

## Comparative Genomic Hybridizations Reveal Genetic Regions within the *Mycobacterium avium* Complex That Are Divergent from *Mycobacterium avium* subsp. *paratuberculosis* Isolates†

Michael L. Paustian,<sup>1\*</sup> Vivek Kapur,<sup>2</sup> and John P. Bannantine<sup>1</sup>

National Animal Disease Center, U.S. Department of Agriculture-Agricultural Research Service, Ames, Iowa,<sup>1</sup>  
and Biomedical Genomics Center, University of Minnesota, St. Paul, Minnesota<sup>2</sup>

Received 5 November 2004/Accepted 22 December 2004

*Mycobacterium avium* subsp. *paratuberculosis* is genetically similar to other members of the *Mycobacterium avium* complex (MAC), some of which are nonpathogenic and widespread in the environment. We have utilized an *M. avium* subsp. *paratuberculosis* whole-genome microarray representing over 95% of the predicted coding sequences to examine the genetic conservation among 10 *M. avium* subsp. *paratuberculosis* isolates, two isolates each of *Mycobacterium avium* subsp. *silvaticum* and *Mycobacterium avium* subsp. *avium*, and a single isolate each of both *Mycobacterium intracellulare* and *Mycobacterium smegmatis*. Genomic DNA from each isolate was competitively hybridized with DNA from *M. avium* subsp. *paratuberculosis* K10, and open reading frames (ORFs) were classified as present, divergent, or intermediate. None of the *M. avium* subsp. *paratuberculosis* isolates had ORFs classified as divergent. The two *M. avium* subsp. *avium* isolates had 210 and 135 divergent ORFs, while the two *M. avium* subsp. *silvaticum* isolates examined had 77 and 103 divergent ORFs. Similarly, 130 divergent ORFs were identified in *M. intracellulare*. A set of 97 ORFs were classified as divergent or intermediate in all of the nonparatuberculosis MAC isolates tested. Many of these ORFs are clustered together on the genome in regions with relatively low average GC content compared with the entire genome and contain mobile genetic elements. One of these regions of sequence divergence contained genes homologous to a mammalian cell entry (*mce*) operon. Our results indicate that closely related MAC mycobacteria can be distinguished from *M. avium* subsp. *paratuberculosis* by multiple clusters of divergent ORFs.

*Mycobacterium avium* subsp. *paratuberculosis* is a gram-positive, acid-fast bacillus that is the causative agent of Johne's disease, a chronic infection primarily of ruminant animals characterized by inflammation of the digestive tract leading to nutrient malabsorption and eventually death. Diagnosis of *M. avium* subsp. *paratuberculosis* infections is confounded by the genetic similarity between *M. avium* subsp. *paratuberculosis* and nonpathogenic environmental mycobacteria, especially other members of the *Mycobacterium avium* complex (MAC). MAC bacteria, comprising three *M. avium* subspecies and *Mycobacterium intracellulare*, possess a high degree of genetic similarity but are capable of infecting a diverse range of host species. *M. avium* subsp. *paratuberculosis* is traditionally distinguished from other mycobacteria by its dependence on exogenous mycobactin, although intermittent mycobactin dependence has been reported among *M. avium* subsp. *paratuberculosis* isolates and within other species (6, 17). Previous studies have utilized DNA hybridizations to demonstrate a close genetic relationship between *M. avium* subsp. *paratuberculosis* and *Mycobacterium avium* subsp. *avium*, although they can normally be separated into two distinct populations (12, 24, 25, 30). Recent work comparing the whole-genome sequences of *M. avium* subsp. *paratuberculosis* and *M. avium*

subsp. *avium* isolates has revealed that they are greater than 97% identical at the nucleotide level within certain segments of genomic DNA (5).

Genetic analyses of conserved *M. avium* subsp. *paratuberculosis* sequences have been able to separate clinical isolates into distinct populations that often correlate with the host species from which they were cultured. Motiwala and coworkers examined a large number of *M. avium* subsp. *paratuberculosis* isolates and found that those obtained from cattle could generally be separated into a distinct group compared to isolates from other species, including sheep and humans (19). These findings were based primarily on comparisons of the integration loci of the IS900 insertion sequence (IS) and agree with the findings of Whittington and coworkers, who used polymorphisms in IS1311 to separate sheep and cattle isolates into separate populations (29). While the use of conserved ISs has been able to resolve the population structure of *M. avium* subsp. *paratuberculosis* isolates, they do not provide information on the conservation of individual protein coding sequences between isolates. A hybridization-based approach will allow us to identify significant variations in the overall sequence of related genetic features but will be insensitive to very small changes such as point mutations.

The objective in our present study was to determine the genetic variability between isolates of *M. avium* subsp. *paratuberculosis*, as well as to identify genomic differences between *M. avium* subsp. *paratuberculosis* and other MAC mycobacteria. To accomplish this, we have designed and constructed a whole-genome DNA microarray representing more than 95%

\* Corresponding author. Mailing address: National Animal Disease Center, USDA-ARS, 2300 North Dayton Ave., Ames, IA 50010. Phone: (515) 663-7979. Fax: (515) 663-7458. E-mail: mpaustia@nadc.ars.usda.gov.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

TABLE 1. Mycobacterial isolates used in this study

Isolate	Organism	Host	Location <sup>a</sup>	Reference or source
MAA104	<i>M. avium</i> subsp. <i>avium</i>	Human		L. Bermudez
6004	<i>M. avium</i> subsp. <i>avium</i>	Chicken		ATCC 35719
6010	<i>M. intracellulare</i>	Swine		ATCC 35773
6006	<i>M. avium</i> subsp. <i>silvaticum</i>	Roe deer	Denmark	V1-72 (13)
6058	<i>M. avium</i> subsp. <i>silvaticum</i>	Wood pigeon	France	ATCC 49884
2009	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine	USA (PA)	This study
4138	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Goat	Argentina	This study
10006	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine	USA (PA)	This study
19698	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine		ATCC 19698
1091	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine	Nova Scotia, Canada	This study
6079	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Mink	USA (ND)	This study
4090	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine	USA (PA)	This study
Kay	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine	USA (IA)	This study
1434	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Sheep	USA (NE)	This study
K10	<i>M. avium</i> subsp. <i>paratuberculosis</i>	Bovine	USA (WI)	16
19420	<i>M. smegmatis</i>			ATCC 19420

<sup>a</sup> Abbreviations: USA, United States; PA, Pennsylvania; ND, North Dakota; IA, Iowa; NE, Nebraska; WI, Wisconsin.

of the open reading frames (ORFs) identified from the genome sequence of *M. avium* subsp. *paratuberculosis* K10 (16).

#### MATERIALS AND METHODS

##### Construction of a whole-genome *M. avium* subsp. *paratuberculosis* microarray.

A library of PCR products representing all coding sequences was constructed with primers designed by Primer3 software (18) to amplify fragments of  $\leq 500$  bp from each ORF identified from the genome sequence of *M. avium* subsp. *paratuberculosis* K10 (NCBI accession no. AE016958) with the use of purified genomic DNA as a template. PCRs consisted of 1 $\times$  PCR Buffer II (Perkin-Elmer, Boston, Mass.), 0.2 mM deoxynucleoside triphosphate mix, 3 mM MgCl<sub>2</sub>, 5% dimethyl sulfoxide, 0.04 U of Taq Gold (Perkin-Elmer)/ $\mu$ l, and 0.6  $\mu$ M primers. PCR products amplified from genomic DNA were diluted 1:200 in distilled water and used as the template for a subsequent 50- $\mu$ l PCR in order to minimize genomic DNA carryover. The conditions used for both rounds of PCR were an initial incubation for 7 min at 94°C and then 35 cycles of 94°C for 45 s, 57°C for 60 s, and 72°C for 90 s followed by a final hold for 7 min at 72°C. The PCR products were checked for quality on agarose gels, at which point it was determined that 4,110 (95%) of the total ORFs had been successfully amplified. They were then purified with Montage 96-well filter plates (Millipore, Bedford, Mass.) and resuspended in 3 $\times$  SSC containing 1.5 M betaine (1 $\times$  SSC is 150 mM NaCl, 15 mM sodium citrate, pH 7.0). The resulting products were arrayed in triplicate onto homemade poly-L-lysine-coated glass slides with a MicroGrid II Compact robot (Genomic Solutions, Ann Arbor, Mich.) along with control samples that included spotting buffer alone and *Arabidopsis thaliana* sequences. This resulted in an array containing over 13,000 spots with an average diameter of 170  $\mu$ m.

**Comparative genomic hybridizations.** Mycobacteria were cultured in Middlebrook H9 broth (pH 6.0) supplemented with oleic acid-albumin-dextrose-catalase (Becton Dickinson Microbiology, Sparks, Md.), and 0.05% Tween 80. Cultures of *M. avium* subsp. *paratuberculosis* were further supplemented with ferric mycobactin J (2 mg/liter; Allied Monitor Inc., Fayette, Mo.). Genomic DNA was extracted from *M. avium* subsp. *paratuberculosis* K10 and 15 mycobacterial isolates (Table 1) with Genomic-tip 100/G anion-exchange columns (Qiagen, Valencia, Calif.) as previously described (3) with the modification that D-cycloserine was not added as part of the extraction procedure. Purified genomic DNA was then randomly sheared by nebulization on ice at 10 lb/in<sup>2</sup> for 1.5 min, resulting in an average fragment size of 800 bp. Aliquots of sheared genomic DNA (4  $\mu$ g) were labeled by incorporating an aminoallyl-dUTP nucleotide (Sigma, St. Louis, Mo.) into cDNA by using Klenow enzyme (U.S. Biochemicals, Cleveland, Ohio) primed with random hexamers (Amersham Biosciences, Piscataway, N.J.) overnight. The amine-modified cDNA was purified, resuspended in 0.13 M sodium bicarbonate, and labeled by the addition of either Alexa Fluor 555 or Alexa Fluor 647 succinimidyl ester dye (Molecular Probes, Eugene, Oreg.) followed by incubation for 2 h at room temperature. Labeled cDNA from experimental mycobacterial isolates was then purified and mixed with alternately labeled *M. avium* subsp. *paratuberculosis* K10 cDNA in a final volume of 45  $\mu$ l containing 3 $\times$  SSC, 0.22% sodium dodecyl sulfate (SDS), and 34  $\mu$ g of salmon sperm DNA (Invitrogen, Carlsbad, Calif.). This hybridization solution was incubated at 100°C for 2 min, applied to the *M. avium* subsp. *paratuberculosis* K10 microarray, and al-

lowed to hybridize overnight at 65°C. The arrays were washed sequentially for 3 min at room temperature in 300-ml volumes of 0.5 $\times$  SSC-0.01% SDS, 0.5 $\times$  SSC, 0.1 $\times$  SSC, and 0.01 $\times$  SSC and then dried by centrifugation and scanned with a ScanArray 4000 confocal laser scanner (Perkin-Elmer). Each mycobacterial isolate was hybridized against *M. avium* subsp. *paratuberculosis* K10 at least twice in a dye-flip experimental design.

**Microarray data analysis.** Raw intensity measurements for each spot on the microarray were extracted from scanned images by using ScanArray Express software (Perkin-Elmer) and adjusted with local background subtraction and LOWESS normalization (9). Poorly detected spots were removed by filtering out those in which the measured spot pixels did not have a total median intensity of at least 2,800 or in which more than 50% of the pixels from the *M. avium* subsp. *paratuberculosis* K10 control sample were within 2 standard deviations of the local background. At this point, any ORFs not represented by at least two acceptable spots on a hybridized microarray were discarded from further analyses. The median intensity of the remaining spots was determined for each ORF, and then ratios of the spot intensities for experimental and control *M. avium* subsp. *paratuberculosis* K10 DNA samples were calculated and log transformed for analysis. The software program Genomotyping Analysis (14) was utilized to grade the ORFs on a linear scale from -0.5 (0% estimated probability of being present) to 0.5 (100% estimated probability of being present). Briefly, this software uses the distribution of measured ratios from a microarray hybridization to determine the probability of an individual ORF being present in a genomic DNA sample. Grades from duplicated experiments were combined, and ORFs were classified as divergent (total grade of  $< -0.75$ ) or present (total grade of  $> 0.75$ ), with the remaining ORFs listed as having intermediate divergence. Results were reported only if data were obtained from both hybridizations.

**IS1311 PCR restriction digest.** The IS1311 IS was amplified via PCR from genomic DNA under the standard conditions described above and then subjected to restriction digest with HinfI as described by Whittington and coworkers (29). Briefly, a PCR product representing IS1311 was amplified using genomic DNA from the mycobacterial isolates examined in this study under the standard thermocycling conditions described above. Six microliters of each PCR product was then incubated at 37°C for 1.5 h with 5 U of HinfI, followed by a 20-min deactivation at 80°C. The resulting product was resolved on a 1.5% agarose gel and visualized with ethidium bromide staining under UV light.

**Comparative sequence analysis with *M. avium* subsp. *avium* 104.** The complete but unannotated genome sequence of *M. avium* subsp. *avium* 104 was obtained from The Institute for Genomic Research (www.tigr.org). The BLASTN algorithm (1) was used to query the *M. avium* subsp. *avium* 104 genome sequence with a library of all the sequences present on the *M. avium* subsp. *paratuberculosis* K10 microarray. The resulting output was filtered so that only matches of at least 100 bp with a maximum size difference of 10% between the query and target sequences were reported.

**Southern blotting.** Four micrograms of genomic DNA from isolates MAA104, 6004, 6058, K10, 19698, 6079, 4138, and 6010 was digested with BamHI (New England Biolabs), separated by electrophoresis on an 0.6% agarose gel, and transferred to a nylon membrane (Ambion) by standard procedures (23). Digoxigenin-labeled probes of approximately 500 bp were generated for MAP0005, MAP0854, and MAP3815 by using the DIG High Prime DNA Labeling and

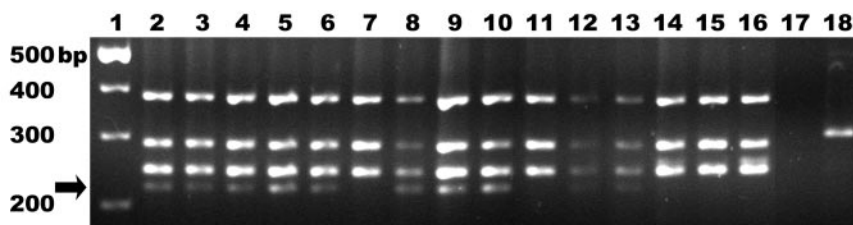


FIG. 1. Restriction digest of IS1311 PCR products. Lane 1, 100-bp DNA standard; lane 2, *M. avium* subsp. *paratuberculosis* K10; lane 3, *M. avium* subsp. *paratuberculosis* 19698; lane 4, *M. avium* subsp. *paratuberculosis* 2009; lane 5, *M. avium* subsp. *paratuberculosis* 4138; lane 6, *M. avium* subsp. *paratuberculosis* 10006; lane 7, *M. avium* subsp. *avium* 6004; lane 8, *M. avium* subsp. *paratuberculosis* 6079; lane 9, *M. avium* subsp. *paratuberculosis* 1434; lane 10, *M. avium* subsp. *paratuberculosis* Kay; lane 11, *M. avium* subsp. *silvaticum* 6058; lane 12, *M. avium* subsp. *paratuberculosis* 4090; lane 13, *M. avium* subsp. *paratuberculosis* 1091; lane 14, *M. intracellulare* 6010; lane 15, *M. avium* subsp. *silvaticum* 6006; lane 16, *M. avium* subsp. *avium* 104; lane 17, no-DNA control; lane 18, *M. smegmatis* 19420.

Detection Kit II (Roche Applied Science). These probes were individually hybridized to the membrane according to the manufacturer's instructions. Hybridized blots were exposed to Biomax MR film (Kodak) for 30 to 60 min.

**Identification of sequences unique to *M. avium* subsp. *paratuberculosis*.** The 4,350 annotated ORFs present in the *M. avium* subsp. *paratuberculosis* genome were compared to the complete genome sequence of *M. avium* subsp. *avium* 104 by using the TBLASTX algorithm (1). With the use of expected value cutoffs of  $1.0 \times 10^{-10}$  for *M. avium* subsp. *avium* and  $1.0 \times 10^{-6}$  for GenBank, there were 32 ORFs that had no significant matches to any of the sequences currently available in these databases.

## RESULTS

**Isolate typing.** A *Hinf*I restriction digest of the IS1311 IS from each isolate revealed that all of the *M. avium* subsp. *paratuberculosis* isolates examined contained a 220-bp product indicating the presence of a previously described polymorphism found exclusively in nonsheep isolates of *M. avium* subsp. *paratuberculosis* (29) (Fig. 1). None of the nonparatuberculosis MAC isolates contained the polymorphism.

**Comparative sequence analysis with *M. avium* subsp. *avium* 104.** The raw genome sequence of *M. avium* subsp. *avium* 104 was compared to the *M. avium* subsp. *paratuberculosis* K10 probe sequences and matched with comparative genomic hybridization data to evaluate the performance of the *M. avium* subsp. *paratuberculosis* K10 microarray. By using the procedures described in Materials and Methods, microarray probe

sequence matches to *M. avium* subsp. *avium* 104 were obtained for 3,709 of the 4,350 annotated *M. avium* subsp. *paratuberculosis* K10 ORFs, 3,497 (94%) of which were also "well-measured" in comparative genomic hybridizations. Of the 210 ORFs classified as divergent in *M. avium* subsp. *avium* 104 by the *M. avium* subsp. *paratuberculosis* K10 microarray (Table 2; see also Table S1 in the supplemental material), 197 (94%) had no significant matches to the genome sequence of *M. avium* subsp. *avium* 104 by BLAST analysis. The remaining 13 ORFs were either lower-quality matches to the *M. avium* subsp. *avium* 104 genome (<92% identical;  $n = 6$ ) or potential false negatives (>98% identical;  $n = 7$ ). A total of 2,932 ORFs were identified as present in *M. avium* subsp. *avium* 104 by use of the *M. avium* subsp. *paratuberculosis* K10 microarray, 2,782 (95%) of which were successfully matched to *M. avium* subsp. *avium* 104 genomic sequences at greater than 95% identity. Weaker matches of 90 to 93% identity were found for three additional ORFs that had been classified as present. The remaining sequences identified as present via microarray analysis ( $n = 147$ ) were either not matched or only partially matched (greater than 10% size difference as described in Materials and Methods) to corresponding *M. avium* subsp. *avium* 104 sequences. There were 355 *M. avium* subsp. *avium* 104 ORFs classified as having intermediate divergence. Of these ORFs, 309 (87%) were at least 95% identical to *M. avium* subsp. *paratuberculosis*

TABLE 2. Classification of *M. paratuberculosis* K10 ORFs compared with other mycobacterial isolates

Isolate	Organism	No. of ORFs in classification:			No. of ORFs measured	% of total K10 ORFs
		Divergent	Intermediate	Present		
MAA104	<i>M. avium</i> subsp. <i>avium</i>	210	355	2,932	3,497	80
6004	<i>M. avium</i> subsp. <i>avium</i>	135	251	2,765	3,151	72
6010	<i>M. intracellulare</i>	130	473	2,994	3,597	83
6006	<i>M. avium</i> subsp. <i>silvaticum</i>	77	96	3,038	3,211	74
6058	<i>M. avium</i> subsp. <i>silvaticum</i>	103	304	2,750	3,157	73
2009	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	49	2,715	2,764	64
4138	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	92	3,332	3,424	79
10006	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	458	2,666	3,124	72
19698	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	4	3,796	3,800	87
1091	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	90	2,720	2,810	65
6079	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	4	2,958	2,962	68
4090	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	398	2,679	3,077	71
Kay	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	4	3,056	3,060	70
1434	<i>M. avium</i> subsp. <i>paratuberculosis</i>	0	0	3,091	3,091	71
19420	<i>M. smegmatis</i>	380 <sup>a</sup>	931	521	1,832	42

<sup>a</sup> *M. smegmatis* ORFs were classified by using a 1.5-fold cutoff for the average ratio of hybridization intensities.



K10 sequences, while the remaining ORFs were either lower-quality matches to *M. avium* subsp. *avium* 104 sequences ( $n = 12$ ; 4%), not detected, or unmatched to microarray probes ( $n = 34$ ; 10%). From this result, we can conclude that a majority of the ORFs classified as intermediate by microarray analysis are most likely present in the genomes examined in this study.

***M. avium* subsp. *paratuberculosis* isolates.** Genomic DNA from nine clinical isolates of *M. avium* subsp. *paratuberculosis* was competitively hybridized against *M. avium* subsp. *paratuberculosis* K10 DNA on a whole-genome DNA microarray (Table 2). Overall, the *M. avium* subsp. *paratuberculosis* isolates were very similar to K10. None of the *M. avium* subsp. *paratuberculosis* isolates examined had any ORFs classified as divergent. The type strain 19698 was the most similar to K10, with 88% of the total K10 ORFs being detected by the array and 99% of these identified as being present. A core set of 1,230 ORFs were classified as present among all of the *M. avium* subsp. *paratuberculosis* isolates. An analysis of the proteins encoded by these conserved ORFs revealed that the distribution of predicted functions was similar to the whole genome, indicating that no functional categories were preferentially represented.

Notably, *M. avium* subsp. *paratuberculosis* 10006 and 4090 were isolated from the same dairy herd in Pennsylvania, and, although they lacked divergent sequences, 458 and 398 ORFs of intermediate divergence were identified in their respective genomic sequences. A total of 152 ORFs were classified as intermediate in both isolates (see Table S2 in the supplemental material). These ORFs include hypothetical proteins ( $n = 52$ ), proteins involved in energy metabolism ( $n = 22$ ), and cell envelope proteins ( $n = 17$ ). It is intriguing that two isolates from a common geographic area contained a significantly higher number of ORFs classified as intermediate than did the other *M. avium* subsp. *paratuberculosis* isolates examined in this study.

**Other mycobacterial isolates.** DNA from five MAC isolates was hybridized against *M. avium* subsp. *paratuberculosis* K10 DNA to establish the degree of genome variability that exists between these closely related mycobacteria (Table 2). Additionally, the type strain of *Mycobacterium smegmatis* (ATCC 19420) was hybridized against the *M. avium* subsp. *paratuberculosis* K10 microarray to examine the genetic conservation between less closely related mycobacteria.

***M. avium* subsp. *silvaticum*.** The two *M. avium* subsp. *silvaticum* isolates examined, 6006 and 6058, were found to have 77 and 103 divergent genes, respectively (see Tables S3 and S4 in the supplemental material). Of all the nonparatuberculosis MAC isolates, 6006 had the fewest ORFs classified as divergent or intermediate compared to K10. Among the 77 divergent ORFs reported in isolate 6006, 45 encoded hypothetical proteins, while 16 ORFs represented IS elements, including multiple copies of *ISMAP02*, *ISmav2* (AF286339), and *IS900*. Additionally, two members of a putative cation transport system (MAP3731c and MAP3732c) and cell invasion proteins (MAP2189 and MAP2192) were also found to be divergent. The divergent ORFs identified in *M. avium* subsp. *silvaticum* isolate 6058 included eight encoding oxidoreductase or oxygenase enzymes, six nonribosomal peptide synthesis enzymes (including two involved in mycobactin biosynthesis), and four cell invasion proteins (MAP2189, MAP2190, MAP2193, and MAP2194). Twenty-five IS elements or transposases were

found to be divergent in 6058; they included *IS1311* (U16276), *ISmav2*, *IS900*, *ISMAP02*, and *IS1601* (AF060182). There were 17 ORFs found to be divergent in both *M. avium* subsp. *silvaticum* isolates, including *IS900*, *ISmav2*, *ISMAP02*, transposases ( $n = 3$ ), and hypothetical proteins ( $n = 5$ ).

***M. avium* subsp. *avium*.** *M. avium* subsp. *avium* isolate 6004 had 135 ORFs classified as divergent from K10 (see Table S5 in the supplemental material). Consistent with the other isolates examined, many of the divergent ORFs ( $n = 57$ ) encode hypothetical proteins. A total of 33 divergent ORFs contained IS-related elements, including *IS1311*, *IS900*, *ISmav2*, *IS1601-B*, and *ISMAP02*. Two ORFs encoding proteins involved in mycobactin synthesis were found to be divergent (*mbtD* and *mbtG*), as well as five outer membrane proteins. Five adjacent ORFs encoding cell invasion proteins (MAP2189 to MAP2193) were divergent in isolate 6004, as well as the two putative cation transport proteins (MAP3731c and MAP3732c) that were also identified as divergent in *M. avium* subsp. *silvaticum* 6006.

As described above, *M. avium* subsp. *avium* isolate 104 had 210 ORFs classified as divergent from *M. avium* subsp. *paratuberculosis* K10, 110 of which were shared with *M. avium* subsp. *avium* 6004. Similar to the other nonparatuberculosis isolates, the largest groups of divergent genes encoded hypothetical proteins ( $n = 110$ ) and IS elements ( $n = 29$ ). Other groups of functionally related divergent genes encoded cell envelope proteins ( $n = 11$ ), transcriptional regulators ( $n = 8$ ), and PPE family proteins ( $n = 3$ ). Clusters of genes divergent only in isolate MAA104 included MAP0956 to MAP0965, which encode multiple hypothetical proteins and putative glycosyltransferases. The presence of glycosyltransferases suggests that some of these ORFs may participate in cell membrane biosynthesis.

***M. intracellulare*.** *M. intracellulare* isolate 6010 was found to have 130 ORFs divergent from K10 (see Table S6 in the supplemental material). Of these divergent ORFs, 27 contained IS elements or transposases, including *IS900*, *ISmav2*, *IS1601-B*, and *ISMAP02*. Six divergent ORFs encoded outer membrane or secreted proteins, while 71 divergent ORFs encoded hypothetical proteins or proteins with unknown function. Notably, a cluster of ORFs from MAP1231 to MAP1237 were classified as divergent in *M. intracellulare* 6010 and *M. avium* subsp. *avium* 104 and include the region previously identified by Tizard and coworkers as a low-GC genetic island present only in *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* (28). This region was classified as present in all of the *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* isolates as well as *M. avium* subsp. *avium* 6004.

**Common divergent regions.** A core set of 97 ORFs were classified as divergent or intermediate in all of the nonparatuberculosis MAC genomes (Table 3). These ORFs included 25 IS features, among which were *IS900*, *IS1311*, *ISmav2*, and *ISMAP02*, as well as sequences similar to *IS1110*, *IS1547*, *IS1601-B*, *IS6110*, and *ISmav2*. As observed in the individual analyses, a large number ( $n = 42$ ) of the shared divergent ORFs encoded hypothetical proteins.

Interestingly, many of the common regions of divergence among the nonparatuberculosis MAC isolates were grouped into clusters of adjacent ORFs (Table 3). Notably, four of the seven *IS1311* copies present in the *M. avium* subsp. *paratuberculosis* genome were found within four independent regions of

TABLE 3. Ninety-seven ORFs identified as divergent or intermediate in the nonparatuberculosis MAC isolates MAA104, 6004, 6006, 6010, and 6058

Region	ORF	Description <sup>a</sup>	Functional category
	MAP0034	<i>IS900</i>	IS element
MAP_RD1	MAP0097c	Acyl desaturase, N terminal	Lipid biosynthesis
MAP_RD1	MAP0098c	Transcriptional regulator, TetR family	Conserved hypothetical
MAP_RD1	MAP0100	GPTase-activating protein	Conserved hypothetical
MAP_RD1	MAP0101	Conserved hypothetical protein	Conserved hypothetical
MAP_RD1	MAP0102	Hypothetical protein	Unknown
MAP_RD1	MAP0103c	Hypothetical protein	Unknown
MAP_RD1	MAP0106c	Integrase-recombinase	Phage-related function
	MAP0159c	<i>IS900</i>	IS element
	MAP0253	<i>ISmav2</i>	IS element
	MAP0284c	Hypothetical protein	Unknown
	MAP0338c	Similar to <i>ISLp1</i> (GEN: AF386079)	IS element
	MAP0388	Diarylpropane peroxidase	Conserved hypothetical
	MAP0428	Similar to <i>ISmav2</i> (GEN: AF286339)	IS element
	MAP0589c	Similar to <i>ISmav2</i> (GEN: AF286339)	IS element
	MAP0664c	<i>ISmav2</i>	IS element
	MAP0832c	Similar to <i>IS1601-B</i>	IS element
MAP_RD2	MAP0850c	Similar to <i>ISmav2</i> (GEN: AF286339)	IS element
MAP_RD2	MAP0851	No hits	Unknown
MAP_RD2	MAP0856c	No hits	Unknown
MAP_RD2	MAP0859c	<i>Mycobacterium</i> phage L5	Phage-related function
MAP_RD2	MAP0866	Integrase	Phage-related function
	MAP1048c	Transposase (GEN: AF373840)	IS element
	MAP1287	Similar to <i>IS1601-B</i>	IS element
	MAP1432	REP family protein	REP13E12 family
	MAP1723	Putative secreted protein	Cell envelope
	MAP1822c	Hypothetical protein	Unknown
	MAP1824c	Transposase (GEN: X58441)	IS element
	MAP1946	Hypothetical protein	Unknown
	MAP1952c	Similar to <i>IS1547</i>	IS element
	MAP2050	Similar to <i>IS1601-B</i>	IS element
MAP_RD3	MAP2148	Integrase-recombinase	Phage-related function
MAP_RD3	MAP2153	No hits	Unknown
MAP_RD3	MAP2155	Similar to <i>IS6110</i>	IS element
MAP_RD3	MAP2156	Putative transposase (GEN: X52471)	IS element
MAP_RD3	MAP2158	No hits	Unknown
MAP_RD4	MAP2179	No hits	Unknown
MAP_RD4	MAP2181c	No hits	Unknown
MAP_RD4	MAP2183c	Linalool 8-monooxygenase	Conserved hypothetical
MAP_RD4	MAP2184c	Short-chain alcohol dehydrogenase family	Energy metabolism
MAP_RD4	MAP2185c	Proline-rich protein precursor	Unknown
MAP_RD4	MAP2188c	Acyl-CoA synthase	Degradation
MAP_RD4	MAP2189	Cell invasion protein	Virulence
MAP_RD4	MAP2190	Part of <i>mce4</i> operon	Virulence
MAP_RD4	MAP2192	Part of <i>mce4</i> operon	Virulence
MAP_RD4	MAP2193	Part of <i>mce3</i> operon	Virulence
MAP_RD4	MAP2302	<i>ISmav2</i>	IS element
	MAP2416c	Similar to <i>ISLp1</i> (GEN: AF386079)	IS element
	MAP2427c	Probable hydrolase	Degradation
	MAP2566	Similar to <i>ISLp1</i> (GEN: AF386079)	IS element
MAP_RD5	MAP2755	Hypothetical protein	Unknown
MAP_RD5	MAP2756c	Hypothetical protein	Unknown
MAP_RD5	MAP2758	Hypothetical protein	Unknown
MAP_RD5	MAP2759	Conserved hypothetical protein	Conserved hypothetical
MAP_RD5	MAP2764c	No hits	Unknown
MAP_RD5	MAP2765c	No hits	Unknown
MAP_RD5	MAP2767c	No hits	Unknown
MAP_RD5	MAP2768c	PhiRV2 integrase	Phage-related function
MAP_RD5	MAP2963c	No hits	Unknown
	MAP2964c	Hypothetical protein	Unknown
	MAP3078c	Transposase (GEN: AF373840)	IS element
	MAP3357c	Similar to <i>ISLp1</i> (GEN: AF386079)	IS element
	MAP3436c	Hypothetical protein	Unknown
	MAP3437c	No hits	Unknown
	MAP3467c	Similar to <i>ISLp1</i> (GEN: AF386079)	IS element
MAP_RD6	MAP3726	Iron ABC transporter	Transport-binding proteins
MAP_RD6	MAP3727	ABC transporter protein	Cell envelope
MAP_RD6	MAP3729	Dioxygenasedioxygenase	Energy metabolism
MAP_RD6	MAP3731c	ABC transporter ATP-binding subunit	Transport-binding proteins
MAP_RD6	MAP3732c	Cobalt permease and related transporters	Transport-binding proteins

Continued on following page

TABLE 3—Continued

Region	ORF	Description <sup>a</sup>	Functional category
MAP_RD6	MAP3733c	Hypothetical protein	Unknown
MAP_RD6	MAP3734c	ABC transporter ATP-binding protein	Cell envelope
MAP_RD6	MAP3736c	Heavy metal tolerance protein	Conserved hypothetical
MAP_RD6	MAP3738c	Conserved hypothetical protein	Conserved hypothetical
MAP_RD6	MAP3739c	Possible sugar transporter	Cell envelope
MAP_RD6	MAP3741	Putative nonribosomal peptide synthase	Nonribosomal peptide synthesis
MAP_RD6	MAP3742	Putative nonribosomal peptide synthase	Nonribosomal peptide synthesis
MAP_RD6	MAP3743	Hypothetical protein	Unknown
MAP_RD6	MAP3745	Thioesterase	Conserved hypothetical
MAP_RD6	MAP3747c	Conserved hypothetical protein	Conserved hypothetical
MAP_RD6	MAP3748c	Similar to IS1110	IS element
MAP_RD6	MAP3751	Conserved large membrane protein	Cell envelope
MAP_RD6	MAP3753	Hypothetical protein	Unknown
MAP_RD6	MAP3754	Similar to transposase of IS3 family	IS element
MAP_RD6	MAP3755	Similar to transposase	IS element
MAP_RD6	MAP3762c	Possible rhamnosyl-glycosyl-transferase	Miscellaneous transferases
MAP_RD6	MAP3764c	Polyketide synthase	Nonribosomal peptide synthesis
MAP_RD6	MAP3766	Membrane protein	Conserved hypothetical
MAP_RD6	MAP3770	Conserved hypothetical protein	Conserved hypothetical
MAP_RD6	MAP3772c	Conserved hypothetical protein	Conserved hypothetical
MAP_RD6	MAP3775c	Putative ABC transporter ATP-binding protein	Transport-binding proteins
MAP_RD7	MAP3815	No hits	Unknown
MAP_RD7	MAP3816	Hypothetical protein	Unknown
MAP_RD7	MAP3817c	No hits	Unknown
MAP_RD7	MAP3818	Cytochrome P450	Cytochrome P450 enzymes
MAP_RD7	MAP4270	Multicopy phage resistance protein Mpr	Conserved hypothetical
MAP_RD7	MAP4280	Cholesterol oxidase	Degradation

<sup>a</sup> Abbreviations: GEN, GenBank sequence accession number; CoA, coenzyme A.

divergence. Furthermore, four of these seven divergent regions contained at least one ORF encoding a putative phage-related sequence.

The divergent clusters comprised of MAP0849 to MAP0866 (MAP\_RD2) and MAP2750 to MAP2768 (MAP\_RD5) include ORFs that predominantly encode hypothetical proteins. Additionally, a whole-genome alignment of *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* reveals that a majority of these divergent regions are bordered by inversions in the genome sequences of these two isolates.

The average G+C content of the MAP\_RD5 gene cluster is similar to the K10 whole-genome level of 69.3%, while the MAP\_RD2 cluster is significantly lower at 60.6%. Another region of relatively low average G+C content that correlates with a cluster of divergent genes is present between MAP2148 and MAP2158 (MAP\_RD3). This 12,000-bp region has an average G+C content of 60.6% and contains ISs and ORFs of unknown function, including two sequences previously identified as specific to *M. avium* subsp. *paratuberculosis* (MAP2149c and MAP2154c) (4, 20). A 9,500-bp region between MAP3749 and MAP3756c has an average G+C content of 60.5% and was also part of a shared divergent ORF cluster (MAP\_RD6). This region encodes two putative membrane proteins, two transposases, and two proteins involved in energy metabolism. Additionally, the MAP\_RD6 region also contains seven genes encoding putative ABC transport proteins. The presence of mobile genetic elements as well as the low G+C content found within some of these divergent regions suggests that they may have been acquired via horizontal gene transfer.

The MAP\_RD4 region contains five adjacent divergent ORFs (MAP2189 to MAP2194) that are homologous to mammalian cell entry (*mce*) genes previously identified as important for invasion and survival within macrophages (2). None of

the other *mce* gene clusters present in the *M. avium* subsp. *paratuberculosis* genome were found to be divergent in any of the isolates. These genes appear to be ideal candidates for future investigations into the pathogenesis of *M. avium* subsp. *paratuberculosis*.

***M. smegmatis*.** *M. smegmatis* is a fast-growing mycobacterium distantly related to the slow-growing MAC mycobacteria. Genomic DNA from *M. smegmatis* ATCC 19420 was hybridized against the *M. avium* subsp. *paratuberculosis* K10 microarray and analyzed with the GACK software program as described for the other mycobacteria. Initially, only six *M. smegmatis* ORFs were classified as divergent by the Genomotyping Analysis software, while a manual examination of the data indicated that there should be significantly more divergent ORFs. This discrepancy was determined to be due to the wide distribution of hybridization ratios resulting from a large number of divergent sequences. A conservative cutoff of 1.5-fold was therefore used to sort the ORFs based on the average ratio of hybridization intensities. According to this cutoff classification system, 380 ORFs were identified as divergent in *M. smegmatis*, while 521 ORFs were present and 931 were intermediate (Table 2). Unlike the nonparatuberculosis MAC mycobacteria, the divergent *M. smegmatis* ORFs were evenly distributed across the entire genome. Among the divergent ORFs were potential virulence factors including nine *mce* genes, seven PPE family genes, and 29 cell envelope-associated proteins (see Table S7 in the supplemental material). Notably, many of the proteins encoded by these divergent ORFs fall into the same functional categories as genes identified as divergent between *Mycobacterium tuberculosis* H37Rv and *M. tuberculosis* BCG Pasteur (reviewed in reference 8). The large number of divergent *M. smegmatis* ORFs is reflective of the phylogenetic distance separating slow- and fast-growing mycobacteria and may also

TABLE 4. PCR amplification of ORFs from shared regions of divergence and adjacent flanking ORFs

Region	ORF	Observation of PCR product <sup>b</sup>						
		6006	6058	6004	6010	MAA104	19698	K10
— <sup>a</sup>	MAP0092	+++	+++	+++	+++	+++	+++	+++
MAP_RD1	MAP0096c	—	—	—	+	+	+++	+++
MAP_RD1	MAP0106c	+++	+++	+++	+++	+++	+++	+++
	MAP0107	+++	+++	+++	+++	+++	+++	+++
	MAP0845	+++	+++	+++	+++	+	+++	+++
MAP_RD2	MAP0850c	+++	+++	+++	+	+	+++	+++
MAP_RD2	MAP0866	—	—	+	+	+++	+++	+++
	MAP0867c	+++	+++	+++	+++	+++	+++	+++
	MAP2146c	+	+	+	+	+	+	+
MAP_RD3	MAP2158	+	+	+	+	+	+++	+++
	MAP2164	+++	+++	+++	+++	+++	+++	+++
MAP_RD4	MAP2189	+	+	+	+	+++	+++	+++
	MAP2198	+++	+++	+++	+++	+++	+++	+++
	MAP2749c	+++	+++	+++	+++	+++	+++	+++
MAP_RD5	MAP2756c	—	—	—	+	+	+++	+++
MAP_RD5	MAP2765c	—	—	—	+	+++	+++	+++
	MAP2770	+++	+++	+++	+++	+++	+++	+++
	MAP3715	—	+	—	—	—	+++	+++
MAP_RD6	MAP3726	—	—	—	+	+	+++	+++
MAP_RD6	MAP3751	+	—	+	+	+++	+++	+++
MAP_RD6	MAP3775c	—	—	—	+	+	+++	+++
	MAP3783	+++	+++	+++	+++	+++	+++	+++

<sup>a</sup> —, ORF adjacent to divergent region and predicted to be present by microarray hybridization.

<sup>b</sup> +++, product observed; —, no product observed; +, weak product observed.

reveal genomic features that contribute to pathogenicity. Our results also illustrate the limitations of comparing highly divergent genomes with DNA microarrays.

**PCR analysis.** Genomic DNA from a limited number of mycobacterial isolates was used as the template for PCRs under the conditions described for the construction of the *M. avium* subsp. *paratuberculosis* K10 microarray. Primers designed to amplify ORFs from the shared divergence regions as well as ORFs flanking these regions were used to confirm the microarray findings (Table 4). With the exception of MAP0106c and MAP0850, all of the ORFs selected from the shared divergent regions either were not detected by PCR or had a low yield of product when genomic DNA from the nonparatuberculosis ORFs was used as template. The presence of small amounts of PCR products for ORFs from the divergent regions indicates that some of these regions represent nucleotide sequence variations, while the absence of PCR products reveals that some regions classified as divergent may be at least partially due to whole-gene deletions. One of the ORFs immediately flanking MAP\_RD6 (MAP3715) was not detected by PCR for several isolates, indicating that this region may be larger than originally determined by microarray analysis.

**Southern blotting.** Genomic DNA from four *M. avium* subsp. *paratuberculosis* and four nonparatuberculosis isolates was subjected to restriction digestion and used for Southern blotting (Table 5). All eight isolates were detected with a probe for MAP0005, which carries the highly conserved gene *gyrB*. Only the *M. avium* subsp. *paratuberculosis* isolates were detected by probes for MAP0854 and MAP3815, which encode hypothetical proteins and were classified as divergent by microarray analysis for all of the nonparatuberculosis isolates with the exception of 6006, for which no data were reported. These results further support the findings of the microarray-based analyses.

**Unique sequences.** Several *M. avium* subsp. *paratuberculosis* ORFs that have no homology to any sequences currently available in public databases have been previously described (4, 16, 20). We updated this subset of unique sequences by running TBLASTX searches against the complete genome of *M. avium* subsp. *avium* 104 and all of the sequences in GenBank as described in Materials and Methods. Of the 32 ORFs identified as unique to *M. avium* subsp. *paratuberculosis* by TBLASTX analysis, comparative genomic hybridizations revealed that 30 were classified as divergent in *M. avium* subsp. *avium* 104, one ORF was intermediate, and the remaining ORF had no data reported (Fig. 2). In the other nonparatuberculosis isolates examined there were from one to five unique ORFs identified as being present, including MAP1718c and MAP3383, which were both classified as present in multiple nonparatuberculosis isolates. None of the *M. avium* subsp. *paratuberculosis*-specific sequences were identified as divergent in the *M. avium* subsp. *paratuberculosis* isolates. These data reveal the utility of using

TABLE 5. Southern hybridization of purified mycobacterial genomic DNA with probes derived from *M. avium* subsp. *paratuberculosis* K10 genomic DNA

Isolate	Organism	Result with probe:		
		MAP0005	MAP0854	MAP3815
MAA104	<i>M. avium</i> subsp. <i>avium</i>	+	—	—
6004	<i>M. avium</i> subsp. <i>avium</i>	+	—	—
6058	<i>M. avium</i> subsp. <i>silvaticum</i>	+	—	—
K10	<i>M. avium</i> subsp. <i>paratuberculosis</i>	+	+	+
19698	<i>M. avium</i> subsp. <i>paratuberculosis</i>	+	+	+
6079	<i>M. avium</i> subsp. <i>paratuberculosis</i>	+	+	+
4138	<i>M. avium</i> subsp. <i>paratuberculosis</i>	+	+	+
6010	<i>M. intracellulare</i>	+	—	—



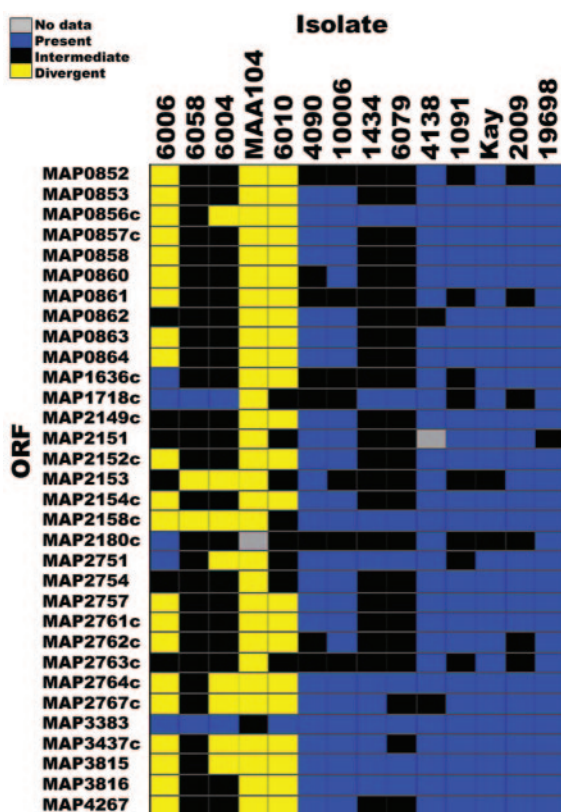


FIG. 2. Microarray hybridization results for *M. avium* subsp. *paratuberculosis*-specific ORFs identified by TBLASTX searches of GenBank and the genome sequence of *M. avium* subsp. *avium* 104.

DNA hybridization studies to further establish the specificity of sequences identified as unique by database searches.

## DISCUSSION

MAC represents an intriguing group of mycobacteria whose members are genetically very similar and yet have extremely variable pathogenicity and host range. Improved diagnostic tests for pathogenic isolates are confounded by the abundance of nonpathogenic mycobacteria present in the environment. In order to better understand the genetic conservation among isolates of *M. avium* subsp. *paratuberculosis* and between other MAC bacteria, we have utilized a whole-genome cDNA microarray based on the sequence of *M. avium* subsp. *paratuberculosis* K10 to compare genomic content at the level of individual genes via competitive DNA hybridization.

The *M. avium* subsp. *paratuberculosis* isolates examined in this study showed a high degree of genetic conservation compared to isolate K10, as no ORFs were classified as divergent in the *M. avium* subsp. *paratuberculosis* isolates. Although this result is not altogether unexpected, it is notable that the isolates were cultured from several different host species and geographic locations. Our findings have provided a more rigorous confirmation of previous reports indicating that bovine *M. avium* subsp. *paratuberculosis* isolates form a distinct subgroup based on conserved genomic features, as it appears that this conservation extends down to the gene level for a majority of the ORFs identified in the genome sequence. Based on

these results, future studies examining additional *M. avium* subsp. *paratuberculosis* isolates from a variety of host species and geographic locations are warranted to determine if this level of conservation encompasses a larger population.

Many of the ORFs classified as divergent in the nonparatuberculosis MAC group were shared among isolates MAA104, 6004, 6010, 6006, and 6058 and could be grouped into clusters of adjacent ORFs on the genome. In four of these clusters, phage integrase genes were found to be present, indicating that these regions of divergence may have been acquired by *M. avium* subsp. *paratuberculosis* K10 via phage-mediated recombination. Additionally, four of these gene clusters contained large regions in which the average GC content was significantly lower than the genome average. The low GC content of these gene clusters and their divergence from other mycobacterial genomes support the theory that the genetic material in these regions may have been acquired via horizontal gene transfer. Although not specifically shown by this study, some of these regions may be deleted from the MAC isolates relative to *M. avium* subsp. *paratuberculosis*, possibly giving investigators clues to identifying genes involved in the pathogenicity of *M. avium* subsp. *paratuberculosis*. Notably, recent work by Semret and coworkers has revealed that *M. avium* subsp. *avium* isolates contain regions of genetic sequence that are not present in *M. avium* subsp. *paratuberculosis*, findings that are complementary to those presented here (26).

One of the regions of shared divergence in the nonparatuberculosis MAC isolates included four ORFs (MAP2189, MAP2190, MAP2192, and MAP2193) homologous to the *mce* gene family originally identified in *M. tuberculosis* (2). Although the ORFs were present in all of the *M. avium* subsp. *paratuberculosis* isolates, they were found to be divergent or intermediate in all of the nonparatuberculosis MAC isolates. Several other clusters of *mce* genes were classified as present in all of the isolates examined. These findings are similar to those reported by Zumarraga and coworkers, who observed that one of the *mce* gene clusters present in *M. tuberculosis* was missing from the *Mycobacterium bovis* genome (31). It is notable that *M. tuberculosis* and *M. bovis* represent another group of mycobacteria with very similar genomic content but varying pathogenicity and host specificity. The divergence of a specific cluster of *mce*-like genes among MAC isolates and the previously established role of *mce* genes in macrophage invasion and survival suggest that this group of genes may confer a specific advantage in the infection of bovine macrophages, a hypothesis that remains to be directly tested.

Our analysis of *M. avium* subsp. *silvaticum* isolate 6006 yielded several unexpected results. Compared with the other nonparatuberculosis MAC mycobacteria examined, 6006 had nearly half the number genes classified as divergent or intermediate compared to *M. avium* subsp. *paratuberculosis* K10. Additionally, five sequences identified as unique to *M. avium* subsp. *paratuberculosis* were detected in 6006. Notably, while 6006 was isolated from a roe deer, it had been previously classified as atypical or wood pigeon *M. avium* subsp. *avium* (13). This isolate type is commonly cultured from avian species, has been found to share genetic and biochemical properties with both *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* (15, 24, 30), and was later reclassified as a novel subspecies of *M. avium* (27). As Jorgensen and Clausen



note in their original description of this isolate, it is intriguing that a member of this subspecies is capable of infecting a mammalian host. Because *M. avium* subsp. *silvaticum* isolates appear to share conserved genomic features with both *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis*, it is tempting to speculate that *M. avium* subsp. *silvaticum* may represent an intermediate step in the adaptation of avian species to mammalian hosts or vice versa.

The *M. avium* subsp. *paratuberculosis* isolates from a goat, sheep, and mink were not found to be divergent from the K10 bovine isolate. Isolates from cattle and goats have previously been found to be genetically similar, while sheep isolates commonly form a distinct subgroup from other *M. avium* subsp. *paratuberculosis* isolates (7, 19, 21). *M. avium* subsp. *paratuberculosis* infections of mink have not previously been reported. A HinfI restriction digest of the IS1311 IS from each isolate revealed that all of the *M. avium* subsp. *paratuberculosis* isolates examined contained a polymorphism found exclusively in nonsheep isolates of *M. avium* subsp. *paratuberculosis*, while none of the nonparatuberculosis MAC isolates contained this polymorphism. Additional isolates from these animals will need to be examined and compared to these findings, as it appears that they were infected from bovine sources that are not necessarily representative of the isolates most commonly cultured from the respective species.

None of the ORFs identified as unique to *M. avium* subsp. *paratuberculosis* in this study were classified as divergent in the *M. avium* subsp. *paratuberculosis* isolates examined. *M. avium* subsp. *silvaticum* isolate 6006 showed the fewest number of ORFs classified as divergent among the nonparatuberculosis MAC isolates, five of which were identified by a TBLASTX search of publicly available sequence data as specific for *M. avium* subsp. *paratuberculosis*. With the exception of MAP1718c and MAP3383, the remaining 30 paratuberculosis-specific ORFs were classified as divergent or intermediate in the other nonparatuberculosis isolates. Isolate 6006 in particular may represent a broader group of isolates that resist rigid classification as either *M. avium* subsp. *paratuberculosis* or other *M. avium* subspecies. As genomic DNA from additional mycobacterial isolates becomes available, a microarray-based approach will be useful to further confirm the specificity of these sequences because it can quickly and easily establish the presence or absence of all potentially unique sequences in a single experiment.

It should be noted that, due to the use of *M. avium* subsp. *paratuberculosis* K10 as the template for construction of the cDNA microarray, our analyses will not detect genes that are absent in K10 but present in the other isolates examined. Alternate methodologies such as subtractive hybridization will need to be utilized in order to identify divergent sequences in other mycobacteria. Additionally, the presence of small genetic polymorphisms would not be detected by cDNA microarray analysis. The advantage of using cDNA microarrays for this type of comparative genomic analysis is that differences between isolates can be localized to individual genes rather than gross genomic features.

Our comparative genomic analysis of MAC bacteria has revealed that the subset of bovine *M. avium* subsp. *paratuberculosis* isolates examined in this study have a high degree of genetic conservation, with 99% of the ORFs examined being

classified as intermediate or present. The closely related *M. avium* subsp. *avium*, *M. intracellulare*, and *M. avium* subsp. *silvaticum* displayed various amounts of genetic divergence from *M. avium* subsp. *paratuberculosis* K10 and contain core groups of ORFs that appear to diverge across all of the nonparatuberculosis MAC isolates examined. These findings are similar to those reported by Dziejman and coworkers, who compared isolates responsible for causing the sixth, seventh, and eighth *Vibrio cholerae* pandemics by hybridizing genomic DNA from these strains to a whole-genome microarray based on the sequence of the *V. cholerae* El Tor isolate (10). On this basis, these *V. cholerae* strains appear remarkably similar, sharing 99% of their genes with the El Tor isolate. However, other pathogens, such as *Staphylococcus aureus* and *Helicobacter pylori*, show much greater divergence among clinical isolates, with a 12% maximum difference in genome content seen between individual strains of *S. aureus* (11) and an 18% difference between isolates of *H. pylori* (22).

An understanding of the genomic diversity among MAC mycobacteria should provide additional insight into the mechanisms of host specificity. Additionally, it will allow for improvements in diagnostic tests to distinguish between nonpathogenic environmental mycobacteria and pathogenic mycobacteria such as the Johne's disease-causing bacillus.

#### ACKNOWLEDGMENTS

The expert technical assistance of Janis K. Hansen is greatly appreciated. *M. avium* subsp. *avium* 104 was a gift from Luiz Bermudez.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

This work was supported by the USDA's Agricultural Research Service.

#### REFERENCES

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- 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.
- Bannantine, J. P., E. Baechler, Q. Zhang, L. Li, and V. Kapur. 2002. Genome scale comparison of *Mycobacterium avium* subsp. *paratuberculosis* with *Mycobacterium avium* subsp. *avium* reveals potential diagnostic sequences. *J. Clin. Microbiol.* **40**:1303–1310.
- Bannantine, J. P., J. K. Hansen, M. L. Paustian, A. Amonsin, L. L. Li, J. R. Stabel, and V. Kapur. 2004. Expression and immunogenicity of proteins encoded by sequences specific to *Mycobacterium avium* subsp. *paratuberculosis*. *J. Clin. Microbiol.* **42**:106–114.
- Bannantine, J. P., Q. Zhang, L. L. Li, and V. Kapur. 2003. Genomic homogeneity between *Mycobacterium avium* subsp. *avium* and *Mycobacterium avium* subsp. *paratuberculosis* belies their divergent growth rates. *BMC Microbiol.* **3**:10.
- Barclay, R., and C. Ratledge. 1983. Iron-binding compounds of *Mycobacterium avium*, *M. intracellulare*, *M. scrofulaceum*, and mycobactin-dependent *M. paratuberculosis* and *M. avium*. *J. Bacteriol.* **153**:1138–1146.
- Bauerfeind, R., S. Benazzi, R. Weiss, T. Schliesser, H. Willems, and G. Baljer. 1996. Molecular characterization of *Mycobacterium paratuberculosis* isolates from sheep, goats, and cattle by hybridization with a DNA probe to insertion element IS900. *J. Clin. Microbiol.* **34**:1617–1621.
- Brosch, R., A. S. Pym, S. V. Gordon, and S. T. Cole. 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol.* **9**:452–458.
- Cleveland, W. S., and S. J. Devlin. 1988. Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**:596–610.
- Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos. 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. USA* **99**:1556–1561.

11. Fitzgerald, J. R., and J. M. Musser. 2001. Evolutionary genomics of pathogenic bacteria. *Trends Microbiol.* **9**:547–553.
12. Hurley, S. S., G. A. Splitter, and R. A. Welch. 1989. Development of a diagnostic test for Johne's disease using a DNA hybridization probe. *J. Clin. Microbiol.* **27**:1582–1587.
13. Jorgensen, J. B., and B. Clausen. 1976. Mycobacteriosis in a roe-deer caused by wood-pigeon mycobacteria. *Nord. Vetmed.* **28**:539–546.
14. Kim, C. C., E. A. Joyce, K. Chan, and S. Falkow. 2002. Improved analytical methods for microarray-based genome-composition analysis. *Genome Biol.* **3**:research0065.1–research0065.17. [Online.]
15. Levy-Frebault, V. V., M. F. Thorel, A. Varnerot, and B. Gicquel. 1989. DNA polymorphism in *Mycobacterium paratuberculosis*, "wood pigeon mycobacteria," and related mycobacteria analyzed by field inversion gel electrophoresis. *J. Clin. Microbiol.* **27**:2823–2826.
16. Li, L., J. P. Bannantine, Q. Zhang, A. Amonsin, B. J. May, D. Alt, S. Kanjilal, and V. Kapur. Unpublished data.
17. Matthews, P. R., and A. McDiarmid. 1979. The production in bovine calves of a disease resembling paratuberculosis with a *Mycobacterium* sp isolated from a woodpigeon (*Columba palumbus* L.). *Vet. Rec.* **104**:286.
18. Misener, S., and S. A. Krawetz. 2000. *Bioinformatics methods and protocols*. Humana Press, Totowa, N.J.
19. Motiwala, A. S., M. Strother, A. Amonsin, B. Byrum, S. A. Naser, J. R. Stabel, W. P. Shulaw, J. P. Bannantine, V. Kapur, and S. Sreevatsan. 2003. Molecular epidemiology of *Mycobacterium avium* subsp. *paratuberculosis*: evidence for limited strain diversity, strain sharing, and identification of unique targets for diagnosis. *J. Clin. Microbiol.* **41**:2015–2026.
20. Paustian, M. L., A. Amonsin, V. Kapur, and J. P. Bannantine. 2004. Characterization of novel coding sequences specific to *Mycobacterium avium* subsp. *paratuberculosis*: implications for diagnosis of Johne's disease. *J. Clin. Microbiol.* **42**:2675–2681.
21. Pavlik, I., L. Bejckova, M. Pavlas, Z. Rozsypalova, and S. Koskova. 1995. Characterization by restriction endonuclease analysis and DNA hybridization using IS900 of bovine, ovine, caprine and human dependent strains of *Mycobacterium paratuberculosis* isolated in various localities. *Vet. Microbiol.* **45**:311–318.
22. Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. USA* **97**:14668–14673.
23. Sambrook, J., and D. W. Russell. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
24. Saxegaard, F., and I. Baess. 1988. Relationship between *Mycobacterium avium*, *Mycobacterium paratuberculosis* and "wood pigeon mycobacteria." Determinations by DNA-DNA hybridization. *APMIS* **96**:37–42.
25. Saxegaard, F., I. Baess, and E. Jantzen. 1988. Characterization of clinical isolates of *Mycobacterium paratuberculosis* by DNA-DNA hybridization and cellular fatty acid analysis. *APMIS* **96**:497–502.
26. Semret, M., G. Zhai, S. Mostowy, C. Cleto, D. Alexander, G. Cangelosi, D. Cousins, D. M. Collins, D. van Soolingen, and M. A. Behr. 2004. Extensive genomic polymorphism within *Mycobacterium avium*. *J. Bacteriol.* **186**:6332–6334.
27. Thorel, M. F., M. Krichevsky, and V. V. Levy-Frebault. 1990. Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. *Int. J. Syst. Bacteriol.* **40**:254–260.
28. Tizard, M., T. Bull, D. Millar, T. Doran, H. Martin, N. Sumar, J. Ford, and J. Hermon-Taylor. 1998. A low G+C content genetic island in *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *silvaticum* with homologous genes in *Mycobacterium tuberculosis*. *Microbiology* **144**:3413–3423.
29. Whittington, R., I. Marsh, E. Choy, and D. Cousins. 1998. Polymorphisms in IS1311, an insertion sequence common to *Mycobacterium avium* and *M. avium* subsp. *paratuberculosis*, can be used to distinguish between and within these species. *Mol. Cell. Probes* **12**:349–358.
30. Yoshimura, H. H., and D. Y. Graham. 1988. Nucleic acid hybridization studies of mycobactin-dependent mycobacteria. *J. Clin. Microbiol.* **26**:1309–1312.
31. Zumarraga, M., F. Bigi, A. Alito, M. I. Romano, and A. Cataldi. 1999. A 12.7 kb fragment of the *Mycobacterium tuberculosis* genome is not present in *Mycobacterium bovis*. *Microbiology* **145**:893–897.