits an average separation distance from all the other sequences (0.22329) that is second only to the average distance from the *M. tuberculosis* H37Ra sequence (0.26281; the next lowest is that for *M. scrofulaceum* 19981, at 0.17622). The amino acid sequence also deviates strongly from the clustering *M. avium-M. intracellulare-M. scrofulaceum* complex sequences and the *M. tuberculosis* sequence (Fig. 3). The observed differences in the predicted amino acid sequence, the significant nucleic acid sequence divergence, and the presence of a point mutation resulting in a predicted stop codon in the portion of the *mce* gene ORF1 analyzed in this study suggest that the *mce* gene in *M. scrofulaceum* 35785 may be and has been nonfunctional for many generations. It is not known whether the sequence truncation in *M. scrofulaceum* 35785 causes an alteration in this strain's virulence.

The deviations between the *mce* sequence fragments of the strains observed in this study are of the same order of magnitude as those observed for other conserved coding regions (17). The *M. avium-M. intracellulare-M. scrofulaceum* complex causes disease in a wide range of host species, and a large sequence variation might be expected for a gene whose product might be host specific. However, the host range for each of the species within the *M. avium-M. intracellulare-M. scrofulaceum* complex is not highly restrictive (7); thus, a cellular entry factor (1) might be expected to display conservation.

The sources of the infective *M. avium-M. intracellulare-M. scrofulaceum* complex have been inferred from culturing experiments, which are hampered by both the basic problems associated with the culturing of the *M. avium-M. intracellulare-M. scrofulaceum* complex and the fast-growing mycobacteria widely associated with most environmental samples. The identification of a relatively conserved virulence factor in the *M. avium-M. intracellulare-M. scrofulaceum* complex allows for the creation of primer-probe sets that might be used to rapidly identify the presence of these pathogens in environmental as well as clinical samples. The presence of a fragment of any virulence gene cannot conclusively demonstrate that a sample contains infectious material, but it is more suggestive of the presence of infectious material than the presence of a fragment of a gene not associated with virulence, like rRNA sequences. The specific fragment examined in this study was not observed in either the fast-growing isolate or in some of the other less clinically important slowly growing mycobacteria. However, PCR is not an effective method of demonstrating the absence of a gene. It is reasonable to suggest that there are sequences for that portion of the *mce* gene analyzed in this study, which differ sufficiently in their primer binding sites, in the genomes of those strains for which we were unable to produce the target fragment.

We believe that this is the first report of the presence of *mce* ORF1 in mycobacteria other than the MTC. The *M. avium-M. intracellulare-M. scrofulaceum* complex sequences display relatedness to one another and deviation from the sequence of the portion of the MTC *mce* gene analyzed in this study.

We are grateful to Ellen Parker for reviewing the manuscript and to Patrick McNelly for editorial comments.

REFERENCES

1. **Arruda, S., G. Bomfim, R. Knights, T. Huima-Byron, and L. W. Riley.** 1993. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. Science **261:**1454–1457.
2. **Bloom, B. R., and C. J. L. Murray.** 1992. Tuberculosis: commentary on a re-emergent killer. Science **257:**1055–1064.
3. **Bose, M., A. Chander, and R. H. Das.** 1993. A rapid and gentle method for the isolation of genomic DNA from mycobacteria. Nucleic Acids Res. **21:** 2529–2530.
4. **Butler, W. R., S. P. O'Connor, M. A. Yakrus, and W. M. Gross.** 1994. Cross-reactivity of genetic probe for detection of *Mycobacterium tuberculosis* with newly described species *Mycobacterium celatum*. J. Clin. Microbiol. **32:**536–538.
5. **Collins, C. H., J. M. Grange, and M. D. Yates.** 1984. Mycobacteria in water. J. Appl. Bacteriol. **57:**193–211.
6. **Ford, E. G., S. J. Snead, J. Todd, and N. G. Warren.** 1993. Strains of *Mycobacterium terrae* complex which react with DNA probes for *M. tuberculosis* complex. J. Clin. Microbiol. **31:**2805–2806.
7. **Good, R. C.** 1985. Opportunistic pathogens in the genus *Mycobacterium*. Annu. Rev. Microbiol. **39:**347–369.
8. **Goslee, S., and E. Wolinsky.** 1976. Water as a source of potentially pathogenic mycobacteria. Am. Rev. Respir. Dis. **113:**287–292.
9. **Gruft, H., A. Loder, M. Osterhout, B. C. Parker, and J. O. Falkinham III.** 1979. Postulated sources of *Mycobacterium intracellulare* and *Mycobacterium scrofulaceum* infection: isolation of mycobacteria from estuaries and ocean waters. Am. Rev. Respir. Dis. **120:**1385–1388.
10. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73:**237–244.
11. **Jonas, V., M. J. Alden, J. I. Curry, K. Kamisango, C. A. Knott, R. Lankford, J. M. Wolfe, and D. F. Moore.** 1993. Detection and identification of *Mycobacterium tuberculosis* directly from sputum sediments by amplification of rRNA. J. Clin. Microbiol. **31:**2410–2416.
12. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution in protein molecules, p. 21–123. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, New York.
13. **Kazda, J. F.** 1983. The principles of the ecology of mycobacteria, p. 323–341. *In* C. Ratlidge and J. Stanford (ed.), The biology of mycobacteria, vol. 2. Academic Press Inc., London.
14. **Lang, L. A., Y.-L. Tsai, C. L. Mayer, K. C. Patton, and C. J. Palmer.** 1994. Multiplex PCR for the detection of the heat-labile toxin gene and shiga-like toxin I and II genes in *Escherichia coli* isolated from natural waters. Appl. Environ. Microbiol. **60:**3145–3149.
15. **Palmer, C. J., Y.-L. Tsai, C. Paszko-Kolva, C. Mayer, and L. Sangermano.** 1993. Detection of *Legionella* species in sewage and ocean water by polymerase chain reaction, direct fluorescent-antibody, and plate culture methods. Appl. Environ. Microbiol. **59:**3618–3624.
16. **Salo, W. L., A. C. Aufderheide, J. Buikstra, and T. A. Holcomb.** 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian mummy. Proc. Natl. Acad. Sci. USA **91:**2091–2094.
17. **Soini, H., E. C. Bottger, and M. K. Viljanen.** 1994. Identification of mycobacteria by PCR-based sequence determination of the 32-kilodalton protein gene. J. Clin. Microbiol. **32:**2944–2947.
18. **Von Reyn, C. F., R. D. Waddell, T. Eaton, R. D. Arbeit, J. N. Maslow, T. W. Barber, R. J. Brindle, C. F. Gilks, J. Lumio, J. Lahdevirta, A. Ranki, D. Dawson, and J. O. Falkinham III.** 1993. Isolation of *Mycobacterium avium* complex from water in the United States, Finland, Zaire, and Kenya. J. Clin. Microbiol. **31:**3227–3230.