GENETIC ANALYSIS OF *Mycobacterium avium* subspecies *paratuberculosis* REVEALS SEQUENCE AND EPIGENETIC VARIATION AMONG FIELD ISOLATES

A Dissertation

by

BRIAN JAMES O'SHEA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2007

Major Subject: Veterinary Microbiology

GENETIC ANALYSIS OF *Mycobacterium avium* subspecies *paratuberculosis* REVEALS SEQUENCE AND EPIGENETIC VARIATION AMONG FIELD ISOLATES

A Dissertation

by

BRIAN JAMES O'SHEA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Allison Rice-Ficht
	Thomas A. Ficht
Committee Members,	L. Gary Adams
	Patricia Klein
Head of Department,	Gerald Bratton

December 2007

Major Subject: Veterinary Microbiology

ABSTRACT

Genetic analysis of *Mycobacterium avium* subspecies *paratuberculosis* reveals sequence and epigenetic variation among field isolates.

(December 2007)

Brian James O'Shea, B.S., Louisiana State University

Co-Chairs of Advisory Committee: Dr. Allison Rice Ficht
Dr. Thomas A. Ficht

Previous research performed in 1999 by Harris et al. has shown that many varieties of ruminants serve as the host species for *Mycobacterium avium* subspecies *paratuberculosis* (MparaTb) infections. Gene sequencing has supported the contention that organisms isolated from different hosts harbor different gene sequences; this has been exemplified by Amonsin et al. in 2004 with the sequencing of the *mfd* (transcription-repair coupling factor) and by Motiwala et al. in 2005 through sequence analysis of phosphatidylethanolamine-binding proteins which reveal a host-specific correlation of isolates. Some contradicting reports from Bannantine et al. from 2003 have further claimed that MparaTb is a monogenic organism based upon sequence data from regions flanking the origin of replication and the 16s rRNA. One of the drawbacks to the techniques implemented in these reports is the extremely restricted region of the bacterial genome that was analyzed; furthermore, only a select number of isolates were analyzed. In the present studies, amplified fragment length

polymorphism (AFLP) was used as a tool for a genome scale comparison of MparaTb isolates from differing isolation types as well as a comparison of MparaTb isolates to the genetically similar yet avirulent *Mycobacterium avium* subspecies *avium* isolates. AFLP data reveals the MparaTb genome to be much more plastic and polymorphic than previously thought. These polymorphic regions were identified and characterized and are shown to be unique to the organism when compared to an array of Mycobacterial isolates of differing species. These polymorphic regions were also utilized in polymerase chain reaction (PCR) based diagnostic as well as epidemiologic tests. Furthermore, AFLP comparative analysis of intracellular and fecal MparaTb isolates reveals polymorphic regions unique to each isolate type. While these genomic differences are not based upon differences in the genetic code, they are based upon epigenetic modifications such as DNA methylation. These DNA methylation patterns are unique to intracellular MparaTb isolates as opposed to isolates from fecal material. Furthermore, AFLP comparisons of fecal MparaTb isolates that were passaged through the bovine ileum revealed banding pattern differences as compared to the original inoculum.

DEDICATION

To my mom, dad and brother,

Thank you for your patience and support.

ACKNOWLEDGMENTS

I would like to thank my committee members, Dr. Allison Rice-Ficht, Dr. Thomas Ficht, Dr. L. Garry Adams and Dr. Patricia Klein for their support and guidance through the years.

I would also like to thank Dr. Sangeeta Khare for all her assistance throughout my project.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

INTRODUCTION AND LITERATURE REVIEW

BACKGROUND

The disease. MparaTb is the etiologic agent of the severe chronic granulomatous enteritis in ruminants, known as Johne's disease or *paratuberculosis* (4, 20, 21, 44). This debilitating disease can affect domestic livestock and causes an estimated annual agricultural loss of $1.5 billion to the cattle industry in the U.S. alone (20). Although this estimate may seem high, actual losses are suspected to be far greater due to the lack of accurately estimating herd profit levels. MparaTb is estimated to be prevalent in 35% of United States dairy herds (4). Although Johne's primarily affects cattle, MparaTb infects other ruminants including bison and is suspected in human infections leading to Crohn's disease (4, 20, 21). However, inconsistent detection of MparaTb in specimens obtained from Crohn's patients has cast doubt on any direct link (20).

The primary route of infection of MparaTb is fecal-oral transmission from an infected food source. Individuals acquire the bacteria by consuming contaminated food products or food products that have come in contact with contaminated areas. Once ingested, the bacterium travels to the

---

This dissertation follows the style of Journal of Clinical Microbiology.

mucosal-associated lymphoid tissue (MALT) of the small intestine where it is endocytosed by M cells of the Peyer's patches and subsequently phagocytosed by intraepithelial macrophages (20). Interaction with or replication within macrophages leads to an inflammatory immune response and clinical signs of disease including chronic granulomatous enteritis. A corrugated mucosal surface is observed which as the disease progresses results in malabsorption and chronic wasting, the clinical signs of Johne's Disease.

The organism. The *Mycobacterium avium* species is currently divided into three subspecies; *M. avium* subspecies *avium* (*M. avium*), *M. avium* subspecies *silvaticum* (*M. silvaticum*), and *M. avium* subspecies *paratuberculosis* (MparaTb). MparaTb, a Gram positive, acid-fast bacillus, can be differentiated phenotypically from the other two *M. avium* subspecies by its dependence on mycobactin J for iron acquisition in growth, its extremely long doubling time and genetically by the presence of the insertion sequence IS900. Although most MparaTb isolates require mycobactin J for growth, there have been reports of MparaTb isolates that grow in the absence of mycobactin J (1). American Type Culture Center (ATCC) has stated that the widely used MparaTb isolate, ATCC strain 19698, has lost its dependency upon mycobactin J for *in vitro* growth. Also, there have been reports of MparaTb isolates that lack the IS900 insertion sequence (16).

Genome organization. Due to the extremely long doubling time of the organism genetic analysis of paratuberculosis has been quite limited. However, recently the entire sequence analysis of the genome was completed and annotated. Genomic sequence analysis of the MparaTb bovine isolate K10 revealed a genome of over 4.8 million base pairs encoding 4,350 ORF's (36). Comparative analysis of these ORF's revealed over 3,000 genes with sequence homology to *M. tuberculosis* and 161 MparaTb unique regions including 39 previously unknown MparaTb genes. These 161 MparaTb unique regions varied in length up to 15.9kb and could possibly be exploited for their diagnostic/epidemiologic potential.

In addition to sequence analysis of the MparaTb genome, other genetic techniques have been employed to analyze the MparaTb genome. Short sequence repeats (SSR) analysis and single nucleotide polymorphism (SNP) analysis have also been employed to locate and distinguish genetic regions for their possible diagnostic or epidemiologic purposes (3). These techniques are severely restricted analyzing only 6-11 independent sites in a single genome. While SSR does involve sequencing, it does not yield a whole genome scale analysis like genomic sequencing. Genomic sequencing however does not analyze multiple isolates. In an effort to rapidly analyze multiple isolates and to analyze a very large portion of the genome at one time, amplified fragment length polymorphism (AFLP) was used in this study.

Amplified fragment length polymorphism (AFLP) analysis is a whole genome approach to comparative genomics. This technique employs the use of a double restriction enzyme digest followed by ligation of synthesized adapters to these restriction sites and a specific polymerase chain reaction (PCR) to amplify a specific set of these restriction fragments (33). The AFLP technique is demonstrated in the diagram below (Figure 1).

AFLP is an excellent technique to implement a more comprehensive comparison of multiple organisms. While there are other techniques that can be used to track or even discover polymorphic genetic regions such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA markers (RAPD), and short sequence repeat polymorphisms (SSR) (28, 38, 42, 52, 54, 57, 65); AFLP is the best technique to analyze multiple loci in a short period of time.

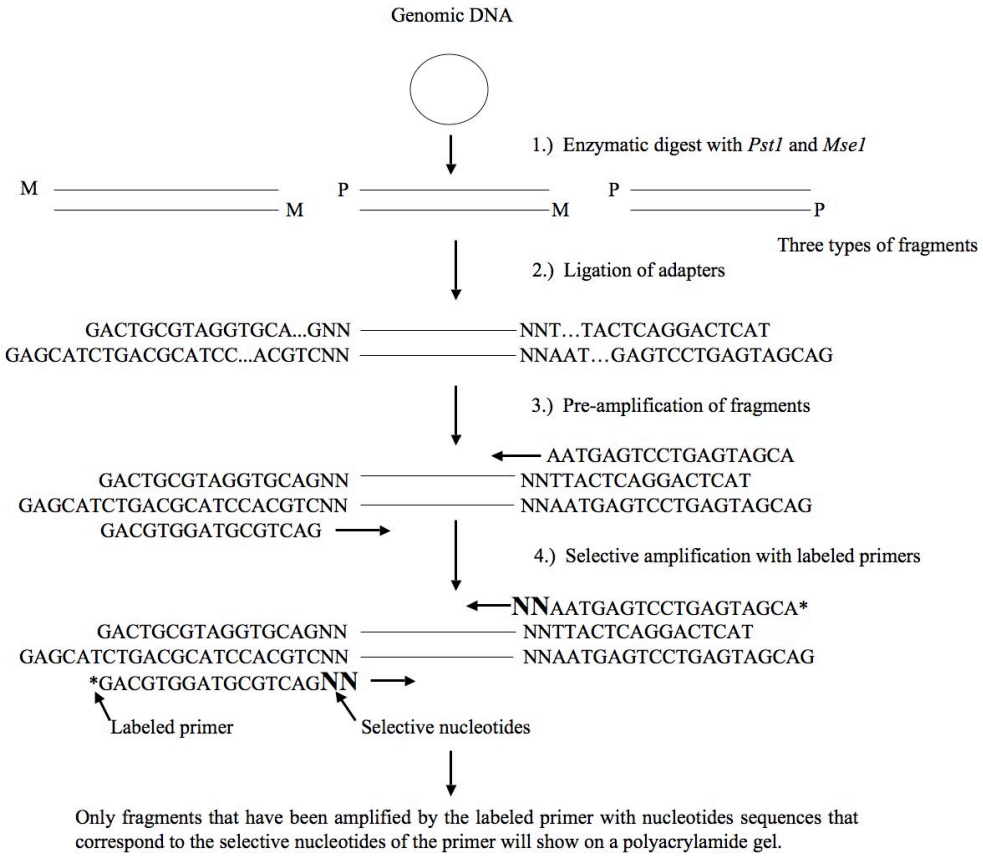


FIGURE 1: Schematic of the AFLP process.
1.) A double restriction enzyme digest is carried out on genomic DNA to reveal three types of fragments. 2.) Adapter oligos are then annealed together to create "sticky ends" and ligated to the cut sites on these genomic fragments. 3.) A pre-amplification is then carried out with PCR primers annealing to the adapters and having the 3' end of the primers stop at the 5' end of the restriction site. 4.) A final selective amplification is then carried out with primers identical to the pre-amplification primers except that they have two additional selective nucleotides on their 3'ends and are labeled with a chromophore on the 5' end.

All of these assays detect genetic polymorphisms based upon a subset of the total DNA sequence variation in the genome. Like AFLP, RFLP detects polymorphic regions based upon restriction size variation (40, 42, 52). AFLP and RAPD detect polymorphisms by DNA sequence variation at primer binding

sites and from DNA length differences between primer binding sites (28, 52, 65). SSR detects polymorphisms in di- and tri-nucleotide repeat sequences by PCR with primers flanking the regions in question followed by direct sequence analysis of the corresponding PCR products (38, 52, 57). All of these analyses only detect a subset of the total polymorphic genetic variation of the corresponding genome.

Powel *et al*. (52) developed a mathematical equation to compare these techniques; this equation was employed to determine that AFLP was the optimal choice for the research project described here. The equation is as follows: $MI=EH_{av}(p)$ where MI is the marker index, E is the multiplex ratio (number of bands simultaneously analyzed per experiment) and $H_{av}(p)$ is the average expected heterozygosity of polymorphic regions. These variables are dependent upon several factors: the number of loci analyzed per lane per gel, the fraction of those loci which are polymorphic, and the average allele frequency within a population. To have a high marker index, such that the test will be effective at detecting polymorphic regions, the test must examine multiple genetic regions in a single test and be able to distinguish the polymorphic regions from the non-polymorphic regions. AFLP does this by detecting a very high frequency of alleles as compared to RFLP and SSR. Powell *et al.* concluded that while AFLP would not be a very effective technique in tracking polymorphic regions from one isolate to the next, it is a very effective technique for discovering these

polymorphic regions based on the massive volume of DNA analyzed per experiment. Thus, based on these characteristics AFLP was selected to detect genomic polymorphisms in our research. PCR and direct DNA sequencing on the other hand would be a more effective but more expensive technique for polymorphic region comparisons from isolate to isolate. To determine if these polymorphic regions were consistent throughout all isolates, PCR and sequence analysis was subsequently applied. By employing the AFLP analysis to MparaTb field isolates from varying host species, the extent of genetic polymorphism occurring within the *M. avium paratuberculosis* subspecies was revealed.

DNA diagnostics. Although there is no current "gold standard" diagnostic test for the presence of MparaTb, the only diagnostic test for the presence of MparaTb DNA is the IS900 PCR. IS900 PCR is based upon the presence of a multi copy insertion element within the bacterial genome (16, 32, 50). IS900 is a 1,451bp insertion element belonging to the IS110 family of insertion elements that has an 82% sequence similarity to the IS1626 insertion sequence of *M. avium avium* (20). While IS900 first appeared to be unique to MparaTb, further studies have shown that not all MparaTb isolates are present for the IS900 sequence and studies have also shown the presence of IS900 in the genomes of non-paratuberculosis soil Mycobacteria (20). The absence of IS900 from a subset of MparaTb isolates and the presence of IS900 in non-

paratuberculosis isolates reduces the effectivness of IS900 as a diagnostic PCR. The development of a new PCR based diagnostic test for the presence of MparaTb is needed.

Culture and ELISA diagnostics. In addition to the IS900 based PCR diagnostic test there is also the widely used direct bacterial culture from fecal samples and the detection of MparaTb antigens through the use of an enzyme-linked immunosorbent assay (ELISA). Although these tests have the potential to become the gold standard, both tests have serious drawbacks that hinder their performance and reliability.

The direct bacterial culture technique is severely hindered by the organism's slow rate of growth with a doubling time reported to be 22-26 hours (5) *in vitro* and as high as 48 hours *in vivo* (19). Confirmation of a negative sample requires a 16-30 week incubation while pinpoint colonies can be visible at 12 weeks in some cases (34). This lengthy doubling time can also results in the spread of the disease through domestic herds while infected cattle await diagnosis. In severe cases, the infected animal may be culled before the test results have been received. Studies have also shown variability in the sensitivity of the fecal culture test from differing host species. In cattle, the sensitivity ranges from 20-73% while in sheep and goats it is much lower at 6-7% (34, 49, 58). This culture technique is also hindered by the bacteria's varying growth requirements. Although it is widely known that MparaTb requires the use of

mycobactin J for growth *in vitro*, many MparaTb field isolates have been found to be mycobactin J independent yielding a false negative result (1). The limited number of bacteria shed in the feces is also a drawback to the direct fecal culture technique. Research has shown that the numbers of bacteria shed in the feces can vary from host to host (with some infected individuals failing to shed any bacteria) and this may prevent a proper diagnosis (11). A low or non-shedder that gives a false negative for a fecal test can be returned to the herd where shedding will increase resulting in transmission from one animal to others.

The ELISA assay for the detection of MparaTb, while being a more rapid test than the direct fecal culture, is hindered by the sensitivity of the antigen used in the test. The current antigen used today for the MparaTb ELISA does not detect all MparaTb isolates, resulting in many false negatives. Reports have given varying sensitivities for the MparaTb ELISA test ranging from 8-63% depending upon the host species tested. Specificity for the MparaTb ELISA ranges between 93-98% for all tested species (34, 49, 56). These reports of variability in the sensitivities of the MparaTb ELISA from host to host further support the hypothesis of a host-specificity for isolates and further indicates the need for a more reliable diagnostic test.

Epigenetic modification. Epigenetic modifications are genomic modifications that do not alter the genetic sequence. One example of an epigenetic modification is the addition of a methyl group to cytosine or adenine residues and commonly referred to as DNA methylation. These minor methylation events can lead to an increase or decrease in the potential pathogenesis of a particular organism (26). These methylation events have been shown to improve adherence to cell surfaces as well as intracellular survival (39).

There are three types of DNA methylation that occur in bacteria. All three of these methylation events either take place on an adenine or cytosine base in the DNA sequence. These three methylation events are N^6-methyladenine (m^6A), 5-methylcytosine (m^5C) and N^4-methylcytosine (m^4C). m^6A and m^5C have long been known to be minor bases in bacterial DNA, but until recently m^4C was only thought to be present in thermophilic bacteria (14, 15, 63). DNA methylated bases occur naturally in bacteria and are a result of post-replicative enzymatic modification of DNA by a specific methyltransferase (63). For m^5C and m^4C specific methytransferases, methylation occurs at the corresponding numbered cytosine position. Specifically, m^5C is methylated at the 5' carbon and m^4C has a methyl group added to the amino group bonded to the 4' carbon. As for m^6A, a methyl group is added to the amino group bonded to the 6' carbon.

While DNA methylation is very common in all bacteria, not all methylations are exhibited in all species of bacteria. In *Mycobacteria* specifically, there have been reports revealing the presence of 5-methylcytosine in the virulent strain of *tuberculosis* H37Rv and its absence in the avirulent strain H37Ra suggesting a role of DNA m^5C methylation as a virulence factor for tuberculosis (59). In the same study *M. smegmatis SN2* revealed the presence of m^5C methylation and all three of these strains contained m^6A (59). Other studies have taken this research a step further and tried to identify the methylase responsible for the m^6A and m^5C methylation of these two *M. tuberculosis* isolates and one *M. smegmatis* isolate. In this report methylation sensitive isoschizomers of known restriction enzymes were used to probe the nature of DNA methylation. These reports found that in all the species examined and at all of the recognition sites with these enzymes, there was no evident methylation. On the other hand, high performance liquid chromatography (HPLC) revealed significant levels of m^6A and m^5C suggesting that methyltransferases other than *Dam* and *Dcm* were responsible for the DNA methylation (24).

A popular technique currently used to determine the methylation status of DNA is methylation specific PCR (MSP) (10). This technique entails the use of bisulfite modification of the DNA to convert cytosine to uracil by replacing the amino group bonded to the 4' carbon of cytosine with a carboxyl group. This

nucleophilic substitution of an amino group will change the cytosine to a uracil. On the other hand, if the cytosine is methylated the bisulfite modification will not be able to cleave the amino group on the 4' carbon, and therefore the cytosine will not be converted to uracil. When sequenced, all uracil bases will yield a sequence of thymine and all methyl-cytosines will sequence as cytosines (8, 25, 63). There have also been reports revealing methylation sites by using direct sequencing analysis with different sequencing dye kits (6). One report states that automated dye terminator sequencing data of m^6A in the template DNA resulted in an increase in the complementary T signal of the sequencing reaction and reactions involving m^5C resulted in a decrease in the complementary G signal (55).

DNA methylation patterns can be inherited from one generation to the next. In *E. coli* for example, GATC sites surrounding the origin of replication and spanning about 250 base pairs are fully methylated before DNA replication, meaning both strands of the sequence (forward and reverse) are methylated. Once replication begins, the strands become hemimethylated such that only one strand of the duplex is methylated, but very quickly the *Dam* methylase methylates the two daughter-stands to create two fully methylated strands. Replication in *E. coli* cannot initiate at the *oriC* until the parental strands are fully methylated. There are many documented GATC regions throughout the genome that are fully methylated in order to distinguish the daughter strand from

the parental strand (9). These methylation patterns are highly conserved within *E. coli* and are passed from generation to generation. Research has also shown that methylation patterns for the pyelonephritis-associated pili (*Pap pili*) in which two GATC sites near the *pap pilin* promoter are reversely methylated depending upon whether the *pap pili* is in the ON-phase or OFF-phase. In the on-phase, the proximal GATC site is methylated, and the distal site is not. The opposite is true for the off-phase of the pap pili (9, 37).

There have been a few documented evaluations of DNA methylase knockout mutants with regard to virulence. In one case, an attenuation in the *Dam* knockout mutants of *Salmonella typhimurium* resulted in defects of protein secretion, cell invasion and cell cytotoxicity(18). Another documented case of methylase knockout mutants determined the role of methylation in virulence of *Shigella flexneri 2a*. In this case, the *Dam* gene was knocked out, and the *Shigella* clone was examined for infectivity and virulence. In contrast to the previous report on *Salmonella*, the *Shigella Dam* mutant retained in *vivo* and *in vitro* virulence. Although *the Shigella Dam* mutant invaded cultured L2 monolayers as efficiently as wild-type cells, its intracellular growth was suppressed up to 7 hours post-invasion. These *Dam* mutants also formed smaller plaques in cell monolayers and produced keratoconjunctivitis in guinea pigs only slightly more slowly than wild type. While the overall infectivity and pathogenicity of the *Dam* mutant was not significantly affected, the spontaneous

mutation rate of the *Dam* mutant was 1000-fold greater than the *Dam* positive wild-type bacteria (27). Another reported case of a methylase knockout mutant found similar results to Honma *et al*. (27) in which the *Dam* knockout mutant was only partially attenuated (43).

AMPLIFIED FRAGMENT LENGTH POLYMORPHISM REVEALS GENOMIC VARIABILITY AMONG *Mycobacterium avium* subspecies *paratuberculosis* ISOLATES

INTRODUCTION

Mycobacterium avium subsp. *paratuberculosis* (MparaTb), a Gram positive, acid-fast bacillus is the etiologic agent of severe chronic granulomatous enteritis in ruminants, known as Johne's disease (4, 20, 21, 41). This debilitating disease can affect all domestic ruminants causing an estimated annual loss of $1.5 billion to the livestock industries (20). Although this estimate may seem high, actual losses are suspected to be far greater due to the lack of accurately estimating herd profit levels. MparaTb is estimated to be prevalent in 35% of United States dairy herds (4). Although Johne's primarily affects cattle, MparaTb infects other ruminants including bison and is suspected in human infections leading to Crohn's disease (4, 20, 21). However, inconsistent detection of MparaTb in specimens obtained from Crohn's patients has cast doubt on a direct link (20).

The primary route of infection of MparaTb is fecal-oral transmission from an infected food source. Once inside an infected ruminant, the bacterium travels to the mucosal-associated lymphoid tissue (MALT) of the distal jejunum and ileum where it is endocytosed by M cells of the Peyer's patches. Endocytosis by

M cells subsequently leads to phagocytosis by intraepithelial macrophages (20). Interaction with or replication within macrophages leads to an inflammatory immune response and clinical signs of disease including a severe chronic granulomatous enteritis resulting a corrugated mucosal surface which, as the disease progresses, results in malabsorption and chronic wasting, the clinical signs of Johne's Disease.

In this study, AFLP analysis was performed on of 20 MparaTb field isolates and these results were compared to AFLP analysis of *M. avium avium* isolates in an effort to identify polymorphic regions present in either of these subspecies. Due to the ability of AFLP analysis to detect single nucleotide polymorphisms in addition to insertions/deletions of the bacterial genomes, the nature and extent of identified polymorphisms was further evaluated. Two of the polymorphisms identified were due to large genomic regions present in MparaTb and missing from *M. avium avium*. Using data on five such deletions PCR primer sets were devised that may be used in combination with IS900 as a diagnostic/epidemiologic tool for MparaTb. When applied to field isolates, ATCC standard strains, and to a standard test set of 25 organisms supplied by the National Veterinary Diagnostics Laboratory, PCR of these polymorphic regions reveals extensive genetic variation. In contrast to the results reported by Kapur *et al*., (4, 5), genomic differences between MparaTb and *M. avium avium* were not consistent throughout genomes of MparaTb field isolates. Genetic variation

at the subspecies level, revealed by AFLP and PCR is common and may have great value in epidemiological studies. The distinction between *Mycobacterium avium* subspecies is much less clear than previously believed.

This study analyses the overall genetic diversity of MparaTb field isolates in comparison to avirulent *M. avium avium*. By analyzing these genetic differences, regions common to MparaTb isolates and absent form *M. avium avium* isolates could be studied for their use in possible diagnostic applications.

MATERIALS AND METHODS

Mycobacterial isolates. Twenty different MparaTb clinical isolates (109, 126, 129, 130, 132, 135, 136, 139, 140, 141, 142, 144, 145, 148, 150, 151, 153, 155, 156 and 160) were obtained from the University of Wisconsin Veterinary Diagnostic Lab (UWDL). One *M. avium avium* clinical human isolate (Mac 104) was obtained from Dr. Raul Barletta at the University of Nebraska. Two strains were obtained from American Type Culture Center (ATCC) and used in this study: *M. avium avium* (ATCC 35716), and *M. avium* subspecies *paratuberculosis* (ATCC 19698). A set of 25 coded fecal samples in the form of a blind fecal test set (consisting of high, moderate, low shedders and negative feces) was generously provided by Janet Peyeur at the National Veterinary Services Laboratory, Ames, IA (NVSL) (Table 4).

Species identification of isolates. All mycobacterial isolates were characterized by growth requirements, colony morphology, acid-fast staining, and IS900 PCR as previously described (32). MparaTb colonies were evident by the sixteenth week of incubation at 37°C. All MparaTb and *M. avium avium* isolates were grown until sufficient growth was evident for DNA extraction. IS900 PCR analysis with primers P90 and P91 coupled with restriction analysis (Table on page 21) was used to confirm identity of MparaTb isolates as previously described (32). *M. avium avium* isolates produced no IS900 amplicons under these PCR conditions.

Preparation of genomic DNA. Genomic DNA was prepared as previously described by Khare *et al*. (32) through a bead beating method; in cases of fecal sample processing, immunomagnetic separation of the organism was employed in conjunction with this method.

AFLP analysis. AFLP analysis was performed on all 20 MparaTb field isolates, Mac 104 and both ATCC strains with all six *PstI* selective primers in conjunction with all 16 possible *MseI* selective primers for a total of 96 unique primer combinations and a total of 2,208 independent AFLP reactions. Generation of AFLP markers was carried out as described previously by Menz *et al*. (44) with the following modifications. Genomic DNA (500 ng) was digested with *MseI* (1.25 U/μg DNA) for 2 hours at 37°C in NE buffer 2 (New England Biolabs, Beverly, MA). After *MseI* digestion, NaCl and Tris-HCl concentrations

were increased to 50μM and 40μM respectively, and the DNA was digested with *PstI* (1.25 U/μg DNA) (New England Biolabs, Beverly MA) for 2 hours at 37°C. *MseI* and *PstI* adapters (Table on page 21) were diluted to 5 μM and 0.5 μM respectively and ligated to restriction sites of the digested DNA through the addition of 1Unit of T4 DNA Ligase (Roche) at 37°C overnight. The DNA was diluted to a final concentration of 1 ng/μl.

Pre-amplification of samples was performed using 5 ng of adapter-ligated template with primers: *PstI* preamp primer, 5'-GACTGCGTAGGTGCAG-3'; *MseI* preamp primer, 5'-ACGATGAGTCCTGAGTAA-3'. Pre-amplification reactions were carried out as previously described in Menz *et a*l. (44)

Selective amplification was performed with two selective nucleotides on the 3' end of the primers. Core sequences of the selective primers were 5'-GACTGCGTAGGTGCAG-3' (*PstI*); 5'-GATGAGTCCTGAGTAA-3' (*MseI*). *PstI* selective primers include the core sequence followed by 3' selective dinucleotides and are termed P-GC, P-GT, P-GA, P-CA, P-CG and P-CT respectively. *MseI* selective primers used *MseI* core sequence with all 16 possible dinucleotide combinations on their 3' ends. These primers were termed M-AA thru M-TT according to the selective dinucleotides at their 3' ends. IRDye™ labeled *PstI* selective primers were obtained from LI-COR (Lincoln, NE); additional methods used in the selective amplification process have been previously reported (33).

Isolation and sequence analysis of unique MparaTb bands.

Selective amplification of samples was performed under the conditions described above (*PstI* selective primers were not labeled). The amplified samples were then electrophoresed on 2% (w/v) Metaphor LE agarose gel for 1.5 hours at 100V and stained with EtBr to visualize. Bands corresponding to the molecular weight observed on the AFLP gels were extracted and purified using QIAquick® Gel Extraction Kit (Qiagen; Valencia, CA). DNA fragments were plasmid-ligated and transformed into Top10 *E. coli* cells using PCR®2.1-TOPO® (Invitrogen; Carlsbad, CA) as described by the manufacturer. Transformants were identified by PCR using primers (M13+10 forward and M13 reverse) flanking the plasmid multiple cloning site, according to the manufacturer (Qiagen). Plasmids yielding amplification products of the predicted size were collected with the QIAprep® Spin Miniprep Kit and subjected to sequence analysis at the Genome Technologies Laboratory at Texas A&M University.

BLAST queries and diagnostic primers. Sequence data was used as a query line in BLAST searches to existing genomes. If the BLAST search results revealed sequence homology to MparaTb and not to *M. avium avium* or (since the genomes have not yet been fully sequenced) if there was no homology to either MparaTb or *M. avium avium* genomes, diagnostic primer sets were then designed based on the internal sequences of these polymorphic regions. Diagnostic PCR primer sets (Table 1) were designed using Mac

Vector© and designed to have a GC content reflecting that of the MparaTb genome. These primers were employed in PCR amplification of genomic DNA to verify the presence or absence of these polymorphic regions.

Table 1: Primers used in AFLP comparison of MparaTb and *M. avium avium.*
Primers used in these experiments include AFLP adapters, diagnostic primers derived from internal sequence analysis of unique MparaTb regions, and IS900 primers used for initial determination of subspecies classification of each isolate. All primers sequences are in the 5' to 3' orientation.

	NAME	5'----SEQUENCE----3'
Adapters	*PstI-1*	GACTGCGTAGGTGCA
	PstI-2	CCTACGCAGTCTACGAG
	MseI-1	GACGATGAGTCCTGAG
	MseI-2	TACTCAGGACTCAT
Diagnostic Primers	F1-300	CGCACCAAAAGGCACACTAATC
	B1-300	CGTCGTCACCATTCAGGAACTC
	F1-440	GGCTCGCCGAACTACTTGT
	B2-440	CTCGAAACAACGGTGACAGA
	F1-255	GCGGAAAGTCACACTGCTGATGC
	B1-255	CGGTCTATTCGATCCCCACCTTG
	F1-600	CGAGTTGTCCTGGGGTTTTGG
	B1-600	CGAAAGTCACCGCATCCACG
	F1-360	CGAGATTCGCAAGCACTTCG
	B1-360	CAAGGCTGGCTGTCAGATGC
Positive Control	F1-PC-190	CAACCTCAAACCCGAATAC
	B1-PC-190	CTTGCTGATGGTGGTCTG
IS900	P90	GAAGGGTGTTCGGGGCCGTC
	P91	GAGGTCGATCGCCCACGTGAC

.

PCR analysis with internal primers. PCR amplification of all 20 MparaTb isolates, ATCC 19698, Mac104, and ATCC 35716 and NVSL test set were performed using diagnostic primers. Except for NVSL test set (performed on purified DNA), all other PCRs were performed using template genomic DNA

on Whatman® FTA® membranes (Whatman Inc.; Clifton, NJ). Genomic DNA for PCR was prepared on FTA cards per the manufacturer's protocols and used directly in PCR reactions as template DNA. PCR was optimized using Epicenter's Fail Safe™ PCR Kit (Epicenter Technologies, Madison WI) per the manufacturer's protocol with all four internal primer sets on ATCC 19698 of MparaTb. Premix G exhibited optimal product formation for all four primer sets. PCR was performed by heating treated and washed FTA® card punches with 7.1 μl ddH_2O for 5 minutes at 94°C then adding 12.9 μl of master mix containing 20 pmol of each primer, 10 μl premix G and 1 unit Taq DNA polymerase. Amplification was performed at 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute for 45 cycles, followed by a final elongation at 72°C for 5 minutes. Samples were analyzed on a 2% (w/v) LE agarose gel with EtBr staining. The same internal primers were also used to characterize the IS900 positive samples in a double blind study of 25 coded fecal samples containing *M. avium* isolates, (Johne's disease fecal check test evaluation-2001).

Bionumerics analysis of AFLP data. The Bionumerics computer analysis program (Applied Maths, Austin Texas) was used for cluster analysis of AFLP generated data of the 20 clinical field isolates. These data were used to create a dendrogram representation of genetic relatedness using the Dice method of scoring bands. The Dice method is based solely on presence or absence of a band. The intensity of a single band was not a factor when

comparing the AFLP images. This method was used to create a composite data set from 4 independent AFLP primer sets to create one dendrogram for these isolates.

RESULTS

Preliminary AFLP for identification of polymorphic regions. In an effort to more rapidly identify polymorphic regions common to MparaTb isolates and absent from *M. avium avium* isolates, a preliminary AFLP screen of 2 MparaTb isolates and 2 *M. avium avium* isolates was conducted. Isolates included MparaTb isolates ATCC 19698 and clinical isolate 109, and *M. avium avium* isolates Mac104 and ATCC 35716. These four isolates were used in conjunction with multiple primer sets for the discovery of bands present in both MparaTb isolates and absent from both *M. avium avium* isolates (Figure 2). Once a region was found only present in the MparaTb isolates, this same primer set was then used in conjunction with all 20 MparaTb clinical isolates and again with both *M. avium avium* isolates for clarification of the presence of the polymorphic region in all MparaTb isolates.

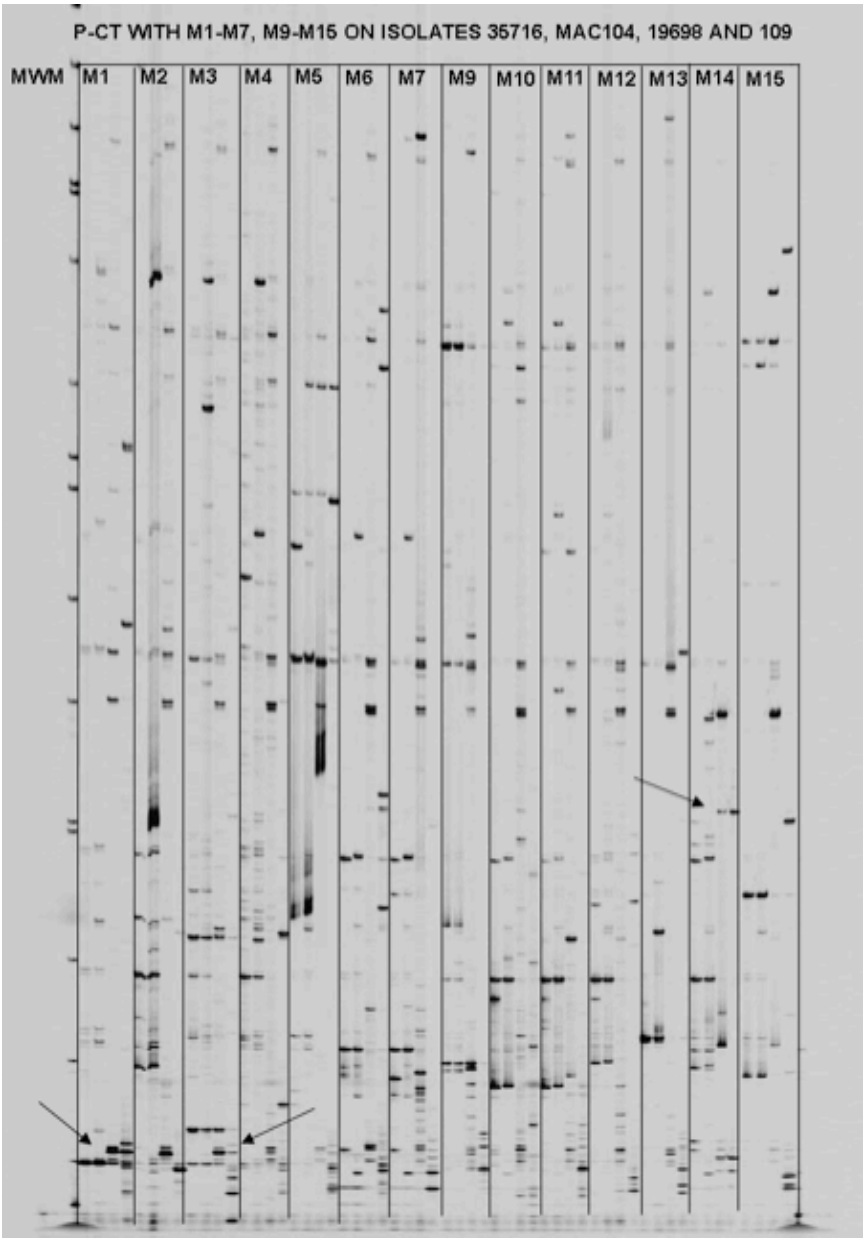


FIGURE 2. Preliminary AFLP screen of MparaTb and *M. avium avium* isolates.

The above image shows a preliminary AFLP screen of one *PstI* primer (*PstI* primer with CT 3' di-nucleotides) in conjunction with fourteen *MseI* primers (from left to right *MseI* primer with 3' di-nucleotides AA, AC, AG, AT, CA, CC, CG, GA, GC, GG, GT, TA, TC and TG). Isolates used are, from left to right in each panel, *M. avium avium* isolates ATCC 35716 and Mac104, and MparaTb isolates ATCC 19698 and clinical isolate 109. Arrows indicate polymorphic regions present in both MparaTb isolates and absent in both *M. avium avium* isolates. These regions warranted further study by conducting AFLP on all isolates with the polymorphic region's corresponding primer set.

Identification of polymorphic MparaTb sequences using AFLP analysis. AFLP analysis revealed 5 regions that were present in the majority of MparaTb isolates and undetected in the *M. avium avium* genome. These regions have the following corresponding AFLP primer combinations: region 1 [P-GC, M-CT (300bp band) (Fig. 3, Panel 2)]; region 2 [P-GC, M-CT (440bp band) (Fig. 3, Panel 2)]; region 3 [P-CT, M-TC (255bp band) (Fig. 3, Panel 5)]; region 4 [P-CG, M-TT (600bp band) (Fig. 3, Panel 8)]; region 5 [P-GG, M-TC (360bp band) (Fig 4, Panel 3)]. The criteria used to warrant a region for further investigation from the AFLP gels was the appearance of the band in a majority of the 20 clinical MparaTb isolates and ATCC 19698 MparaTb isolate, and absence or reduction in band intensity in the two *M. avium avium* isolates. Although putative MparaTb specific bands were observed on some of the polyacrylamide gels down to 50bp, only bands present above 200bp were investigated further. Analysis of the 20 MparaTb isolates by AFLP revealed a consistent banding pattern in as many as 8 repetitions and an apparent heterogeneity was revealed at the subspecies level among these 20 MparaTb field isolates.

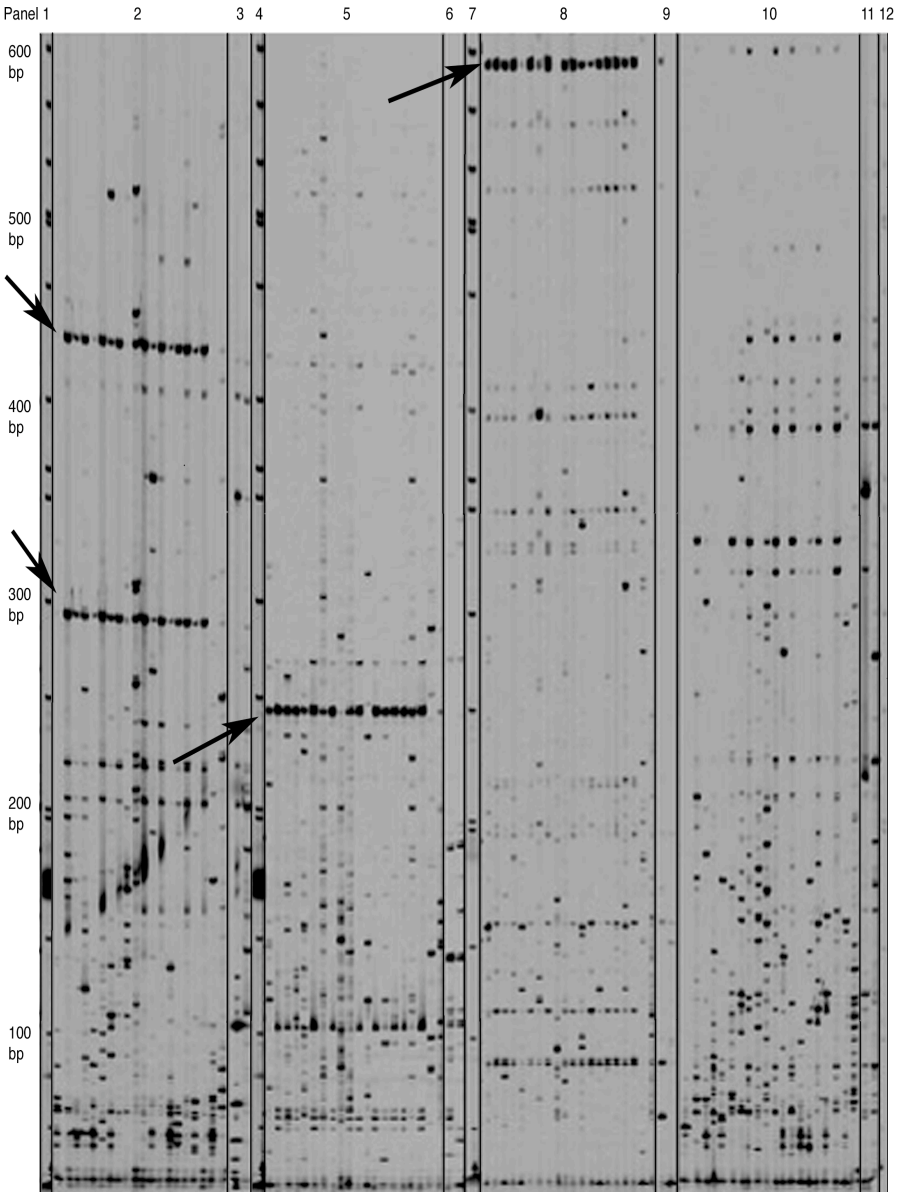


FIGURE 3. AFLP gel showing regions 1-4 and polymorphism between isolates.
Panels: 1, 4, 7, and 12, molecular weight markers; 2, *M. avium* subsp. *paratuberculosis* field isolates 109 to 160 (left to right), respectively, amplified with AFLP primers P-GC and M-CT showing regions 1 (300bp) and 2 (440 bp); 3, *M. avium* subsp. *avium* isolates ATCC 35716 and Mac104, respectively, amplified with AFLP primers P-GC and M-CT; 5, *M. avium* subsp. *paratuberculosis* field isolates 109 to 160 (left to right), respectively, amplified with AFLP primers P-CT and M-TC showing region 3 (255 bp); 6, *M. avium* subsp. *avium* isolates ATCC 35716 and Mac 104, respectively, amplified with AFLP primers P-CT and M-TC; 8, *M. avium* subsp. *paratuberculosis* field isolates 109 to 160 (left to right), respectively, amplified with AFLP primers P-CG and M-TT showing region 4 (600bp); 9, *M. avium* subsp. *avium* isolates ATCC 35716 and Mac 104, respectively, amplified with AFLP primers P-CG and M-TT; 10, *M. avium* subsp. *paratuberculosis* field isolates 109 to 160 (left to right), respectively, amplified with AFLP primers P-GC and M-TT showing a high degree of heterogeneity among isolates; 11, *M. avium* subsp. *avium* isolates ATCC 35716 and Mac 104, respectively, amplified with AFLP primers P-GC and M-TT.

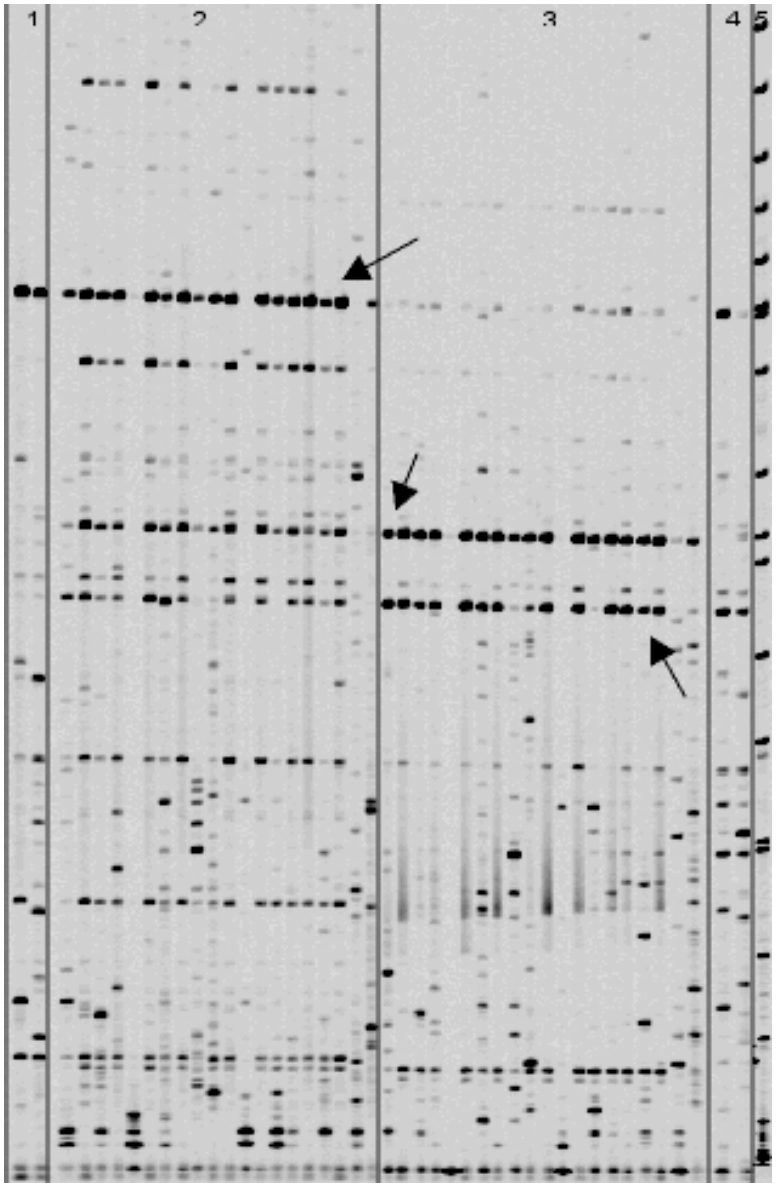


FIGURE 4. AFLP gel showing polymorphic region 5.
Panels 1 and 2 are AFLP primer combination P-GG, M-CC. Panels 3 and 4 are AFLP primer combination P-GG, M-TC. Panels 1 and 4 are *M. avium avium* isolates Mac 104 and ATCC 35716 from left to right. Panels 2 and 3 are, from left to right, MparaTb isolates 109, 126, 129, 130, 132, 135, 136, 139, 140, 141, 142, 144, 145, 148, 150, 151, 153, 155, 156 and 160. Arrows corresponding to AFLP band in panel 2 and lower band in panel 3 illustrates a band present in most of the MparaTb field isolates but also present in both *M. avium avium* isolates. Upper arrow in panel 3 illustrates MparaTb common band corresponding to region 5. Panel 5 is molecular weight marker.

Identification of MparaTb-polymorphic sequences. Preparative amplifications of regions 1-5 were performed using the appropriate AFLP primer sets; and corresponding DNA fragments were cloned and sequenced as described in the Materials and Methods section. Sequence data were used as a query in nucleotide BLAST searches of both the MparaTb and *M. avium avium* genomes. The five MparaTb-polymorphic regions identified through AFLP analysis described above (regions 1-5) corresponded to contigs 161, 210, 185 and 178 of the MparaTb genome, respectively. None of the regions had any significant similarity to the *M. avium avium* genome. These sequences were also used as a query in an amino acid BLAST search (BLASTX) of the NCBI non-redundant (nr) database. region 1 retrieved no significant similarity to any known protein in the nr NCBI database and designated as unidentified region one (UR1). region 2 displayed amino acid sequence homology to the primosomal protein N' (PPN') of *Brucella melitensis* (E value= 0.033). region 3 displayed amino acid sequence homology to the P44k protein of *Rhodococcus erythropolis* (E value= 5e-16). region 4 displayed amino acid sequence homology to the putative polyketide type 1 synthase (PPKS) of Streptomyces (E value= 0.043). The final region described above, region 5, had amino acid homology to a putative replicase of *Rhodococcus erythropolis* (E value= 0.011).

Table 2. MparaTb polymorphic region primer information.
Table describing five MparaTb unique polymorphic regions and their corresponding AFLP primer combination used to discover these regions, forward and reverse internal diagnostic PCR primers and their corresponding diagnostic PCR product size.

region	AFLP Primer Combination	Forward Primer	Reverse Primer	Product Size
region 1	P-GC M-CT	F1-300	B1-300	158bp
region 2	P-GC M-CT	F1-440	B2-440	226bp
region 3	P-CT M-TC	F1-255	B1-255	112bp
region 4	P-CG M-TT	F1-600	B1-600	291bp
region 5	P-GG M-TC	F1-360	B1-360	206bp

PCR with internal primers derived from polymorphic MparaTb genomic regions. Internal primer sets were investigated for the purpose of validating their use in conjunction with IS900 PCR in identifying and tracking *M. avium* subspecies *paratuberculosis* isolates. The PCR when combined with magnetic bead isolation (7) provided a powerful tool for rapid analysis. Confirmation of polymorphic regions in MparaTb genomes was performed through PCR analysis with primers derived from sequence data internal to these five regions. Field isolate #153 was not used in this part of the study due to loss of culture. All internal primers are listed in Table 1 and region information is listed in Table 2. PCR results can be seen in Table 3. The primer sets for regions 1-3 and 5 amplified their respective products in 19 of 19 MparaTb field isolates and ATCC 19698 strain of MparaTb. region four's primer set (F1-600 and R1-600) amplified its respective product from 18 of 19 clinical isolates and from the ATCC 19698 MparaTb isolate. Amplification of regions 1-3 and 5 were

Table 3: Results of PCR on all isolates with diagnostic primers.
Clinical isolates 109-160, *M. avium avium* isolates (ATCC 35716, Mac104) and ATCC 19698 were all used as template DNA in PCR reactions with diagnostic primer sets for all five MparaTb unique regions. A "+" denotes appropriate PCR product was apparent, a "-" denotes no product.

Isolate #	Source[a]	Diagnostic Primer Set[b]				
		region 1	region 2	region 3	region 4	region 5
109	W	+	+	+	+	+
126	W	+	+	+	+	+
129	W	+	+	+	+	+
130	W	+	+	+	+	+
132	W	+	+	+	-	+
135	W	+	+	+	+	+
136	W	+	+	+	+	+
139	W	+	+	+	+	+
140	W	+	+	+	+	+
141	W	+	+	+	+	+
142	W	+	+	+	+	+
144	W	+	+	+	+	+
145	W	+	+	+	+	+
148	W	+	+	+	+	+
150	W	+	+	+	+	+
151	W	+	+	+	+	+
155	W	+	+	+	+	+
156	W	+	+	+	+	+
160	W	+	+	+	+	+
ATCC 35716	A	-	-	-	+	-
Mac 104	N	-	-	-	-	-
ATCC 19698	A	+	+	+	+	+

[a]*Isolate source: W, Univ. of Wisconsin Vet. Diagnostic Lab; N, Univ. of Nebraska Barletta Lab; A, ATCC*

[b]*Outcome of PCR using primers derived from sequences of unique bands present on AFLP gels: +, presence of PCR product; -, absence of PCR product*

negative for both *M. avium avium* isolates reported here. region four's primer set revealed amplification from ATCC 35716 of *M. avium avium* while Mac 104 was negative leading to the conclusion that region 4 was not MparaTb specific. These primer sets correctly identified all isolates as MparaTb or *M. avium avium*, in agreement with IS900 analysis, with the exception of the Mac 104 isolate.

The positive control primers, F1-PC-190 and B1-PC-190 (Table 1) were designed to amplify a 190bp region that is retained in both the *M. avium avium* and MparaTb genomes. PCR amplification of this region from both *M. avium avium* isolates was used to confirm the presence of genomic DNA and serve as a positive control.

PCR analysis of NVSL isolates with internal primers. Due to previous PCR results suggesting genomic heterogeneity at region 4 among MparaTb isolates obtained from UWDL, PCR analysis was performed using a standard test set of isolates obtained from the NVSL with the MparaTb-internal primer sets derived through AFLP analysis (Table 1). These MparaTb isolates were obtained for the purpose of conducting a double blind diagnostic study for the detection of MparaTb in stool samples via PCR amplification of the IS900 insertion sequence. A portion of this DNA was used in our research to further evaluate our diagnostic primers. PCR analysis of these 25 isolates was optimized and performed with each of the diagnostic primer sets (regions 1-4). region 5 was not used to test these NVSL isolates due to loss of culture.

Seventeen of the 25 NVSL isolates used in the study were identified as MparaTb using IS900 analysis (S. Khare, 2002) that were later confirmed by NVSL. Six of the 25 isolates were confirmed MparaTb negative and 2 of the 25 were considered to be invalid by NVSL in the double blind study due to inconsistencies in results from participating labs. NVSL considered an isolate to be invalid if there was less than a 70% consensus of participating labs. PCR amplification with diagnostics primer sets for regions 1-4 designed in this study were performed on these 25 NVSL isolates. region 1 primer set amplified its corresponding region in 15 of 17 MparaTb isolates and one of the invalid samples. region 2 primer set amplified its corresponding region in 11 of the 17 MparaTb isolates. region 3 primer set amplified its corresponding region in 12 of the 17 MparaTb isolates. Finally, region 4 primer set amplified its corresponding region in only 7 of the 17 MparaTb isolates.

Table 4: Results of PCR on USDA isolates with diagnostic primers.
Results of PCR with region 1-4 diagnostic primer sets on all 25 isolates provided by USDA NVSL. USDA deemed samples 12 and 13 invalid for the fact that less than 70% consensus of laboratories participating in blind study had same results. A "+" denotes PCR product was apparent; "-" denotes no PCR product.

Isolate #	Type of isolate[a]	Diagnostic Primer Sets[b]			
		region 1	region 2	region 3	region 4
1	MparaTb	+	+	+	+
2	NEG	-	-	-	-
3	MparaTb	+	+	-	+
4	MparaTb	+	+	+	-
5	MparaTb	+	+	+	+
6	NEG	-	-	-	-
7	MparaTb	+	-	-	-
8	MparaTb	+	+	+	+
9	MparaTb	+	-	-	-
10	MparaTb	+	+	+	+
11	NEG	-	-	-	-
12	Invalid	+	-	-	-
13	Invalid	-	-	-	-
14	MparaTb	+	+	+	-
15	MparaTb	-	-	-	-
16	NEG	-	-	-	-
17	MparaTb	+	+	+	-
18	MparaTb	+	+	+	+
19	MparaTb	+	+	+	-
20	MparaTb	+	-	-	-
21	NEG	-	-	-	-
22	MparaTb	+	-	+	-
23	NEG	-	-	-	-
24	MparaTb	-	-	+	-
25	MparaTb	+	+	+	+

[a]Type of isolate: Results of the double blind study as provided by NVSL

[b]Diagnostic primer sets: Primer sets developed from sequence data of unique AFLP bands; "+" is a positive PCR product; "-" is a negative PCR product

One of the 17 MparaTb isolates showed no amplification with any primer set and none of the MparaTb-specific primer sets described here produced amplification products from the negative samples. These data suggested that there was a high degree of genomic heterogeneity among these 17 MparaTb isolates. All NVSL PCR results are provided in Table 4.

Bionumerics analysis of AFLP gels. Bionumerics cluster analysis of the previous AFLP gels revealed a high degree of heterogeneity among the twenty MparaTb clinical field isolates (Figure 5). These banding patterns were based and scored upon the presence or absence of the bands at corresponding molecular weights. If an isolate had a band at a certain corresponding molecular weight and another isolate had a more prominent band at the same molecular weight, these two bands are scored with the same weight as each other. Band intensity was not a factor in the creation of the cluster analysis dendrograms. While a few isolates clustered together, the overall dendrogram revealed that the MparaTb genome to be much more diverse than previously thought.

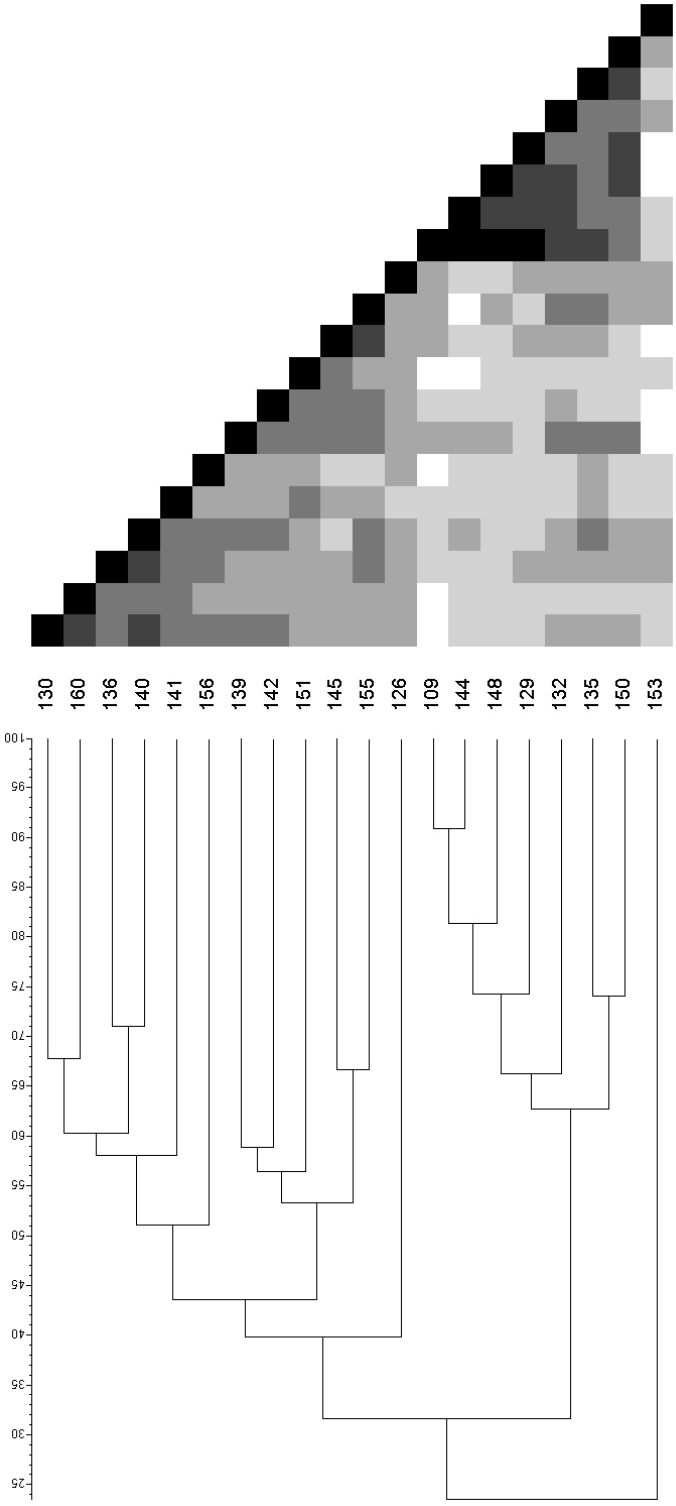


FIGURE 5. Cluster analysis comparison of MparaTb and *M. avium avium* AFLP gels.
A composite data set of AFLP primer sets showing a genetic diversity among MparaTb isolates. Dendrogram was created using the Dice method of clustering which does not account for polymorphic band intensity.

DISCUSSION

AFLP was proven to be a very powerful tool for the analysis of *M. avium* genomes from a variety of sources. While others have suggested that the species is monomorphic, the AFLP data developed in these experiments clearly identified extensive polymorphisms. Extensive AFLP analysis of samples from the UWDL and ATCC strains illustrated a degree of polymorphism uncharacteristic of monomorphic species such as *Bacillus anthracis* (5,6). Several of these observed polymorphisms were verified through cloning, sequence analysis and PCR. The genetic diversity of these clinical field isolates was further evident through Bionumerics analysis of the AFLP gels. . Bionumerics analysis provided strong evidence that the genome of MparaTb is much more diverse than previously believed. However, detailed conclusions about these data were not drawn due to the lack of information about the isolates used in the study. All that was known about these MparaTb isolates was that these isolates were obtained from the University of Wisconsin Diagnostic Lab and that they were of bovine origin. If more information was available (region of isolation, age of bovine host, herd size, etc.), a more detailed conclusion about these data could be drawn. But, the fact that the cluster analysis revealed a much more diverse genome than previously thought was considered to be a new discovery of potential significance.

To assure reproducibility, AFLP patterns of clinical isolates were verified through duplication of AFLP analysis as many as 8 times on single isolates. It was also evident even from PCR analysis that there was heterogeneity among the standard set of NVSL test isolates (Table 4). While the UWDL isolates were relatively homogeneous with respect to PCR of the five regions examined, the more diverse NVSL test set was polymorphic. These data suggest that the host source of the isolate may be a factor in the presence or absence of these five polymorphic regions. Due to the lack of history of all isolates used in this study this could not be investigated further. Approaches by Kapur and others have suggested a more restricted genomic variation among isolates when using PCR directed against specific sequences mined from the NCBI database (4, 5). These sequences were selected based on the sequence of the single MparaTb and the single *M. avium avium* genomes used to construct the sequence databases. The analysis of clinical isolates presented here strongly suggests that the complement of organisms in the environment is richer, and the genomes are more plastic than that represented by the NCBI database sequences. The analysis presented here complements the PCR diagnostic work of others and reveals the ability to detect and utilize polymorphic regions for epidemiology as well as diagnostics.

Initially our approach was directed to commonalities among the MparaTb isolates and to the isolation of DNA sequence that could be used in conjunction

with IS900 for diagnostics and epidemiology. As the study progressed, a large degree of heterogeneity based on AFLP analysis was discovered. As the analysis progressed, even the PCR primer sets designed to be MparaTb-specific detected polymorphism to a limited extent within the UWDL isolates and a much greater extent in the NVSL test set. When applying the PCR tools combined with magnetic bead isolation techniques, it was possible to isolate, identify and fingerprint individual field isolates in a very short time (fewer than 72 hours) (Table 4). In cases where AFLP may be required to track clinical isolates, culture will still be required. Our goal was to identify enough PCR primer sets to allow multiplex identification and to enable isolate tracking and identification using a 72 hour turnaround time. Compared with the weeks required for culture diagnostics, this would enable a highly improved herd management strategy and the reduction in spread of this chronic and insidious pathogen.

Scrutiny of these MparaTb polymorphic regions revealed putative functions relating to housekeeping genes such as integrase/recombinase, PPN', P44k and PPKS. Although the functions of these genes range from the integration and recombination of genomic DNA (integrase/recombinase) (30), priming of lagging strand DNA during replication (PPN') (64), cobalt binding (P44k), and the synthesis of polyketides with any number of biochemical uses (PPKS) (35, 53, 62), no genes with sequence homology to these regions have been identified within the *M. avium avium* genome. When examining the *M.*

avium avium genome for genes of similar function, a gene was found with an 85% amino acid homology to an integrase/recombinase of *M. bovis* (E= e-140) that exists elsewhere within the genome. Whether or not this sequence represents the presence of a recombinase that could provide the functions of the recombinase located within the MparaTb genome was not clear. Of the other genes (PPN', P44k, PPKS), no significant amino acid homology existed; a search of the *M. avium avium* annotated sequence likewise did not reveal any obvious candidates to fulfill these functions. Some of the genes are hypothesized to be 'missing' from the *M. avium avium* genome that may contribute to the attenuated phenotype observed with *M. avium avium* isolates. Conversely, these data suggest the possibility that the presence of these genes within the MparaTb genome may lead to an increased viability and virulence of MparaTb within the host. The fact that a wide heterogeneity of these regions occurred within the array of natural isolates used in this study suggests a basis for the variation in growth rates, pathogenesis and any number of variant phenotypes.

EVALUATION OF A PCR BASED DIAGNOSTIC ASSAY AND A NEW MOLECULAR EPIDEMIOLOGY TOOL BASED UPON PARATUBERCULOSIS POLYMORPHIC REGIONS

INTRODUCTION

Mycobacterium avium subspecies *paratuberculosis* (MparaTb), a Gram positive, acid-fast bacillus, can generally be differentiated phenotypically from *M. avium* subspecies avium by its dependence on mycobactin J for iron acquisition in growth, its extremely long doubling time and genetically by the presence of the insertion sequence IS900. Although most MparaTb isolates require mycobactin J for growth, there have been reports of MparaTb isolates that grow in the absence of mycobactin J (1). Also, there have been reports of MparaTb isolates that lack the IS900 insertion sequence (16).

MparaTb is the etiologic agent of the severe chronic granulomatous enteritis of ruminants, known as Johne's disease or paratuberculosis (4, 20, 21, 44). Although Johne's Disease primarily affects cattle, MparaTb infects other ruminants including bison and is suspected in human infections associated with Crohn's disease (4, 20, 21). However, inconsistent detection of MparaTb in specimens obtained from Crohn's patients has cast doubt on any direct link (20).

Although there is no current "gold standard" diagnostic test for the presence of MparaTb, the most widely used tests are direct bacterial culture

from fecal samples and the detection of MparaTb antigens through the use of an enzyme-linked immunosorbent assay (ELISA). While these tests have the potential to become the gold standard, both of these tests have serious drawbacks that hinder their performance and reliability.

The direct bacterial culture technique is severely limited by the organism's slow rate of growth with a doubling time reported to be 22-26 hours (5) *in vitro* and as high as 48 hours *in vivo* (19). Confirmation of a negative sample requires a 16-30 week incubation while pinpoint colonies can be visible at 12 weeks in some cases (34). This lengthy doubling time can also promote the spread of the disease throughout domestic herds while infected cattle await diagnosis. In severe cases, the infected animal may be culled before the test results have been received. Studies have also shown variability in the sensitivity of the fecal culture test from differing host species. In cattle, the sensitivity ranges from 20-73% while in sheep and goats it is much lower at 6-7% (34, 49, 58). This culture technique is also hindered by the bacteria's varying growth requirements. Although it is widely known that MparaTb requires the use of mycobactin J for growth *in vitro*, several MparaTb field isolates have been found to be mycobactin J independent yielding a false negative result (1). The limited number of bacteria shed in the feces is also a problematic with the direct fecal culture technique. Research has shown that the numbers of MparaTB shed through the feces can vary from host to host (with some infected individuals

failing to shed any bacteria), preventing a proper diagnosis (11). A low-shedder or non-shedder that gives a false negative for a fecal test may be returned to the herd where shedding will increase and transmission to other animals occurs. ELISA assay for the detection of MparaTb, while a more rapid test than the direct fecal culture, is hindered by the lack of sensitivity provided by the antigen used in the test. The current antigen used for the MparaTb ELISA fails to detect all MparaTb isolates, resulting in numerous false negatives. Reports have given varying sensitivities for the MparaTb ELISA test ranging from 8-63% depending upon the host species being tested. Specificity for the MparaTb ELISA ranges between 93-98% for all tested species (34, 49, 56). The reports of variability in the sensitivities of the MparaTb ELISA from host to host further support the hypothesis of host-specificity of isolates and further underscores the critical need for a more reliable diagnostic test. Amplified fragment length polymorphism was applied to discover regions unique to the MparaTb genome which could be exploited as the basis of PCR based diagnostic assays against an array of non-paratuberculosis mycobacteria.

MATERIALS AND METHODS

Bacterial isolates. MparaTb isolates (Table 6) were received from University of Wisconsin Veterinary Diagnostic Lab (UWDL), National Veterinary Services Laboratory (NVSL), American Type Culture Collection (ATCC) and the

Texas A&M Veterinary Medical Diagnostic Laboratory (TVMDL). See Table 6 for isolate name and corresponding source.

Non-paratuberculosis isolates. Non-paratuberculosis isolates were received from University of Nebraska (UN), University of Wisconsin Diagnostic Lab (UWDL) and American Type Culture Collection (ATCC). See Table 6 for corresponding isolate names and source.

Identification of mycobacterial isolates to species level. MparaTb isolates were characterized by growth requirements, acid-fast staining and IS900 PCR as previously described (51). All MparaTb isolates were mycobactin J dependent except for isolate #5 from TVMDL, which demonstrated bacterial growth with and without mycobactin J present. Non-MparaTb mycobacterial isolates were characterized to the species level through HPLC of mycolic acid and cross-referenced to the CDC database for similarities to known Mycobacteria (Dr. Michael Collins, University of Wisconsin, Veterinary Medicine). All non-MparaTb mycobacteria were negative for amplicons under IS900 PCR conditions.

Preparation of genomic DNA. Genomic DNA was prepared as previously described (32), through the use of a bead beating method. For the processing of fecal samples, immunomagnetic separation of the organism was used in conjunction with this method.

PCR analysis with diagnostic primers. Isolation, sequencing and primer design of MparaTb unique AFLP region was carried out as previously described (51). PCR amplification of all 55 *M. avium* subsp. *paratuberculosis* isolates and 65 non-MparaTb *Mycobacterium* spp. isolates was performed with primers internally derived from the corresponding AFLP region. Except for the NVSL test set which was performed with purified DNA, all PCR assays were performed with template genomic DNA immobilized on Whatman FTA membranes (Whatman Inc., Clifton, N.J.). Genomic DNA for PCR was prepared on FTA cards as previously described (51) and used directly as PCR template DNA. PCR was optimized with Epicenter's Fail Safe PCR kit (Epicenter Technologies, Madison, Wis.) in accordance with the manufacturer's protocols with all five internal primer sets on *M. avium* subsp. *paratuberculosis* ATCC 19698. Premix G provided optimal product formation for all five primer sets. PCR reaction conditions were those previously reported (51). For all MparaTb isolates IS900 PCR was used as a positive control for confirmation of the presence of DNA template suitable for PCR.

For the non-MparaTb mycobacterial isolates, PCR of the mycobacterial heat shock protein Hsp65 was used as a positive control (12, 60, 61). Genomic template DNA was prepared on Whatman FTA cards as previously described and PCR conditions were optimized with Epicenter's Fail Safe PCR Kit to use premix I. The following PCR reagents were added to a treated and washed 2

Table 5. PCR primers used in diagnostic and primer walk study.
All primers are shown in 5' to 3' orientation; PCR product sizes are indicated for each primer pair.

	Name	5'----Sequence----3'	Product Size bp
Diagnostic Primers	F1-300	CGCACCAAAAGGCACACTAATC	158
	B1-300	CGTCGTCACCATTCAGGAACTC	
	F1-440	GGCTCGCCGAACTACTTGT	226
	B2-440	CTCGAAACAACGGTGACAGA	
	F1-255	GCGGAAAGTCACACTGCTGATGC	112
	B1-255	CGGTCTATTCGATCCCCACCTTG	
	F1-600	CGAGTTGTCCTGGGGTTTTGG	291
	B1-600	CGAAAGTCACCGCATCCACG	
	F1-360	CGAGATTCGCAAGCACTTCG	206
	B1-360	CAAGGCTGGCTGTCAGATGC	
5KB Primer Walk	PW6	CGATCGGCTTTACACCAC	1101
	PW1106	AGAAGCCGGCAACACTCGCC	
	PW952	GCTTAGGAGGAGGTGGCTGATAATC	1534
	PW2485	CCCCAACCCTGTTCGGTATG	
	PW2465	GCATACCGAACAGGGTTGGG	1277
	PW3741	GTGGCACTGGTGGACATTTTG	
	PW3706	GTCCATCGCACACCTCAAAATG	1655
	PW5360	CGCTCAGCATCATCGTGAAGTAG	
HSP 65	Tb11	ACCAACGATGGTGTGTCCAT	439
	TB12	CTTGTCGAACCGCATACCCT	
15KB Primer Walk	PWF1	CATACACAGACAAAGTGCCCCC	1568
	PWR1	GTAGTAAAGGTTGGCGGCGAC	
	PWF2	TCGACGCCCATCAAGATTTG	2233
	PWR2	GATAGTCGTGGACGGTCAGCAG	
	PWF3	GGTCGTCGTCAGATTACTTGGTCTC	1993
	PWR3	CAGTGATTCGGGCTGACTACGG	
	PWF4	CATATTAGTGCGTGAGCTGACAGG	2236
	PWR4	CATCCGCTTACGCAGTTCGA	
	PWF5	CCGTCCACTGCTGTGAAACG	2448
	PWR5	CCTCTACCTGGGGCTTTGAATAC	
	PWF6	TTGCAGTCGGTCACCCATTC	1782
	PWR6	CGTTACTGATGCTGGCTCCG	
	PWF7	CTACAGATGCGTTACTGATGCTGG	774
	PWR7	CATTTGTGAGGACATTCGGTCG	
	PWF8	CTGCCACCGCAAGGTTTATG	626
	PWR8	GCTTGTTGCGACAGTGGCA	
	PWF9	GCAATCGGTAAGGGTCAACAGC	1272
	PWR9	GCTTAGTTCGCCGCTTGAATG	
	PWF10	CGGCATACGAGAATGGAGTGAAG	1536
	PWR10	CGTAATTGCGACCCATTCGTG	
	PWF11	AACGGCAGTCCTTGGAGACG	2009
	PWR11	CAAAAATGGTGAGTGCCCCC	
	PWF12	CGTGTTTCCATCTGCGAACG	811
	PWR12	GTGAAGGACGAGACGCCGTA	

mm FTA biopsy punch for the use of genomic template DNA: 7.6 μl double-distilled H_2O, 20 pmol each primer, 10 μl premix I and 1 U of Taq DNA polymerase for a total reaction volume of 20 μl. Thermocycler conditions were performed as follows: initial denaturation for 5 minutes at 94°C followed by 45 cycles of 94°C for 30 seconds, 59°C for 30 seconds and 72°C for 1 minute. A final elongation was conducted at 72°C for 10 minutes. The corresponding 439-bp amplicons were visualized on 2% w/v LE agarose gel with ethidium bromide staining. Isolates showing no amplification under these conditions were excluded from further analysis.

Characterization of a 5kb and 15kb deletion in *M. avium avium* corresponding to MparaTb region 2 and region 5. Primer pairs PW6-PW1106, PW952-PW2485, PW2465-PW3741 and PW3706-PW5360 (Table 5) amplify PCR products spanning the 5,145bp deletion of 1101bp, 1534bp, 1277bp and 1655bp respectively. These primer combinations were designed so that their PCR products overlap to avoid missing any smaller deletions (Figure 6). PCR amplification was performed using DNA from MparaTb (ATCC 19698) and *M. avium avium* (ATCC 35716) isolates. PCR products contiguously covering the region (primer walk) were amplified at 94°C for 30 seconds, 56°C for 30 seconds, 72°C for 1 minute 20 seconds through 40 cycles and a final elongation of 5 minutes at 72°C. PCR reactions utilized 4 ng genomic template

DNA, 30 ng each primer, in a final volume of 20 μl per manufacturer's instructions (Promega). Samples were analyzed on 1.5% (w/v) Metaphor ME agarose.

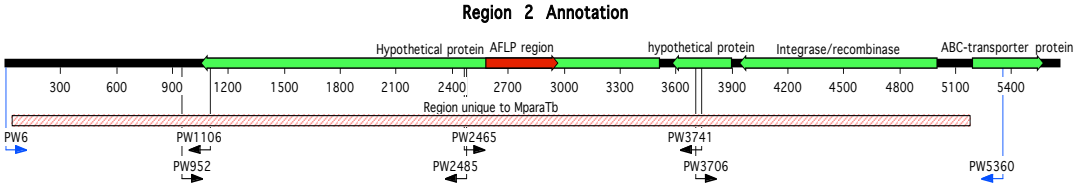


Figure 6. Schematic of 5kb region present only in the MparaTb genome.
Schematic of 5kb region showing directional overlapping primer pairs used in a primer walk across the 5kb region. Primers PW6 and PW5360 have 100% homology to MparaTb and *M. avium avium* genomes. Solid red arrow corresponds to region identified through AFLP analysis. Solid green arrows correspond to ORF's present within the 5kb deletion region. These regions include hypothetical proteins and an open reading frame with homology to an integrase/recombinase gene.

PCR primer pairs (Table 5) for amplification of a PCR primer walk across a 15kb region were designed to amplify overlapping regions across the 15kb region (Figure 7). PCR was conducted on 4ng genomic DNA templates with 100 ng each primer, 10 μl Epicenter Fail Safe premix G, 5.2 μl double-distilled H_2O and 1 U Taq DNA polymerase in a final volume of 20 μl. Thermocycler conditions were performed with all primer sets as follows: initial denaturation of 94°C for 5 minutes followed by 40 cycles of 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 2.5 minutes. A final elongation step was conducted for 5 minutes at 72°C. Corresponding amplicons were visualized on 1% LE agarose with ethidium bromide staining.

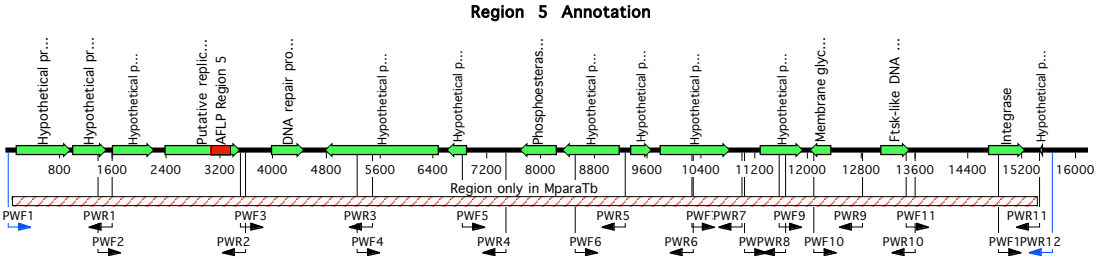


Figure 7. Schematic of 15kb region present only in the MparaTb genome. Schematic of 15kb region showing directional overlapping primer pairs used in primer walk across 15kb region. Primers PWF1 and PWR12 have 100% homology to MparaTb and *M. avium avium* genomes. Solid red arrow corresponds to region identified through AFLP analysis. Solid green arrows correspond to ORF's present within the region. These regions correspond to 10 hypothetical proteins, in addition to putative replication protein, DNA repair protein, phosphoesterase from Mycobacterial phage L5, membrane glycoprotein, Ftsk-like DNA segregation ATPase, and an integrase protein.

PCR of truncated 202bp product of 15kb region in *M. avium avium* was performed to confirm the absence of the 15kb region from the *M. avium avium* genome. PCR was conducted on 4ng genomic DNA templates with 200 ng each primer, 10 μl Epicenter Fail Safe premix G, 1.6 μl double-distilled H_2O and 1 U Taq DNA polymerase in a final volume of 20 μl. Thermocycler conditions were carried out as follows: initial denaturation of 94°C for 5 minutes followed by 45 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute. A final elongation step was conducted for 5 minutes at 72°C. Corresponding amplicons were visualized on 2% LE agarose with ethidium bromide staining.

Multiplex PCR. Multiplex PCR analysis was carried out on UWDL isolates 132 and 136. Genomic template DNA was prepared on Whatman FTA cards as previously described and the following PCR reagents were added to a

treated and washed 2 mm FTA biopsy punch for the use of genomic template DNA: 7.6 ul double-distilled H_2O, 1.6 ul each primer at a concentration of 100 ng/ul, 2 ul magnesium free buffer (Promega), 2 ul 25mM MgCl2 (Promega), 1.0 ul dNTPs at 10mM concentration (Promega), and 1 *U* of Taq DNA polymerase for a total reaction volume of 20 *ul*. Thermocycler conditions were performed as follows: initial denaturation for 5 minutes at 94°C followed by 45 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute. A final elongation was conducted at 72°C for 10 minutes. The corresponding amplicons were visualized on 4% w/v LE agarose gel with ethidium bromide staining.

RESULTS

Characterization of 5kb and 15kb deletions of *M. avium avium* genome. Entire MparaTb contigs including portions homologous to the MparaTb-polymorphic sequenced regions (regions 1-5) were examined using BLAST search to determine sequence homology with the *M. avium avium* genome. The MparaTb-specific regions 2 and 5 originate from a 31.6kb contig and a 36kb contig respectively which were used as a query in a BLAST search against *M. avium avium*. The results of these BLAST searches revealed numerous *M. avium avium* contigs with 99% homology to the query sequences. When these *M. avium avium* contigs were aligned using overlapping sequences of the contigs, a 5,145bp region and a 15,372bp region of MparaTb were

revealed; these regions have no homologue in the *M. avium avium genome*. In order to determine the extent of these deletions, primer walks were performed over region 2 and region 5 and their flanking sequences. All primer pairs for the 5kb deletion produced the expected products from the MparaTb isolate (Figure 8A, lanes 2-5). However, no products were identified from *M. avium avium* spanning an approximately 5,200bp region (Figure 8A, lanes 2-5). Amplification of the region conserved in both the *M. avium avium* and MparaTb genomic databases using primers F1-PC-190 and B1-PC-190 produced the expected 190bp product validating the presence of DNA from the *M. avium avium* genome (Figure 8A, lane 6) in the PCR analysis. These results confirmed that region 2 was part of a much larger genomic deletion in *M. avium avium*. This region contained several unidentified ORFs; one near the 3' end of the 5.1kb region had sequence homology to bacterial genes responsible for integrase/recombinase function (E value = 4e-17). The identified integrase/recombinase gene was contained in contig 210 of the MparaTb genome and was distinct from that referred to by Bannantine *et al* (4).

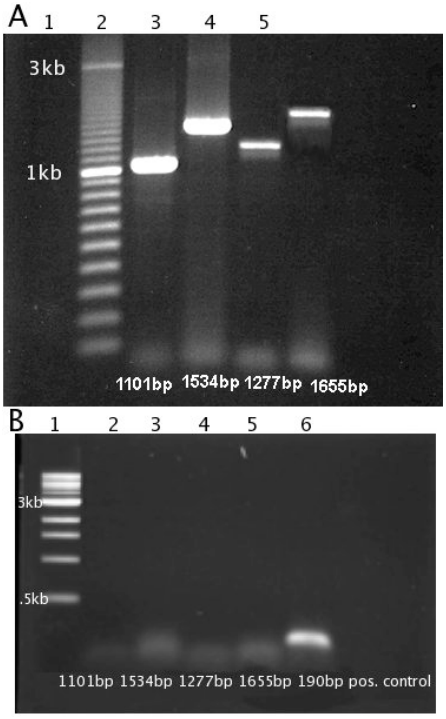


Figure 8. Primer-walk across 5,145bp region in MparaTb isolate ATCC 19698.

A. 1% agarose gel electrophoresis of PCR products. Lane 1: 100bp molecular weight marker. Lane 2: 1,101bp product using primers PW6 and PW1106. Lane 3: 1,534bp product using primers PW952 and PW2485. Lane 4: 1,277bp product using primers PW2465 and PW3741. Lane 4: 1,655bp product using primers PW3706 and PW5360.

B. 1% agarose gel electrophoresis of PCR products. Primer walk on *M. avium avium* isolate ATCC 35716. Lane 1: 1kb molecular weight marker. Lane 2: PCR with primers PW6 and PW1106. Lane 3: PCR with primers PW952 and PW2485. Lane 4: PCR with primers PW2465 and PW3741. Lane 5: PCR with primers PW3706 and PW5360. Lane 6: PCR with primers F1-PC-190 and B1-PC-190 showing 190bp product as a positive control.

Twelve overlapping primer sets revealed their corresponding amplicons in the MparaTb strain but yielded no PCR product in the *M. avium avium* strain (Figure 9), confirming the presence of this region in MparaTb and its absence in the *M. avium avium*. Furthermore, the forward primer in the first region of the primer walk (PWF1) and the reverse primer in the last region of the primer walk

(PWR12) (Figure 7) annealed to the flanking regions of the deletion. These two primers were homologous in MparaTb and *M. avium avium* genomes. PCR performed with these primers revealed a corresponding 202bp amplicon in *M. avium avium* but yielded no product in MparaTb (Figure 10). PCR primer walk results for the 15kb deletion region from the *M. avium avium* genome confirmed that this large region was absent from the avium subspecies genome (Figure 10). These data further confirmed the absence of the 15kb region from the *M. avium avium* genome.

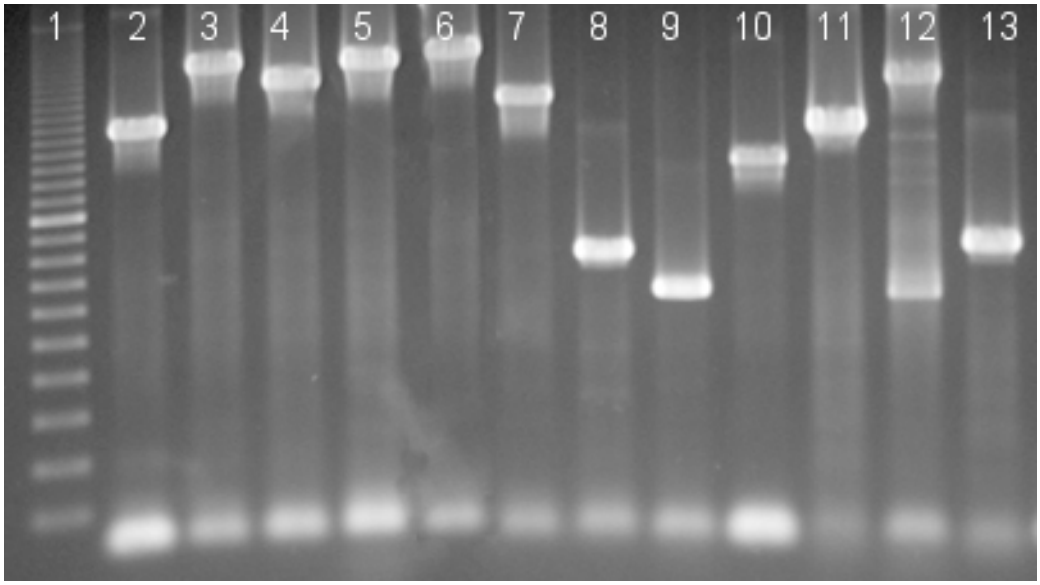


Figure 9. Primer walk across 15kb region in MparaTb genome.
1% agarose gel electrophoresis of MparaTb ATCC 19698 PCR products. Lane 1: 100bp molecular weight marker; Lane 2: primer set PWF1-PWR1; Lane 3: primer set PWF2-PWR2; Lane 4: primer set PWF3-PWR3; Lane 5: primer set PWF4-PWR4; Lane 6: primer set PWF5-PWR5; Lane 7: primer set PWF6-PWR6; Lane 8: primer set PWF7-PWR7; Lane 9: primer set PWF8-PWR8; Lane 10: primer set PWF9-PWR9; Lane 11: primer set PWF10-PWR10; Lane 12: primer set PWF11-PWR11; Lane 13: primer set PWF12-PWR12. *M. avium avium* isolate 35716 was absent for all primer sets (data not shown).

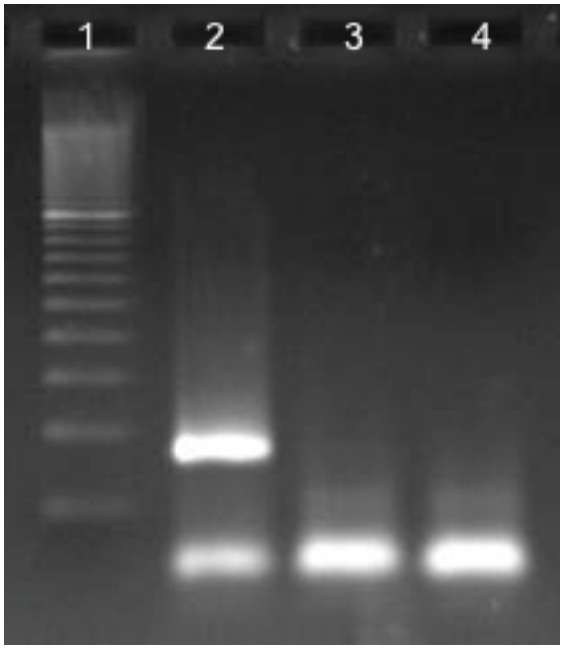


Figure 10. PCR analysis of *M. avium avium* genomic region corresponding to the 15kb region of MparaTb.
2% agarose gel electrophoresis of *M .avium avium* truncated region. PCR products of *M. avium avium* isolate 35716, MparaTb isolate 19698, and negative control (Lanes 2-4 respectively). PCR reveals truncated product (202bp) from the *M. avium avium* isolate and no product from the MparaTb isolate. Lane 1 is 100bp molecular weight marker.

PCR analysis with diagnostic primers. In an effort to determine the overall polymorphic characteristics of the five diagnostic MparaTb unique regions described earlier, and in order to study these diagnostic primers sets for their usefulness as a MparaTb specific diagnostic test, PCR analysis was carried out. All five diagnostic primer sets were applied to an array of MparaTb clinical field isolates from varying hosts and on an array of non-*paratuberculosis* isolates. These PCR primer sets were used to analyze 55 MparaTb isolates [19 clinical MparaTb isolates from the University of Wisconsin Diagnostic Lab

(UWDL), 5 ATCC MparaTb human isolates, 1 ATCC MparaTb bovine isolate, 17 NVSL MparaTb bovine isolates, 6 clinical Texas bovine MparaTb isolates, and 7 Texas clinical bovine isolates from the Texas Veterinary Medical Diagnostic Lab (TVMDL) and 65 non-*paratuberculosis* isolates [1 ATCC *M. avium avium* isolate 35716, 1 *M. avium avium* human isolate Mac104, 63 non-*paratuberculosis* isolates of varying species from UWDL (Table 6)]. All MparaTb isolates were subspecies differentiated by growth characteristics, IS900 PCR, and dependence of mycobactin J. All MparaTb isolates used were positive for IS900 PCR except for isolate #6 from TVMDL. All MparaTb isolates were dependent upon mycobactin J for *in vitro* growth except isolate #5 from TVMDL. All non-*paratuberculosis* isolates used were PCR positive for the corresponding 439bp amplicon of the Hsp65 PCR product with primers Tb11 and Tb12 as described earlier (12, 13, 60, 61). region 5 PCR was not performed on NVSL isolates due to loss of culture.

Table 6. Diagnostic PCR results.
Table showing diagnostic PCR results of regions 1-5 on 55 Mparatb isolates and 65 non-*paratuberculosis* Mycobacterial isolates. Isolate name, subspecies designation, isolate source, mycobactin J dependence and PCR results for Hsp65, IS900 and diagnostic regions 1-5 are indicated. "+" indicates PCR was positive for primer set or dependence on mycobactin J. "-" indicates PCR was negative for respective primer set. "N/A" indicates test was not carried out. Noteworthy results are highlighted in yellow

Isolate	Species (ssp)	Hsp65	Mycobactin J	region 1	region 2	region 3	region 4	region 5	IS900
UWDL 19 Clinicals									
109	MparaTb	N/A	+	+	+	+	+	+	+
126	MparaTb	N/A	+	+	+	+	+	+	+
129	MparaTb	N/A	+	+	+	+	+	+	+
130	MparaTb	N/A	+	+	+	+	+	+	+
132	MparaTb	N/A	+	+	+	+	-	+	+
135	MparaTb	N/A	+	+	+	+	+	+	+
136	MparaTb	N/A	+	+	+	+	+	+	+
139	MparaTb	N/A	+	+	+	+	+	+	+
140	MparaTb	N/A	+	+	+	+	+	+	+
141	MparaTb	N/A	+	+	+	+	+	+	+
142	MparaTb	N/A	+	+	+	+	+	+	+
144	MparaTb	N/A	+	+	+	+	+	+	+
145	MparaTb	N/A	+	+	+	+	+	+	+
148	MparaTb	N/A	+	+	+	+	+	+	+
150	MparaTb	N/A	+	+	+	+	+	+	+
151	MparaTb	N/A	+	+	+	+	+	+	+
155	MparaTb	N/A	+	+	+	+	+	+	+
156	MparaTb	N/A	+	+	+	+	+	+	+
160	MparaTb	N/A	+	+	+	+	+	+	+
ATCC MparaTb									
ATCC 49164	MparaTb	N/A	+	+	+	+	+	+	+
ATCC 19851	MparaTb	N/A	+	+	+	+	+	+	+
ATCC 43544	MparaTb	N/A	+	+	+	+	+	+	+
ATCC 43545	MparaTb	N/A	+	+	+	+	+	+	+
ATCC 43015	MparaTb	N/A	+	+	+	+	+	+	+
ATCC 19698	MparaTb	N/A	+	+	+	+	+	+	+

Table 6 continued.

Isolate	Species (ssp)	Hsp65	Myco. J	region 1	region 2	region 3	region 4	region 5	IS900
NVSL Isolates									
MparaTb	MparaTb	N/A	N/A	+	+	+	+	N/A	+
NEG	MparaTb	N/A	N/A	-	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	-	+	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	+	N/A	+
NEG	MparaTb	N/A	N/A	-	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	+	N/A	+
MparaTb	MparaTb	N/A	N/A	+	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	+	N/A	+
NEG	MparaTb	N/A	N/A	-	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	-	N/A	+
MparaTb	MparaTb	N/A	N/A	-	-	-	-	N/A	+
NEG	MparaTb	N/A	N/A	-	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	+	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	-	-	-	N/A	+
NEG	MparaTb	N/A	N/A	-	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	-	+	-	N/A	+
NEG	MparaTb	N/A	N/A	-	-	-	-	N/A	+
MparaTb	MparaTb	N/A	N/A	-	-	+	-	N/A	+
MparaTb	MparaTb	N/A	N/A	+	+	+	+	N/A	+
Texas 6 Clinicals									
T12	MparaTb	N/A	+	+	+	+	+	+	+
T136	MparaTb	N/A	+	+	+	+	+	+	+
T139	MparaTb	N/A	+	+	+	+	+	+	+
T140	MparaTb	N/A	+	+	+	+	+	+	+
T141	MparaTb	N/A	+	+	+	-	+	+	+
T143	MparaTb	N/A	+	+	+	+	+	+	+
TVMDL Isolates									
#5	MparaTb	N/A	-	+	+	+	+	+	+
#6	MparaTb	N/A	+	-	-	-	-	+	-

Table 6 continued.

Isolate	Species (ssp)	Hsp65	Myco. J	region 1	region 2	region 3	region 4	region 5	IS900
301	MparaTb	N/A	+	+	+	+	+	+	+
302	MparaTb	N/A	+	+	+	+	+	+	+
304	MparaTb	N/A	+	+	+	+	+	+	+
305	MparaTb	N/A	+	+	+	+	+	+	+
306	MparaTb	N/A	+	+	+	+	+	+	+
Non-MparaTb Isolates									
ATCC 35716	avium	+	-	-	-	-	+	-	-
Mac 104	avium	+	-	-	-	-	-	-	-
2	MAC	+	-	-	-	-	-	-	-
8	scrofulaceum	+	-	-	-	-	-	-	-
10	scrofulaceum	+	-	-	-	-	-	-	-
11	scrofulaceum	+	-	-	-	-	-	-	-
22	avium	+	-	-	-	-	-	-	-
27	scrofulaceum	+	-	-	-	-	-	-	-
28	scrofulaceum	+	-	-	-	-	-	-	-
32	MAC	+	-	-	-	-	-	-	-
35	scrofulaceum	+	-	-	-	-	-	-	-
38	scrofulaceum	+	-	-	-	-	-	-	-
39	mucogenicum	+	-	-	-	-	-	-	-
40	M. species	+	-	-	-	-	-	-	-
42	nonchromogenicum	+	-	-	-	-	-	-	-
44	avium	+	-	-	-	-	-	-	-
46	intracellulare	+	-	-	-	-	-	-	-
47	scrofulaceum	+	-	-	-	-	-	-	-
48	asiaticum	+	-	-	-	-	-	-	-
49	asiaticum	+	-	-	-	-	-	-	-
50	MAC atypical	+	-	-	-	-	-	-	-
51	MAC	+	-	-	-	-	-	-	-
53	scrofulaceum	+	-	-	-	-	-	-	-
54	MAC	+	-	-	-	-	-	-	-
55	scrofulaceum	+	-	-	-	-	-	-	-
56	MAC	+	-	-	-	-	-	-	-
57	scrofulaceum	+	-	-	-	-	-	-	-
58	asiaticum	+	-	-	-	-	-	-	-

Table 6 continued.

Isolate	Species (ssp)	Hsp65	Myco. J	region 1	region 2	region 3	region 4	region 5	IS900
59	scrofulaceum	+	-	-	-	-	-	-	-
60	asiaticum	+	-	-	-	-	-	-	-
61	asiaticum	+	-	-	-	-	-	-	-
62	asiaticum	+	-	-	-	-	-	-	-
66	MAC	+	-	-	-	-	-	-	-
67	scrofulaceum	+	-	-	-	-	-	-	-
68	scrofulaceum	+	-	-	-	-	-	-	-
69	scrofulaceum	+	-	-	-	-	-	-	-
70	kansasii	+	-	-	-	-	-	-	-
71	MAC	+	-	-	-	-	-	-	-
72	MAC	+	-	-	-	-	-	-	-
73	scrofulaceum	+	-	-	-	-	-	-	-
74	scrofulaceum	+	-	-	-	-	-	-	-
75	scrofulaceum	+	-	-	-	-	-	-	-
76	celatum	+	-	-	-	-	-	-	-
77	kansasii	+	-	-	-	-	-	-	-
78	celatum	+	-	-	-	-	-	-	-
79	celatum	+	-	-	-	-	-	-	-
80	scrofulaceum	+	-	-	-	-	-	-	-
81	celatum	+	-	-	-	-	-	-	-
82	MAC	+	-	-	-	-	-	-	-
83	scrofulaceum	+	-	-	-	-	-	-	-
84	MAC	+	-	-	-	-	-	-	-
85	scrofulaceum	+	-	-	-	-	-	-	-
86	scrofulaceum	+	-	-	-	-	-	-	-
87	gordonae	+	-	-	-	-	-	-	-
89	celatum	+	-	-	-	-	-	-	-
92	avium	+	-	-	-	-	-	-	-
93	avium	+	-	-	-	-	-	-	-
94	MAC	+	-	-	-	-	-	-	-
95	kansasii	+	-	-	-	-	-	-	-
96	celatum	+	-	-	-	-	-	-	-
97	intracellulare	+	-	-	-	-	-	-	-
98	terrae complex	+	-	-	-	-	-	-	-
99	avium	+	-	-	-	-	-	-	-
100	avium	+	-	-	-	-	-	-	-

.

PCR results of region 1 (Table 7A) revealed appropriately sized amplicons in 19 of 19 UWDL isolates, 6 of 6 ATCC MparaTb isolates, 6 of 6 Texas clinical isolates, 15 of 17 NVSL isolates, and 6 of 7 TVMDL isolates. These data disclosed that while region 1 is highly conserved, it is not specific for all MparaTb isolates. region 1 produced no amplification with any non-MparaTb mycobacteria.

PCR results of region 2 (Table 7B) revealed appropriately sized amplicons in 19 of 19 UWDL isolates, 6 of 6 ATCC MparaTb isolates, 6 of 6 Texas clinical isolates, 11 of 17 NVSL isolates, and 6 of 7 TVMDL isolates. These data revealed region 2 to be less conserved than region 1, but was still present in a high percentage of MparaTb isolates (48 of 55). region 2 produced no amplification with any non-MparaTb mycobacteria.

PCR results of region 3 (Table 7C) revealed appropriately sized amplicons in 19 of 19 UWDL isolates, 6 of 6 ATCC MparaTb isolates, 5 of 6 Texas clinical isolates, 12 of 17 NVSL isolates, and 6 of 7 TVMDL isolates. These data again revealed region 3 to be less conserved than region 1, but was still present in a high percentage of MparaTb isolates (48 of 55). region 3 produced no amplification with any non-MparaTb mycobacteria.

PCR results of region 4 (Table 7D) revealed appropriately sized amplicons in 18 of 19 UWDL isolates, 6 of 6 ATCC MparaTb isolates, 6 of 6

Texas clinical isolates, 7 of 17 NVSL isolates, and 6 of 7 TVMDL isolates. Similar to regions 1, 2 and 3, region 4 was less conserved, but was present in a high percentage of MparaTb isolates (43 of 55). region 4 produced amplification with one non-MparaTb mycobacterial isolate (*M. avium avium* isolate ATCC 35716).

PCR results of region 5 (Table 7E) revealed appropriately sized amplicons in all 19 UWDL isolates, all 6 ATCC MparaTb isolates, all 6 Texas clinical isolates and 7 of 7 TVMDL isolates, but region 5 primer set was not evaluated in the NVSL isolates due to loss of culture. These data revealed that region 5 was the most highly conserved region, being present in every MparaTb isolate (37 of 37). Region 5 produced no amplification with any non-MparaTb mycobacteria.

Table 7. Diagnostic PCR for MparaTb regions 1-5 with sensitivities and specificities for each.
Positive and negative test results, false positive and false negative test results, sensitivities and specificities are indicated for each primer set. PCR analysis carried out for region 1 (7A), region 2 (7B0, region 3 (7C), region 4 (7D), and region 5 (7E).

A.

Evaluation of MparaTb Diagnostic Primer Sets			
	True		
region 1 Primer Set	+	-	total
+	52	0	52
-	3	65	68
total	55	65	120
Sensitivity	94.55%		
Specificity	100.00%		

B.

Evaluation of MparaTb Diagnostic Primer Sets			
	True		
region 2 Primer Set	+	-	total
+	48	0	48
-	7	65	72
total	55	65	120
Sensitivity	87.27%		
Specificity	100.00%		

Table 7 continued.

C.

Evaluation of MparaTb Diagnostic Primer Sets			
	True		
region 3 Primer Set	+	-	total
+	48	0	48
-	7	65	72
total	55	65	120
Sensitivity	87.27%		
Specificity	100.00%		

D.

Evaluation of MparaTb Diagnostic Primer Sets			
	True		
region 4 Primer Set	+	-	total
+	43	1	44
-	12	64	76
total	55	65	120
Sensitivity	78.18%		
Specificity	98.46%		

E.

Evaluation of MparaTb Diagnostic Primer Sets			
	Truth		
region 5 Primer Set	+	-	total
+	37	0	37
-	0	59	59
total	37	59	96
Sensitivity	100.00%		
Specificity	100.00%		

These data were analyzed in 2X2 tables for their specificities and sensitivities for each primer set for their possible application as a PCR diagnostic test for detecting and diagnosing MparaTb. The specificity of a diagnostic test can be determined by the following calculation: Sp = 1-(false positives/(false positives+test negatives)), and the sensitivity was determined by the following calculation: Se = test positives/(test positives+false negatives).

Results of the analysis indicated that region 5 had a 100% sensitivity and specificity within the sampled population, but all primer sets employed exhibited minimum sensitivity of 78% and specificity of 98%. These data provide evidence that these PCR primer sets could be used for detection and diagnostic purposes. Primer sets that yielded less than 100% specificity and sensitivity could also be used for a molecular epidemiology assay for the purposes of herd tracking.

Multiplex PCR of diagnostic regions. region 1-4 primer sets with the same melting point were tested in a multiplex PCR reaction for detection of all four amplicons. Primer set for region 5 was not used due to its presence in all MparaTb isolates tested. The multiplex PCR data revealed that it was possible to amplify all four diagnostic regions simultaneously with no adverse effects on PCR conditions or amplicons.

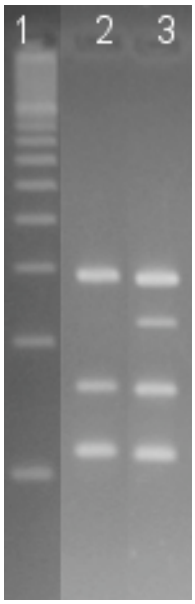


Figure 11. Diagnostic multiplex PCR gel.
Agarose gel showing amplification of regions 1-4 diagnostic primers in a multiplex PCR. Lane 1: 100bp molecular weight marker; Lane 2: UWDL isolate 132; Lane 3: UWDL isolate 136. Refer to table 5 for product sizes.

Multiplex PCR was tested on UWDL isolates 132 and 136 for the presence of regions 1-4. Isolate 132 was used due to the absence of region 4 in previous diagnostic PCR and isolate 136 was used for presence of all four diagnostic regions. Multiplex PCR results produced appropriately sized amplicons for all four primer sets with isolate 136 (Figure 11). Also, isolate 132 produced amplification for regions 1-3 and an absence of PCR product for region 4, confirming with previous results revealing the absence of region 4 from UWDL isolate 132. These data confirm that the primer sets developed in this study can be used for an effective and reliable multiplex PCR reaction for MparaTb field isolates.

If a positive result from one PCR primer set determines a positive result for the presence of MparaTb, and there was only one isolate (NVSL check set) that was negative for all 5 primer sets, and that there was only one false positive (UWDL isolate 132), then the sensitivity of the multiplex diagnostic PCR is 98.18% and the specificity of the test is 98.46%. Since region 4 was the only region to show any false positives, region 4 could be eliminated from the multiplex PCR and the specificity would increase to 100%. For epidemiologic and isolate tracking, inclusion of region 4 within the multiplex PCR would be useful due to its highly polymorphic characteristics.

Table 8. Multiplex PCR results.
Table of results of multiplex PCR with regions 1-5 on all Mycobacterial isolates used in study. Table shows 55 true positives and 65 true negatives of which 54 were positive and 64 were negative respectively. Sensitivity and specificity of multiplex PCR test is 98.18% and 98.46% respectively.

Evaluation of MparaTb Diagnostic Primer Sets			
	TRUE		
Multiplex PCR (regions 1-5)	+	-	total
+	54	1	55
-	1	64	65
total	55	65	120
Sensitivity	98.18%		
Specificity	98.46%		

DISCUSSION

In this study, five polymorphic genomic regions of the MparaTb genome were analyzed by PCR for detection, diagnosis and molecular epidemiology with 120 mycobacterial isolates from differing host species. These results confirmed earlier results that the MparaTb genome is much more diverse than previously believed. These results are in direct contradiction of previous findings that the genome of MparaTb is monogenic and invariant from isolate to isolate (5). Although, there have been scientific reports claiming that MparaTb isolates exhibit genetic diversities from isolate to isolate (3, 47), these reports are based upon short sequence repeats (SSR) which samples a small portion of the genome. Results reported here were based upon AFLP analysis of 55 isolates that samples a much larger portion of the genome and confirmed the polymorphic nature of the MparaTb genome.

Similar to previous reports (47), the present research confirmed genetic heterogeneity among MparaTb isolates. Unlike prior studies, the results reported here found that these genetic differences were not based upon single nucleotide polymorphisms, but were instead based upon deletions of large large regions of the genome. In particular, the polymorphic regions identified through AFLP analysis were part of larger 5 kb and 15kb deletions from the *M. avium avium* genome. This was further supported by comparison of genomic sequencing of the MparaTb and *M. avium avium* genomes and confirmed by

primer walk PCR analysis. Within this 5kb region, there were genes responsible for integrase/recombinase functions found to be present in the MparaTb genome and suggesting a potential means for the evolutionary separation of the MparaTb and *M. avium avium*. The larger 15kb region present in the MparaTb genome was absent from the *M. avium avium* genome. Again these data were discovered through sequence comparison and confirmed by PCR primer walk across the 15kb region. The 15kb region also contained genes required for integrase functions at the 3' end of the region (Figure 7) further supporting the hypothesis that these MparaTb specific regions may be linked to the evolutionary separation of the MparaTb and *M. avium avium*.

To evaluate amplification of five AFLP derived MparaTb specific regions, the primer sets were tested in mycobacterial isolates consisting of 55 MparaTb and 65 non-MparaTb mycobacterial isolates. The primer sets were effective in amplifying their corresponding regions in 94.55% of the MparaTb isolates for region 1, 87.27% of MparaTb isolates for regions 2 and 3, and 78.18% of the isolates for region 4. region five provided the highest level of sensitivity at 100%. All regions tested were negative for all non-MparaTb isolates tested except region 4, which revealed amplification from one non-MparaTb isolate (ATCC 35716). While regions 1-4 may be less than ideal for diagnostic purposes, they may be ideal for molecular epidemiology applications due to polymorphic differences among MparaTb isolates. In contrast region 5 was an

ideal diagnostic candidate due to its presence in 100% of all MparaTb isolates and absence from 100% of all non-MparaTb isolates.

It is also noteworthy that all five regions amplified their corresponding region in isolate #5 from TVMDL that was negative for mycobactin J dependence. In this case our diagnostic test would detect an MparaTb isolate that would not be detected using the current diagnostic culture method. Mycobactin J independent MparaTb isolates are becoming more frequent. ATCC reports that even the highly studied ATCC 19698 MparaTb isolate has lost its dependence upon mycobactin J for *in vitro* growth.

Isolate #6 from TVMDL was shown to be IS900 negative and yet typed as MparaTb by the NVSL laboratory. In this instance, the region 5 primer set detected this isolate where the current IS900 PCR based diagnostic would not have. These data illustrate that the use of a region 5-based PCR would be an excellent choice for an alternative to the IS900 based PCR diagnostic test.

The PCR based tests designed in this study were used to develop a multiplex PCR test. By creating a multiplex PCR based test and by defining an MparaTb positive result as an isolate that yields a positive result in at least one of the five diagnostic regions, then the diagnostic multiplex PCR test yielded a sensitivity of 98.18% and a specificity of 98.46%, and That by excluding region 4 from the multiplex PCR, the specificity could increase to 100%. However, region 4 was the most polymorphic, so for molecular epidemiologic applications,

inclusion of region 4 within the PCR reaction may be beneficial. The multiplexed PCR coupled with the immunomagnetic bead isolation of MparaTb should dramatically decrease the 16-week lag period of current diagnostics to a 72-hour turn around time.

AMPLIFIED FRAGMENT LENGTH POLYMORPHISM REVEALS ISOLATION SPECIFIC EPIGENETIC DISTINCTIONS BETWEEN *Mycobacterium avium* subspecies *paratuberculosis* ISOLATES OF VARYING INVASION TYPES

INTRODUCTION

Mycobacterium avium subspecies *paratuberculosis* (MparaTb) is the etiologic agent of a chronic granulomatous enteritis of ruminants known as Johne's disease (20). MparaTb has also been suspected to be involved in the chronic inflammatory bowel disorder in humans known as Crohn's disease (7, 17, 48). Although mostly prevalent in the bovine host species, MparaTb has been found to be an infectious agent of numerous mammals. MparaTb infections occur through the fecal oral route of contaminated food sources and suspected to occur in human cases through the contamination of milk sources (50).

There have been reports of distinct differences and strain diversity among MparaTb isolates from bovine and human hosts (3, 47). These data suggest that genetic differences exist between MparaTb isolates from different host species, but these conclusions were based upon genetic analysis of very small regions of the genome. Previously reported AFLP data suggest heterogeneity among MparaTb isolates based on a whole genome wide analysis of bovine isolates, thus it is hypothesized that while genetic polymorphisms exist between isolates of the same host species, AFLP analysis will reveal distinct genetic

differences between MparaTb isolates from bovine hosts as compared to isolates of human origin.

These genetic differences may be based upon genetic insertion/deletion events unique to each host isolate or more subtle genetic variations such as single nucleotide polymorphisms (SNP) or even epigenetic variations to the bacterial genome. Previous data has confirmed that AFLP analysis is capable of detecting large polymorphic differences between isolates (51), and that AFLP can also differentiate genetic differences due to SNPs at the AFLP restriction sites. Furthermore, the highly sensitive AFLP technique can distinguish between isolates with epigenetic differences at the AFLP restriction sites if the restriction sites are differentiated due to methylation differences. Multiple studies have found that DNA methylation plays a key role in the up regulation and down regulation of the corresponding gene products and can even influence the virulence of the bacterial organism (2, 6, 9, 23, 25, 29, 39, 43).

In this study, AFLP profiles of MparaTb isolates from human and bovine hosts will be demonstrated to exhibit different and distinct banding patterns. Host specific AFLP regions will be investigated for their genetic differences and the effects of the corresponding gene on virulence.

MATERIALS AND METHODS

Bacterial isolates. Four MparaTb human isolates were obtained from the American Type Culture Collection (ATCC 49164, 43544, 43545, and 43015). Six Texas bovine isolates from clinical cases were obtained from Dr. Michael Collins (University of Wisconsin Diagnostic Laboratory): isolates T12, T136, T139, T140, T141, and T143. Two MparaTb bovine isolates were obtained from ATCC: ATCC 19698 and 19851. All Texas clinical bovine isolates and ATCC 19698 were isolated from fecal material while all human isolates and ATCC 19851 were isolated from tissue. All isolates were identified as MparaTb by culture requirements, colony morphology and IS900 PCR presence as previously described (51).

Amplified fragment length polymorphism. AFLP analysis was carried out as previously described (51) on all isolates with the following exceptions: isolate ATCC 19698 was used in duplicate as an internal control. In this case, ATCC 19698 bacterial broth culture was divided into two separate equal aliquots and the entire AFLP procedure was performed separately on each aliquot.

Cluster analysis. The Bionumerics computer analysis program (Applied Maths, Austin Texas) was used for cluster analysis of AFLP generated data. The Dice method of scoring bands was used to create a composite data set from 4 independent AFLP primer sets to develop a summary dendrogram .

PCR confirmation of host specific regions. AFLP generated regions unique to one host type were analyzed through agarose gel electrophoresis by ethidium bromide staining. regions were excised, purified using the Qiagen Gel Purification Kit, and subsequently cloned into *E. coli* chemically competent cells using the TOPO 2.1 vector per manufactures protocols (Invitrogen). Plasmid DNA of positive transformants were purified with Qiagen's Miniprep Kit per manufactures protocols (Qiagen) and sent for sequence analysis (Lone Star Labs, Houston Texas). DNA sequence analysis of MparaTb regions consisted of queries in BLAST searches for sequence alignments against known genomes and for possible sequence homology to known organisms. Sequences were also aligned to the K10 MparaTb known genomic sequence and external PCR primers were derived (Figure on page 81). PCR primers were designed to amplify regions flanking AFLP derived regions and also AFLP restriction sites for evidence of single nucleotide polymorphism (SNP). PCR conditions used to amplify MparaTb regions with external primers were as follows: 20ul reaction consisting of 100ng each primer, 10ul Fail Safe premix G (Epicenter), 1 unit Taq DNA polymerase (Epicenter), 7.6ul DNA allusions from FTA storage cards (Whatman) per manufactures protocols. Thermocycler protocols were as follows: initial denaturation of 94°C for 5 minutes followed by 45 cycles of 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute. A final elongation

was carried out for 10 minutes at 72ºC. Amplicons were viewed on 2% (w/v) SeaKem LE agarose with ethidium bromide staining.

Data mining of host specific regions. PCR products corresponding to host specific regions were used as queries in BLAST searches of the NCBI database and as alignments to the known MparaTb K10 genomic sequence. regions 17 and 20 were also aligned to each other using MacVector data analysis software to locate evidence consensus sequences used for possible recognition sites for DNA methyltransferases.

Isolation of total RNA from bovine and human MparaTb. Total RNA was isolated from cultures of MparaTb grown in 7H9 Middlebrook media supplemented with mycobactin J and glycerol. MparaTb was precipitated, media removed, and ice-cold Trizol reagent (500 μl) added to the MparaTb culture. Tubes were frozen and thawed several times by immersing the tubes completely in liquid nitrogen and warming to room temperature by immersion in water bath and then bead-beating for 45 seconds at speed 6. After the tubes cooled, 500 μl of Trizol reagent and 200 μl of chloroform were added, and the sample was vortexed and incubated at RT for 15 minutes. The tubes were centrifuged at 12,000 rpm for 20 minutes and the upper layer, containing the RNA, was transferred to a new tube, mixed with an equal volume of isopropanol and stored at -20ºC for overnight. RNA was then pelleted by centrifugation at 12,000 rpm.

The RNA pellet was washed in 80% ethanol and re-suspended in DNAse RNAse free water.

DNAse treatment. Contaminant genomic DNA was removed by RNAse-free DNAse I treatment (Ambion) according to the manufacture's instructions, and samples were stored at -80ºC until used.

Reverse transcription. DNAse treated RNA was reverse transcribed into cDNA using 2 μg of total RNA in a 100 μl reaction volume containing 2.5 μM random hexamer and TaqMan® Reverse Transcription Reagents (Applied Biosystems, Foster City, CA), according to the manufacturer's instructions. The thermal cycle parameters were 10 minutes at 25ºC, 30 min at 48ºC and 5 min at 95ºC. cDNAs were stored at -80°C until use.

Real time PCR. Real time PCR was performed with SYBR Green PCR Master Mix (Applied Biosystems., Foster City, CA). All primers were designed with Primer Express Software (Applied Biosystems, Foster City, CA) and were synthesized by Sigma-Genosys (The Woodlands, TX). The list of primers used in this study include as follows:

region 17 forward 5'- CGACGCACAGCAAACGATTCAC -3' and

reverse 5'- TCCCAGCACTTCCTGACGGC -3' amplifying a 127bp product.

region 17 was also amplified with primers

forward 5'- GCAATAGGCGTTGACGGCG -3' and

reverse 5'- GCGGCAAGCAGATGTCGGAC -3' to amplify a 107bp product. region 20 was amplified with primers forward 5'-CGTTGATACCGAGGGGATTTCG -3' and reverse to amplify an 80bp region. region 20 was also amplified with primers forward 5'-AGGAACAGCACGGTCGGC -3' and reverse 5'- CGAGCGGGAATACCACGAG -3' to amplify a 120bp product. Dual primer sets were designed and tested against each region for reproducibility and verification of results.

AFLP analysis of cell-tissue associated isolates. Male Holstein calves, 3-4 weeks of age and weighing 45–55 kg, were fed milk replacer twice daily and water *ad libitum*. The calves were clinically healthy before the experiment. Calves were fasted for 24 hours prior to the surgery, anesthetized and maintained analgesic for the course of the 12 hours experiment. In brief, for the non-survival ligated loops of ileum and jejunum, anesthesia was induced with Propofol (Abbot Laboratories, Chicago, IL) followed by placement of an endotracheal tube and maintenance with isoflorane (Abbot Laboratories, Chicago, IL) for the duration of the experiment. The abdominal wall was clipped, prepared aseptically with chlorhexidine and isopropanol prior to opening. Sterile drapes were used as a barrier. The abdominal wall was opened and the entire length of the Peyer's patch of the distal jejunum and ileum exteriorized. The lumen of the ileum and distal jejunum proximal to the Peyer's patch were flushed with saline to remove any remaining intestinal digest and manually propelled into

the cecum. Six to ten cm loops of the distal jejunum and ileum were ligated with umbilical tape leaving about 1-2 cm area for interloop. Loops were injected with $3X10^9$ CFU of MparaTb ATCC 19698. Sterile PBS was inoculated into the control loops. The loops were replaced into the abdominal cavity and the incision temporarily closed with Backhaus towel clamps. 2.5% dextrose and 0.45% normal physiological saline was used to maintain circulating blood volume at 5 ml/kg/hr, IV. At 0.5, 1, 2, 4, 8 and 12 hours after bacterial inoculation, loops were excised. Samples for bacteriologic culture, and DNA extraction were collected as described below. Electrocautery was used to control hemorrhage after excision of loops. Throughout the experimental procedure, the calves were monitored for vital signs (blood pressure, heart rate, hydration status, anesthesia depth and temperature). The calves were euthanized with a rapid overdose (single bolus at 60mg/lb IV) of pentobarbital sodium after the final 12-hour loops were excised.

A 6 mm biopsy punch was used to collect two tissue samples from Peyer's patches for bacteriology. Intestinal tissue samples were washed three times in PBS, weighed, homogenized in PBS, serially diluted, and plated onto Herrold Egg Yolk Media containing ANV (Becton Dickinson and Company, Sparks, MD) for incubation at 37°C. The cultures were observed visually weekly for any contamination and the final counts of colony forming units were recorded on week 16.

The colonies were selected and grown in 7H9 broth containing Mycobactin J, OADC and glycerol. DNA was prepared from this bacterial suspension for host-passaged AFLP studies.

RESULTS

AFLP analysis of bovine and human isolates. In an effort to determine the polymorphic differences between MparaTb isolates of differing hosts, AFLP analysis was performed on 4 human ATCC MparaTb isolates and compared to AFLP analysis of 6 Texas Brahman cattle clinical isolates and 1 ATCC bovine isolate 19851. Duplicate ATCC 19698 was used for the normalization of the AFLP protocol. AFLP primers and adapters were used as previously described (51). By comparing these AFLP data, 6 regions only common to MparaTb isolates of intracellular origin (Figure 12) were identified. These regions corresponded to the following primer combinations: region 17 P-GG, M-AC (240bp); region 20 P-CG, M-AC (330bp); region 21 P-CG, M-AC (335bp); region 22 P-GC, M-AC (215bp); region 23 P-GC, M-AC (395bp); region 24 P-GC, M-AC (410bp); and region 25 P-GC, M-AC (490bp). These AFLP data suggested a genetic separation of MparaTb isolates based upon the host species in which the bacterium infects.

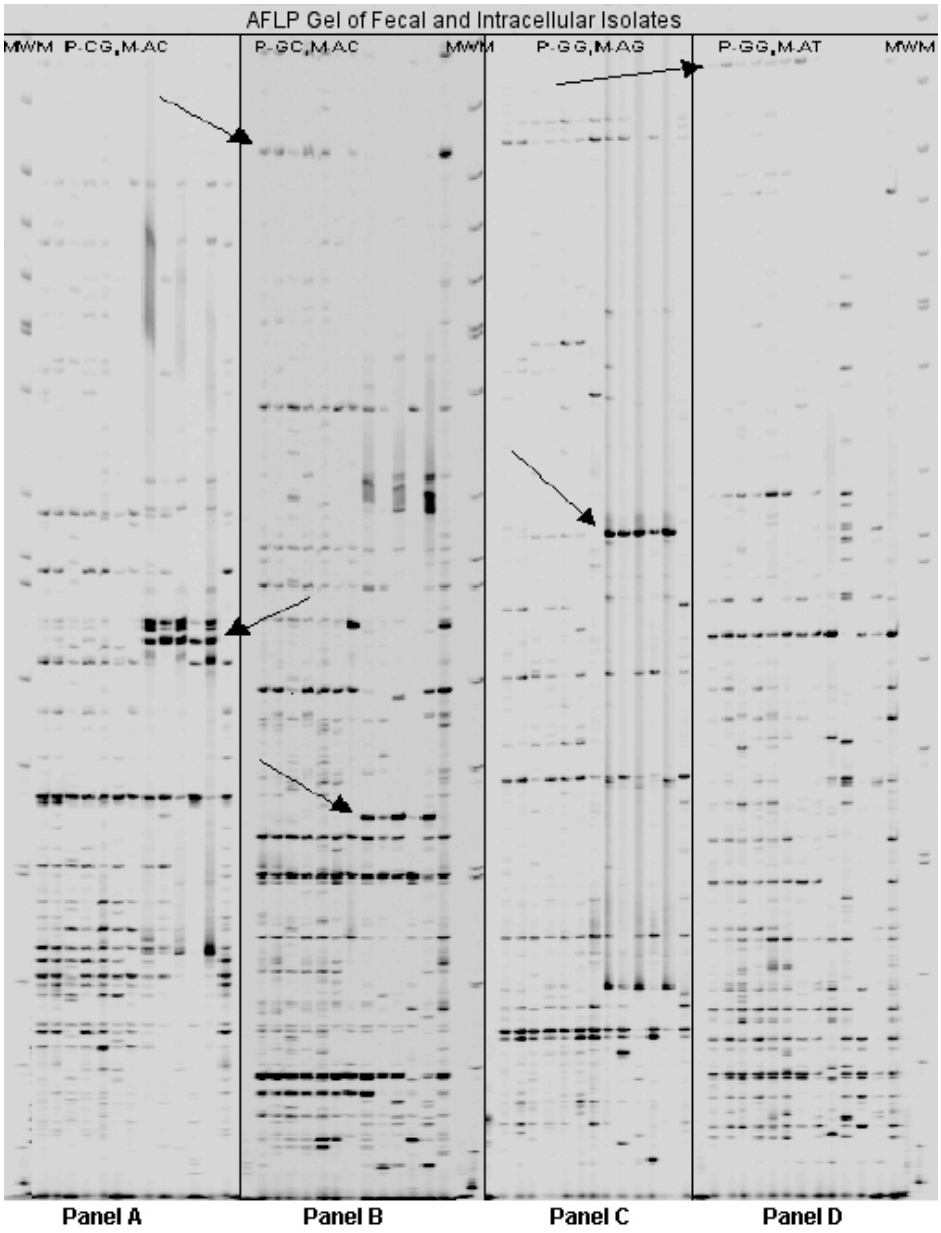


Figure 12. Composite AFLP gel comparison of fecal and intracellular MparaTb isolates. Far left and far right lanes of image are molecular weight markers. Isolates in each panel from left to right are as follows: fecal isolates: T12, T136, T139, T140, T141, T143; ATCC 19698 internal control; intracellular isolates: 19851, 43015, 43544, 43545, 49164; duplicate ATCC 19698 internal control. Panel A: AFLP primer set P-CG, M-AC. Arrow corresponds to region only found in intracellular isolates. Panel B: AFLP primer set P-GC, M-AC. Top arrow corresponds to region present in fecal isolates only; lower arrow corresponds to region present in intracellular isolates only. Panel C: AFLP primer set P-GG, M-AG. Arrow corresponds to region present in intracellular isolates only. Panel D: AFLP primer set P-GG, M-AT. Arrow corresponds to region present in fecal isolates only.

Cluster analysis. In an effort to determine the overall genetic divergence of fecal and intracellular MparaTb isolates, a cluster analysis was performed using Bionumerics data analysis software employing the Dice method of unweighted pair group method with arithmetic mean (UPGMA). This analysis scored AFLP bands according to their size apparent on polyacrylamide gels, not taking into account the corresponding bands apparent intensity. A composite data set analyzing 4 independent AFLP primer sets for each isolate was used to create a dendrogram for the 6 fecal, 5 intracellular and 2 ATCC 19698 internal controls (Figure 13). These data revealed a very tight (97% similarity) clustering of the internal control ATCC 19698 isolates, confirming the very high reproducibility of the AFLP technique. The cluster analysis also revealed clustering of the fecal isolates as well as a clustering of the intracellular isolates. These data also provided evidence of multiple strains of MparaTb within a single bovine herd. Isolates 139, 140, 141 and 143 were all isolated from individual cows in herd 32 from Texas. The cluster analysis revealed a high (89%) similarity of isolates 140 and 141, but 139 and 143 only had a 74% similarity to each other and a 77% similarity to the other herd 32 isolates, suggesting for the possibility of 3 distinct MparaTb isolates within herd 32.

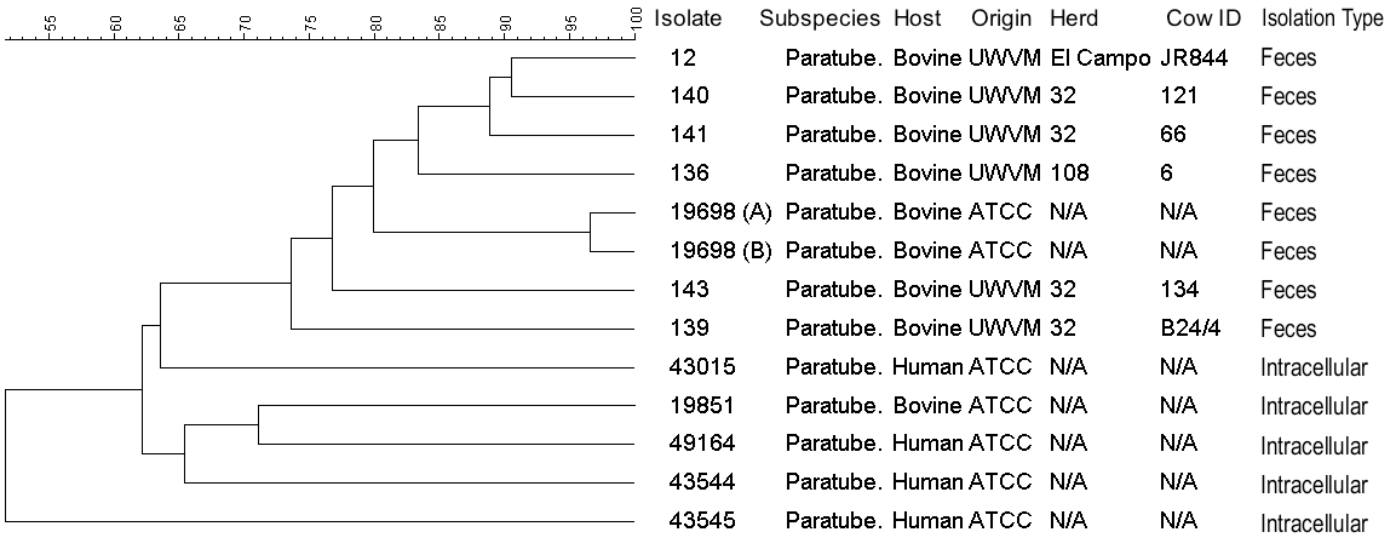


Figure 13. Cluster analysis of fecal and intracellular MparaTb AFLP data.
Cluster analysis reveals from left to right: comparative dendrogram of bovine and human isolates with corresponding percent similarities, isolate number, subspecies designation, host species, origin of sample, herd, cow ID and isolation type.

PCR confirmation of host specific regions. Unique convensional AFLP bands were extracted from agarose gels, cloned into TOPO PCR 2.1 (Invitrogen), and sequenced as described in Materials and Methods. Sequence analysis of each of these products revealed 100% sequence homology to the MparaTb K10 sequence strain. While this evidence was inconsistent with the variation observed through AFLP, the possibility of SNPs at the AFLP restriction sites were then investigated. Unlike previous research (51) approaches, primer pairs flanking the AFLP regions were created in order to sequence the entire region including the restriction sites and flanking regions (Figure 14).

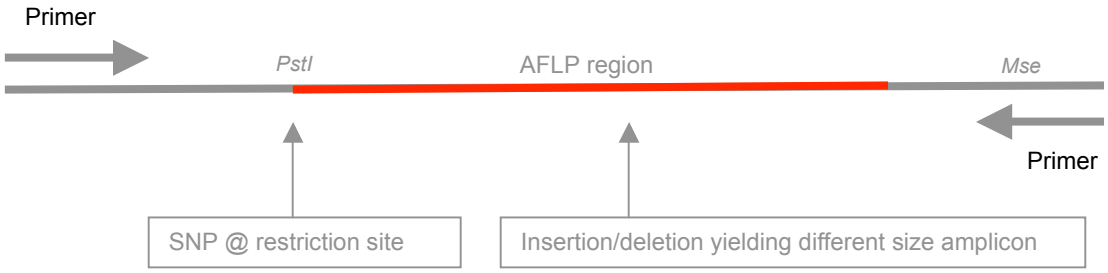


Figure 14. Graphical representation of external primer design.
The above figure shows a graphical representation of the design of external primers to AFLP derived host specific regions. AFLP region (shown in red) was aligned to MparaTb K10 sequence and primers were derived in the regions flanking the AFLP region (grey regions). External primers were derived to research the possibility of SNP occurrence at restriction sites (left box) and the possibility of insertion or deletion events internal to the AFLP region (right box) that may lead to banding patterns differences apparent on AFLP gels.

PCR results of numerous regions present in only intracellular isolates and regions present in only fecal isolates yielded identical amplicons in all isolates tested regardless of origin. These data suggested that there was no sequence

variation event in any of the regions tested. These corresponding PCR products were studied for SNP analysis at AFLP restriction sites that could lead to different banding patterns on AFLP gels. All fecal and intracellular specific regions were amplified with flanking primers and PCR product were cloned into TOPO 2.1 (Invitrogen). Plasmid was purified from positive transformants and cloned regions submitted for sequence analysis. Sequence analysis of all regions tested revealed no SNP occurrence at either the *PstI* or *MseI* restriction sites. These data suggested an unknown epigenetic trait for the AFLP differences observed between intracellular and fecal MparaTb isolates.

Data mining of isolation type specific regions. After eliminating insertions/deletions and single nucleotide polymorphisms at restriction sites as explanations for AFLP differences between intracellular and fecal derived MparaTb isolates, the sequence analysis of these regions was investigated. Under further examination, two AFLP intracellular specific regions, region 17 (hypothetical protein) and region 20 (fructose-6-phosphate amidotransferase), appeared to have a *PstI* restriction site internal to the AFLP derived region. Sequence analysis of amplicons from fecal isolates with these primer sets (region 17 forward 5'- GCAACGATTGTCCCAAACCC -3 reverse 5'- ACAGCACCGACGACGCATTC -3' to amplify a 474bp product; region 20 forward 5'- GCCCAGATCACCGCCAACTAC -3' reverse 5'- CGCATACTTCCTCCCGAACG -3' to amplify a 744bp product) revealed the

same internal *PstI* sites. In an effort to determine if these internal *PstI* sites were cut in the fecal isolates during the AFLP process, we designed AFLP primer sets to amplify a smaller truncated region in the fecal isolates as indicated in figure 15.

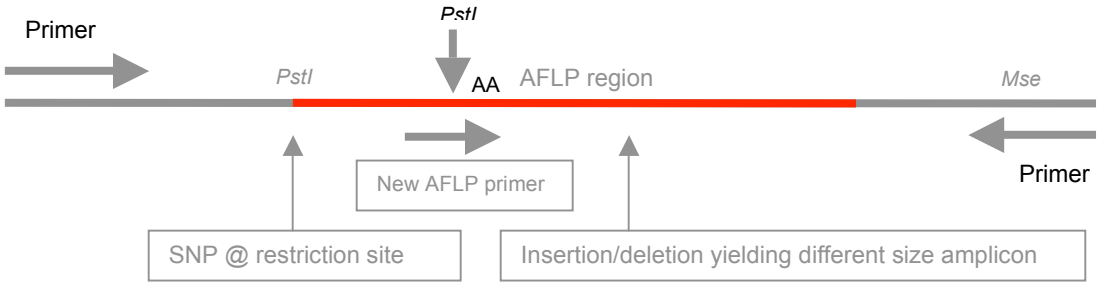


Figure 15. Graphical representation of internal *PstI* site of region 20.
Sequence analysis of region 20 revealed an internal *PstI* site that was not cut in intracellular isolates on AFLP gels. New AFLP primers were designed with corresponding 3' di-adenine for amplification from fecal MparaTb isolates.

AFLP amplification of fecal and intracellular MparaTb isolates with internally derived AFLP *PstI* primer with 3' di-adenine amplified the corresponding sized amplicon from all fecal isolates and corresponding amplicons were absent from all intracellular isolates (Figure 16).

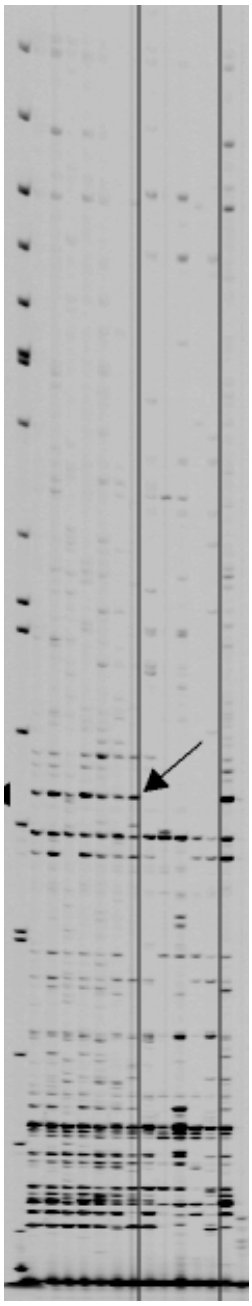


Figure 16. Methylation specific AFLP gel.
AFLP gel with primer sets P-AA, M-AC. Isolates from left to right are: molecular weight marker, Fecal isolates (T12, T136, T139, T140, T141, T143) ATCC 19698, Intracellular isolates (19851, 43015, 43544, 43545, 49164), 19698. Arrow indicates truncated AFLP region present in all fecal isolates and absent from all intracellular isolates corresponding to the 3' end of region 20.

Corresponding amplicons from fecal isolates were extracted from gels and subjected to sequencing as previously described. Sequence analysis confirmed 100% sequence homology when compared with the truncated portion of region 20. These data confirmed the presence of an internal *PstI* restriction site present in all fecal and intracellular isolates of MparaTb tested, but only digested in the AFLP process applied to isolates of fecal origin. A possible explanation for this occurrence could be an intracellular specific epigenetic trait that was only displayed once the host cell internalizes the bacterium. Alternatively, the invasive bacteria bearing the epigenetic modification was not detectable in the original inoculum, but selected during infection. Once internalized, an epigenetic trait becomes apparent inhibiting *PstI* enzymatic restriction of certain *PstI* sites and yielding the AFLP genetic differences observed in the intracellular MparaTb isolates.

In an effort to identify and characterize an epigenetic trait that may act as a restriction inhibitor of the internal *PstI* sites in intracellular isolates, the internal *PstI* sites of regions 17 and 20 were investigated in detail. DNA methylation is one such DNA modification. By aligning the internal *PstI* sites of region 17 and 20, a consensus sequence 22 base pairs upstream of the internal *PstI* sites was identified common to both regions (Figure 17). The positioning of consensus residues relative to the resistant *PstI* site suggests that it could serve as a binding site for a putative methylase. Other DNA methylases such as *BsgI* and

EcoP15I also demonstrate a binding site around 22 base pairs upstream of the target.

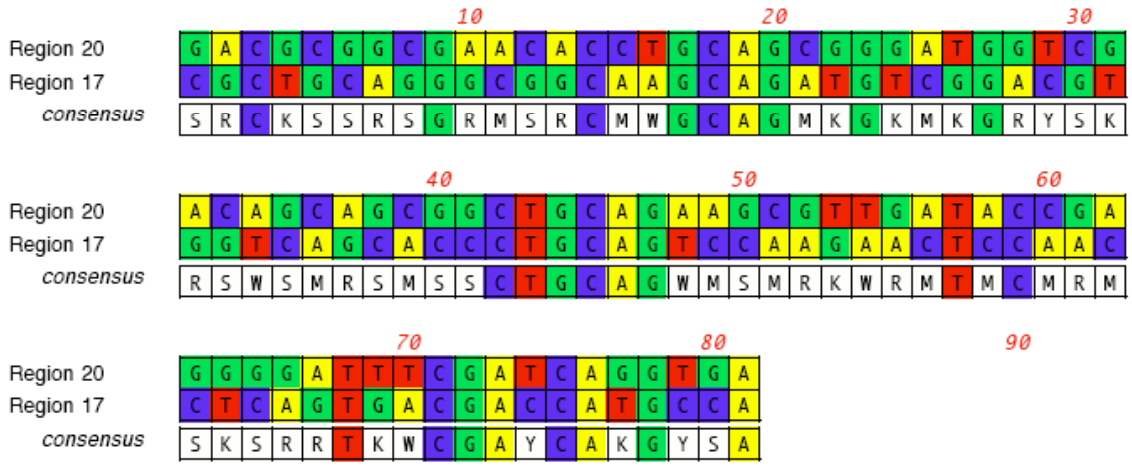
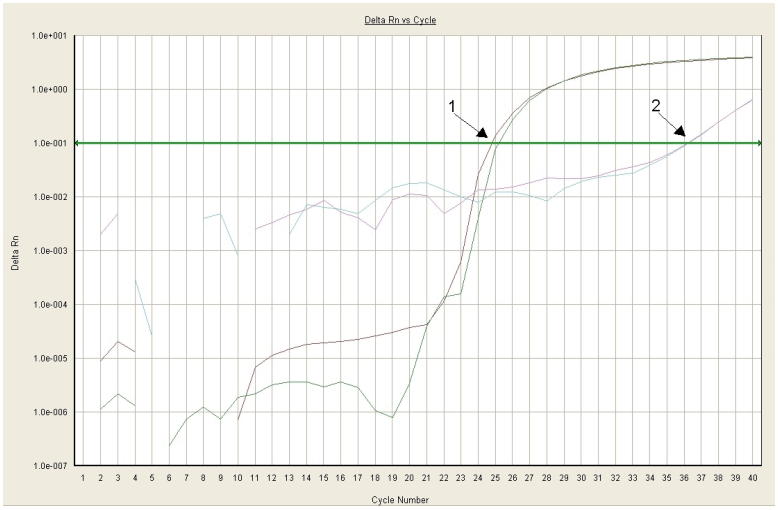


Figure 17. Regions 17 and 20 alignment.
Alignment of sequences of regions 17 and 20 with internal *PstI* sites aligned at position 42. region 17 AFLP digested *PstI* site at position 3. region 20 AFLP digested *PstI* site at position 15. Consensus sequence shows evidence of four base pair region common to both regions at position 17. Internal AFLP *PstI* primer with 3' di-adenine is located at position 48 of region 20.

rtPCR of gene products of regions 17 and 20. In order to determine if epigenetic traits such as methylation of internal *PstI* sites of regions 17 and 20 have an effect on the regulation of transcription of their corresponding gene products, real time PCR analysis of mRNA of MparaTb isolates of fecal and intracellular origins was performed. Preliminary data suggested a dramatic increase in the gene product levels of region 20 from intracellular isolates as compared with those from fecal isolates (Figure 18A and B). These results are

currently being confirmed with additional primer sets for this region and rtPCR primer sets for region 17 are currently being tested against intracellular and

A.



B.

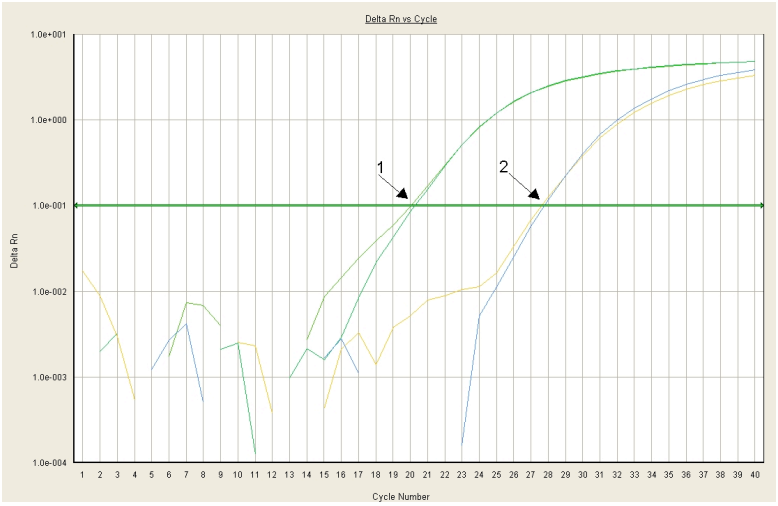


Figure 18. Real-time PCR delta Rn vs. cycle number graphs for T140 and ATCC 19851 with region 20 primers.
A. Real-time PCR data for fecal isolate T140. Arrow 1 represents threshold value for 16sRNA corresponding to cycle 25. Arrow 2 represents threshold value for region 20 corresponding to cycle 36.
B. Real-time PCR data for intracellular isolate ATCC 19851. Arrow 1 represents threshold value for 16sRNA corresponding to cycle 20. Arrow 2 represents threshold value for region 20 corresponding to cycle 28.

fecal isolates. These data reveals an increase in the gene product level for region 20 with intracellular isolate 19851. ATCC 19851 required 28 cycles to reach the delta Rn threshold level as compared to 36 cycles with isolate T140. The 8 cycle increase in threshold level from intracellular to fecal isolates leads to a dramatic increase in gene product levels of region 20. These data provide evidence that the epigenetic characteristic unique to MparaTb isolates of intracellular origin has an effect on the regulation of transcription of gene products.

AFLP of CTA isolates. Intracellular ATCC 19698 was isolated from bovine ileal loops at 30 minutes, 1, 2, 4, 8 and 12-hour increments. DNA from these isolates was analyzed by AFLP and their AFLP banding patterns were compared with those of the 19698 un-inoculated controls (figure 19). The AFLP banding patterns demonstrated clear differences between the un-inoculated control organism and all of the intracellular CTA time points. These banding pattern differences further supported our previous claims of an epigenetic trait, unique to intracellular MparaTb, which inhibits digestion by *PstI* causing AFLP banding pattern differences. These epigenetic differences were more than likely induced by host selection of a sub-population of MparaTb. This is evident also by the fact that the un-inoculated isolate has a different banding pattern than the time-points (arrows in panels A-C figure 19). Moreover, this is further supported by the fact that all of the intracellular CTA time-points, each representing an

independent DNA isolation, had very similar banding patterns. These banding pattern differences were not due to DNA rearrangement due to the very short period of time within the host cell (30 minutes). All of the time-points fall far short of the 24 hours required for the MparaTb generation time.

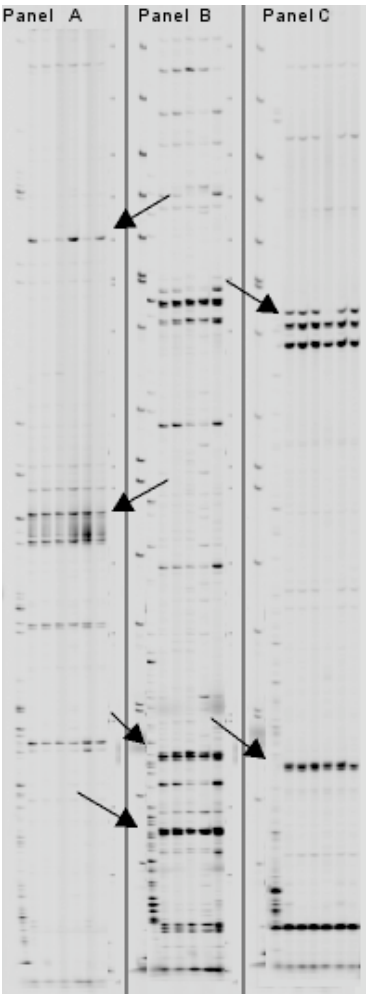


Figure 19. AFLP gel of CTA isolates.
Composite gel image of CTA time-points of 30 minutes, 1, 2, 4, 8, and 12-hour increments. Panel A: AFLP primer set P-GC, M-CG, Isolate designations from left to right: inoculum, 30 minutes, 1, 2, 4, 8, 12-hour, molecular weight marker. Panel B: AFLP primer set P-CT, M-TG, Isolate designation from left to right: molecular weight marker, inoculum, 30 minutes, 1, 2, 4, 8, 12-hour. Panel C: AFLP primer set P-CT, M-GG, Isolate designation from left to right: molecular weight marker, inoculum, 30 minutes, 1, 2, 4, 8, 12-hour. Arrows designate regions present in all CTA time-point isolates and absent from inoculum.

DISCUSSION

In this study, the ability of AFLP to discriminate between MparaTb isolates of varying host species and between isolates of intracellular and fecal origin was confirmed. These discriminating genetic variations were evident upon examination of AFLP amplified DNA and were unique to isolates of specific invasions, whether isolates were intracellular or fecal origin (Figure 12). These intracellular/extracellular genetic variations were not confined to one invasion type of MparaTb. On the contrary, invasion specific regions were found to be present in fecal and intracellular isolates alike. These data confirmed previously published results that suggest a differentiation among MparaTb isolates from differing hosts (3, 22). While our findings confirm with the results of Amonsin *et al.*, these earlier publications were based upon multilocus short sequence repeat (MLSSR) analysis of MparaTb isolates which only compares a very small region of the bacterial genome. The techniques used in this study incorporated a whole genome approach to genetic comparisons and found many different and distinct regions.

As verification for the AFLP data that suggested a genetic separation of intracellular and extracellular (fecal) MparaTb isolates, cluster analysis was performed using the UPGMA method of the Bionumerics program. This analysis confirmed the AFLP observations that fecal and intracellular isolate banding patterns of MparaTb were genetically distinct by clustering MparaTb fecal

isolates together. MparaTb intracellular isolates were also clustered together but distinct from one another. The cluster analysis also demonstrated the reproducibility of AFLP by the clustering of the duplicate sample of ATCC 19698 to a similarity of 97%. It is also of note that the ATCC 19698 isolate itself fell in the middle of the fecal cluster of isolates. This was not unanticipated since ATCC 19698 was also a fecal isolate of MparaTb. Another apparent observation from the cluster analysis was the uniqueness of three different strains of bovine MparaTb from herd 32. Cluster analysis revealed that isolates 140 and141 clustered closely together while isolates 139 and 143 diverged from one another and only show a 74% similarity to each other and a 77% similarity to the other herd 32 isolates. Taken together these data suggested that herd 32 was infected with 3 strains of MparaTb. Cluster analysis further revealed a clustering of the MparaTb human isolates as a less similar group of isolates when compared with the bovine isolates. These data were reasonable in view of the fact that the human MparaTb isolates were isolated from regions all over the world over a period of decades, and the bovine isolates were isolated recently exclusively from herds in Texas. These data validated previous work in which multiplex PCR of IS900 integration loci (MPIL) revealed human MparaTb isolates exhibit a greater genetic diversity when compared with bovine isolates (46). These data also provided evidence of possible evolutionary changes over time for the human isolates.

The genetic diversity of MparaTb isolates from fecal and intracellular samples as revealed through AFLP analysis identified genetic regions specific to one invasion type or the other. In this study, the genetic reasons behind these invasion specific variations were investigated. Sequence analysis of the intracellular specific regions revealed sequence homology to the MparaTb K10 isolate. In an effort to determine if these isolation specific regions were based upon insertion/deletion events internal to the AFLP regions, or if they were based upon more subtle characteristics, primers flanking these AFLP regions were designed to perform sequence analysis and to detect SNP events at the AFLP restriction sites. All regions targeted had PCR products of identical sizes when applied to both invasion types of MparaTb isolates. These led to the conclusion that insertion/deletion events did not play a role in AFLP banding polymorphisms. Sequence analysis of these amplicons further revealed no SNP events at any of the AFLP restriction sites. However, upon examination of regions 17 and 20, an internal *PstI* restriction site refractory to restriction during the AFLP process was apparent. region 17 and 20 corresponded to AFLP bands unique to intracellular isolates. Their apparent uniqueness was not however based upon any differences in the DNA sequence. Therefore the possibility of an epigenetic event such as DNA methylation at this internal *PstI* site was investigated as a possible explanation for the unique AFLP banding patterns. An AFLP primer with 3' di-adenine was created for the amplification of

the truncated 3' region of region 20 (Figure 15). These AFLP data revealed the presence of an amplicon of the corresponding size present in the fecal isolates and absent from the intracellular isolates. These data were confirmed through sequence analysis of the fecal isolate region. These data further suggested the possibility of an intracellular isolate specific methylation of the internal *PstI* site of region 20. The MparaTb K10 genome has been searched for known methylases with no success. Support for the hypothesis of an intracellular specific methylation of the internal region 20 *PstI* site was found in the presence of a consensus sequence 22 base pairs upstream of both internal *PstI* sites in region 17 and 20. Although the methyltransferase has not yet been identified, a putative enzyme recognition sequence was observed with the following sequence 5'- GCAGnnGnnnGnnnnnnnnnnnnnnCTGCAG -3'. The presence of a methylase recognition sequence 22 base pairs upstream of the methylation site is not uncommon. For example, the restriction enzyme *BsgI* of *Bacillus sphaericus* recognizes a site 14 base pairs away from the restriction site (5'-GTGCAGNNNNNNNNNNNNN,NN^ -3'). Additionally the restriction enzyme *EcoP15I* of *E. coli* P15 binds 25 base pairs upstream of the restriction site (5'-CAGCAGNNNNNNNNNNNNNNNNNNNNNNN^NN -3'). In the case of *BsgI* as with the MparaTb site, the recognition sequence does not constitute a perfect palindrome. These examples demonstrate the possibility of an intracellular

specific methylase as the explanation for the differences evident in the AFLP comparisons of intracellular and fecal derived MparaTb banding patterns.

That DNA methylation plays an important role in the regulation of gene products in all types of cells (2, 9, 23, 29, 39, 43) is widely observed. In the study described here, real time PCR analysis was performed to quantitate the gene transcription levels of hypothetical proteins corresponding to regions 17 and 20 respectively. These data revealed a dramatic increase in the gene transcript levels of region 20 in the intracellular derived MparaTb isolates as compared to those of fecal origin. These data suggest that the methylation status of the internal *PstI* restriction site of region 20 plays a role in the up-regulation of fructose-6-phosphate amidotransferase. These data are intriguing due to previous studies that have demonstrated an overexpression of fructose-6-phosphate amidotransferase (involved in the formation of glucosamine-6-phosphate) therefore providing a resistance to methylmercury exposure in *Saccharomyces cerevisiae* (45). Methylmercury, an organometallic cation, is ingested and readily absorbed by the GI tract. Methylmercury largely found bound to free cysteine or with proteins and peptides containing this amino acid, the complex is recognized by amino acid transporting proteins as methionine (31). It is possible that the methylation of fructose-6-phosphate causes an up-regulation of the corresponding gene product resulting in an enhanced ability to survive in the presence of this toxin.

Although the confirming evidence for a role of *PstI* methylation in intracellular survival or virulence of MparaTb has yet to be shown, the suggestions of such an effect is increasing. First, AFLP differences between intracellular and extracellular bacteria were not due to genetic sequence variations. The best explanation for other genetic influence or epigenetic variation that could inhibit the *PstI* digestion in one isolate and not the other is DNA methylation. DNA methylation has been proven to play a role in the variation of transcription and therefore gene product levels in bacteria. Data from the present investigation provide good evidence that gene product levels from intracellular isolates for region 20 are greater than those for region 20 of fecal isolates. These data coupled with the discovery of a consensus sequence that could possibly be a recognition sequence for a DNA methyltransferase, supports the conclusion that the methylation profiles of intracellular and fecal derived MparaTb isolates are unique and that these methylation patterns likely influence the transcription levels of the methylated genes. These conclusions are also supported by the AFLP analysis of CTA isolates that revealed banding pattern differences of intracellular passaged MparaTb as compared to the original inoculum. These data further suggest the presence of an epigenetic trait expressed only in intracellular MparaTb bacteria. These data coupled with the AFLP patterns of intracellular and fecal bacteria that revealed a PstI restriction site that was not digested in all of the intracellular bacteria, and that was

digested in all of the fecal isolates supports the conclusion that the epigenetic trait common to intracellular MparaTb inhibits PstI digestion. Furthermore, the real-time PCR data of region 20 suggest an up-regulation of the corresponding gene product linked to this intracellular epigenetic trait.

SUMMARY AND CONCLUSIONS

The genetic variation of 20 clinical MparaTb isolates was compared with two *M. avium avium* isolates, using amplified fragment length polymorphism (AFLP). The overall resolution of the AFLP gels is extremely high, and AFLP is able to identify polymorphic regions from subspecies to subspecies, isolate to isolate within a subspecies, and from isolate to isolate within a subspecies from different hosts. Bionumerics analysis of AFLP gels confirmed the MparaTb genome to be very diverse from isolate to isolate and that the genome was much more plastic than previously believed. AFLP analysis also revealed many regions to be present in one subspecies; thus five of regions were present exclusively in MparaTb isolates. Sequence analysis of these five regions revealed their absence from the *M. avium avium* genome; thus these polymorphic regions were not based upon SNPs or methylation events. PCR analysis of these regions with internal primers performed on all 20 MparaTb and 2 *M. avium avium* isolates revealed that four of the 5 regions were unique to MparaTb and only one region (region 4) amplified its corresponding region in one of the *M. avium avium* isolates. To confirm that these AFLP regions were unique to MparaTb, 55 additional MparaTb isolates and 65 non-paratuberculosis isolates were collected and analyzed. PCR of the 5 AFLP identified regions was performed on all 120 mycobacterial isolates and these findings confirmed that 4 of 5 regions tested occur exclusively in MparaTb. Region 4 amplified its

corresponding region in one of the non-paratuberculosis isolates even when another 63 isolates were added. With respect to diagnostics, the specificity of these regions to MparaTb was very high, while the sensitivity of these regions varied from 78.18% for region 4 to 100% for region 5, illustrating a very broad range of heterogeneity. These data are in agreement with our previous AFLP data that also demonstrated a high degree of genomic variation among MparaTb isolates.

In an effort to understand the nature of these unique MparaTb regions, results of sequence information of the contigs of these regions originating from the entire *M. avium avium* sequence database were analyzed. In the sequence alignment, 2 of the 5 regions examined were part of much larger regions that were not present in from the *M. avium avium* genome. PCR primer walk primers were designed across these 5kb and 15kb regions that confirmed their presence in the MparaTb genome and absence from the *M. avium avium* genome. These data further confirm that the prior understanding of the genetic relatedness of MparaTb and *M. avium avium* was incorrect, thus these closely related mycobacteria were not separated by small genetic differences as previously thought, but were separated instead by expansive deletions in the genomes when compared side by side. Sequence analysis coupled with BLAST alignments to known databases revealed that genes present in these large MparaTb specific regions have sequence homology to known integrase and

recombinase genes. This information implies a possible explanation for the genetic divergence of the two subspecies based upon the ability of MparaTb to manipulate its genome through the activity of integrase/recombinase genes.

Once the genetic heterogeneity of MparaTb isolates was established, efforts were focused on studying invasion specificity of MparaTb isolates and the predilection of MparaTb to infect certain hosts. In order to accomplish this investigation, AFLP data from intracellular MparaTb isolates were compared to data from fecal MparaTb isolates. After analyzing the AFLP data, it was noted that genetic regions existed in intracellular isolates that were absent from fecal isolates and vice versa. Sequence analysis of these invasion specific regions confirmed that there was no obvious insertion or deletion event driven for example by transposition responsible for the genetic differences. This also was the case with the comparison of MparaTb and *M. avium avium*. Thus, SNP analysis was performed on the AFLP restriction sites as a possible explanation for the genetic differences, but SNPs were not detected. The investigation then proceeded to evaluate a potential epigenetic event occurring at one or more of the AFLP restriction sites that would inhibit the restriction analysis of isolates from one invasion type but not the other. A hypothesis supported by sequence analysis of region 20 (intracellular specific) revealed the location of a *PstI* restriction site internal to the AFLP region. This restriction site was present in the sequence analysis of both intracellular and fecal isolates alike but only

region 20 was recognized by the enzyme. Logically, the next step was to investigate the possibility of a methylation event at this internal *PstI* restriction site in intracellular isolates and not in the fecal isolates. This could serve to explain the AFLP differences between the two strains. AFLP analysis with the *PstI* primer with corresponding 3' di-adenines confirmed its corresponding amplicon in fecal isolates and its absence in intracellular isolates. These data, along with the sequence analysis of this amplicon, confirmed the existence of an intracellular MparaTb isolate specific epigenetic trait as the explanation for the genetic differences in the banding patterns on the AFLP gels. The effect of the intracellular specific epigenetic trait on the gene product levels of the affected genes was determined through real-time PCR analysis of mRNA derived from infected tissue. Both intracellular and fecal MparaTb isolates were analyzed with primers specific for region 17 and 20 genes. These data provided promising preliminary evidence for a dramatic increase in the gene product levels of fructose-6-phosphate amidotransferase (region 20) in intracellular isolates of MparaTb as compared with fecal isolates. Because methylation is known to play an important role in the transcription levels of virulence factor genes in bacteria causing up-regulation in the methylated state (9). Fructose-6-phosphate amidotransferase is involved in the formation glucose-6-phosphate. Up regulation of this gene in yeast was found to be responsible for the resistance to methylmercury *in vitro*. The occurrence of an intracellular isolate

specific epigenetic trait, such as the methylation of the internal *PstI* site of region 20, could possibly lead to the prolonged survival or even resistance to environmental toxins inside the host cell.

The research described herein is expected to change the dogma of MparaTB mycobacterial genetics. The MparaTb genome is not the static genome as once previously thought, but it is genetically variable from isolate to isolate. Furthermore, the genetic variation among MparaTb isolates has demonstrated invasion specificity based upon specific epigenetic traits such as DNA methylation. MparaTb is not as genetically similar to *M. avium avium* as previously thought, because MparaTb is genetically distinct from *M. avium avium* based upon the presence of large 5-15kb regions throughout the MparaTB genome that are absent from *M. avium avium*. Furthermore, these large regions reveal genes with integrase/recombinase functions, suggesting a mechanism for the genetic divergence of MparaTb from *M. avium avium*. The genetic differences between MparaTb and *M. avium avium* also serve as the basis for a differential PCR diagnostic assay with 98% sensitivity and 100% specificity for MparaTb, a potentially powerful tool for molecular epidemiology.

REFERENCES

1. **Aduriz, J. J., R. A. Juste, and N. Cortabarria.** 1995. Lack of mycobactin dependence of mycobacteria isolated on Middlebrook 7H11 from clinical cases of ovine paratuberculosis. Vet Microbiol **45:**211-7.

2. **Allamane, S., P. Jourdes, D. Ratel, J. M. Vicat, I. Dupre, M. Laine, F. Berger, A. L. Benabid, and D. Wion.** 2000. Bacterial DNA methylation and gene transfer efficiency. Biochem Biophys Res Commun **276:**1261-4.

3. **Amonsin, A., L. L. Li, Q. Zhang, J. P. Bannantine, A. S. Motiwala, S. Sreevatsan, and V. Kapur.** 2004. Multilocus short sequence repeat sequencing approach for differentiating among *Mycobacterium avium subsp. paratuberculosis* strains. J Clin Microbiol **42:**1694-702.

4. **Bannantine, J. P., E. Baechler, Q. Zhang, L. Li, and V. Kapur.** 2002. Genome scale comparison of *Mycobacterium avium subsp. paratuberculosis* with *Mycobacterium avium subsp. avium* reveals potential diagnostic sequences. J Clin Microbiol **40:**1303-10.

5. **Bannantine, J. P., Q. Zhang, L. L. Li, and V. Kapur.** 2003. Genomic homogeneity between *Mycobacterium avium subsp. avium* and *Mycobacterium avium subsp. paratuberculosis* belies their divergent growth rates. BMC Microbiol **3:**10.

6. **Bart, A., M. W. van Passel, K. van Amsterdam, and A. van der Ende.** 2005. Direct detection of methylation in genomic DNA. Nucleic Acids Res **33:**e124.

7. **Barta, Z., I. Csipo, G. Mekkel, M. Zeher, and L. Majoros.** 2004. Seroprevalence of *Mycobacterium paratuberculosis* in patients with Crohn's Disease. J Clin Microbiol **42:**5432; author reply 5432-3.

8. **Cao, X., and S. E. Jacobsen.** 2002. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A **99 Suppl 4:**16491-8.

9. **Casadesus, J., and D. Low.** 2006. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev **70:**830-56.

10. **Cottrell, S. E., J. Distler, N. S. Goodman, S. H. Mooney, A. Kluth, A. Olek, I. Schwope, R. Tetzner, H. Ziebarth, and K. Berlin.** 2004. A real-time PCR assay for DNA-methylation using methylation-specific blockers. Nucleic Acids Res **32:**e10.

11. **Crossley, B. M., F. J. Zagmutt-Vergara, T. L. Fyock, R. H. Whitlock, and I. A. Gardner.** 2005. Fecal shedding of *Mycobacterium avium subsp. paratuberculosis* by dairy cows. Vet Microbiol **107:**257-63.

12. **Devallois, A., K. S. Goh, and N. Rastogi.** 1997. Rapid identification of mycobacteria to species level by PCR-restriction fragment length polymorphism

analysis of the hsp65 gene and proposition of an algorithm to differentiate 34 mycobacterial species. J Clin Microbiol **35:**2969-73.

13. **Devallois, A., M. Picardeau, K. S. Goh, C. Sola, V. Vincent, and N. Rastogi.** 1996. Comparative evaluation of PCR and commercial DNA probes for detection and identification to species level of *Mycobacterium avium* and *Mycobacterium intracellulare*. J Clin Microbiol **34:**2756-9.

14. **Ehrlich, M., M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, and C. W. Gehrke.** 1985. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. Nucleic Acids Res **13:**1399-412.

15. **Ehrlich, M., G. G. Wilson, K. C. Kuo, and C. W. Gehrke.** 1987. N4-methylcytosine as a minor base in bacterial DNA. J Bacteriol **169:**939-43.

16. **Englund, S., G. Bolske, and K. E. Johansson.** 2002. An IS900-like sequence found in a *Mycobacterium sp.* other than *Mycobacterium avium subsp. paratuberculosis*. FEMS Microbiol Lett **209:**267-71.

17. **Francois, B., R. Krishnamoorthy, and J. Elion.** 1997. Comparative study of *Mycobacterium paratuberculosis* strains isolated from Crohn's disease and Johne's disease using restriction fragment length polymorphism and arbitrarily primed polymerase chain reaction. Epidemiol Infect **118:**227-33.

18. **Garcia-Del Portillo, F., M. G. Pucciarelli, and J. Casadesus.** 1999. DNA adenine methylase mutants of *Salmonella typhimurium* show defects in

protein secretion, cell invasion, and M cell cytotoxicity. Proc Natl Acad Sci U S A **96:**11578-83.

19. **Hamilton, H. L., D. M. Follett, L. M. Siegfried, and C. J. Czuprynski.** 1989. Intestinal multiplication of *Mycobacterium paratuberculosis* in athymic nude gnotobiotic mice. Infect Immun **57:**225-30.

20. **Harris, N. B., and R. G. Barletta.** 2001. *Mycobacterium avium subsp. paratuberculosis* in Veterinary Medicine. Clin Microbiol Rev **14:**489-512.

21. **Harris, N. B., Z. Feng, X. Liu, S. L. Cirillo, J. D. Cirillo, and R. G. Barletta.** 1999. Development of a transposon mutagenesis system for *Mycobacterium avium subsp. paratuberculosis*. FEMS Microbiol Lett **175:**21-6.

22. **Harris, N. B., J. B. Payeur, V. Kapur, and S. Sreevatsan.** 2006. Short-sequence-repeat analysis of *Mycobacterium avium subsp. paratuberculosis* and *Mycobacterium avium subsp. avium* isolates collected from animals throughout the United States reveals both stability of loci and extensive diversity. J Clin Microbiol **44:**2970-3.

23. **Heithoff, D. M., R. L. Sinsheimer, D. A. Low, and M. J. Mahan.** 1999. An essential role for DNA adenine methylation in bacterial virulence. Science **284:**967-70.

24. **Hemavathy, K. C., and V. Nagaraja.** 1995. DNA methylation in mycobacteria: absence of methylation at GATC (Dam) and CCA/TGG (Dcm) sequences. FEMS Immunol Med Microbiol **11:**291-6.

25. **Herman, J. G., J. R. Graff, S. Myohanen, B. D. Nelkin, and S. B. Baylin.** 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A **93:**9821-6.

26. **Heusipp, G., S. Falker, and M. A. Schmidt.** 2007. DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol **297:**1-7.

27. **Honma, Y., R. E. Fernandez, and A. T. Maurelli.** 2004. A DNA adenine methylase mutant of *Shigella flexneri* shows no significant attenuation of virulence. Microbiology **150:**1073-8.

28. **Huber, B. S., D. V. Allred, J. C. Carmen, D. D. Frame, D. G. Whiting, J. R. Cryan, T. R. Olson, P. J. Jackson, K. Hill, M. T. Laker, and R. A. Robison.** 2002. Random amplified polymorphic DNA and amplified fragment length polymorphism analyses of *Pasteurella multocida* isolates from fatal fowl cholera infections. J Clin Microbiol **40:**2163-8.

29. **Ivascu, C., R. Wasserkort, R. Lesche, J. Dong, H. Stein, A. Thiel, and F. Eckhardt.** 2007. DNA methylation profiling of transcription factor genes in normal lymphocyte development and lymphomas. Int J Biochem Cell Biol.

30. **Keim, P., A. M. Klevytska, L. B. Price, J. M. Schupp, G. Zinser, K. L. Smith, M. E. Hugh-Jones, R. Okinaka, K. K. Hill, and P. J. Jackson.** 1999. Molecular diversity in *Bacillus anthracis*. J Appl Microbiol **87:**215-7.

31. **Kerper, L. E., N. Ballatori, and T. W. Clarkson.** 1992. Methylmercury transport across the blood-brain barrier by an amino acid carrier. Am J Physiol **262:**R761-5.

32. **Khare, S., T. A. Ficht, R. L. Santos, J. Romano, A. R. Ficht, S. Zhang, I. R. Grant, M. Libal, D. Hunter, and L. G. Adams.** 2004. Rapid and sensitive detection of *Mycobacterium avium subsp. paratuberculosis* in bovine milk and feces by a combination of immunomagnetic bead separation-conventional PCR and real-time PCR. J Clin Microbiol **42:**1075-81.

33. **Klein, P. E., R. R. Klein, S. W. Cartinhour, P. E. Ulanch, J. Dong, J. A. Obert, D. T. Morishige, S. D. Schlueter, K. L. Childs, M. Ale, and J. E. Mullet.** 2000. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res **10:**789-807.

34. **Kostoulas, P., L. Leontides, C. Enoe, C. Billinis, M. Florou, and M. Sofia.** 2006. Bayesian estimation of sensitivity and specificity of serum ELISA and faecal culture for diagnosis of *paratuberculosis* in Greek dairy sheep and goats. Prev Vet Med **76:**56-73.

35. **Langfelder, K., B. Jahn, H. Gehringer, A. Schmidt, G. Wanner, and A. A. Brakhage.** 1998. Identification of a polyketide synthase gene (pksP) of *Aspergillus fumigatus* involved in conidial pigment biosynthesis and virulence. Med Microbiol Immunol (Berl) **187:**79-89.

36. **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 U S A **102:**12344-9.

37. **Lim, H. N., and A. van Oudenaarden.** 2007. A multistep epigenetic switch enables the stable inheritance of DNA methylation states. Nat Genet **39:**269-75.

38. **Litvintseva, A. P., R. Thakur, R. Vilgalys, and T. G. Mitchell.** 2006. Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans var. grubii* (serotype A), including a unique population in Botswana. Genetics **172:**2223-38.

39. **Low, D. A., N. J. Weyand, and M. J. Mahan.** 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect Immun **69:**7197-204.

40. **Mariette, S., D. Chagne, C. Lezier, P. Pastuszka, A. Raffin, C. Plomion, and A. Kremer.** 2001. Genetic diversity within and among *Pinus pinaster* populations: comparison between AFLP and microsatellite markers. Heredity **86:**469-79.

41. **Marsh, I. B., and R. J. Whittington.** 2001. Progress towards a rapid polymerase chain reaction diagnostic test for the identification of *Mycobacterium avium* subsp. *paratuberculosis* in faeces. Mol Cell Probes **15:**105-18.

42. **Martinelli, A., P. Hunt, R. Fawcett, P. V. Cravo, D. Walliker, and R. Carter.** 2005. An AFLP-based genetic linkage map of *Plasmodium chabaudi chabaudi*. Malar J **4:**11.

43. **Mehling, S., H. Lavender, and S. Clegg.** 2007. A Dam methylation mutant of *Klebsiella pneumoniae* is partially attenuated. FEMS Microbiol Lett **000:**1 - 7.

44. **Menz, M. A., R. R. Klein, J. E. Mullet, J. A. Obert, N. C. Unruh, and P. E. Klein.** 2002. A high-density genetic map of Sorghum bicolor (L.) Moench based on 2926 AFLP, RFLP and SSR markers. Plant Mol Biol **48:**483-99.

45. **Miura, N., S. Kaneko, S. Hosoya, T. Furuchi, K. Miura, S. Kuge, and A. Naganuma.** 1999. Overexpression of L-glutamine:D-fructose-6-phosphate amidotransferase provides resistance to methylmercury in *Saccharomyces cerevisiae*. FEBS Lett **458:**215-8.

46. **Motiwala, A. S., M. Strother, A. Amonsin, B. Byrum, S. A. Naser, J. R. Stabel, W. P. Shulaw, J. P. Bannantine, V. Kapur, and S. Sreevatsan.** 2003. Molecular epidemiology of *Mycobacterium avium* subsp. *paratuberculosis*: evidence for limited strain diversity, strain sharing, and identification of unique targets for diagnosis. J Clin Microbiol **41:**2015-26.

47. **Motiwala, A. S., M. Strother, N. E. Theus, R. W. Stich, B. Byrum, W. P. Shulaw, V. Kapur, and S. Sreevatsan.** 2005. Rapid detection and typing of

strains of *Mycobacterium avium* subsp. *paratuberculosis* from broth cultures. J Clin Microbiol **43:**2111-7.

48. **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:**1039-44.

49. **Nielsen, S. S., C. Gronbaek, J. F. Agger, and H. Houe.** 2002. Maximum-likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of *paratuberculosis*. Prev Vet Med **53:**191-204.

50. **O'Mahony, J., and C. Hill.** 2002. A real time PCR assay for the detection and quantitation of *Mycobacterium avium* subsp. *paratuberculosis* using SYBR Green and the Light Cycler. J Microbiol Methods **51:**283-93.

51. **O'Shea, B., S. Khare, K. Bliss, P. Klein, T. A. Ficht, L. G. Adams, and A. C. Rice-Ficht.** 2004. Amplified fragment length polymorphism reveals genomic variability among *Mycobacterium avium* subsp. *paratuberculosis* isolates. J Clin Microbiol **42:**3600-6.

52. **Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski.** 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding **2:**225-238.

53. **Proctor, R. H., A. E. Desjardins, R. D. Plattner, and T. M. Hohn.** 1999. A polyketide synthase gene required for biosynthesis of fumonisin mycotoxins in *Gibberella fujikuroi* mating population A. Fungal Genet Biol **27:**100-12.

54. **Rademaker, J. L., B. Hoste, F. J. Louws, K. Kersters, J. Swings, L. Vauterin, P. Vauterin, and F. J. de Bruijn.** 2000. Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. Int J Syst Evol Microbiol **50 Pt 2:**665-77.

55. **Rao, B. S., and A. Buckler-White.** 1998. Direct visualization of site-specific and strand-specific DNA methylation patterns in automated DNA sequencing data. Nucleic Acids Res **26:**2505-7.

56. **Robbe-Austerman, S., I. A. Gardner, B. V. Thomsen, D. G. Morrical, B. M. Martin, M. V. Palmer, C. O. Thoen, and C. Ewing.** 2006. Sensitivity and specificity of the agar-gel-immunodiffusion test, ELISA and the skin test for detection of paratuberculosis in United States Midwest sheep populations. Vet Res **37:**553-64.

57. **Roy, J. K., M. S. Lakshmikumaran, H. S. Balyan, and P. K. Gupta.** 2004. AFLP-based genetic diversity and its comparison with diversity based on SSR, SAMPL, and phenotypic traits in bread wheat. Biochem Genet **42:**43-59.

58. **Sockett, D. C., T. A. Conrad, C. B. Thomas, and M. T. Collins.** 1992. Evaluation of four serological tests for bovine *paratuberculosis*. J Clin Microbiol **30:**1134-9.

59. **Srivastava, R., K. P. Gopinathan, and T. Ramakrishnan.** 1981. Deoxyribonucleic acid methylation in mycobacteria. J Bacteriol **148:**716-9.

60. **Taylor, T. B., C. Patterson, Y. Hale, and W. W. Safranek.** 1997. Routine use of PCR-restriction fragment length polymorphism analysis for identification of mycobacteria growing in liquid media. J Clin Microbiol **35:**79-85.

61. **Telenti, A., F. Marchesi, M. Balz, F. Bally, E. C. Bottger, and T. Bodmer.** 1993. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J Clin Microbiol **31:**175-8.

62. **Tsai, H. F., Y. C. Chang, R. G. Washburn, M. H. Wheeler, and K. J. Kwon-Chung.** 1998. The developmentally regulated alb1 gene of *Aspergillus fumigatus*: its role in modulation of conidial morphology and virulence. J Bacteriol **180:**3031-8.

63. **Vilkaitis, G., and S. Klimasauskas.** 1999. Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. Anal Biochem **271:**116-9.

64. **Zavitz, K. H., and K. J. Marians.** 1992. ATPase-deficient mutants of the *Escherichia coli* DNA replication protein PriA are capable of catalyzing the assembly of active primosomes. J Biol Chem **267:**6933-40.

65. **Zhivotovsky, L. A.** 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecol **8:**907-13.

VITA

BRIAN JAMES O'SHEA

ADDRESS

440 Reynolds Medical Building. Room 420

College Station, Texas 77843-1114

EDUCATION

Ph.D. Veterinary Microbiology, Texas A&M University- December 2007

B.S. Microbiology, Louisiana State University- May 2000