JOURNAL OF BACTERIOLOGY, Jan. 2006, p. 711–723 Vol. 188, No. 2
0021-9193/06/$08.00+0 doi:10.1128/JB.188.2.711–723.2006

Whole-Genome Plasticity among *Mycobacterium avium* Subspecies: Insights from Comparative Genomic Hybridizations†

Chia-wei Wu,[1] Jeremy Glasner,[1] Michael Collins,[2] Saleh Naser,[3] and Adel M. Talaat[1]*

Departments of Animal Health and Biomedical Sciences[1] and Pathobiological Sciences,[2] University of Wisconsin—Madison, 1656 Linden Drive, Madison, Wisconsin 53706, and Department of Molecular Biology and Microbiology, University of Central Florida, Orlando, Florida[3]

Received 26 August 2005/Accepted 20 October 2005

Infection with *Mycobacterium avium* subsp. *paratuberculosis* causes Johne's disease in cattle and is also implicated in cases of Crohn's disease in humans. Another closely related strain, *M. avium* subsp. *avium*, is a health problem for immunocompromised patients. To understand the molecular pathogenesis of *M. avium* subspecies, we analyzed the genome contents of isolates collected from humans and domesticated or wildlife animals. Comparative genomic hybridizations indicated distinct lineages for each subspecies where the closest genomic relatedness existed between *M. avium* subsp. *paratuberculosis* isolates collected from human and clinical cow samples. Genomic islands ($n = 24$) comprising 846 kb were present in the reference *M. avium* subsp. *avium* strain but absent from 95% of *M. avium* subsp. *paratuberculosis* isolates. Additional analysis identified a group of 18 *M. avium* subsp. *paratuberculosis*-associated islands comprising 240 kb that were absent from most of the *M. avium* subsp. *avium* isolates. Sequence analysis of DNA regions flanking the genomic islands identified three large inversions in addition to several small inversions that could play a role in regulation of gene expression. Analysis of genes encoded in the genomic islands reveals factors that are probably important for various mechanisms of virulence. Overall, *M. avium* subsp. *avium* isolates displayed a higher level of genomic diversity than *M. avium* subsp. *paratuberculosis* isolates. Among *M. avium* subsp. *paratuberculosis* isolates, those from wildlife animals displayed the highest level of genomic rearrangements that were not observed in other isolates. The presented findings will affect the future design of diagnostics and vaccines for Johne's and Crohn's diseases and provide a model for genomic analysis of closely related bacteria.

DNA rearrangements are responsible for genomic diversity in microbial systems and usually contribute to the fitness of a pathogen in specific microenvironments (24). Some of this variability leads to adaptation to a specific microenvironment, while other rearrangements are the products of the coexistence of recombinogenic microbes in an environment supportive of genetic exchange. For a group of closely related organisms such as members of *Mycobacterium avium* complex (MAC), including *M. avium* subspecies *avium*, *M. avium* subspecies *paratuberculosis*, and *M. intracellulare*, it is intriguing to investigate the genome contents of each organism and its relationship to the host microenvironment where these organisms evolve. Both *M. avium* subspecies *avium* and *M. intracellulare* are opportunistic pathogens widely distributed in the environment and can cause disseminated tuberculosis in immunocompromised patients (28, 44). In contrast, *M. avium* subsp. *paratuberculosis* is an obligate pathogen of ruminants causing Johne's disease characterized by chronic enteritis, with severe economic losses for the dairy industry (22). Recent reports also implicated *M. avium* subsp. *paratuberculosis* in cases of Crohn's disease in humans (27) in which patients suffer from chronic enteritis and intestinal pathology that is reminiscent to Johne's disease in cattle. Under laboratory growth conditions, *M. avium* subsp. *paratuberculosis* is a slow-growing mycobacterium that usually depends on the presence of mycobactin-J for in vitro growth, a criterion differentiating isolates of that subspecies from *M. avium* subspecies *avium* isolates. Additionally, *M. avium* subspecies *avium* isolates display more colony polymorphism than *M. avium* subsp. *paratuberculosis* isolates when grown on solid medium (10). Using comparative genomic hybridizations, we examined several isolates belonging to the MAC group to better understand the changes responsible for adaptation to different microenvironments and to identify possible genomic rearrangements that could explain their divergent phenotypes.

Several analyses were attempted to examine diversity among members of MAC strains. Using sequence analysis of the *dnaJ* gene to assess genetic diversity among *M. avium* subspecies *avium* strains indicated a limited diversity among animal and human isolates (25). However, experiments examining restriction fragment length polymorphism in the *hsp65* gene showed greater variability and suggested that there are distinct lineages of strains that infect animals and strains that infect humans (29). On a genome-wide level, long oligonucleotide microarrays identified large sequence polymorphisms in comparisons of *M. avium* subspecies *avium* and *M. avium* subsp. *paratuberculosis*, including polymorphisms affecting the mycobactin biosynthesis pathway (36), despite the presence of >98% identity between both genomes at the nucleotide level (31). A more recent study of genomic differences between *M. avium* subsp. *paratuberculosis* and *M. avium* subspecies *avium* confirmed this polymorphism among *M. avium* subspecies *avium* strains (32).

* Corresponding author. Mailing address: Department of Animal Health and Biomedical Sciences, University of Wisconsin—Madison, 1656 Linden Drive, Madison, WI 53706-1581. Phone: (608) 262-2861. Fax: (608) 262-7420. E-mail: atalaat@wisc.edu.

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

TABLE 1. Strains used in this study

Species	Strain	Host	Sample origin	Location
M. avium subsp. *paratuberculosis*	k10	Cow	Feces	Wisconsin
	ATCC19698	Cow	Feces	Unknown
	JTC33666	Turkomen markhor (goat)	Feces	California
	JTC33770	Cow	Feces	Wisconsin
	CW303	Cow	Feces	Wisconsin
	1B	Human	Ileum	Florida
	3B	Human	Ileum	Florida
	4B	Human	Ileum	Florida
	5B	Human	Ileum	Florida
	DT3	British red deer	Feces	Unknown
	DT9	African eland	Feces	Unknown
	DT12	Chinese Reeve's muntjac (deer)	Ileum	Unknown
	DT19	White rhino	Feces	Unknown
	JTC1281	Oryx	Lymph node	Florida
	JTC1282	Cow	Lymph node	Wisconsin
	JTC1283	Cow	Feces	Georgia
	JTC1285	Goat	Feces	Virginia
	JTC1286	Cow	Ileum	Wisconsin
M. avium subsp. *avium*	104	Human	Blood	Unknown
	T93	Cow	Feces	Texas
	T99	Cow	Feces	Texas
	T100	Cow	Feces	Texas
	DT30	Angolan springbok	Feces	Unknown
	DT44	Formosan Reeve's muntjac (deer)	Lymph node	Unknown
	DT78	Water buffalo	Ileum	Unknown
	DT84	Lowland wisent	Lymph node	Unknown
	DT247	Cuvier's gazelle	Lymph node	Unknown
	JTC956	Ankoli	Feces	Florida
	JTC981	Bongo	Feces	Florida
	JTC982	Nyala	Feces	Florida
	JTC1161	Cow	Feces	Florida
	JTC1262	Bison	Lymph node	Montana
	JTC33793	Dama gazelle	Feces	Indiana
M. intracellulare	$mc^2$76	Human	Sputum	Unknown

The genome sequences of both *M. avium* subspecies *avium* (http://www.tigr.org) and *M. avium* subsp. *paratuberculosis* (20) are currently available, which allowed us to provide a higher-resolution analysis of *M. avium* subspecies genomes.

The main objective in the present investigation was to identify genomic rearrangements among subspecies of *M. avium* to provide insights into the evolution of strains with distinct host preference and disease etiologies. We employed high-density oligonucleotide microarrays covering the entire *M. avium* genome to profile the genome contents of isolates from both animal and human sources. Both *M. avium* subspecies *avium* and *M. avium* subsp. *paratuberculosis* isolates clustered into distinct lineages regardless of the source of samples. This comparative genomic analysis provided the most comprehensive list of genomic island (GI) polymorphisms among different subspecies of *M. avium*. We used the identified islands to examine the genome synteny (gene order) of *M. avium* subspecies *avium* strains, which revealed several areas of genomic inversions that could play a role in antigenic variations. The presented findings will impact our understanding of microbial evolution, especially for pathogens from a closely related progenitor. The results also will help define a better set of diagnostics and vaccine candidates for use against pathogenic subspecies of *M. avium*.

MATERIALS AND METHODS

Bacterial strains. Mycobacterial isolates ($n = 34$) examined in this report were collected from different human and domesticated or wildlife animal specimens representing different geographical regions within the United States (Table 1). *Mycobacterium avium* subsp. *paratuberculosis* strain k10 (14), *M. avium* subsp. *avium* strain 104 (*M. avium* 104) (43), and *M. intracellulare* were obtained from Raul Barletta (University of Nebraska). *M. avium* subsp. *paratuberculosis* ATCC 19698 and other animal isolates used throughout this study were obtained from the Johne's Testing Center, University of Wisconsin—Madison, while the *M. avium* subsp. *paratuberculosis* human isolates were obtained from Saleh Naser (University of Central Florida). All strains were grown in Middlebrook 7H9 broth (Difco, Sparks, MD) supplemented with 0.5% glycerol, 0.05% Tween 80, and 10% ADC (2% glucose, 5% bovine serum albumin fraction V, and 0.85% NaCl) at 37°C (7). For *M. avium* subsp. *paratuberculosis* strains, 2 μg/ml of mycobactin-J (Allied Monitor, Fayette, MO) was also added for optimal growth.

Microarray design. Throughout this study, we used oligonucleotide microarrays synthesized in situ on glass slides by use of a maskless array synthesizer (1). Probe sequences were chosen from the complete genome sequence of *M. avium* subspecies *avium* 104. Preliminary sequence data for *M. avium* subspecies *avium* strain 104 were obtained from The Institute for Genomic Research through the website at http://www.tigr.org, and we predicted open reading frames (ORFs) by use of GeneMark (21). For every ORF, 18 pairs of 24-mer sequences were selected as probes. Each pair of probes consists of a perfect match (PM) probe along with a mismatch (MM) probe with mutations at the 6th and 12th positions of the corresponding PM probes. A total of ~185,000 unique probe sequences were synthesized on derivatized glass slides by NimbleGen System, Inc. (Madison, WI) (37).

Genomic DNA extraction and labeling. Genomic DNA (gDNA) was extracted using a modified cetyltrimethylammonium bromide-based protocol (40) followed by two rounds of ethanol precipitation. For each hybridization, 10 μg of genomic DNA was digested with 0.5 U of RQ1 DNase (Promega, Madison, WI) until the fragmented DNA was in the range of 50 to 200 bp (examined on a 2% agarose gel). The reaction was stopped by adding 5 μl of DNase stop solution and incubating at 90°C for 5 min. Digested DNA was purified using YM-10 microfilters (Millipore, Billerica, MA). Genomic DNA hybridizations were prepared by an end-labeling reaction. Biotin was added to purified mycobacterial DNA fragments (10 μg) by use of terminal deoxynucleotide transferase (Promega) in the presence of 1 μM biotin-N6-ddATP (PerkinElmer Life Sciences Inc., Boston, MA) at 37°C for 1 h. Before hybridization, biotin-labeled gDNA was heated to 95°C for 5 min followed by 45°C for 5 min and centrifuged at 14,000 rpm for 10 min before addition to the microarray slide (1). After microarray hybridization for 12 to 16 h, slides were washed in nonstringent (6× SSPE [1× SSPE is 0.18 NaCl, 10 mM NaH_2PO_4, and 1 mM EDTA {pH 7.7}] and 0.01% Tween 20) and stringent (100 mM MES, 0.1 M NaCl, 0.01% Tween 20) buffers for 5 min each, followed by fluorescent detection by addition of Cy3 streptavidin (Amersham Biosciences Corp., Piscataway, NJ). Washed microarray slides were dried by argon gas and scanned with an Axon GenPix 4000B laser scanner (Axon Instruments, Union City, CA) at 5-μm resolution. Replicate microarrays were hybridized for every genome tested in this study. Two hybridizations of the same genomic DNA with high reproducibility (correlation coefficient >0.9) were allowed for downstream analysis.

Data analysis and prediction of genomic deletions. The images of scanned microarray slides were analyzed using specialized software (NimbleScan) developed by NimbleGen System Inc. The average signal intensity of an MM probe was subtracted from that of the corresponding PM probe. The median value of all PM-MM intensities for an ORF was used to represent the signal intensity for the ORF. The median intensity value for each slide was normalized by multiplying each signal by a scaling factor that was 1,000 divided by the average of all median intensities for that array. To compare hybridization signals generated from each of the genomes to that of *M. avium* subsp. *avium* strain 104, the normalized data from replicate hybridizations were then exported to an R language program with EBarrays package version 1.1, which employs a Bayesian statistical model for pair-wise genomic comparisons using a log-normal-normal model (19). Genes with a probability of differential expression (PDE) larger than 0.5 were considered significantly different between the genomes of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis*. The hybridization signals corresponding to each gene of all investigated genomes were plotted according to the genomic location of *M. avium* subsp. *avium* strain 104 by use of GenVision software (DNASTAR Inc., Madison, WI). The same data set was also analyzed using MultiExperiment Viewer 3.0 (13) to identify common cluster patterns among mycobacterial isolates.

PCR verification and sequence analysis. To confirm the results predicted by microarray hybridizations, we employed a three-primer PCR protocol to amplify the regions flanking predicted genomic islands. For every island, one pair of primers (F and R1) was designed upstream of the target region and a third primer (R2) was designed downstream of the same region. The primers were designed so that expected lengths of the products were less than 1.5 kb between F and R1 and less than 3 kb between F and R2 when amplified from the genomes with the deleted island. Each PCR mixture contained 1 M betaine, 50 mM potassium glutamate, 10 mM Tris-HCl (pH 8.8), 0.1% Triton X-100, 2 mM magnesium chloride, 0.2 mM deoxynucleoside triphosphates, 0.5 μM of each primer, 1 U *Taq* DNA polymerase (Promega), and 15 ng genomic DNA. The PCR cycling conditions were 94°C for 5 min followed by 30 cycles of 94°C for 1 min, 59°C for 1 min, and 72°C for 3 min. All PCR products were examined using 1.5% agarose gels and stained with ethidium bromide. To further confirm sequence deletions, amplicons flanking deleted regions were sequenced using a standard BigDye Terminator v. 3.1 (Applied Biosystems, Foster City, CA) and compared to the genome sequence of *M. avium* subsp. *paratuberculosis* or *M. avium* subsp. *avium* by use of BLAST analysis (2).

RESULTS

Microarray analysis of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes. The main goal of this study was to investigate the genomic rearrangements among *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* isolates from various hosts to understand their adaptive evolution in the host microenvironments. We began the analysis using five mycobacterial isolates and DNA microarrays and expanded our analysis to include an additional 29 isolates employing a more affordable technology of PCR followed by direct sequencing. All of the isolates were collected from human and domesticated or wildlife animal sources and had been previously identified at the time of isolation by use of standard culturing techniques for *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis*. The identity of each isolate was confirmed further by acid-fast staining and positive PCR amplification of IS*900* sequences from all isolates of *M. avium* subsp. *paratuberculosis* (15). Additionally, the growth of all *M. avium* subsp. *paratuberculosis* isolates was mycobactin-J dependent while that of all *M. avium* subsp. *avium* isolates was not. Before starting the microarray analysis, we also performed an *hsp65* PCR typing protocol (38) to ensure the identity of each isolate. The PCR typing protocol agreed with the results of an earlier characterization of all mycobacterial isolates used throughout this study (Fig. 1A).

To investigate the extent of variation among *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* isolates on a genome-wide scale, we used oligonucleotide microarrays designed from the *M. avium* subsp. *avium* strain 104 genome sequence. The GeneMark algorithm was used to predict potential ORFs (21) in the raw sequences of the *M. avium* genome obtained from TIGR. A total of 4,987 ORFs were predicted for *M. avium* subsp. *avium* compared to 4,350 ORFs predicted in *M. avium* subsp. *paratuberculosis* (31). Relaxed criteria (i.e., determinations of sequences at least 100 bp in length with a maximal permitted overlap of 30 bases between ORFs) for assigning ORFs were chosen to allow the use of a comprehensive representation of the genome to construct DNA microarrays. In similarity to the characteristics seen with other bacterial genomes, the average ORF length was ~1 kb. Using the ASAP comparative genomic software suite (16), the ORFs shared by *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* had an average identity of 98%, a result corroborated by others (4). BLAST analysis of the ORFs from both genomes showed that about 65% (n = 2,557) of the *M. tuberculosis* genes have a significant match ($E < 10^{-10}$) in the other genome. This preliminary analysis of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes can be downloaded from the ASAP web site (http://www.genome.wisc.edu/tools/asap.htm) (see tables in the supplementary material). To test the reliability of genomic DNA extraction protocols and array hybridizations, the signal intensities of replicate hybridizations of the same mycobacterial genomic DNA were compared using scatter plots. ORFs with positive hybridization signals in at least 10 probe pairs were normalized and used for downstream analysis to ensure the inclusion of only ORFs with reliable signals. In all replicates, independently isolated hybridized samples of gDNA had high correlation coefficients ($r >$ 0.9) (Fig. 1B).

To investigate the genomic relatedness among isolates compared to relatedness to the *M. avium* subsp. *avium* 104 strain, we employed a hierarchical cluster analysis to assess the similarity of the hybridization signals among isolates on a genome-wide level. *M. avium* subsp. *avium* isolates were more similar to each other than to the *M. avium* subsp. *paratuberculosis* isolates (Fig. 1C). Within the *M. avium* subsp. *paratuberculosis* cluster, the human and the clinical animal isolates were far more similar to each other than to the ATCC 19698 reference strain,

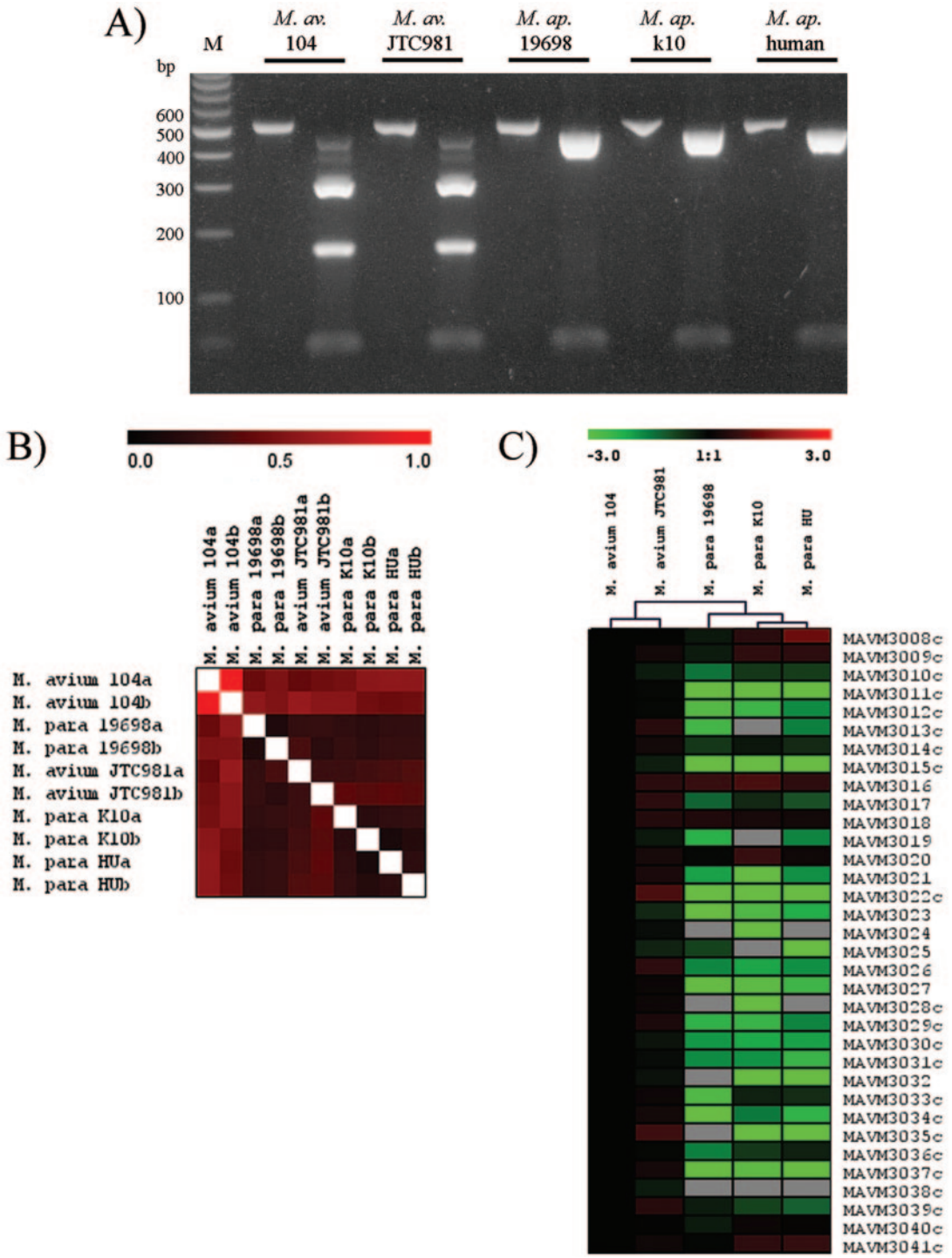

FIG. 1. Comparative genomic hybridization of *M. avium* subspecies by use of DNA microarrays. A) PCR confirmation of the identity of the examined genomes. An ethidium bromide-stained agarose gel (2%) shows the PstI digestion pattern (lane 2 of each set) of the PCR amplicons (lane 1 of each set) seen when different templates (name on top of each set) were used to amplify a 500-bp fragment of the *hsp65* gene. A 100-bp marker (Promega) is included. B) Reproducibility of the genomic microarray hybridizations. The overall Pearson's correlation values are plotted for biological replicates (denoted as a or b) of all mycobacterial genomic DNA used for microarray analysis. The black-to-red scale bar represents low to high correlation levels. Note that all replicates have *r* values close to 1. M. avium, *M. avium* subsp. *avium*; M. para, *M. avium* subsp. *paratuberculosis*. C) A dendrogram displaying the overall genomic hybridization signals generated from biological replicates of different mycobacterial isolates from animal or human (HU) sources. The name of each sample is indicated in the dendrogram tree. An example of the hierarchical cluster analysis of the hybridization signals from a region encompassing *M. avium* subsp. *avium* GIs 16 and 17 is chosen. The green-to-red color bar represents low to high log ratios of the hybridization signals generated from each genome relative to that of *M. avium* subsp. *avium* 104.

implying a closer relatedness between human and clinical isolates of *M. avium* subsp. *paratuberculosis*. Interestingly, despite the high degree of similarity between genes shared among isolates, hundreds of genes appeared to be missing from different genomes relative to *M. avium* genome. Most of the genes were found in clusters in the *M. avium* subsp. *avium* 104 genome, the reference strain used for designing the microarray chip (see supporting data). Consequently, regions absent from *M. avium* subsp. *avium* 104 but present in other genomes could not be identified in this analysis.

Large genomic deletions among *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* isolates. To better analyze the hybridization signals generated from examined genomes, a Bayesian statistical principle (EBarrays package) (19) was used to compare the hybridization signals generated from different isolates to the signals generated from the *M. avium* subsp. *avium* strain 104 genome. The Bayesian analysis estimates the likelihood of observed differences in ORF signals for each gene between each isolate and the *M. avium* subsp. *avium* 104 reference strain. Initial analysis of these data identified a large number of differences among isolates, including many ORFs scattered throughout the genome (Fig. 2A). PCR analysis of the deletions in few single genes did not confirm the microarrays data (data not shown), most likely because of the low cutoff value (PDE > 0.5) that we used for making decisions on deleted genes. Instead of increasing the PDE value, with the consequent missing of gene deletions, we chose to focus our analysis on the deletions that occurred in consecutive ORFs to better characterize large genomic regions that could contribute to a specific phenotype or pathotype. Additionally, we decided to use PCR and sequencing to confirm all deletions identified by microarrays where possible. When regions included three or more consecutive ORFs, they were defined as a GI regardless of the size. Applying such criterion for GIs, 24 islands were present in *M. avium* subsp. *avium* strain 104 but absent from all *M. avium* subsp. *paratuberculosis* isolates, regardless of the source of the *M. avium* subsp. *paratuberculosis* isolates (animal or human). The GIs ranged in size from 3 to 196 kb (Table 2), with a total of 846 kb encoding 759 ORFs. Interestingly, a clinical strain of *M. avium* subsp. *avium* (JTC981) was also missing seven GIs (nearly 518 kb) in common with all *M. avium* subsp. *paratuberculosis* isolates, in addition to the partial absence of five other GIs. This variability indicated a wide spectrum of genomic diversity among *M. avium* subsp. *avium* strains that was not evident among *M. avium* subsp. *paratuberculosis* isolates.

To confirm the absence of GI regions from isolates, we employed a strategy based on PCR amplification of the flanking regions of each GI followed by sequence analysis to confirm the missing elements. Because the size of most of the genomic island regions exceeds the amplification capability of a typical PCR, we designed three primers for each island, including one forward and two reverse primers (Fig. 2B). This strategy was successfully applied with 21 genomic islands, while amplification from the rest of the islands ($n = 3$) was not possible due to extensive genomic rearrangements. Overall, the PCR and sequencing verified the GI content as predicted by comparative genomic hybridizations (Table 2). The success of this strategy in identifying island deletions provided us with a robust protocol to examine several clinical isolates that could not otherwise be analyzed using the costly DNA microarrays.

Bioinformatic analysis of genomic islands. While we were working on this project, the genome sequence of *M. avium* subsp. *paratuberculosis* was completed and published (20). We reasoned that pair-wise BLAST analysis of the genome sequences of *M. avium* subsp. *avium* strain 104 and *M. avium* subsp. *paratuberculosis* strain k10 could further refine the ability to detect genomic rearrangements, especially for regions present in the *M. avium* subsp. *paratuberculosis* k10 genome but deleted from the *M. avium* subsp. *avium* 104 genome. The pair-wise comparison allowed us to better analyze the flanking sequences for each GI and to characterize the mechanism of genomic rearrangements among examined strains. As expected, BLAST analysis (E scores >0.001 and $<25\%$ sequence alignment between ORFs) correctly identified the deleted GIs in which ORFs of *M. avium* subsp. *avium* were missing from *M. avium* subsp. *paratuberculosis*, as detected by using the comparative genomic hybridization protocol. ORFs in a large proportion of each genome ($>75\%$) are likely orthologous ($>25\%$ sequence alignment of the ORF length and $>90\%$ sequence identity at the nucleotide level). This high degree of similarity between orthologues indicates a fairly recent ancestor. Looking for consecutive ORFs from *M. avium* subsp. *paratuberculosis* that do not have a BLAST match in *M. avium* subsp. *avium* identified sets of ORFs representing 18 GIs comprising 240 kb that are present only in the *M. avium* subsp. *paratuberculosis* genome (Table 3), among which seven islands were identified before (32).

Genes encoded within *M. avium* subsp. *avium*- and *M. avium* subsp. *paratuberculosis*-specific islands were analyzed using the BLASTP algorithm and the GenPept database (19 October 2004 release) to identify their potential functions. The BLAST results allowed the assignment of signature features to each island. As detailed in Table 3 and Table 4, with the presence of a large number of ORFs encoding mobile genetic elements (e.g., insertion sequences and prophages), several ORFs encode transcriptional regulatory elements, especially from the TetR family of regulators (23). The polymorphism in TetR regulators could be attributed to the fact that their sequences allow them to be amenable to rearrangements. Alternatively, it is possible that the bacteria are able to differentially acquire specific groups of genes suitable for a particular microenvironment.

Further analysis of the GIs identified islands in both *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* (such as MAV-7, MAV-12, and MAP-13) encoding different operons of the *mce* (mammalian cell entry) sequences that were shown to participate in the pathogenesis of *M. tuberculosis* (3, 8). Another island (MAV-17) encodes the *drrAB* operon for antibiotic resistance (11), which is a well-documented problem for treating *M. avium* subsp. *avium* infection in HIV patients (30). Interestingly, the GC percentages of the majority of *M. avium* subsp. *paratuberculosis*-specific islands (11/18) were at least 5% less than the average GC percentages of the *M. avium* subsp. *paratuberculosis* genome (69%) compared to only 3 GIs (out of 24) specific for the *M. avium* subsp. *avium* genome (Table 4) with lower-than-average GC percentages. The implication of this variation is discussed below.

Genomic deletions among field isolates of *M. avium* subsp. *avium*. Microarrays and PCR analysis of five mycobacterial isolates identified the presence of variable GIs between the *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis*

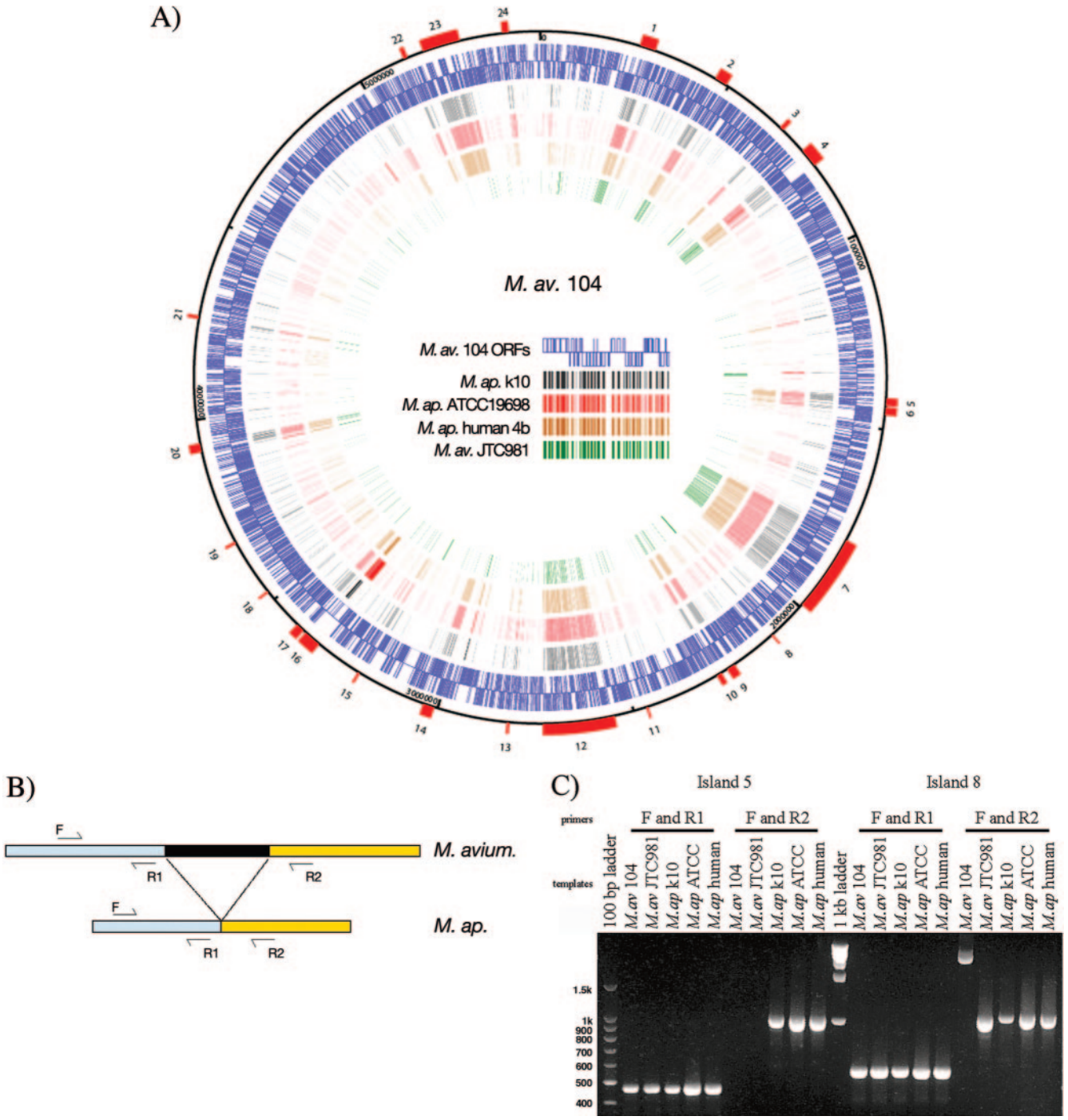

FIG. 2. Identification of genomic islands in the *M. avium* subsp. *avium* genome. A) A genome map based on *M. avium* subsp. *avium* sequences, displaying GIs deleted from the examined strains as predicted by DNA microarrays. Inner circles denote the microarray hybridization signals for each examined genome (see legend in center of panel). The outermost red boxes denote the locations of all GIs associated with *M. avium* subsp. *avium*. *M. av*., *M. avium* subsp. *avium*; *M. ap.*, *M. avium* subsp. *paratuberculosis*. B) A diagram illustrating the PCR and sequence-based strategy implemented to verify genomic deletions. Three primers for each island were designed, including a forward (F) and two reverse primers. C) PCR confirmation of genomic deletions. An ethidium bromide-stained agarose gel (1%) displays amplicons from two GIs by use of DNA templates extracted from five different isolates of *M. avium* subsp. *avium*. The first and middle lanes are occupied by the 100 bp and 1 kb DNA markers (Promega). Note that GI 8 was only 3 kb, i.e., within the amplification range for *M. avium* subsp. *avium* as well as *M. avium* subsp. *paratuberculosis* isolates.

TABLE 2. List of genomic regions that displayed different hybridization signals determined using DNA microarrays designed from the genome of the *M. avium* subsp. *avium* 104 strain

Island	Coordinate (bp)[a]		*M. avium* subsp. *avium* strain or isolate[b]				PCR and sequence confirmation
	Start	End	k10	19698	Human	JTC981	
1	254,394	294,226	−	−	−	−	Yes
2	461,414	492,800	−	−	−	−	Yes
3	666,033	675,725	−	−	−	−	Yes
4	747,095	794,450	−	−	−	−	Yes
5	1,421,722	1,439,626	−	−	−	+	Yes
6	1,444,205	1,463,365	−	−	−	+	Yes
7	1,795,281	1,991,691	−	−	−	+/−	Yes
8	2,097,907	2,100,883	−	−	−	−	Yes
9	2,220,320	2,241,163	−	−	−	+/−	Yes
10	2,259,120	2,271,610	−	−	−	−	Yes
11	2,462,693	2,466,285	−	−	−	+	Yes
12	2,549,555	2,730,999	−	−	−	−	ND[c]
13	2,815,625	2,821,149	−	−	−	+	Yes
14	3,008,716	3,036,980	−	−	−	+	Yes
15	3,214,820	3,219,550	−	−	−	+	ND
16	3,340,393	3,384,549	−	−	−	+	Yes
17	3,392,586	3,413,804	−	−	−	+	ND
18	3,523,417	3,527,334	−	−	−	+/−	Yes
19	3,670,518	3,675,686	−	−	−	+	Yes
20	3,917,752	3,939,034	−	−	−	+/−	Yes
21	4,254,594	4,261,488	−	−	−	+/−	Yes
22	5,122,371	5,132,301	−	−	−	+	Yes
23	5,174,641	5,270,187	−	−	−	+	Yes
24	5,378,903	5,395,102	−	−	−	+	Yes

[a] Coordinates of start and end of island based on the genome sequence of *M. avium* subsp. *avium* strain 104.

[b] + or − denotes the presence or absence of genomic regions in the examined genomes, while +/− denotes an incomplete deletion.

[c] ND, not done.

genomes. To analyze the extent of such variations among clinical isolates circulating in both human and animal populations, we used PCR and a sequencing-based strategy to examine 28 additional *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* isolates collected from different geographical locations within the United States (Table 1). An additional isolate of *M. intracellulare* was included as a representative strain that belongs to the MAC group but is not a subspecies of *M. avium*. For PCR amplification, we examined GIs spatially scattered throughout the *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes (Table 5 and Table 6) to identify any potential rearrangements in all quarters of the genome. Because of the wide-spectrum diversity observed among *M. avium* genomes, four GIs (MAV-3, MAV-11, MAV-21, and MAV-23) were chosen to assess genomic rearrangements in clinical isolates. Alternatively, because of the limited diversity observed among *M. avium* subsp. *paratuberculosis* genomes, a total of six *M. avium* subsp. *paratuberculosis*-specific GIs (MAP-1, MAP-3, MAP-5, MAP-12, MAP-16, and MAP-17) were chosen for testing genomic rearrangements. As suggested from the initial comparative genomic hybridization results, clinical isolates of *M. avium* subsp. *paratuberculosis* showed a limited diversity with respect to the existence of *M. avium* subsp. *avium*-specific islands (DT9 clinical isolate from a red deer), indicating the clonal nature of this organism (Table 5). In contrast, *M. avium* subsp. *avium* isolates showed a different profile from those of both *M. avium* subsp. *avium* 104 and *M. avium* JTC981, indicating extensive variability within *M. avium* isolates. A similar pattern of genomic rearrangements was observed when *M. avium* subsp. *paratuberculosis*-specific GIs were analyzed using *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* isolates (Table 6). Interestingly, most of the *M. avium* subsp. *paratuberculosis* clinical isolates with GI deletions were from wildlife animals, suggesting that strains circulating in wildlife animals could provide a potential source for genomic rearrangements in *M. avium* subsp. *paratuberculosis*.

Combined with the hierarchical cluster analysis employed on the whole genome hybridizations, PCR and sequence analyses provided more evidence that genomic diversity is quite extensive among *M. avium* subsp. *avium* strains but much less limited in strains of *M. avium* subsp. *paratuberculosis*. Unfortunately, analysis of GIs was not conclusive when *M. intracellulare* was used, suggesting more rearrangements in the *M. intracellulare* than in

TABLE 3. Characteristics of *M. avium* subsp. *paratuberculosis*-specific (MAP) genomic islands deleted in the *M. avium* subsp. *avium* genome

Island	No. of ORFs	GC (%)	Island type	Size (bp)	Signature feature(s)
MAP-1	17	63.90	I	19,343	Transposition and TetR family transcriptional regulator genes
MAP-2	3	60.43	I	3,858	Conserved hypothetical proteins
MAP-3	3	66.16	I	2,915	Formate dehydrogenase alpha subunit
MAP-4	17	60.66	I	16,681	Transposition, unknown genes, and a possible prophage
MAP-5	12	69.56	I	14,191	Transposition and oxidoreductase genes, PPE family domain protein
MAP-6	6	57.73	II	8,971	Variable genes such as *drrC*
MAP-7	6	67.26	II	6,914	Transcriptional regulator *psrA* and biosynthesis genes
MAP-8	8	61.59	II	7,915	TetR family transcriptional regulator and unknown genes
MAP-9	10	65.49	II	11,202	Transposition, metabolic and TetR family transcriptional regulator genes
MAP-10	3	66.68	II	2993	Biosynthesis of cofactors, prosthetic groups, and carriers, transcriptional regulator, TetR family domain protein
MAP-11	4	62.89	I	2,989	Serine/threonine protein kinase and glyoxalase genes
MAP-12	11	61.08	I	11,977	Transposition, iron metabolism genes, and a prophage
MAP-13	19	66.01	II	19,977	TetR family transcriptional regulator and *mce* family proteins
MAP-14	19	65.76	II	19,315	Possible prophage and unknown proteins
MAP-15	3	62.93	I	4,143	Unknown proteins and a prophage function genes
MAP-16	56	64.32	I	79,790	Transposition and iron regulatory genes
MAP-17	5	61.60	I	3,655	Unknown proteins and a multicopy phage resistance gene
MAP-18	3	60.36	I	3,512	Hypothetical proteins
Total	204			239,969	

TABLE 4. Characteristics of *M. avium* subsp. *avium*-specific (MAV) genomic islands

Island	No. of ORFs	% GC	Island type	Size (bp)	Signature feature(s)
MAV-1	38	68.93	I	39,833	Eukaryotic genes with an integrase gene
MAV-2	32	65.87	I	31,387	Transposition and *M. tuberculosis* genes
MAV-3	10	63.34	I	9,693	Insertion sequence and *M. tuberculosis* or *M. avium* genes
MAV-4	53	66.83	I	47,356	PPE family and eukaryotic genes
MAV-5	16	64.10	I	17,905	Transposition and insertion sequences genes
MAV-6	23	68.80	I	19,161	Transposition, transcriptional regulator and heavy metal resistance genes
MAV-7	187	65.50	II	196,411	Transposition, transcriptional regulators, cell entry and iron regulation genes
MAV-8	3	65.18	I	2,977	Transposition and transcriptional regulator genes
MAV-9	15	62.43	I	20,844	Transposition and type III restriction system endonuclease genes
MAV-10	12	63.87	I	12,491	Transposition genes
MAV-11	5	65.45	I	3,593	Reductases and hypothetical proteins
MAV-12	168	65.05	II	181,445	Transposition, transcriptional regulators and cell entry genes
MAV-13	7	67.78	II	5,525	Transcriptional regulator
MAV-14	26	67.32	I	28,265	Transposition and *M. tuberculosis* genes
MAV-15	3	64.12	II	4,731	*Streptomyces* and *M. leprae* genes
MAV-16	6	69.64	I	44,157	Transposition and *Pst* genes
MAV-17	20	65.23	II	21,219	Transposition and *drrAB* genes (antibiotic resistance)
MAV-18	4	68.13	I	3,918	Transcriptional regulator and *Streptomyces* genes
MAV-19	4	65.30	I	5,169	Transposition genes
MAV-20	15	63.93	I	21,283	Transposition, transcriptional regulator and membrane-protein genes of *M. tuberculosis*
MAV-21	8	65.93	I	6,895	Transposition and antigen genes
MAV-22	9	67.71	I	9,931	Transcriptional regulator and metalloprotease genes
MAV-23	77	64.08	I	95,547	Transposition, transcriptional regulators, secreted proteins and cell entry genes
MAV-24	18	70.25	I	16,200	Hypothetical and unknown proteins from *M. tuberculosis* and *Streptomyces* spp.
Total	759			845,936	

the *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes.

Large DNA fragment inversions within the genomes of *M. avium* subspecies. Because of the high similarity among the genomes of *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* reported earlier (4), we expected considerable conservation in the synteny between genomes (gene order) within *M. avium* subsp. *avium* strains. To test our hypothesis, we used the order of GIs as markers for conserved gene order and the overall genome structure between *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* genomes. To our surprise, when the GIs associated with both genomes were aligned, three large genomic fragments with sizes of 54.9 kb, 863.8 kb, and 1,969.4 kb were identified as inverted relative to each other (Fig. 3). The largest inverted region is flanked by MAV-4 and MAV-19, the second inversion is flanked by MAV-21 and MAV-24, near the origin of replication in both genomes, and the smallest inversion is flanked by MAV-1 and MAV-2. Because the bioinformatics analysis used raw genome sequences, we used a PCR and sequencing approach to substantiate the genomic inversions in seven mycobacterial isolates (three isolates of *M. avium* subsp. *avium* and four isolates of *M. avium* subsp. *paratuberculosis*). As predicted from the initial sequence analysis, primers flanking the junction sites of the inverted regions gave the correct DNA fragment sizes and orientations consistent with the sequences of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes. Inversions were also analyzed in *M. intracellulare* with inconclusive results (data not shown). It is possible that genomic variations could be the reason for unsuccessful amplification of target sequences from *M. intracellulare*. More sequence analysis is needed to accurately investigate the inversions in *M. intracellulare*.

Further analysis identified several other smaller inversions that are present between *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* and scattered throughout the large inversions (data not shown). The presence of such inversions could reflect active changes in controlling gene expression (5), indicating the ability of the organism to adapt to different microenvironments.

DISCUSSION

Recent technological advances in the field of DNA microarrays combined with the availability of completed genome sequences have had a paramount impact on the field of comparative genomics. After a long period of slow progress, several microarray platforms were developed specifically to address questions related to the genome and transcriptome of mycobacterial infectious agents, including *M. tuberculosis*, *M. avium* subsp. *avium*, and *M. avium* subsp. *paratuberculosis* (32, 36, 41). In this report, we took advantage of DNA microarrays based on the genome sequence of *M. avium* and bioinformatic comparisons of the genome sequences of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* to provide a comprehensive view of the genomic rearrangements within *M. avium*. Our

TABLE 5. PCR identification of selected MAV island regions from 29 clinical isolates of *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* collected from different states

Species and clinical isolate	Genomic island[a]			
	MAV-3	MAV-11	MAV-21	MAV-23
M. avium subsp. *paratuberculosis*				
JTC33666	−	−	−	−
JTC33770	−	−	−	−
CW303	−	−	−	−
1B	−	−	−	−
3B	−	−	−	−
4B	−	−	−	−
5B	−	−	−	−
DT3	−	−	−	−
DT9	+	NA	−	−
DT12	−	−	−	−
DT19	−	−	−	−
JTC1281	−	−	−	−
JTC1282	−	−	−	−
JTC1283	−	−	−	−
JTC1285	−	−	−	−
JTC1286	−	−	−	−
M. avium subsp. *avium*				
T93	+	−	−	−
T99	+	−	−	−
T100	+	+	−	−
DT30	−	+	+	+
DT44	−	+	+	+
DT78	−	+	+	+
DT84	−	+	−	+
DT247	−	+	+	+
JTC956	NA	NA	NA	−
JTC982	NA	+	NA	+
JTC1161	+	+	−	−
JTC1262	+	−	−	−
JTC33793	+	+	+	+

[a] + or − denotes the presence or absence of genomic regions, while NA denotes no amplification of small or large DNA fragments.

TABLE 6. PCR identification of selected MAP island regions from 29 clinical isolates of *M. avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* collected from different states

Species and clinical isolate	Genomic island[a]					
	MAP-1	MAP-3	MAP-5	MAP-12	MAP-16	MAP-17
M. avium subsp. *paratuberculosis*						
JTC33666	+	+	+	+	+	+
JTC33770	+	+	+	+	+	+
CW303	+	+	+	+	+	+
1B	+	+	+	+	+	+
3B	+	+	+	+	+	+
4B	+	+	+	+	+	+
5B	+	+	+	+	+	+
DT3	−	+	+	+	+	+
DT9	−	+	+	+	+	+
DT12	+	+	+	+	+	+
DT19	+	+	+	+	+	+
JTC1281	−	+	+	+	+	+
JTC1282	−	+	+	+	+	+
JTC1283	−	+	+	+	+	+
JTC1285	−	−	+	+	+	−
JTC1286	+	+	+	+	+	+
M. avium subsp. *avium*						
T93	−	−	−	−	−	−
T99	−	NA	+	−	+	+
T100	+	NA	+	+	−	+
DT30	−	−	−	−	−	−
DT44	−	−	−	−	−	−
DT78	−	−	+	−	−	+
DT84	−	−	−	−	−	−
DT247	−	−	+	−	−	−
JTC956	NA	−	NA	−	+	+
JTC982	−	−	+	−	−	−
JTC1161	−	−	+	NA	+	+
JTC1262	−	−	−	−	−	−
JTC33793	−	−	−	−	−	−

[a] + or − denotes the presence or absence of genomic regions, while NA denotes no amplification of small or large DNA fragments.

analysis identified a total of 24 GIs present in *M. avium* subsp. *avium* but absent from 95% of the *M. avium* subsp. *paratuberculosis* isolates examined so far. An additional 18 islands specific to *M. avium* subsp. *paratuberculosis* that were absent from *M. avium* subsp. *avium* were also identified. The arrangements in these islands were verified by PCR amplification and sequencing. Previous studies analyzing polymorphism among *M. avium* subsp. *avium* strains (32, 36) reported only a proportion of the islands identified in this study (Table 7). This reflects the different levels of sensitivity in technologies used to interrogate *M. avium* genomes. Use of long oligonucleotide microarrays had identified only 14 genomic regions, which were all identified by our analysis (36). By use of PCR-based microarrays, seven regions of deletions were identified as present in *M. avium* subsp. *paratuberculosis* and absent or divergent in other *M. avium* strains (32). In our hands, BLAST analysis identified an additional 11 regions that were specific for *M. avium* subsp. *paratuberculosis*. In the short oligonucleotide arrays employed in this study, every ORF is represented by 18 pairs of probes spanning the whole ORF; thus, the analysis may be less sensitive to cross-hybridization artifacts that could obscure detection of islands observed when long oligonucleotide (36) or PCR (32) microarrays were used. Unfortunately, the short, tiled oligonucleotide DNA microarrays are costly to produce.

Despite the overall identity between *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* (up to 98%) on the nucleotide level, the hierarchical cluster analysis of the hybridization signals was able to identify separate lineages for *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* isolates. Overall analysis of the variations in GIs among isolates identified more widespread plasticity among *M. avium* subsp. *avium* isolates that was not detected in *M. avium* subsp. *paratuberculosis* isolates, implying that *M. avium* subsp. *avium* is more polymorphic than *M. avium* subsp. *paratuberculosis*, a conclusion that was drawn from a morphological analysis of *M. avium* subsp. *avium* colonies (10) and is now supported by our genomic analysis. Despite this genomic polymorphism, an extensive study of clinical isolates of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* was able to identify diagnostic DNA targets for each organism (35). Nonetheless, the genome of the human isolate of *M. avium* subsp. *paratuberculosis* from a Crohn's disease patient was closely related to that of an isolate from a cow with a clinical case of Johne's disease. This result was consistent with our PCR analysis of additional human isolates and in complete agreement with previous reports of studies employing short sequence repeats of *M. avium* subsp. *paratuberculosis* (15). However, wildlife animals could provide a reservoir for genomic diversity in *M. avium* subsp.

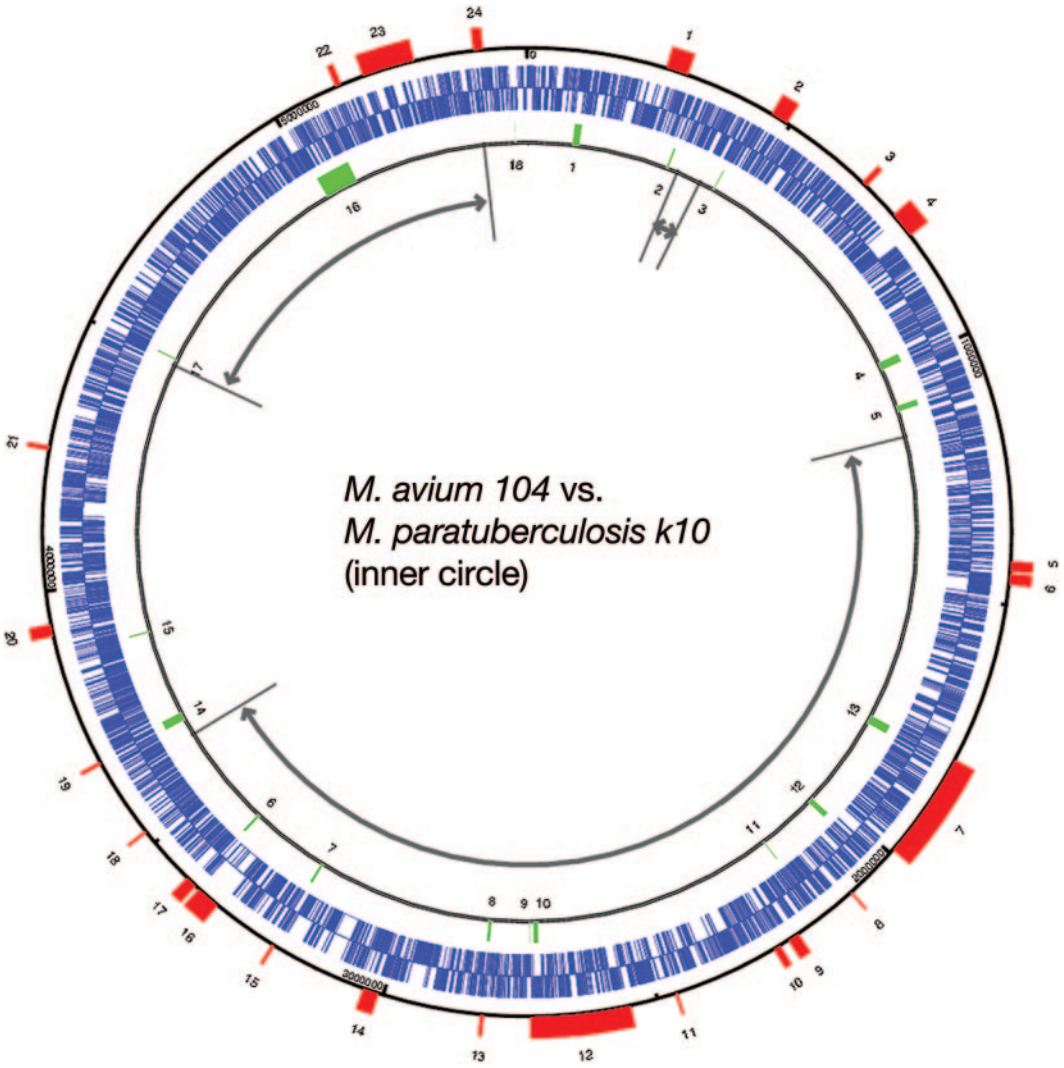

FIG. 3. Synteny of *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes. The locations of genomic islands present in *M. avium* subsp. *avium* (red boxes) or in *M. avium* subsp. *paratuberculosis* (green boxes) genomes are drawn to scale on the circular map of *M. avium* subsp. *avium* (outer circle) as well as the map of *M. avium* subsp. *paratuberculosis* (inner circle). The sequences of *M. avium* subsp. *paratuberculosis* k10 (query sequence) were compared with the whole genome sequence of the *M. avium* subsp. *avium* 104 ORFs (target sequence) by use of the BLAST algorithm; cutoff values of $E > 0.001$ and alignment percentages of $<25\%$ for the whole gene were accepted as indications for gene deletion. Blue short bars represent predicted ORFs in forward (outermost) or reverse (innermost) orientations. Large gray arrows indicate sites of genomic inversions.

paratuberculosis. Because of the implications of such findings for strategies to control Johne's disease, it is essential to analyze more strains isolated from variable sources, including wildlife animals, on a genome-wide level before synthesizing conclusions.

An interesting finding in our analysis of the *M. avium* subsp. *avium* genome is the high level of polymorphism observed in TetR family of transcriptional regulators. Some members of this family of regulators are involved in antibiotic resistance as well as transcription repression (17, 33). Mycobacterial species are notorious for resisting common chemotherapies, especially members of *M. avium* complex infecting AIDS patients (30). The process of active recruitment of GIs encoding the TetR genes could represent a mechanism that *M. avium* subsp. *avium* strains employ to resist levels of antibiotics once introduced to their microenvironment. Alternatively, when the antibiotics are not present, organisms may lose the TetR sequences. The mechanisms giving rise to genomic diversity in different microenvironments may differ, as evidenced by the differences in GC content identified between the *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes. The presence of GIs with a lower GC percentage in *M. avium* subsp. *paratuberculosis* may reflect a propensity for this organism to acquire genetic elements from the bacterium-rich intestinal microenvironment through lateral gene transfer mechanisms, as opposed to acquisition from other *M. avium* strains with similar GC percentages. The more typical GC content of *M. avium* subsp. *avium*-specific islands may reflect limitations on sources or mechanisms for acquisition of genetic materials from more-diverse organisms. Another example of divergence

TABLE 7. Matching of genomic islands identified by our analysis to the large sequence of polymorphisms identified by Semert et al. (36)

MAV island (this study)	Polymorphism[a]
MAV-1	LSP 3
MAV-2	LSP 7
MAV-3	LSP12
MAV-4	LSP 5
MAV-5	LSP 14
MAV-6	LSP 13
MAV-7	LSP 4
MAV-8	ND
MAV-9	LSP 10
MAV-10	ND
MAV-11	ND
MAV-12	LSP 1
MAV-13	ND
MAV-14	ND
MAV-15	ND
MAV-16	ND
MAV-17	LSP 9
MAV-18	ND
MAV-19	ND
MAV-20	LSP 2
MAV-21	ND
MAV-22	LSP 8
MAV-23	LSP 6
MAV-24	ND

[a] LSP, large sequence of polymorphisms; ND, not detected.

between *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* in pathogenesis is the polymorphism observed in GIs encoding different types of *mce* operons. The *mce* genes are a group of four operons that were shown to contribute to the entry of *M. tuberculosis* to mammalian cells (3, 8). Definitely, examples for genomic plasticity among *M. avium* members need to be studied in detail to delineate the role of genomic exchange on microbial fitness.

Throughout our analysis of standard and clinical isolates of subspecies of *M. avium* we identified two main types of genomic rearrangements. The first source of rearrangements in the examined isolates is insertions and/or deletions of genomic islands that could be necessary for pathogen survival within a particular microenvironment. The second source for large-scale rearrangements is genomic inversion, with its implications for regulation of the expression of key antigens. Mechanisms for the latter include homologous recombination, as suggested before for *Lactococcus lactis* (12), and could be supported by the presence of prophage sequences in the flanking sequences, as suggested for *Streptococcus pyogenes* (26). On the other hand, for the rearrangements introduced by the GIs, detailed analysis of their sequences and the flanking DNA regions has resulted in classifying these islands into two categories. A type I island is simply an additional fragment of *M. avium* subsp. *avium*- or *M. avium* subsp. *paratuberculosis*-specific DNA sequence that is present in the genome of one but not the other. Most of these GIs contain mobile genetic elements (45), suggesting that horizontal gene transfer events led to the insertion or deletion of the GIs. Genes encoded in type I GIs included transposases from different insertional sequence families (e.g., IS*117*, IS*1601*, IS*200*), integrases, and plasmid transfer proteins (Table 3 and Table 4).

In *M. avium* subsp. *paratuberculosis*-specific islands, some of the type I GIs (MAP-12, MAP-13) included prophage sequences, a unique feature that was not detected in MAV GIs. All these mobile genetic elements can play a role in genomic rearrangements through simple transposition and integration and could play a role in the inversion of the largest genomic DNA fragment. In one of the type I GIs (MAV-9), a type III restriction enzyme system was found, which could be associated with island integration or deletion from the ancestral organism (39). Such patterns of rearrangement are well documented for other bacteria such as *Escherichia coli* and *Streptomyces* spp. (42, 45). Insertion or deletion of GIs frequently involves large DNA fragments, as previously described in the case of *Streptomyces glaucescens* (6). We observed that the median size of type I MAV GIs is 21 kb, which is four times larger than the median size of the rest of the GIs (4.7 kb). In the other type of GI (type II), unique DNA fragments are present in *M. avium* subsp. *avium* or *M. avium* subsp. *paratuberculosis* genomes at the corresponding breaking points of each island (complex genomic island). For GIs belonging to type II, transposition-related genes were found in fewer islands than in those belonging to type I, indicating a potential difference in the mechanisms responsible for introduction of these islands. In these cases it is possible that homologous recombination is responsible for their introduction when DNA fragments exchange between homologous sites of the genome following crossover and resolution events. Taken together, our data suggest that some GIs belonging to type II could be responsible for unique mechanisms of pathogenicity islands involved in virulence. This hypothesis is supported by the presence of lower GC percentages in *M. avium* subsp. *paratuberculosis* GIs near tRNA genes, a hallmark of pathogenicity islands (18). This particular type of GI could provide advantages for *M. avium* subsp. *paratuberculosis* with respect to persistence inside the host microenvironment.

Finally, the comparative genomic analysis of *M. avium* subsp. *avium* versus *M. avium* subsp. *paratuberculosis* identified two large fragments of genomic inversions. Previously, genetic inversions were believed to be used as a mechanism for regulating gene activity, such as in the case of type I fimbriae expression in *E. coli* (34). In another system, 12 genomic inversions were detected in *Bacteroides fragilis*, an opportunistic pathogen that colonizes the intestine (9). It was suggested that such extensive inversions could contribute to the reversible phase and antigenic variations. Because of the very large sizes of inversions detected and despite the overall sequence identity between the *M. avium* subsp. *avium* and *M. avium* subsp. *paratuberculosis* genomes, we predict a substantial difference in the expression profiles between both strains, especially for genes encoded in the inverted regions. The implications of such inversions for the antigenic variations among *M. avium* subspecies remain to be investigated on both the transcriptome and proteome levels. So far, we have confirmed the inversions in seven isolates; additional isolates could be examined to investigate the extent and distribution of such inversions among isolates from different hosts.

The presented analysis of genomic rearrangements among *M. avium* genomes supported the notion of the emergence of distinct lineages of opportunistic and pathogenic strains of mycobacteria. The presented findings provide a wealth of in-

formation for developing novel diagnostics and chemotherapies that could differentially target specific members within MAC. Additional observations of large genomic inversions among *M. avium* subspecies genomes suggest that *M. avium* subsp. *avium* strains might undergo antigenic variation. Comparative genomic analysis of other species within MAC (e.g., *M. avium* subsp. *silvaticum*, *M. intracellulare*) or closely related to *M. avium* (e.g., *M. scrofulaceum*) will help to select the most promising targets for evolutionary characterization.

ACKNOWLEDGMENTS

We acknowledge Gireesh Rajashekara and Christine Tavano for reading the manuscript. We also thank Shelly Immel, Gail Thomas, and Becky Manning for technical help with mycobacterial clinical isolates.

Sequencing of *M. avium* subsp. *avium* strain 104 was accomplished with support from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. Research in the AMT laboratory is supported by the Animal Formula Fund (WIS04794) and the National Research Initiative of the U.S. Department of Agriculture Cooperative State Research, Education and Extension Service (grant WIS04823 and Johne's Disease Integrated Program [2004-35605-14243]).

REFERENCES

1. **Albert, T. J., J. Norton, M. Ott, T. Richmond, K. Nuwaysir, E. F. Nuwaysir, K. P. Stengele, and R. D. Green.** 2003. Light-directed 5′→3′ synthesis of complex oligonucleotide microarrays. Nucleic Acids Res. **31:**e35.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **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.
4. **Bannantine, J. P., E. Baechler, Q. Zhang, L. 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.
5. **Bentley, S. D., and J. Parkhill.** 2004. Comparative genomic structure of prokaryotes. Annu. Rev. Genet. **38:**771–792.
6. **Birch, A., A. Hausler, C. Ruttener, and R. Hutter.** 1991. Chromosomal deletion and rearrangement in *Streptomyces glaucescens*. J. Bacteriol. **173:** 3531–3538.
7. **Braunstein, M., S. S. Bardarov, and W. R. Jacobs.** 2002. Genetic methods for deciphering virulence determinants of *Mycobacterium tuberculosis*. Methods Enzymol. **358:**67–99.
8. **Casali, N., M. Konieczny, M. A. Schmidt, and L. W. Riley.** 2002. Invasion activity of a *Mycobacterium tuberculosis* peptide presented by the *Escherichia coli* AIDA autotransporter. Infect. Immun. **70:**6846–6852.
9. **Cerdeno-Tarraga, A. M., S. Patrick, L. C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M. A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M. T. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch, J. Woodward, B. Barrell, and J. Parkhill.** 2005. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. Science **307:**1463–1465.
10. **Chacon, O., L. E. Bermudez, and R. G. Barletta.** 2004. Johne's disease, inflammatory bowel disease, and *Mycobacterium paratuberculosis*. Annu. Rev. Microbiol. **58:**329–363.
11. **Choudhuri, B. S., S. Bhakta, R. Barik, J. Basu, M. Kundu, and P. Chakrabarti.** 2002. Overexpression and functional characterization of an ABC (ATP-binding cassette) transporter encoded by the genes *drrA* and *drrB* of *Mycobacterium tuberculosis*. Biochem. J. **367:**279–285.
12. **Daveran-Mingot, M. L., N. Campo, P. Ritzenthaler, and P. Le Bourgeois.** 1998. A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. J. Bacteriol. **180:**4834–4842.
13. **Dudoit, S., R. C. Gendeman, and J. Quackenbush.** 2003. Open source software for the analysis of microarray data. BioTechniques **34:**S45–S51.
14. **Foley-Thomas, E. M., D. L. Whipple, L. E. Bermudez, and R. G. Barletta.** 1995. Phage infection, transfection and transformation of *Mycobacterium avium* complex and *Mycobacterium paratuberculosis*. Microbiology **141:**1173–1181.
15. **Ghadiali, A. H., M. Strother, S. A. Naser, E. J. B. Manning, and S. Sreevatsan.** 2004. *Mycobacterium avium* subsp. *paratuberculosis* strains isolated from Crohn's disease patients and animal species exhibit similar polymorphic locus patterns. J. Clin. Microbiol. **42:**5345–5348.
16. **Glasner, J. D., P. Liss, G. Plunkett, A. Darling, T. Prasad, M. Rusch, A. Byrnes, M. Gilson, B. Biehl, F. R. Blattner, and N. T. Perna.** 2003. ASAP, a systematic annotation package for community analysis of genomes. Nucleic Acids Res. **31:**147–151.
17. **Godsey, M. H., E. E. Zheleznova-Heldwein, and R. G. Brennan.** 2002. Structural biology of bacterial multidrug resistance gene regulators. J. Biol. Chem. **277:**40169–40172.
18. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. **54:**641–679.
19. **Kendziorski, C. M., M. A. Newtone, H. Lan, and M. N. Goululd.** 2003. On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. Stat. Med. **22:**3899–3914.
20. **Li, L., J. P. Bannantine, Q. Zhang, A. Amonsin, B. J. May, D. Alt, N. Banerji, S. Kanjilal, and V. Kapur.** 2005. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. Proc. Natl. Acad. Sci. USA **102:**12344–12349.
21. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115.
22. **Manning, E. J. B., and M. T. Collins.** 2001. *Mycobacterium avium* subsp. *paratuberculosis*: pathogen, pathogenesis and diagnosis. Rev. Sci. Tech. **20:** 133–150.
23. **Martinez-Bueno, M., A. J. Molina-Henares, E. Pareja, J. L. Ramos, and R. Tobes.** 2004. BacTregulators: a database of transcriptional regulators in bacteria and archaea. Bioinformatics **20:**2787–2791.
24. **Mira, A., L. Klasson, and S. G. E. Andersson.** 2002. Microbial genome evolution: sources of variability. Curr. Opin. Microbiol. **5:**506–512.
25. **Morita, Y., S. Maruyama, H. Kabeya, A. Nagai, K. Kozawa, M. Kato, T. Nakajima, T. Mikami, Y. Katsube, and H. Kimura.** 2004. Genetic diversity of the *dnaJ* gene in the *Mycobacterium avium* complex. J. Med. Microbiol. **53:**813–817.
26. **Nakagawa, I., K. Kurokawa, A. Yamashita, M. Nakata, Y. Tomiyasu, N. Okahashi, S. Kawabata, K. Yamazaki, T. Shiba, T. Yasunaga, H. Hayashi, M. Hattori, and S. Hamada.** 2003. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. Genome Res. **13:**1042–1055.
27. **Naser, S. A., G. Ghobrial, C. Romero, and J. F. Valentine.** 2004. Culture of *Mycobacterium avium* subspecies *paratuberculosis* from the blood of patients with Crohn's disease. Lancet **364:**1039–1044.
28. **Ohta, H.** 2003. Disseminated *Mycobacterium avium* complex (MAC) in a patient with acquired immunodeficiency syndrome (AIDS). Ann. Nucleic Med. **17:**114.
29. **Oliveira, R. S., M. P. Sircili, E. M. D. Oliveira, S. C. Balian, J. S. Ferreira-Neto, and S. C. Leão.** 2003. Identification of *Mycobacterium avium* genotypes with distinctive traits by combination of IS*1245*-based restriction fragment length polymorphism and restriction analysis of *hsp65*. J. Clin. Microbiol. **41:**44–49.
30. **Opravil, M.** 1997. Epidemiological and clinical aspects of mycobacterial infections. Infection **25:**56–59.
31. **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.
32. **Paustian, M. L., V. Kapur, and J. P. Bannantine.** 2005. Comparative genomic hybridizations reveal genetic regions within the *Mycobacterium avium* complex that are divergent from *Mycobacterium avium* subsp. *paratuberculosis* isolates. J. Bacteriol. **187:**2406–2415.
33. **Perez-Rueda, E., and J. Collado-Vides.** 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res. **28:**1838–1847.
34. **Schembri, M. A., D. W. Ussery, C. Workman, H. Hasman, and P. Klemm.** 2002. DNA microarray analysis of fim mutations in *Escherichia coli*. Mol. Genet. Genomics **267:**721–729.
35. **Semret, M., D. C. Alexander, C. Y. Turenne, P. de Haas, P. Overduin, D. van Soolingen, D. Cousins, and M. A. Behr.** 2005. Genomic polymorphisms for *Mycobacterium avium* subsp. *paratuberculosis* diagnostics. J. Clin. Microbiol. **43:**3704–3712.
36. **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.
37. **Singh-Gasson, S., R. D. Green, Y. J. Yue, C. Nelson, F. Blattner, M. R. Sussman, and F. Cerrina.** 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat. Biotechnol. **17:**974–978.
38. **Smole, S. C., F. McAleese, J. Ngampasutadol, C. F. von Reyn, and R. D. Arbeit.** 2002. Clinical and epidemiological correlates of genotypes within the *Mycobacterium avium* complex defined by restriction and sequence analysis of *hsp65*. J. Clin. Microbiol. **40:**3374–3380.
39. **Suyama, M., and P. Bork.** 2001. Evolution of prokaryotic gene order: genome rearrangements in closely related species. Trends Genet. **17:**10–13.
40. **Talaat, A. M., S. T. Howard, I. W. Hale, R. Lyons, H. Garner, and S. A.**

Johnston. 2002. Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. Nucleic Acids Res. **30:**E104.
41. **Talaat, A. M., R. Lyons, S. T. Howard, and S. A. Johnston.** 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc. Natl. Acad. Sci. USA **101:**4602–4607.
42. **Wren, B. W.** 2000. Microbial genome analysis: insights into virulence, host adaptation and evolution. Nat. Rev. Genet. **1:**30–39.
43. **Yakrus, M. A., and R. C. Good.** 1990. Geographic distribution, frequency, and specimen source of *Mycobacterium avium* complex serotypes isolated from patients with acquired immunodeficiency syndrome. J. Clin. Microbiol. **28:**926–929.
44. **Yamakita, N., and K. Yasuda.** 2005. Pulmonary *Mycobacterium avium* complex infection in patients with panhypopituitarism not receiving hormone replacement therapy. Mayo Clin. Proc. **80:**291.
45. **Ziebuhr, W., K. Ohlsen, H. Karch, T. Korhonen, and J. Hacker.** 1999. Evolution of bacterial pathogenesis. Cell. Mol. Life Sci. **56:**719–728.