The complete genome sequence of Mycobacterium avium subspecies paratuberculosis

Lingling Li*†‡, John P. Bannantine†‡, Qing Zhang*†‡, Alongkorn Amomsin*†‡, Barbara J. May*†‡, David Alt§, Nilanjanana Banerji*†‡, Sagarika Kanjilal*†‡, and Vivek Kapur*†‡¶

*Department of Microbiology, †Biomedical Genomics Center, and §Department of Medicine, University of Minnesota, St. Paul, MN 55108; and ¶Department of Animal Disease Center, U.S. Department of Agriculture–Agriculture Research Service, Ames, IA 50010

Communicated by Harley W. Moon, Iowa State University, Ames, IA, July 13, 2005 (received for review March 18, 2005)

We describe here the complete genome sequence of a common clone of Mycobacterium avium subspecies paratuberculosis (Map) strain K-10, the causative agent of Johne's disease in cattle and other ruminants. The K-10 genome is a single circular chromosome containing 3,024,781 base pairs and encodes 4,350 predicted ORFs, 45 tRNAs, and one rRNA operon. In silico analysis identified >3,000 genes with homologs to the human pathogen, M. tuberculosis (Mtb), and 161 unique genomic regions that encode 39 previously unknown Map genes. Analysis of nucleotide substitution rates with Mtb homologs suggest overall strong selection for a vast majority of these shared mycobacterial genes, with only 68 ORFs with a synonymous to nonsynonymous substitution ratio of >2. Comparative sequence analysis reveals several noteworthy features of the K-10 genome including: a relative paucity of the PE/PPE family of sequences that are implicated as virulence factors and known to be immunostimulatory during Mtb infection; truncation in the EntE domain of a salicyl-AMP ligase (MbtA), the first gene in the mycobactin biosynthesis gene cluster, providing a possible explanation for mycobactin dependence of Map; and Map-specific sequences that are likely to serve as potential targets for sensitive and specific molecular and immunologic diagnostic tests. Taken together, the availability of the complete genome sequence offers a foundation for the study of the genetic basis for virulence and physiology in Map and enables the development of new generations of diagnostic tests for bovine Johne's disease.

Mycobacterium avium subspecies paratuberculosis (Map) is an extremely slow-growing, acid-fast, mycobactin-dependent multispecies pathogen. Infection with this bacterium leads to a chronic granulomatous enteritis in cattle and other wild and domestic ruminants, termed Johne's disease (1). Clinical signs of Johne's disease include diarrhea, weight loss, decreased milk production, and mortality. A recent study estimated that 21% of United States dairy herds are infected, resulting in considerable economic losses to the dairy industry totaling more than $200 million per annum (2, 3). A major concern with Johne's disease is the ease with which the bacterium is spread. Subclinically or clinically infected animals shed Map in feces and milk, enabling dissemination to susceptible calves, the environment, and in retail milk (4–6). Map-containing milk may be of particular concern because the bacterium has been suggested as a possible cause of Crohn's disease in humans (7).

The detection and diagnosis of Map-infected animals poses great difficulties. Map culture can require up to 16 weeks or more, and serological tests lack sensitivity because of the conversion occurring relatively late during the course of the disease (8). Also, previously developed PCR-based approaches (e.g., IS900) have been shown to lack specificity (9). This result is, in part, due to the high levels of genetic similarity of Map with other mycobacteria, in particular, Mycobacterium avium (Mav) (10). Previous studies from our laboratories and elsewhere show >95% nucleotide sequence similarity between many strains of Map and Mav (11–13). Therefore, it is widely recognized that the development of rapid, sensitive, and specific assays to identify infected animals is essential to the formulation of rational strategies to control the spread of Map.

As a first step toward elucidating the molecular basis of Map's physiology and virulence, and providing a foundation for the development of the next generation of Map diagnostic tests and vaccines, we report the complete genome sequence of a common clone of Map, strain K-10.

Materials and Methods

Bacterial Strains. We chose to sequence Map K-10, a bovine clinical isolate, because it is a virulent, low passage clinical strain that was isolated from a dairy herd in Wisconsin by investigators at the U.S. Department of Agriculture National Animal Disease Center in the mid-1970s. In addition, K-10 is amenable to genetic manipulation by transposon mutagenesis (14). The organism was grown in Middlebrook 7H9 broth supplemented with OADC (Difco), Tween 80, and mycobactin J (Allied Monitor, Fayette, MO).

Sequence Analysis. A random shotgun approach was adopted to sequence the genome of Map K-10 (15). To create a small (1.8- to 3.0-kb) insert library, genomic DNA was initially isolated by using a chloroform/cetyltrimethylammonium bromide-based method, as described (16). The DNA was sheared by nebulization (www.genome.ou.edu) and 1.8- to 3.0-kb fragments were cloned into pUC18 for isolation and sequencing. Approximately 30,240 clones and 3,000 PCR fragments were sequenced from both ends via Dye-terminator chemistry on Applied Biosystems 3700 and 3100 sequencing machines. A total of 66,129 sequences were used to generate the final assembly, representing a 6-fold coverage of the genome. Sequence assembly was performed with PHREDPHRAP (http://genome.washington.edu), and the ∼400 gaps that remained at the end of the shotgun phase were closed by primer walking and multiplex random PCR.

Informatics. ORFs were predicted by both ARTEMIS and GLIMMER and verified manually in ARTEMIS. BLAST analysis of the ORFs was performed at the Computational Biology Center at the

Abbreviations: Map, Mycobacterium avium subspecies paratuberculosis; Mav, Mycobacterium avium; Mtb, Mycobacterium tuberculosis.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AE016958).

1L.L., J.P.B., Q.Z., A.A., S.K., and V.K. have a financial conflict of interest that results from a patent application that has been filed on some DNA sequences that are disclosed and discussed in the manuscript. The patent applications that have been filed are jointly owned by the University of Minnesota and the U.S. Department of Agriculture. As named inventors, these authors may potentially financially benefit from the commercialization of the technology. In addition, some of the technology disclosed in this paper has also been licensed to ANDX, Inc., a University of Minnesota based startup company focusing on the development of molecular diagnostic assays, in which S.K. and V.K. have financial interests and are cofounders.

To whom correspondence should be addressed. E-mail: vkapur@umn.edu.

© 2005 by The National Academy of Sciences of the USA
University of Minnesota (www.cbc.umn.edu), and transfer RNAs were predicted with TRNASCAN-SE (17–20). Comparative genomic analysis was performed with *Mycobacterium tuberculosis* (Mtb), strain H37Rv (GenBank accession no. NC_000962) and *Mav* strain 104 (The Institute for Genomic Research, www.tigr.org) (21, 22).

Results and Discussion

Characteristics of the Map Genome. The analysis showed that *Map* K-10 has a single circular sequence of 4,829,781 base pairs, with a G+C content of 69.3% (Table 1 and Fig. 1). The putative origin of replication was identified based on the presence of *dnaA* boxes, characteristic oligomer skew, and G-C skew between the putative genes *dnaA* and *dnaN* (18, 23). The initiation codon for the *dnaA* gene was chosen as the start point for numbering the genome (Fig. 1). The G+C content is relatively constant throughout the genome. The analysis also revealed only a few genomic regions with lower G+C content corresponding to prophages or coding RNA sequences (Fig. 1).

As in the three other mycobacterial genomes sequenced to date, *Mtb*, *Mycobacterium leprae*, and *Mycobacterium bovis*, a single *rrn* operon (16S-23S-5S) was identified in K-10 along with 50 additional genes coding for functional RNA molecules. The *rrn* operon is located ~2.75 Mb from the putative oriC on the opposite strand. This is ~1.3 Mb further from the oriC than what is described in *Mtb* (22).

Repetitive DNA in Map. Approximately 1.5% (or 72.2 kb) of the *Map* genome is comprised of repetitive DNA like insertion sequences, multigene families, and duplicated housekeeping genes. The analysis also identified 17 copies of the previously described insertion sequence IS900, seven copies of IS1311, and three copies of ISMav2 in the K-10 genome (Fig. 2). A total of 16 additional *Map* insertion sequence elements were identified in the analysis, totaling 19 different insertion sequences with 58 total copies in the K-10 genome. Although many of these newly

Table 1. Summary of the complete genome of *M. paratuberculosis* K-10 and the comparison with other *Mycobacterium* species

<table>
<thead>
<tr>
<th>Property</th>
<th>Map</th>
<th>Mav</th>
<th>Mtb</th>
<th>M. bovis</th>
<th>M. leprae</th>
<th>M. smegmatis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size, bp</td>
<td>4,829,781</td>
<td>5,475,738</td>
<td>4,411,532</td>
<td>4,345,492</td>
<td>3,268,203</td>
<td>6,988,209</td>
</tr>
<tr>
<td>G+C content, %</td>
<td>69.30</td>
<td>68.99</td>
<td>65.61</td>
<td>65.63</td>
<td>57.79</td>
<td>67.40</td>
</tr>
<tr>
<td>Protein coding, %</td>
<td>91.30</td>
<td>NA</td>
<td>90.80</td>
<td>90.59</td>
<td>49.50</td>
<td>92.42</td>
</tr>
<tr>
<td>ORFs</td>
<td>4,350</td>
<td>NA</td>
<td>3,959</td>
<td>3,935</td>
<td>1,604</td>
<td>6,897</td>
</tr>
<tr>
<td>Gene density, bp per gene</td>
<td>1,112</td>
<td>NA</td>
<td>1,114</td>
<td>1,099</td>
<td>2,037</td>
<td>1,013</td>
</tr>
<tr>
<td>Average gene length, bp</td>
<td>1,015</td>
<td>NA</td>
<td>1,012</td>
<td>995</td>
<td>1,011</td>
<td>936</td>
</tr>
<tr>
<td>tRNAs</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>rRNA operon</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Li et al. PNAS August 30, 2005 vol. 102 no. 35 12345

![Fig. 1.](image1.png) Circular representation of the Map K-10 genome. From inside: red arrow, rRNA operon; dark purple histogram, GC content; multicolored histogram, MAP ORFs coded according to functional classification (small molecule metabolism, blue; macromolecule metabolism, red; cell processes, purple; other processes, yellow; hypotheticals, green; unknowns, orange). The outer colored histogram indicates the same direction of transcription as the origin of replication. The inner colored histogram indicates the opposite direction of transcription as the origin of replication. Black arrows, 45 tRNAs. Outer circle, scale. The figure was generated with GENESCE software (DNAstar, Madison, WI).

![Fig. 2.](image2.png) Linear representation of repeat and unique regions within the Map K-10 genome. From the top: Blue histogram, VNTR/DR repeats; red histogram, SSR repeats; green histogram, unique regions; multicolored histogram, insertion sequences (IS900, blue; MAP01, aqua; MAP02, gray; MAP03, pink; MAP04, green; MAP05, purple; MAP06, black; MAP07, orange; MAP08, red; MAP09, light pink; MAP10, light purple; MAP11, beige; MAP12, brown; MAP13, fuschia; MAP14, light aqua; MAP15, light gray; MAP16, dark green; IS1311, maroon; Mav2, light blue; REP, light green). Purple histogram, GC content. The figure was generated with GENESCE software (DNAstar, Madison, WI).
discovered Map IS elements are homologs with previously described insertion sequences in Mtb, Mav, M. bovis, and Mycobacterium marinum, the analysis also revealed several insertion sequences with no identifiable homologs in other mycobacteria. For example, IS_MAP02, present in six copies in the K-10 genome (Fig. 2), has no identified homolog in other mycobacteria and only very low levels of homology (28% identity) with a transposase described in Legionella pneumophila. Similarly, IS_MAP04, present in four copies in the K-10 genome, has no homologs in other mycobacteria but is similar to insertion sequences found in Arthrobacter nicotinovorans and Streptomyces coelicolor. These newly discovered IS elements are of particular interest for their use as specific potential diagnostic targets due to their absence in other mycobacteria. In addition, the analysis identified 12 homologs to the REP13E12 family in K-10 (Fig. 2); this is a ~1,400-bp repeated insertion sequence that was first described in the Mtb genome (24, 25).

It is believed that insertion sequences preferentially integrate within intergenic regions so as to avoid the disruption of essential genes (26). Consistent with this hypothesis, the majority of the IS elements found in K-10 appear to be clustered within intergenic regions. For example, MAP0028c and MAP0029c, MAP0849c and MAP0850c, and MAP2155, MAP2156, and MAP2157 are clustered within 5 kb of each other in noncoding regions of the chromosome (Fig. 2). The analysis also shows that insertion sequences in Map are absent from the region flanking 32 kb of either side of the oriC. A similar observation was made for the Mtb genome; however, in the case of Mtb, this distance is considerably greater at 600 kb (24). It is thought that there may be detrimental effects to chromosomal replication when insertion sequences found in Arthrobacter nicotinovorans and Streptomyces coelicolor. These newly discovered IS elements are of particular interest for their use as specific potential diagnostic targets due to their absence in other mycobacteria. In addition, the analysis identified 12 homologs to the REP13E12 family in K-10 (Fig. 2); this is a ~1,400-bp repeated insertion sequence that was first described in the Mtb genome (24, 25).

Protein-Encoding Genes. The K-10 genome contains 4,350 ORFs with lengths ranging from 114 bp (a ribosomal subunit encoding gene) to 19,155 bp (a peptide synthetase), which, in sum, account for 91.5% of the entire genome. A total of 52.5% of the genes are transcribed with the same polarity as that of DNA replication, a fraction that is slightly lower than the 59% observed in Mtb. Interestingly, there appears to be a higher bias toward transcription in the same polarity as replication in some other organisms (for e.g., 75% in Bacillus subtilis) (26, 27). Although it is tempting to speculate that this bias may, in part, contribute to the slow growth in Mtb and the even slower growth in Map, it is important to note that only 55% of Escherichia coli genes are transcribed in the same polarity as replication, suggesting that gene location in relation to the origin of replication cannot fully explain Map’s slow growth in laboratory culture (28).

The analysis showed that a total of 60% of the putative proteins in Map had homologs to other microbial proteins with known functions and 25% were homologous to hypothetical proteins (Table 2 and Fig. 4, which are published as supporting information on the PNAS web site). A total of 39 predicted proteins are unique to Map, with no identifiable homologs in the current databases. Of the predicted proteins, ~75% had homology to those identified in Mtb (22). Interestingly, the functional redundancy caused by gene duplication that was previously observed in Mtb (~52% of genes are functionally redundant) exists to an even greater extent in Map (29). Functional redundancy, based on amino acid homology comparisons, is particularly high among genes involved in lipid metabolism and oxidoreduction; for instance, there are 254 predicted genes functioning as oxidoreductases and oxygenases, compared to 171 in Mtb. The Map genome encodes the complete set of enzymes for many metabolic pathways including glycolysis, the pentose phos-
transport of iron (34). Homologs to the mbtA-J cluster were identified in the Map genome. However, a direct comparison of the Mbta-J cluster in Map with those of Mav and Mtb show significant differences in primary structure of this region (Fig. 3 and Table 5, which is published as supporting information on the PNAS web site). First, in Mtb, mbtl (ipkE) and mbtJ (lipK) are adjacent to each other and immediately downstream of the mycobactin biosynthesis gene cluster (Fig. 3) (34). However, in Map and Mav, there is a 6.6- and 5.7-kb gap, respectively, between the two genes. Furthermore, there is 19.3-kb gap in Map and 197.3-kb gap in Mav between mbtl and mbta of mbt cluster confirming a preliminary observation made by DNA microarray analysis (55). Because Mav can successfully produce mycobactin, the distance in genes is not likely to be the limiting factor that affects the production of mycobactin in Map. Second, we identified frame-shift mutations in both mbtB and mbtE in Map when compared with Mtb. The mbtB frameshift also appears in Mav, suggesting that this too is not the limiting factor affecting mycobactin production in Map (Fig. 3). In addition, even though the gene is divided, all functional domains identified in Mtb appear to be present in the mbtB gene product in Map, suggesting that its function could still be maintained. A similar frameshift exists in the mbtE gene in Map as well, but the functional domains of MbtE that are present in Mtb are present in Map as well. For each of mbtC, mbtD, and mbtF, the gene products include the common domains that are found in corresponding proteins in Mtb.

The major difference between Map, Mav, and Mtb in the mbtA-J cluster was in the mbtA gene. Gene mbtA is shorter in Map: encoding a 400-aa protein, compared with 565- and 551-residue polypeptides in Mtb and Mav, respectively. As a result of this truncation, Mbta has only 337 residues in Map with homology to the N-terminal of the EntE domain, and lacks >200 residues of the EntE C terminus that are presumably important for protein function. Because Mbta is thought to initiate mycobactin production, the truncation observed in this key gene suggests that the entire cascade leading to mycobactin production may be attenuated in Map (34). Therefore, it is tempting to speculate that the truncated EntE domain in the mbtA gene product in Map might be the limiting factor in mycobactin production, a hypothesis that remains to be formally tested.

**Immunological and Virulence-Related Insights from the Map Genome.** Despite intensive research efforts, there is still little information regarding the molecular basis for Map pathogenesis. Hence, we paid particular attention to the identification of genes with a potential role in pathogenesis in the Map genome and note several interesting observations. There is a paucity in the number of the PE/PPE family of proteins that are thought to play an important role in mycobacterial infection from both an antigenic as well as an immunologic standpoint. These proteins are acidic and glycine-rich proteins, and are identified by their specific domains (Pro-Pro-Glu and Pro-Glu, respectively) that frequently contain polymorphic repetitive sequences (PGRSS) and multiple copies of major polymorphic tandem repeats, respectively (22, 36–39). These proteins are thought to be expressed on the cell surface and provide the antigenic variation that elicits varying immunological responses in Mtb depending on the type of PE/PPE protein expressed on the cell surface (22, 40). Genome-scale comparisons between two isolates of Mtb show that regions of the genome encoding PE/PPE proteins have a higher single-base substitution frequency, further supporting the hypothesis that they are recognized by the immune system and hence subject to positive Darwinian selection (40). Although these families of proteins comprise 10% of the Mtb genome, there were only six PE homologs and 36 PPE homologs in Map (comprising 1% of the genome) compared to 38 and 68, respectively, in Mtb (22). Within the PE family, there is no intact PE/PGRS subfamily of proteins identified in the K-10 genome, although this subfamily has been identified in other mycobacteria including M. bovis and M. marinum (41). Interestingly, this subfamily of proteins is also absent in Mav and M. leprae (42). Although the exact significance of this observation is unknown, it may suggest a more limited, less variable, and different immune response toward Map as compared with Mtb. This observation also leads to the tempting speculation that antimicrobial agents and vaccines directed against these major virulence factors may be more likely to be effective against Map as compared with Mtb.

Pathogens often express proteins that alter the effects of the host’s immune response so as to evade destruction. Because mycobacteria are facultative pathogens and are assumed to selectively express specific genes to allow for survival inside the host macrophage, much attention has been directed toward the characterization of virulence genes that are important for the entry and persistence in the host (43). One such gene, the mammalian cell entry (mce) gene, has been identified in Mtb (44) and was shown to enhance E. coli’s ability to survive in macrophages (44). Four copies of the mce gene are present in Mtb (22). The analysis of the complete sequence of the Map genome revealed eight homologs of the mce gene. The gene has also been identified in different mycobacterial species, including Mav, M. bovis, and Mycobacterium smegmatis. The wide distribution of the mce operon in pathogenic and nonpathogenic mycobacteria implies that the mere presence of these genes does not endow a bacterium with the ability to cause disease. However, the role of this operon in virulence may be determined by its expression under specific conditions (43, 45, 46).
Mycobacteria were originally classified as such by the presence of mycolic acids (47, 48). Not only do mycobacteria produce this type of lipid, but these organisms are also known for their ability to produce and use a vast array of other lipophilic molecules (47). Importantly, these diverse structures that are located primarily on the cell wall are thought to play a role in pathogenesis in many mycobacterial species by their ability to allow entry into host cells or suppress or evade host immune defense mechanisms (49, 50). Increased survival of mycobacteria may also be enabled, in part, by their ability to preferentially use fatty acids instead of carbohydrates for basic metabolic needs (51–53). Our analysis shows that there are ~80 more genes in Map (n = 266) that are predicted to be involved in lipid metabolism than there are in Mtb. Although this difference in number of lipid metabolism and biosynthesis related genes is due primarily to genetic redundancy in Map, there are some noteworthy differences. For example, Map contains a gene (MAP3194) encoding hydroxymethylglutaryl-CoA lyase, an enzyme that is found in other bacteria as well as in humans, and catalyzes the last step of ketogenesis and leucine catabolism (54). The enzyme may play a role in fatty acid biosynthesis by altering what is produced and distributed to the cell membrane (54). This difference between Map and Mtb indicates there may be variation in lipid metabolism and biosynthesis that may play a role in antigens present/absent on the cell surface, thus, affecting host immune defense mechanisms.

Map lacks one of the largest operons in Mtb involved in the production of phenolpthiocerol, a polyketide. In the Mtb genome, this operon (pppABCDE) is immediately upstream of another gene cluster (mas), which encodes an enzyme responsible for the synthesis of mycrococeric acid. In combination, these two products form the abundant cell-wall-associated molecule phthiocerol dimycocerosate (DIM) (22). Importantly, it has been found that Mtb isolates lacking this cell wall lipid are attenuated in virulence (55–57). Map contains no homologs to either the pppABCDE or the mas gene homologs. Instead, the analysis identified 35 other genes with a possible role in polyketide synthesis (including 12 involved in mycobactin biosynthesis; Table 2). Included among these are chalcone synthase-like genes. These polyketides are found primarily in plants, but a four-gene cluster (pkx7-10) has been identified in Mtb with high similarity to a chalcone synthase-like genes (22). Similarly, a four-gene cluster was identified in Map (MAP1369, MAP1370, MAP1371, and MAP1372). These polyketides are believed to be involved in the production of DIM in Mtb as well as in its virulence (57). Mutations in pkx10 showed attenuation of the virulent Mtb H37Rv isolate upon infection of alveolar macrophages (57). Therefore, although Map lacks the two major gene clusters required for the production of the polyketide DIM in Mtb, the analysis identified other genes that may play a role in DIM and polyketide synthesis in Map. These results suggest that there are likely to be considerable differences in the presence or expression of various lipids on the cell surface of Map that may in turn have a major influence on the virulence and host specificity of this bacterium. In addition, the relative lack of functional redundancy in this pathway in Map suggests that the inactivation of the putative DIM polyketide biosynthesis pathway related genes in Map may be a promising approach to the development of attenuated or vaccine strains in this organism.

Map Diagnostics and Strain Differentiation. The identification of unique sequences within the Map genome has already provided a foundation on which to design and implement better diagnostic assays for Map detection. Our analysis has identified ~161 unique sequence regions in the Map genome, with the longest region being 15.9 kb in length (Fig. 1). More importantly, our preliminary studies show that these unique sequences have considerable potential for the development of more specific and sensitive diagnostic assays for detection of Map infection with both molecular and immunoassay based approaches (58–61). Our more recent studies, enabled by the identification of the unique regions in the Map genome, have resulted in the development of highly sensitive real-time PCR-based approaches for the sensitive and specific detection of Map directly from bovine feces (N.B., L.L., A.A., J.P.B., V.K., and S.K., unpublished data). Studies are underway to prepare and express the putatively unique Map genes as well as the genes encoding cell surface proteins for construction of a partial protein array to evaluate the humoral and cell-mediated immunostimulatory capabilities of these recently discovered unique proteins. Therefore, the combination of genomic information, molecular tools, and immunological assays will provide key insights to the host immune response to Map infection. Overall, the elucidation of all of the unique sequences as well as those that may be associated with the cell surface of Map provides a strong foundation on which to develop the next generation of specific and sensitive diagnostic assays for Map.

Short sequence repeats (SSRs) or variable number tandem repeat (VNTR) sequences have been used as markers for differentiation and subtyping strains of several bacterial species including Mtb, Yersinia pestis, and Bacillus anthracis (62–65). SSRs consist of simple homopolymeric tracts of single nucleotide (mononucleotide repeat) or multimeric tracts (of homogeneous or heterogeneous repeat), such as di- or trinucleotide repeats, which can be identified as VNTRs in the genome of the organism (66). The variability of the repeats is believed to be caused by slipped-strand mispairing, the genetic instability of polynucleotide tracts, especially poly(G-T), and DNA recombination between homologous repeat sequences. In preliminary bioinformatic analyses, we had identified 185 mono-, di-, and trinucleotide repeat sequences dispersed throughout the Map genome, of which 78 were perfect repeats (67). Comparative nucleotide sequencing of the 78 loci in six Map isolates from different host species and geographic locations identified a subset of 11 polymorphic SSRs with an average of 3.2 alleles per locus (67), and has provided the foundation for the development of highly discriminatory and powerful multilocus SSR (MLSSR)-based typing approach for strain differentiation among isolates of Map (68). In the current investigation, we identified an additional 362 sequences representing either direct or indirect sequence repeats of length distribution of 6–74 bp with a repeat number ranging from 2 to 16, and with a mutual homology of 67–100% (Fig. 2). Based on our recent success using MLSSR for Map strain differentiation and in understanding the molecular epidemiology of the organism, it is likely that these repeat elements in the Map genome will provide additional strain differentiation capabilities and may enable rapid and facile discrimination of epidemiologically and geographically distinct strains of isolates of Map by using nonsequencing based approaches as well (69).

It is noteworthy that the availability of the full complement of all of the genes in Map provides an opportunity to perform a complete metabolic reconstruction to enable a better understanding of the natural physiology of this organism and the metabolic requirements for growth, and thereby enable faster growth of Map in laboratory culture.

Concluding Comments. In summary, the complete genome sequencing and comparative genomics analyses described herein have provided key insights and a strong foundation for future investigations on the genetics, evolution, natural physiology, and virulence of this important animal pathogen. Furthermore, the results of our studies provide the foundation for the development of the next generation of diagnostic and strain differentiation approaches, and provide a framework for the application of genomics and proteomics based approaches for the development of vaccines to prevent and control Johne’s disease in domestic livestock.
Funding for this project was provided, in part, by grants from the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service National Research Initiative competitive grants program as well as the Agricultural Research Service (to V.K. and J.P.B.).