Quantitative Detection and Discrimination of Pathogenic Mycobacteria Using LATE-PCR and Lights-On/Lights-Off Probes

Master's Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Biology
Lawrence Wangh, Advisor

In Partial Fulfillment
of the Requirements for

Master’s Degree

By
Skye Fishbein

May 2012
Copyright

Skye Fishbein
© 2012
Acknowledgements:

I extend my gratitude to Dr. Barry Kreiswirth and his group at PHRI for providing me with mycobacterial genomic isolates to test in my assay. I would like to first thank my advisor Dr. Lawrence Wangh, for providing me with support both inside and outside the laboratory. I would also like to thank my supervisor John Rice, for his unyielding patience, comic relief, and dedication to the success of my project. Additionally, I thank Steve Van Hooser for sharing all of his quantitative analysis techniques with me, along with teaching me MatLab. I thank the Wangh lab as a whole, for helping me to understand and correctly employ the lab’s technologies. Finally, I thank my family for their support throughout my career, that has led me to the completion of such a thesis.
ABSTRACT

Quantitative Detection and Discrimination of Pathogenic Mycobacteria Using LATE-PCR and Lights-On/Lights-Off Probes

A thesis presented to the Department of Biology

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Skye Fishbein

Tuberculosis and HIV/AIDS are two of the most common infectious diseases in humans worldwide. As a result of these global epidemics, there has also been a rise in opportunistic mycobacterial infections. The genus mycobacteria contains pathogenic species that should be accurately identified to curb disease mortality. To this end, a two-color, single tube mycobacterial species identification assay has been designed using LATE-PCR and Lights-On/Lights-Off probe technologies. This assay uniquely identifies both tuberculous and non-tuberculous mycobacteria based on their individual fluorescence profiles. Quantitative classification was accomplished, using Euclidean distance metrics, with high rates of true positive classification for all mycobacterial species tested. The PCR assay and quantitative classification has been automated for efficient acquisition of results. Future efforts will focus on further verification of the assay’s capacity, and expansion of the mycobacterial species tested.
Table of Contents

Introduction........................................................................................................p1-13

Materials and Methods.......................................................................................p14-24

Results..................................................................................................................p25-55

Discussion...........................................................................................................p56-66

References..........................................................................................................p67-69

Appendix..............................................................................................................p70
Tables

Table 1: Primer Design ........................................................................................................p 14

Table 2: Probe Design ........................................................................................................p 17

Table 3: Mycobacterial Isolates ........................................................................................p 18

Table 4.1: Predicted Probe Melting Temperature: 16s .............................................p 31

Table 4.2: Predicted Probe Melting Temperature: gyrB .............................................p 31

Table 5: Probability of Species Classification ...............................................................p 55
Figures

Figure 1: Mycobacterial Phylogeny ................................................................. p2

Figure 2: 16s rRNA Gene Region Probe Scheme ......................................... p26

Figure 3: gyrB Gene Region Probe Scheme ................................................. p29

Figure 4: Fluorescence Normalization Scheme ............................................ p33

Figure 5: First Derivative Melt Curves: Mycobacterial Isolates ................. p36-42

Figure 6: Second Derivative Melt Curves: Mycobacterial Isolates ............ p44-50

Figure 7: Principal Component Analysis of Fluorescence Data ................. p52

Figure 8: Principal Component Analysis of Composite Fluorescence Data .. p54
Chapter 1: Introduction

Mycobacteria – a dangerous genus:

Tuberculosis and HIV/AIDS are two of the most common infectious diseases in humans worldwide. One third of the world’s population is believed to be infected with TB, including an estimated 8.8 million new infections in 2010 and 1.1 million deaths [1]. Ninety-five percent of tuberculosis infections occur in underdeveloped, rural regions of the world[2]. However, an increase in human immigration from TB-endemic regions of the world has prompted an increase in TB incidence in developed countries [3]. At the clinical level, inadequate case control in many TB-endemic regions has contributed to the mortality of the disease [4].

The genus mycobacterium contains relevant human pathogens, responsible for tuberculosis, leprosy, and other lung disease. Named for a characteristic mycolic acid-rich cell envelope, genus Mycobacterium is a part of the Actinobacteria phylum. The mycobacterium is a gram-positive, rod-shaped bacillus, with a unique lipid envelope responsible for the organism’s pathogenicity and host inflammatory response[5]. Due to the mycobacterial cell wall hydrophobicity, these microbes are easily transmitted through aerosol particles, Mycobacterial are transmitted from human, animal, and environmental reservoirs[6]. Upon infection, the mycobacterium invades the alveolar macrophage, and can dwell in a granuloma within the human host. A latent infection can continue for years[7]. In culture,
mycobacteria stain acid-fast, and have a diversity of carotenoid pigmentation. The mycobacteria that cause tuberculosis are members of the *Mycobacterium tuberculosis* complex (MTBC), while the other members of the genus are considered the non-tuberculous mycobacteria (NTM).

The incidence of NTM infection has increased with the rise of the HIV epidemic. In the US and Europe, it is estimated that forty to fifty percent of AIDS patients have contracted some form of NTM disease[8]. Non-tuberculous mycobacterial disease affects immunocompromised patients with cancer, cystic fibrosis, and lupus, among other autoimmune diseases [9]. Nontuberculous mycobacteria are not novel species; rather, the increase in incidence reflects the advancements in mycobacterial identification techniques[10].

Accurate identification of the species causing a mycobacterial infection can add precision to drug therapy, lessening the likelihood of mortality. To reduce the global TB burden, the WHO advises “earlier and improved detection” via the development of novel diagnostic methods[1]. An increasing occurrence of non-tuberculous mycobacterial disease encourages the development of mycobacterial genotyping techniques that would improve species-specific case management.

A Systematics Overview of Mycobacteria:

Mycobacteria cause a range of diseases, but a definitive phylogeny for this genus remains elusive. Originally classified by their rate of colony formation in culture, mycobacteria were divided into: rapid growers and slow growers [11]. Human mycobacterial infection usually results from infection by slow-growing
mycobacteria, although some rapid-growing mycobacteria have been identified in immunocompromised patients [12]. Phylogenetic trees have been constructed based on housekeeping genes, such as ribosomal RNA genes (16s), and heat-shock protein genes (hsp65). These phylogenies group some species into pathogenic complexes, such as the *Mycobacterium tuberculosis* complex (MTBC) and the *Mycobacterium avium* complex (MAC)[13]. Pathogenicity outside of the MTBC, in MAC and rapid grower *M. fortuitum*, is a result of horizontal gene transfer earlier in the evolution of mycobacteria [14].

The gold standard for genotyping mycobacteria is sequencing of the 16s rRNA gene. Two hypervariable regions in this gene contain polymorphic sequences that differentiate many environmental mycobacteria from one another[13]. All members of the MTBC have identical 16s gene sequences, attesting to the restricted genomic evolution within this group of species.
The MTBCs

The group of tubercle bacilli comprising the MTBC include *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. microti*, *M. canettii*, *M. caprae*, and *M. pinnipedi*. Some species are obligate human pathogens, and all members share greater than ninety-nine percent of their genome. They are therefore regarded as subspecies [15]. The closely-related complex of mycobacteria show little gene transfer [16], but display a range of host preference. *M. tuberculosis*, *M. africanum* and *M. canettii* cause tuberculosis in human hosts. *M. bovis* infects both humans and cattle species. More recently, bovine tuberculosis, caused by *M. bovis*, has been observed in free-ranging wildlife in parts of Africa. This tuberculosis can be transmitted to humans by the consumption of contaminated milk from cattle [17]. *M. microti* primarily causes disease in vole populations [18]. Among the uncommon members of the MTBC, *M. pinnipedi* causes tuberculosis in seal populations, and *M. caprae* causes a form of tuberculosis in goats [19]. For clinical identification of disease-causing MTBC species, we focus on the identification of the major human pathogens: *M. tuberculosis*, *M. africanum*, and *M. bovis*.

Analysis of MTBC genomic sequence data has revealed numerous genes harboring SNPs unique to single members of the complex. SNPs, in both drug resistance genes and housekeeping genes, discriminate members of the MTBC from one another [20]. Comparison of IS6110 insertion sequences between MTBC members reveal variation within the complex [21]. The presence of such insertion
sequences has provided some insight into the evolution of the *Mycobacterium tuberculosis* complex. These mobile elements have been relevant to epidemiological contact tracing and construction of phylogenetic trees, down to the level of geographic strains [22]. Given the global migration patterns of people, the construction of a phylogenetic tree, which reflects microbial geographic speciation, is imperative.

Tubercle bacilli have existed in human populations since the evolution of *homo sapiens* in Africa and subsequent migration out of Africa, beginning about 60,000 years ago [23]. The global diversity observed in the seven major clades of *Mycobacterium tuberculosis* reflects these routes of human migration [23]. *M. africanum* infection is confined to populations of West Africa, suggesting some host genetic element necessary for infection [24]. Human *M. bovis* infection surfaces in rural settings, where the increased proximity of human to animal increases the likelihood for zoonotic transmission of bovine tuberculosis [25]. *M. microti* has caused infection in immunocompromised patients, and is believed to be endemic to parts of Europe [26]. Geographic diversity in the MTBC may be a result of adaptations to the host, but functional diversity from small genetic variation produces different anti-biotic resistance profiles.

MTBC genetic variation is also evident in different drug susceptibility profiles [27]. Due to its greater affinity for the human host, *M. tuberculosis* demonstrates increased susceptibility to rifampin and streptomycin, anti-microbial drugs commonly used by humans. In addition, among the *M. tuberculosis* strain lineages, phylogenetic patterns create specific drug resistance profiles [15]. *M. bovis* is
naturally resistant to the first-line TB drugs pyrazinamide and ethambutol [25]. In the context of public health, the WHO recommends the genotyping of MTBC infection, due to the subspecies diversity that equips mycobacteria of the MTBC with distinct drug resistance and transmission phenotypes [1].

The NTMs

Crippling the immune system makes a person susceptible to potentially severe NTM disease. Moreover, growth of the HIV pandemic has resulted in a concomitant rise in the number of opportunistic mycobacterial infections. *M. avium* is the most frequent cause of NTM infection in HIV-afflicted individuals. Also a result of the increase in HIV infection, *M. kansasii* infection occurrences have risen [28]. Outside of the HIV patient, mycobacteria causes other types of opportunistic infections.

Immunocompromised patients are at risk for nosocomial infections due to non-tuberculous mycobacteria. *M. abscessus* infections are observed in cystic fibrosis patients [29]. Rapid growers *M. fortuitum* and *M. chelonae* commonly cause opportunistic infections in cancer patients [9]. Environmental mycobacteria present in water can also infect a susceptible person, contributing to the span of opportunistic mycobacterial infections.

The clinical presentation of NTM disease is pleiotropic, manifesting in both pulmonary and extrapulmonary infection. Commonly, *M. avium* lung disease resembles a TB-like pulmonary infection [8]. Lymphadenitis, infection of the lymph nodes, is observed in cases of both MAC infection and *M. szulgai* infection [30]. *M.*
fortuitum, and M. marinum cause skin and subcutaneous infections in bone and soft tissue, presenting as a bulbous nodule on the tissue. In addition, the presence of NTM species, such as M. kansasii and M. xenopi, in hospital water supply can result in surgical wound or skin infections [31]. Overall, pathogenic environmental mycobacteria cause a variety of disease that requires specific treatment, depending on the species of NTM and the place of infection.

Anti-microbials used to treat M. tuberculosis are 100 times less effective on environmental mycobacteria. The typical environment of these NTMs has crafted their innate resistance to toxins produced by other saprophytic, environmental microbes; some of these toxins are used for TB treatment [28]. The treatment of NTM disease can consist of TB regimens of antibiotics, as in the case of M. kansasii, or can be variable like that of M. avium, requiring clathromycin and azithromycin [31]. Meanwhile, M. fortuitum is susceptible to oral macrolide and quinolone drug therapies [7]. Given the prospect of treating an unknown mycobacterial infection, species-specific information strengthens case management, and averts the emergence of drug resistance.

**Conventional Molecular Diagnostics**

1. **Visualization Diagnostics**

Pulmonary TB/NTM infections produce ambiguously similar clinical phenotypes in patients and can persist, complicating treatment. Therefore, detection of mycobacterial infection must be rapid, sensitive, and informative. Acid fast staining of sputum smears, invented by Robert Koch
150 years ago, remains the most widely used method of TB diagnosis in the world, especially in low-income, high burden countries [32]. Methods of NTM diagnosis and corresponding drug susceptibility tests depend on traditional microbiological culturing techniques. These methods are slow, due to the growth rate of mycobacteria [11]. Novel detection systems such as high-performance liquid chromatography have been employed for identification of species, but are costly to implement [33]. More recent molecular assays use sequencing of 16s ribosomal RNA gene for diagnosing NTM infections [8]. The high sequence homology of the MTBC has prompted the development of additional diagnostic method for members of the M. tuberculosis complex. A number of genes, including gyrB, rpoB, and 23 rRNA, have been used to detect and differentiate members of the MTBC [34]. Initially, restriction fragment length polymorphism (RFLP) technique was used to differentiate MTBC species based on the gyrB polymorphisms, but the protocol is time consuming [20]. These older detection methods generally lack the rapidity of recent PCR diagnostic technology.

2. **PCR Diagnostics**

Conventional PCR methods have been conceived in the last 10 years to identify both MTBC and NTM species. PCR assays generate double-stranded amplicons for genotypic analysis, enabling species identification more rapidly than sequencing, culturing, and gel visualization methods. Specifically, commercial diagnostic assays employ quantitative PCR to analyze and differentiate between tuberculous mycobacterial species using
the melting temperature of the product template as a distance measure between species [35]. A mycobacterial PCR assay using line probes was designed to discriminate between MTBC and some relevant NTMs, but produced results with a low sensitivity for NTM identification [34]. Aside from mycobacterial species identification, an Xpert MTB/RIF diagnostic assay for identification of \textit{M. tuberculosis} drug resistance has been developed, and is being implemented globally[36]. While PCR-based methods provide the user with fast results, conventional PCR can be nonspecific, unable to visualize single-nucleotide polymorphisms between organisms.

\textbf{New Non-Symmetric Methods of PCR Analysis:}

Linear-after-the-exponential PCR utilizes unequal primer concentrations to generate a single-stranded DNA template for probe hybridization. Exponential amplification persists until the limiting primer concentration is depleted. Then, linear amplification of the excess primer strand proceeds, generating a larger concentration of single-stranded product. In addition to primer concentration differences, the engineered melting temperature (T\textsubscript{m}) of the limiting primer is slightly higher than that of the excess primer. The primer characteristics in LATE-PCR make template production specific and efficient[37]. Although production of single-strand is asymmetric in nature, asymmetric PCR lacks the thermal flexibility of LATE-PCR, where primer annealing and probe detection are separate. Generally, LATE-PCR thermocycling proceeds as denaturation of strands, annealing of primers, target extension, and probe detection of singe-stranded DNA[38].
The separation of primer annealing and probe detection in LATE-PCR enables optimal hybridization of fluorescent probes. Double-stranded amplicons generated via conventional PCR are characterized in real-time by hybridization to fluorescently labeled oligonucleotide probes, during the annealing step of the thermocycle. In contrast, the single-stranded amplicons generated in a LATE-PCR amplification are analyzed at end-point by hybridization to Lights-On/Lights-Off probes at temperatures below the annealing temperature. Along with Lights-On/Lights-Off probes, the generation of the single-strand product simplifies the construction of multiplex assays for detection of disease/species genotype [39].

Lights-On/Lights-Off probe technology utilizes two kinds of oligonucleotide probes to generate a complex fluorescence signal. Each signal is created by a set of these probes, with a Lights-On probe that generates a fluorescence signal when it binds the target, and an adjacent Lights-Off probe that quenches the fluorescence signal when it binds the target. Engineered to a target sequence, the Lights-On probe hybridizes to that sequence at a higher temperature than the off probe. Multiple pairs of these probes, in cohesion with molecular beacons, can coat a long sequence of single-stranded DNA in a single color[39]. In a species identification assay, this probe technology hybridizes uniquely to each target.

Species identification requires the use of mismatch tolerant probes to distinguish one species, or sequence variant, from another. To discriminate multiple sequence variants, Lights-On/Lights-Off probes are designed to hybridize one sequence, but cover a region of the specific gene that contains some of polymorphic sequences within the genus. The majority of sequence variants with a reasonable
number of mismatches can be detected by a set of Lights-On/Lights-Off probes. A single mismatched nucleotide lowers the melting temperature ($T_m$) of the probe, and can be observed in the corresponding fluorescence signal. Hybridization to a mismatched sequence can produce deviations from the target-probe $T_m$ and differences in the amplitude of the fluorescence signal. The sensitivity and flexibility of Lights-On/Lights-Off probe technology creates distinct fluorescence profiles for target sequences, which are highly reproducible.

With any genotyping technology, the characterization of genotype should be reproducible across assays. The use of LATE-PCR and Lights-On/Lights-Off probe technology generates fluorescent contours for a single target sequence that are preserved from one PCR reaction to the next. Additionally, a variation in template concentration does not alter the fluorescence contour significantly. With reliable fluorescence profiles coordinated to species-specific sequences, genotypic assays can be easily conducted using LATE-PCR and Lights-On/Lights Off probes. LATE-PCR fluorescence data is high-resolution melt fluorescence data available for quantitative analysis in a detection system.

**Mathematical Tools for Data Analysis**

Multivariate analysis in the realm of species identification generates information from many forms of experimental data. Multivariate analysis techniques are employed to assess information content in highly complex data. Microorganism classification has been accomplished using a combination of infrared spectroscopy and a form of multivariate analysis called principal component
analysis (PCA) [40]. Along with other forms of multivariate analysis, PCA visualizes the richness of the variation within the data, in an accessible manner. Principle component analysis and linear regression analysis have been used to visualize microarray data reflecting bacterial gene composition, in a similar method of genotypic identification [41]. With applications in physics, neuroscience and, now biology, multivariate analysis characterizes data involved in species/genotype identification.

Principal component analysis, a type of multivariate analysis, visualizes the most discriminatory features of the data. From a mean-subtracted covariance matrix of data, principle component analysis decomposes data into its feature components, or eigenvectors for each data dimension [42]. The eigenvectors and corresponding to the dimensions of extracted variance and the eigenvalues reflect the magnitude of the variance along that eigenvector [43]. These characteristic vectors define a vector space where the data is decomposed and visualized in its most informative dimensions, formed by an eigenbasis of the eigenvectors with the highest eigenvalues. In the score plots, the number of clusters reflects the maximum number of groups that the data can be subdivided into [44]. Principal component analysis visualizes any quantifiable variance within a given set of data, and demonstrates the potential for objective classification by clustering data.

In order to genotype fluorescence signals, the probe dynamics should be highlighted in the fluorescence signal. High-resolution melting analysis (HRMA), used in PCR genotyping assays, handles the raw fluorescence signal, a set of fluorescence measurements taken at each temperature cycle along the temperature
range. Quantitative normalization methods emphasize the relevant DNA-fluorescence activity by converting the fluorescence signal to a fluorescence melt curve[45]. In some diagnostic settings, further subtraction of a wildtype vector highlights genotypic differences in PCR fluorescence[46]. The transformation of PCR data into usable genotypic information can be accomplished by measuring the relative distance between two visualized genotypes.

The basic principle behind performing pattern recognition analysis relies on finding the dissimilarity between any two objects of data, either by computing distance or by comparing characteristics of the data objects. The most popular measure of the difference between two continuous curves from the same n-dimensional space is the Euclidean distance metric. Derived from the distance of a line connecting two points, the Euclidean distance is the sum-squared difference between two points and their coordinates. In the case of multi-dimensional data, the Euclidean distance between two vectors is a distance vector with the same number of dimensions as the data [47]. This metric is useful for biological data that exists in multiple dimensions such as temperature or time [44]. Unlike the select dimensions of variance visualized in principal component analysis, all dimensions of the data are used in the computation of Euclidean distance between two data objects.

**Summary**

The incidence of non-tuberculous mycobacterial disease aggravates both the TB and HIV/AIDS epidemics and causes human morbidity. In order to curb this trend, diagnostic developments must discriminate between relevant mycobacterial
species, both MTBCs and NTMs. LATE-PCR is a rapid, highly sensitive diagnostic method of DNA target amplification. When used with Lights-On/Lights-Off probes, multiple mycobacterial genes can be probed to produce species-specific fluorescent signatures. Discriminatory features in the composite fluorescence signatures of pathogenic mycobacterial species are quantified in both visualization techniques and distance measurements between species fluorescence curves. Objective classification by quantitative methods provides valuable genotypic information to the naïve health care worker, faced with a patient burdened by mycobacterial disease. The mycobacterial discrimination system could produce rapid, informative results for clinical use in a setting where mycobacterial diseases are prevalent. Species-specific information could enhance treatment of tuberculous/non-tuberculous disease, contributing to the reduction of the global disease burden of TB and NTM disease.
Chapter 2: Materials and Methods

To identify and discriminate mycobacteria in a LATE-PCR reaction, two gene targets were chosen for amplification and probe detection: 16s rRNA and gyrB. The 16s rRNA gene was amplified and probed to discriminate between clinically relevant NTM species. On the other hand, the gyrB gene was selected to distinguish between members of the pathogenic *Mycobacterium tuberculosis* complex.

Primer Design

Primers were designed to amplify mycobacterial gene products in a LATE-PCR amplification system. Melting temperatures \(T_m\) were generated by a program called VisualOmp Software (DNA Software, Inc, Ann Arbor, MI), where simulation used primer and template concentrations to calculate primer \(T_m\). Two sets of primers were created to amplify two separate mycobacterial gene products used for detection by probe analysis.

1. **16s**
   
   Primers were placed at highly-conserved regions of this gene, so as to amplify a fragment of hypervariable region A among the majority of mycobacterial species potentially in human sample. A limiting primer and excess primer pair was created to amplify a 210-bp sequence of the mycobacterial 16s rRNA gene. The gene product contains hypervariable
region I, and primers anneal to flanking sequences of the hypervariable region. Primers were designed by a former research technician Bonnie Ronish, and were ordered from Sigma Rich (St. Louis). Primer sequences are shown in Table 1.

2. **gyrB**

A limiting primer and excess primer were engineered to amplify a 170-bp fragment of the DNA gyrase B in some mycobacteria. Primers were made complementary to members of the MTBC. Primer sequences are shown in Table 1, with primers generated from Sigma Rich (St. Louis).

<table>
<thead>
<tr>
<th>Table 1: Primer Design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
</tr>
<tr>
<td>16s EP</td>
</tr>
<tr>
<td>16s LP</td>
</tr>
<tr>
<td>gyrB EP</td>
</tr>
<tr>
<td>gyrB LP</td>
</tr>
</tbody>
</table>

Primers were designed under LATE-PCR criteria that specifies $T_{mL} > T_{mE}$. In VisualOmp Software, the simulated EP concentration was 1000 nm, and the simulated LP concentration was 50 nm. Melting temperatures were calculated for hybridization to *M. tuberculosis*.

**Probe Design**

In both gene products, probes were created to generate a range of different fluorescent profiles for various mycobacterial species. Probes were designed in VisualOmp Software and ordered from Biosearch Technologies, Inc. (Novato, CA).

1. **16s**

   Initially, a single set of Lights-On/Lights-Off probes was designed to cover the hypervariable region of the gene. Both probes were designed with
quenchers, and the Lights-On probe was designed with a FAM fluorophore. Carbon spacers were added to free 3’ ends to prevent mis-priming. The Lights-On probe was engineered to have a higher $T_m$ for members of the MTBC. Later, a second set of probes was designed with a different fluorophore, Cal Orange (CO560), in order to add resolution to the mycobacterial discrimination outside the MTBC. CO560 probes were most complementary to selected pathogenic NTM sequences. Two sets of Lights-On/Lights-Off probes were created to cover the hypervariable region of the 16s gene, and fluoresce in two colors. Probe sequences are shown in Table 2.

2. **gyrB**

A gyrB probe set was engineered to cover single-nucleotide polymorphisms (SNP) within the *Mycobacterium tuberculosis* complex (MTBC). A set of Lights-On/Lights-Off probes and a molecular beacon were designed with FAM fluorophores. Carbon spacers were added to free 3’ ends to prevent mis-priming. Within the single-stranded amplicon, a high temperature hairpin formed, preventing any low temperature probe annealing. A single oligonucleotide was developed to prevent hairpin formation, and facilitate subsequent fluorescent probe hybridization. Probe sequences are shown in Table 2. The gyrB silent probe has no fluorophore on it. The 16s #3 On probe and the gyrB #2 On probe were created as molecular beacons.
**Table 2: Probe Design**

<table>
<thead>
<tr>
<th>Probe Name</th>
<th>Sequence (5' to 3')</th>
<th>5' Modification</th>
<th>3' Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>16s #1 On</td>
<td>AAAGCGCTTTAGCGGTGGGATGAGCC</td>
<td>BHQ</td>
<td>FAM</td>
</tr>
<tr>
<td>16s #1 Off</td>
<td>GGACCACGGGATGCACTTGTGTGGT</td>
<td>BHQ</td>
<td>carbon-spacer</td>
</tr>
<tr>
<td>16s #2 On</td>
<td>CTTCCCAGTAGGCCAGTTGCAGTTCCAAG</td>
<td>CO560</td>
<td>BHQ</td>
</tr>
<tr>
<td>16s #2 Off</td>
<td>CAGTTCCCAGGGCTTA</td>
<td>None</td>
<td>BHQ</td>
</tr>
<tr>
<td>16s #3 On</td>
<td>GTGGTAGGCCATTACCCCCC</td>
<td>CO560</td>
<td>BHQ</td>
</tr>
<tr>
<td>gyrB #1 On</td>
<td>AGGACGCGAAAGTCCTGTTGCT</td>
<td>BHQ</td>
<td>FAM</td>
</tr>
<tr>
<td>gyrB #1 Off</td>
<td>TGAACAAGGCT</td>
<td>BHQ</td>
<td>carbon-spacer</td>
</tr>
<tr>
<td>gyrB #2</td>
<td>CGTGTAATGAATAGCTGCG</td>
<td>FAM</td>
<td>BHQ</td>
</tr>
<tr>
<td>gyrB silent</td>
<td>CCACTGTGTTGAAGCACAACCCA</td>
<td>None</td>
<td>carbon-spacer</td>
</tr>
<tr>
<td>IC30-84Crev</td>
<td>GGTCAGACTGCACCTCCCAGTGCAGCTG</td>
<td>None</td>
<td>BHQ</td>
</tr>
<tr>
<td>IC30-84C-CO560</td>
<td>CAGCTGACTGGAAGGGTGACGTTGACC</td>
<td>CO560</td>
<td>carbon-spacer</td>
</tr>
</tbody>
</table>

BHQ= black-hole quencher, CO560 = Cal Orange. Probes were tested in VisualOmp at a 500 nm concentration with above modifications.

**LATE-PCR Diagnostic Assay**

The LATE-PCR assay was carried out to amplify and detect clinically relevant mycobacteria in sample. The duplex assay was developed to examine two genes, 16s and gyrB, among 19 mycobacterial isolates. Probe/target hybridization was carried out at endpoint, at temperatures 5 °C lower than the T_m of the limiting primer, as specified by Sanchez et. al (2004).

1. **Assay Composition**

Each PCR tube contained a final volume of 25 uL containing the following components: 1 uM 16s excess primer, 0.5 uM 16s limiting primer, 1 uM gyrB excess primer, 0.5 uM gyrB limiting primer, 1X PCR buffer, 0.4 mM dNTPs, 3.0 mM Mg^{2+},
200 nM of all fluorophore-modified probes, 500 nM of all non-fluorophore-modified probes, 25 nM of IC30-84C-rev, 75 nM of IC30-84C-C0560 and 1.5 units of Platinum Taq Polymerase.

### Table 3: Mycobacterial Isolates

<table>
<thead>
<tr>
<th>Type of Mycobacteria</th>
<th>Species</th>
<th>Number of Replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTBC</td>
<td><em>M. africanum</em></td>
<td>15</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. asiaticum</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. avium</em></td>
<td>3</td>
</tr>
<tr>
<td>MTBC</td>
<td><em>M. bovis</em></td>
<td>15</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. branderi</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. chelonae</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. flavescens</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. fortuitum</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. intracellulare</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. kansasii</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. marinum</em></td>
<td>3</td>
</tr>
<tr>
<td>MTBC</td>
<td><em>M. microti</em></td>
<td>15</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. phlei</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. rhodesiae</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. scrofulaceum</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. simiae</em></td>
<td>3</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. szulgai</em></td>
<td>3</td>
</tr>
<tr>
<td>MTBC</td>
<td><em>M. tuberculosis</em></td>
<td>15</td>
</tr>
<tr>
<td>NTM</td>
<td><em>M. xenopi</em></td>
<td>3</td>
</tr>
</tbody>
</table>

Samples were screened for contamination in the form of *M. tuberculosis*, and all replicates subsequently tested were pure.

Purified mycobacterial DNA was provided by Barry Kreiswirth and Natalia Kurpina of the Public Health Research Institute (PHRI) TB Center and PHRI Hospital Infections Program. They kindly supplied genomic DNA for 19 species of NTMs and MTBCs. Mycobacterial isolates contained unknown concentrations of purified genomic DNA. All of these materials were handled under BSL1 conditions in accord with Brandeis BioSafety Protocol #09013.
uL of mycobacterial sample were added to each reaction. Species samples are shown in Table 3, with the number of PCR replicates generated in the rightmost column.

Some replicates were not used in the final data set due to sample evaporation. Sample evaporation occurs when the lid of the tube is not completely closed, changing the volume and thermodynamics of the reaction. Conventional PCR data is acquired in real time, and does not detect such an artifact. Conversely, LATE-PCR is analyzed at endpoint, and sample evaporation results in a horizontal transformation of the fluorescence signature (Rice, Personal Communication).

3. **Thermocycling Conditions Using Stratagene MX3005P Sequence Detector**

LATE-PCR was carried out on mycobacterial genomic isolates in a Stratagene MX3005P Detector (Stratagene, La Jolla, CA), according to the following thermal profile: 1 cycle at 95°C for 3 minutes, 70 cycles of 10 seconds at 98°C, 40 seconds at 75°C, 1 10-minute hold at 75°C, 1 10-minute hold at 25°C, and a final melt (increasing the temperature) beginning at 25°C, at a rate of 1°C per 33 seconds for 72 cycles. Product formation was confirmed by amplification data, and product content was examined by quantitative analysis.

4. **Assessment of Variation between Assays**

With the generation of fluorescence data, there was a necessity to assess the variance of a single sample’s fluorescence, given the random noise generated by the PCR machine. For a select number of mycobacterial samples, three
different PCR experiments were used to generate replicate fluorescence data for a single species. More PCR replicates were generated for members of the MTBC to assess variance within a species signature between different PCR reactions, and mixtures.

**Quantitative Data Analysis**

A LATE-PCR diagnostic assay was constructed to produce distinct fluorescence patterns for the majority of mycobacterial species tested. The goal of quantitative analysis was to match an unknown fluorescence pattern to a known set of fluorescence standards, thus enabling mycobacterial species identification. Data analysis methods were created and performed in Matlab, The MathWorks Inc.

1. **Data Acquisition**

   In order to simplify the structure of raw fluorescence data, given the two types of fluorescence for each temperature point, a reading function was developed. Stratagene PCR machine output is an Excel table of raw fluorescence columns alongside temperature columns. The reading function integrated the sample name, temperature vector, and corresponding fluorescence vectors into a single standard data object. Samples were integrated into a final set of species standards, containing fluorescence data for each species.

2. **Data Normalization**
Converting a raw fluorescence signal to a melt curve requires data normalizations. In the context of species identification, the curvature of the raw fluorescence signal captures the essence of probe kinetics, and should be revealed in a melt curve. Palais et. al (2009) define the raw fluorescence signal of a DNA sample as an exponential function, with a minimum background fluorescence signal at the lowest temperature and a maximum fluorescence signal at the highest temperature. The raw fluorescence signal is defined as \( R(t, f) \) where the fluorescence exists for a certain number of temperatures, or \( t \), and in a specified number of fluorescent colors, \( f \). Because we are interested in the shape of the underlying curve rather than the overall intensity or abundance of DNA or label, we normalize the raw fluorescence signal such that the normalized fluorescence is: 

\[
\hat{R}(t, f) = \frac{R(t, f) - R(t_{\text{min}}, f)}{R(t_{\text{max}}, f)},
\]

where the minimum fluorescence at \( t_{\text{min}} \) is subtracted from the signal for each color, and the signal is normalized by making the fluorescence at \( t_{\text{max}} \) equal to 1. Assuming that the peaks in fluorescence signal are exponential in nature, the derivative of the signal retains that quality, while random noise and intrinsic background fluorescence is excised by the differentiation of the fluorescence signal. Using discrete differentiation, the second derivative of the normalized fluorescence is: 

\[
\hat{R}''(t, f) = \hat{R}(t+1, f) + \hat{R}(t-1, f) - 2\hat{R}(t, f),
\]

given that the temperature step, \( \Delta t \), is constant. We denote this normalized and twice differentiated curve as the *melt curve* \( M(t, f) \); that is, \( M(t, f) = \hat{R}''(t, f) \). Transforming a raw fluorescence signal into this melt curve signal helps to
highlight the features of a fluorescence pattern that provide the most information in making genotypic discriminations.

3. **Data Visualization**

The melt curves obtained from PCR lie in a high dimensional space, and it is difficult to visualize the similarities and differences among different species on a low dimensional graph. In order to visualize these curves, we reduce the spatial dimensions to those few dimensions that truly contain the majority of variance between genotypes. To do this, principal component analysis (PCA) is performed by eigenvector analysis of the data covariance matrix. Within each color of fluorescence f, the mean for each temperature dimension t is subtracted from each replicate, such that

$$\Delta_i(t, f) = M_{i,j}(t, f) - \overline{M}(t, f).$$

Following mean subtraction, the data matrix D is the concatenation of all deviations from the mean, for each replicate j of each standard i:

$$D = [\Delta_1 \Delta_2 \cdots \Delta_m],$$

where D is an N x M matrix, with N representing the number of temperature dimensions, and M representing the number of total replicates being analyzed. The covariance matrix C examines the relationship between the N dimensions among each of the data objects, such that

$$C = \frac{1}{M} \sum_{i=1}^{N} \Delta_i \Delta_i^T = DD^T,$$

where C is N x N. In order to create the new dimensions upon which data will be visualized, eigenvectors $v_1 v_2 \cdots v_n$ and eigenvalues $\lambda_1 \lambda_2 \cdots \lambda_n$ are extracted. The eigenvectors with the maximum eigenvalues were used to create the basis for the reduced-dimension data.
space. Depending on the number of vectors chosen for the basis of the coordinate plane, the original data is projected using those eigenvectors to generate the M N-dimensional curves in PCA space. PCA visualizes the data in a low-dimensional space, and characterizes the intrinsic variance within the data.

4. Data Similarity

Given a set of standard melt curves representing mycobacterial species and an unknown sample with an obtained melt curve, a metric of similarity must be computed between the standards and the unknown curve. Given a number of fluorescence colors \( f > 1 \) read from a single species sample, channel fluorescence vectors were concatenated into a single fluorescence vector, with the vector containing all colors of fluorescence data. This fluorescent vector will be \( f \) times the length single channel fluorescent vector. Euclidean distance measures are used to compute the most likely match for the unknown sample. In the set of species standards, there exist \( i \) standards, each with \( j \) replicate melt curves, \( M_{i,j}(t,f) \). In order to compare the species melt curves to the unknown melt curve that we will call \( M_{unknown}(t,f) \), the replicates must be averaged into an \( n \)-dimensional mean:

\[
\overline{M}_i(t,f) = \frac{1}{N_j} \sum_{j=1}^{N_j} M_{i,j}(t,f),
\]

where each standard fluorescence curve is the average of its replicates. Finally, the Euclidean distance is computed for each standard, such that:
\[ \text{Dist}(\overline{M}_i(t, f), M_{\text{unknown}}(t, f)) = \left[ \frac{1}{N_f N_t} \sum_{t=1}^{t_{\text{min}}} \sum_{f=1}^{N_f} (\overline{M}_i(t, f) - M_{\text{unknown}}(t, f))^2 \right]^{1/2}, \]

where \(N_f\) is the number of fluorescent colors and \(N_t\) is the number of temperature dimensions with corresponding fluorescence values. The standard \(\overline{M}_i\) that most precisely matches \(M_{\text{test}}\) is the one that minimizes the Euclidean distance between the two fluorescence curves.

5. **Detection Analysis**

In order to examine the capacity of the diagnostic PCR assay to detect mycobacterial species with accuracy, a set of fluorescence curves are generated to test against the set of mycobacterial fluorescence standards. A constant number of artificial replicates were created for each species, by augmenting the specific mean fluorescence vector with random noise as a factor of the covariance within replicates of a single species. For a species melt curve, \(\overline{M}_i(t, f)\) an artificial replicate, \(M_a(t, f)\) was created such that

\[ M_a(t, f) = \overline{M}_i(t, f) + f(a)C, \]

where \(C\) is the covariance of all fluorescence replicates. \(f(x) = ce^{-x^2/2}\) is the probability density function for the normal distribution, and \(a\) is a randomly selected number; the value of \(f(x)\) ranges from 0 to 1. For each species, artificial fluorescence curves were produced and compared to the standard set of fluorescence curves to quantify of diagnostic strength.
Chapter 3: Results

LATE-PCR Gene Product Primer and Probe Design

LATE-PCR and Lights-On/Lights-Off probe technologies were applied to a species identification assay for mycobacteria. The assay was constructed to identify the most pathogenic mycobacteria in two fluorescence colors, using two genomic targets.

1. **16s**

   A gene fragment of the 16s rRNA hypervariable region A(nt.100-308) was amplified with one pair of primers and the intervening hypervariable sequence was targeted with three sets of probes in two fluorescent colors which work together to separate and identify pathogenic non-tuberculous mycobacteria from the uniform MTBC sequence.
Two sets of Lights-On/Lights-Off probes, each in a different fluorescence channel, were used to distinguish the 16s gene region among mycobacterial species, as shown in Figure 1. Two colors of fluorescence probes were used in probe design, so that the quality of all probe-target hybridization signals would be preserved in two dimensions of fluorescence data. 16s #3 On probe is a molecular beacon rather than a Lights-On probe because it is not paired with a Lights-Off probe. FAM channel Lights-On probes in the 16s gene are most complementary to the MTBC sequence while the Cal Orange 560 channel probes are most complementary to sequences of some pathogenic NTMs.

The 16s #1 On/Off probes comprise a set of Lights-On/Lights-Off probes, with mismatches lying under both probes in the FAM channel. The
#1 On probe hybridizes to the gene target of MTBC species at the highest melting temperature ($T_m$) of all mycobacteria tested. Thus, a high temperature signal from this probe demonstrates that the sample contains one of the MTBC species rather than an NTM species. The 16s #1 Off probe covers a region of the gene target with multiple polymorphisms in multiple species. This Off probe was predicted by Visual OMP to have a range of melting temperatures for mycobacterial species, as shown in Table 4.1.

Certain NTM species, including *M. avium*, *M. intracellulare*, and *M. fortuitum* are pathogenic and clinically important. Probes #2 and #3 probes were deliberately designed to detect these species, while additionally characterizing other NTMs potentially in sample. 16s #2 On probe is most complementary to the target variant of *M. avium*, with a $T_m$ of 72.7°C, and 16s #3 On probe is most complementary to the target variant of *M. fortuitum*, with a $T_m$ of 60.3°C. The 16s Off probe #2 is complementary to a conserved target sequence found in the tested NTMs. Yet, because it is an Off probe, it is not observed unless On probe #2 is bound.

2. **gyrB**

Because all members of the MTBC have an identical 16s rRNA sequence, the gyrB gene target was chosen to uniquely identify members within the complex. The gyrB gene was examined by targeting SNPs that have allelic variants among the complex. The gyrB probe set consisted of one set of Lights-On/Lights-Off probes, a molecular beacon, and an unmodified oligonucleotide. Sequences under all probes and beacons on the gyrB are
identical for *M. africanum* and *M. microti*. Currently available genomic sequences for various species of mycobacteria suggests that the *gyrB* gene primers used here will only amplify the target sequence of the MTBC species. In agreement with this prediction none of the 15 NTM species tested in this thesis generated an amplicon with these primers.

The high-GC content of the mycobacterial genome allows for formation of a thermally stable hairpin structure encompassing nucleotides 1295-1320 in the *gyrB* amplicon. Formation of this secondary structure at 58.3°C sterically-hinders hybridization of Lights-On/Lights-Off probe pair for *gyrB*. For this reason an unlabeled oligonucleotide with a predicted Tm of 65.5°C was added to the reaction to bind to and prevent hairpin formation. This oligonucleotide is unlabeled and therefore serves as a “silent” probe. This oligonucleotide has the highest Tm of all of the *gyrB* probes, for all MTBC species, Table 4.2.

Ideally the *gyrB* probes would uniquely identify each of the MTBC species. In Figure 2, the sequences of four members of the MTBC are shown, with each fluorescent probe located on a SNP. *GyrB #1 On probe* complemented the *M. tuberculosis* sequence completely, while mismatching *M. bovis, M. africanum, and M. microti* by one nucleotide. *gyrB #1 Off probe* bound all species at the same Tm. *gyrB #2 On probe* was uniformly complementary to all species of the complex except *M. bovis*, which had a single-nucleotide change in sequence. Probe Tm's predicted that the single
nucleotide mismatches under both probes would produce species-specific fluorescence signals.

Thus, the two probe sets served separate purposes in individually identifying mycobacteria in a PCR reaction. The 16s probe set added a diversity of probe-target melting temperatures for numerous non-tuberculous mycobacteria. The gyrB probe set contributed probe-target hybrids, that would produce discriminatory signals, for members of the *Mycobacterium tuberculosis* complex. Together, these probe sets created a wealth of probe-target hybrids that would produce unique fluorescence profiles for a number of clinically relevant mycobacteria.

**Computation of Mycobacterial Target-Probe Melting Temperatures**

In total, 8 probes were designed to detect and differentiate the 16s and gyrB gene targets using fluorescence increase and decreases in two colors. Table 4.2
lists the Visual OMP predicted $T_m$ of each probe to its target sequence. In some cases, probe-target hybridization results in a positive increase in the fluorescent signal over background. This is the case because the rigid probe-target double-strand separates the fluorophore and the quencher of the On probe relative to the proximity of these two moieties on probe molecules in solution, producing positive fluorescence. However, probe hybridization also results in negative fluorescence, a decrease in signal relative to the background level of fluorescence. This decrease in signal occurs when an Off probe binds to the target at a high $T_m$ of its adjacent On probe and therefore quenches the background signal of the On probe as soon as it hybridizes to the target.

Certain probes were created to hybridize all mycobacterial templates at a similar $T_m$. For example 16s #2 Off probe, read in the CalOrange channel, had a uniform $T_m$ for all mycobacterial species and quenches the the fluorescence of any 16s #2 On probe regardless of species. gyrB #1 Off probe, in the FAM channel, also has the same $T_m$ for all MTBC species and detectable. All probes were built to compound the fluorescence signal.
### Table 4.1: Predicted Probe Melting Temperatures: 16s Templates

<table>
<thead>
<tr>
<th>Species</th>
<th>16s #1 On (°C)</th>
<th>16s #1 Off (°C)</th>
<th>16s #2 On (°C)</th>
<th>16s #2 Off (°C)</th>
<th>16s #3 Beacon (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. asiaticum</em></td>
<td>64.4</td>
<td>54.1</td>
<td>62.9</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td><em>M. avium</em></td>
<td>64.4</td>
<td>54</td>
<td>72.7</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td><em>M. branderi</em></td>
<td>64.4</td>
<td>36.2</td>
<td>67.5</td>
<td>41.2</td>
<td>55.6</td>
</tr>
<tr>
<td><em>M. cheloneae</em></td>
<td>67.3</td>
<td>14.7</td>
<td>54.3</td>
<td>51.2</td>
<td>48.9</td>
</tr>
<tr>
<td><em>M. flavescens</em></td>
<td>64.4</td>
<td>7.8</td>
<td>61.7</td>
<td>51.2</td>
<td>-4.5</td>
</tr>
<tr>
<td><em>M. fortuitum</em></td>
<td>64.4</td>
<td>31.1</td>
<td>54.3</td>
<td>51.2</td>
<td>60.3</td>
</tr>
<tr>
<td><em>M. intracellularare</em></td>
<td>64.4</td>
<td>56.1</td>
<td>72.7</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td><em>M. kansasii</em></td>
<td>64.4</td>
<td>54.1</td>
<td>62.9</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td><em>M. marinum</em></td>
<td>64.4</td>
<td>65.2</td>
<td>67.5</td>
<td>51.2</td>
<td>55.6</td>
</tr>
<tr>
<td><em>M. phlei</em></td>
<td>64.4</td>
<td>14.7</td>
<td>54.3</td>
<td>51.2</td>
<td>55.6</td>
</tr>
<tr>
<td><em>M. rhodesiae</em></td>
<td>64.4</td>
<td>39.6</td>
<td>61.7</td>
<td>51.2</td>
<td>60.3</td>
</tr>
<tr>
<td><em>M. scrofulaceum</em></td>
<td>55.4</td>
<td>19</td>
<td>63.7</td>
<td>51.2</td>
<td>55.6</td>
</tr>
<tr>
<td><em>M. simiae</em></td>
<td>64.4</td>
<td>54.1</td>
<td>68.3</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td><em>M. szulgai</em></td>
<td>64.4</td>
<td>58.5</td>
<td>68.3</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td>MTBC</td>
<td>72.3</td>
<td>65.2</td>
<td>67.5</td>
<td>51.2</td>
<td>45.8</td>
</tr>
<tr>
<td><em>M. xenopi</em></td>
<td>65.5</td>
<td>31.1</td>
<td>66.2</td>
<td>51.2</td>
<td>55.6</td>
</tr>
</tbody>
</table>

### Table 4.2: Predicted Probe Melting Temperatures: gyrB Templates

<table>
<thead>
<tr>
<th>Species</th>
<th>gyrB #1 On (°C)</th>
<th>gyrB #1 Off (°C)</th>
<th>gyrB #2 Beacon (°C)</th>
<th>gyrB silent (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. africanum, M. microti</em></td>
<td>50.5</td>
<td>34.4</td>
<td>43.4</td>
<td>65.9</td>
</tr>
<tr>
<td><em>M. bovis</em></td>
<td>50.5</td>
<td>34.4</td>
<td>26.5</td>
<td>65.9</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>60.6</td>
<td>35.8</td>
<td>44.7</td>
<td>66.7</td>
</tr>
</tbody>
</table>

Probe-target hybrid melting temperatures were calculated in VisualOmp, at a reaction temperature of 70°C. To calculate the melting temperature, conditions were specified as: template concentration at 1 uM, probe concentration at 500nM.

### Melting of Mycobacterial Templates with Lights-On/Lights-Off Probes

A LATE-PCR assay was run with DNA from 19 mycobacterial species, with all reagent concentrations identical in each sample. For all NTM species, samples were run in triplicate. For members of the MTBC, 5 replicates were run for each species in three separate reactions, each with a separate reaction mixture. To detect product
by fluorescent probe activity, the PCR product was melted from 25°C to 95°C.

Multiple reactions were performed to assess run-to-run variation in members of the MTBC. Reaction data was available at the end of a 3 hour PCR run for analysis and classification.

**Isolating A Melt Curve from a Raw Fluorescence Signal**

Raw fluorescence signals are a composite of probe dynamics, background signal, and machine noise. Assumptions about probe thermodynamics, similar to those used for HRMA [45], were applied to mycobacterial raw fluorescence data, in a series of data manipulations. In Figure 3, graphs a-f, reflect the series of transformations performed on raw fluorescence data to obtain a corresponding melt curve. The temperature vector of each replicate from run-to-run varied minutely due to machine noise. In (b), the temperature vectors are interpolated onto an integer-based temperature vector, reflecting the prescribed temperature step in the PCR melting of reaction product. In order to normalize for background signal, two properties of DNA-melting were considered; at the lowest temperature, the fluorescence should be at a minimum; at the highest temperature, the fluorescence should be at a maximum. Background fluorescence was subtracted from the fluorescence vector in (b), and magnitude differences were normalized in (c) to obtain(d). Lastly, the fluorescence vectors are differentiated to obtain the first derivative and second derivative in (e) and (f), reducing the number of points by one point each differentiation. The second derivative of the melt curve was then used to quantitatively classify mycobacterial species.
Figure 3 - Data manipulations were performed in MatLab. The normalization scheme shown for comparing species fluorescence pattern is shown with two single replicates from *M. tuberculosis* and *M. bovis*, using only the FAM channel. In (a), the raw fluorescence is plotted over temperature. In order to equalize temperature scales that are experiment specific, (b) shows the interpolation of temperature scales onto a uniform integer temperature scale. In (c), the lowest fluorescence value is subtracted from all fluorescence values. In (d), all fluorescence values are divided by the highest fluorescence value. In (e) and (f), the y-axis reflects differentiated units of fluorescence, with (e) showing the first derivative of fluorescence and (f) showing the second derivative of fluorescence.
First Derivative Mycobacterial Melt Curves: Observed Probe Dynamics

While FAM channel fluorescence produced different fluorescent curves for all members of the MTBC, the CalOrange channel fluorescence was the same for species within the *Mycobacterium tuberculosis* complex. In Figure 4, members of the MTBC display the greatest magnitude of fluorescence change in the FAM channel. gyrB #1 On probe produces an increase in fluorescence at 55°C for an *M. tuberculosis* sample while producing an increase in fluorescence at 50°C for *M. africanum*, *M. bovis*, and *M. microti* samples. gyrB #2 On probe produces an increase in fluorescence at 35°C for *M. bovis*, but not for the other members of the complex. Yet, fluorescence curves for *M. microti* and *M. africanum* have the same observed probe Tₘₛ in both fluorescence channels.

Non-tuberculous mycobacteria displayed a range of fluorescence profiles in the CO560 channel. Specifically, a number of pathogenic species, such as *M. kansasii*, *M. szulgai*, and *M. xenopi* bound both fluorescence probes, resulting in two maxima in the first derivative curves. With 16s #2 probe complementary to *M. avium*, a large maximum was observed in the first derivative. The same was observed in the case of #3 probe and *M. fortuitum*. Additionally, *M. flavescens* fluorescence displayed binding of 16s #3 probe, while VisualOmp predicted a lack of binding within the PCR melt temperature range. There was a general discrepancy between predicted Tₘₛ and observed Tₘₛ.

In comparison to predicted Tₘₛ, observed probe Tₘₛ differed significantly. VisualOmp predicted Tₘₛ using template/probe sequences, template/probe concentrations, and assay temperature. In the LATE-PCR mycobacterial assay,
multiple primers and probes were present in addition to the template, which was in unknown concentration. Further, VisualOmp did not accurately predict probe T\textsubscript{m}s for targets with more than one mismatched nucleotide, in the case of \textit{M. flavescens}, \textit{M. asiaticum}, \textit{M. kansasii} and others. Regardless of prediction, fluorescent probes in a large temperature range generated a diversity of species fluorescent profiles.
First Derivative Melt Curves: Mycobacterial Isolates
Second Derivative Mycobacterial Melt Curves

While the first derivative creates maximums and minimums in fluorescence, the second derivative clarifies inflection points, maximums, and minimums in fluorescence. In Figure 5, a single maximum was observed in the FAM channel curves of species such as *M. asiaticum*, *M. avium*, *M. intracellulare*, *M. kansasii*, and *M. marinum*. In the second derivative, the location of this peak was an inflection point with a maximum and minimum on either side of the point. In a very different shape, even the flat-line fluorescence profile of *M. phlei* in the first derivative FAM channel changed in the second derivative. Upon differentiation, this curve tightly oscillates about zero from low to high temperature. The second derivative fluorescence accentuates many features of the first derivative fluorescence.

Some magnitude differences in the first derivative resulted in temperature shifts in the second derivative. In the first derivative CO560 channel, *M. avium* and *M. intracellulare* displayed a single maximum with no significant difference between
the location of the peak, but differences between the magnitude of the fluorescence peaks. The second derivative showed a maximum and minimum for both species, but at different temperature points for each species. The second derivative adds discriminating information to the fluorescence profile and to the temperature phase of each fluorescence profile.

Among *M. tuberculosis* replicates, the second derivative characterized a difference in the magnitude of the first derivative minimum in the 50-60°C range. Raw fluorescence curves generated for *M. tuberculosis* represented three different reaction mixtures from 3 separate reactions. While two out of the three sets of *M. tuberculosis* fluorescence curves had 5 uniform replicates, one set had two groups of fluorescence curves, that had different heights from 80-90°C.
Second Derivative Melt Curves: Mycobacterial Isolates

M. tuberculosis

M. africanum

M. microti

Temperature
M. marinum

M. phlei

M. rhodesiae

Temperature
High-Dimensional Visualization of LATE-PCR Fluorescence Data Using PCA

For each channel, second derivative fluorescence vectors were visualized in a high-dimensional space defined by three axes, each representing a direction of maximal variance. Principal component analysis was performed on each channel's fluorescence data for all species replicates. Figure 5 displays the principal component analysis of all mycobacterial fluorescent profiles generated in the FAM channel (a), and the CO560 channel (b). Each replicate melt curve corresponds to a species-specific symbol in the PCA space.

In the FAM channel, members of the MTBC are separated from NTM clusters. One group is composed of *M. tuberculosis* replicates, separate from both the other members of the MTBC and NTM. *M. bovis* and *M. africanum*/*M. microti* are also separate from NTM clusters. Upon rotating the 3-D graph, the FAM channel fluorescence definitively separates the grouping of *M. africanum/M. microti* and *M. xenopi*. 

Figure 5: Orange=CO560 Channel fluorescence; Green=FAM Channel Fluorescence; Black= *M. tuberculosis* reference curve in corresponding channel. Graphs with a gray background are fluorescent curves for species of the *M. tuberculosis* complex. The PCR reaction contained mycobacterial template and the full probe set with 16s and gyrB components. Data was imported from the Stratagene PCR machine to MatLab. Curves were generated in MatLab, after performing the normalization scheme up to the second derivative. Each color curve represents a different well in the PCR machine, with the corresponding species' DNA.
bovis. The group on the right-hand side contains all the NTMs, with tight clustering among species, but little distance between species clusters.

In the CO560 channel, NTM replicates were clustered into species-specific areas of the PCA space. *M. szulgai, M. intracellulare, and M. xenopi* replicates are clustered into their respective groups. Other species have one aberrant replicate with other tightly clustered replicates, such as *M. fortuitum* and *M. avium*. At the center of the CO560 PCA, near the origin, members of the MTBC, *M. simiae*, and *M. branderi* are clustered in a larger group, reflecting the low variation between the fluorescent profiles of those species. The variance in each channel was examined by PCA visualization, but composite fluorescence profiles were used in the classification algorithm.
Simulated Diagnostic Testing

Standard mycobacterial fluorescent profiles were generated in a LATE-PCR Light-On/Lights-Off detection assay. To create a test population of fluorescent
profiles that might appear in sample, fluorescent vectors in each channel were concatenated into one fluorescent vector for each replicate. For each species, noise, proportional to the covariance of the dataset, was injected into the mean fluorescence curve to create an artificial replicate. One thousand replicates were created for each standard species fluorescence profile. Figure 7 depicts the relationship of artificial fluorescence curves to the true mycobacterial fluorescence curves, using artificial *M. tuberculosis* replicates. Rather than performing principle component analysis on each channel of fluorescence data, the fluorescence from each channel was combined into a single composite fluorescence vector for each replicate. This principle component analysis displays the maximum variation present in the composite vectors, and more accurately represents the classification of mycobacterial species. Upon rotation of the three-dimensional space, all 100 replicates displayed are tightly clustered together, overlapping slightly with true *M. bovis* replicates and true *M. africanum* replicates.
Quantitative Classification of Mycobacterial Fluorescence

Artificial replicates were classified by iteratively computing the Euclidean distance between an artificial curve and each standard mycobacteria curve. The mycobacterial species that the test curve was classified as was the species curve closest in distance to the test curve. In order to ascertain the probability of classification, a meter of diagnostic strength, mycobacteria fluorescence classification results were analyzed for true positive, false negative, and false positive classification. Table 5 summarizes the results of the classification test performed between one thousand artificially generated replicates and the standard mycobacterial replicates depicted in Figures 4 and 5.

In classifying MTBC species, the classifier successfully classified the majority of artificial MTBC samples as true positives. *M. tuberculosis* and *M. africanum*/*M.
microti had the highest probability of true positive classification. M. bovis had a false negative probability of less than 1%, and M. africanum had a false positive probability of less than 1%.

Classification of NTM species produced nearly perfect true positive classification probabilities, with trivial false classification probabilities. M. avium and M. intracellulare were the only NTM species incorrectly classified. Upon examination of the classification results, they were misclassified as each other. This method of classification used all dimensions of fluorescence to classify the artificial mycobacterial replicates to the standard mycobacterial profiles, produced in the LATE-PCR fluorescence detection assay.

<table>
<thead>
<tr>
<th>Table 5: Probability of Species Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>%</th>
<th>0.0</th>
<th>&lt;0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. africanum/M. microti</td>
<td>100.0</td>
<td>0.0</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>M. bovis</td>
<td>99.9</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>M. tuberculosis</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. asiaticum</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. avium</td>
<td>99.9</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>M. branderi</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. cheloneae</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. flavescens</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. fortuitum</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. intracellulare</td>
<td>100.0</td>
<td>0.0</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>M. kansasii</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. marinum</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. phlei</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. rhodesiae</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. scrofulaceum</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. simiae</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. szulgai</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>M. xenopi</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The probability of a false negative classification is defined as the probability that a sample with species X is not classified as species X. The probability of a false positive classification is defined as the probability that a sample that is not species X is classified as species X.
Chapter 4: Discussion

LATE-PCR Lights-On/Lights-Off Detection Assay

LATE-PCR and Lights-On/Lights-Off technologies generated species-specific fluorescence profiles in a mycobacterial diagnostic assay. LATE-PCR used unequal concentration of primers, with deliberately different melting temperatures, to amplify single-stranded gene targets. After amplification, these gene products were analyzed at end-point over a wide temperature range, using Lights-On/Lights-Off probes. The probe technology created both positive and negative fluorescence signals that yielded complex PCR fluorescence data. Quantitative analysis and discrimination of species fluorescence data was accomplished through a range of mathematical techniques.

The data shown in Figures 4 and 5 are the resulting fluorescent profiles from testing the 19 mycobacterial samples. Future expansion of this mycobacterial assay could include testing many more species of mycobacteria. In addition, given the clinical application of the assay, its sensitivity to detect product should be tested in the presence of human bodily fluid (such as blood or sputum).

Construction of Target Probe Sets

Lights-On and Lights-Off probes signals generated composite fluorescence data characterized by positive and negative fluorescence contributions to the signal.
Two colors of fluorescence probes were employed to add resolving power to fluorescence data for two gene targets. 16s and gyrB gene targets contained polymorphic sequences, for members of the mycobacteria genus, that were coated with probes to produce variant fluorescence signatures.

1. **16s**

   Hypervariable region A of the 16s rRNA gene was chosen due to its variant nature among mycobacteria, and probes were designed to target a host of polymorphic sequences. Primers for the 16s target were designed to amplify all mycobacterial species, in an effort to uniquely identify a range of pathogenic mycobacteria. The FAM probe set of the 16s gene preferentially hybridized to members of the MTBC, while hybridizing to NTM species at a lower melting temperature. The CO560 probe set was designed to produce diversity of melting temperatures for mycobacteria-probe hybrids. Target sequences were those of clinically relevant non-tuberculous mycobacteria. Probes complemented these sequences to generate distinct fluorescence signals for these species. Probe sets were designed using the sequences of 20 mycobacterial species. If the number of species being examined is expanded, it may prove necessary to either redesign the probes used here or add additional probes.
The 16s rRNA gene, and specifically the hypervariable region, contains a diversity of sequences among members of the mycobacterial genus. LATE-PCR and Light-On/Lights-Off probe technologies have the capacity to produce a catalog of variant fluorescence signatures for all species of a genus. It has been shown that this technology is able to generate species barcodes (L. Rice unpublished), and the 16s rRNA gene contains enough genetic variation to be distinguished in fluorescence profiling. A future assay might focus solely on probing the hypervariable region to identify mycobacterial species outside the *Mycobacterium tuberculosis* complex, given the sensitivity of the LATE-PCR Lights-On/Lights-Off technology in examining genomic products.

2. **gyrB**

   The *gyrB* gene is essential in distinguishing between prevalent members of the MTBC. In a completely different probe design approach, the *gyrB* target region and probes were selected to produce fluorescence data that would be distinctive for each member of the MTBC. Due to the single nucleotide changes detected under *gyrB* probes, predicted $T_m$s agreed with observed $T_m$s. *gyrB* #1 On probe hybridizes *M. tuberculosis* at a higher $T_m$ than it does *M. bovis*, *M. africanum*, or *M. microti*, producing a different signal for 1 out of 4 targeted members of the complex. The oligonucleotide was used to prevent a high temperature hairpin on the template, allowing low $T_m$ probe fluorescence. This feature of the probe set created a distinct low
temperature peak unique to *M. bovis*. With this probe design, each potential
human pathogen has been separately identified based on melting
temperature combinations.

While *M. africanum* and *M. microti* are indistinguishable in target
sequences, these species are genotypically and phenotypically distinct. In the
context of diagnostic detection, *M. africanum* is much more likely to present
in patients, and also has geographic specificity to West Africa. *M. microti*, on
the other hand, rarely presents in humans. These differences decrease the
need for differentiating between the two species in a diagnostic setting.

**LATE-PCR Fluorescence Data**

Two types of melt curves were analyzed during the process of constructing
the mycobacterial diagnostic assay. The first derivative served to highlight the
melting temperature variants among the species in each channel, and further
differentiation to the second derivative removes any machine noise left in the
fluorescence signal. The second derivative fluorescence vectors were used to
computationally visualize and objectively compare data in a species classification
assay.

Initially, a fluorescent probe set was created that would differentiate
between the members of the MTBC. gyrB #1 On probe produces fluorescence
increase, with the location of this peak dependent on the member of the pathogenic
complex. gyrB #2 On probe differentiates the remaining members of the complex
that had not been distinguished with the first On probe. As a relic of initial
strategies, the 16s #1 Lights-On/Lights Off pair could be removed, as it does not add to any further discrimination of the MTBC.

The second set of C0560 probes provided a means for NTM classification. Some important NTM pathogens, as discussed, are: members of the *Mycobacterium avium* complex (*M. avium* and *M. intracellulare*), *M. szulgai*, *M. marinum*, *M. fortuitum*, *M. kansasii*, and *M. xenopi*. By using the second derivative, the two members of MAC have a maximum and minimum shifted from one another, thus are significantly different curves. *M. szulgai*, *M. kansasii*, *M. marinum*, and *M. xenopi* all have two maxima in the first derivative with magnitude differences contributing to significantly different second-derivative fluorescence profiles from one another. *M. fortuitum* has a unique fluorescence profile in the second derivative, where as in the first derivative, it resembles *M. rhodesiae*. Many pathogenic NTMs have unique second derivative profiles, that could result in successful identification in a clinical setting.

Yet, in solution and in the fluorescence profiles, the combination of the two colors on one target (the 16s) may decrease the amount of information and classification that can be derived from fluorescent PCR data. It would make logical sense to separate the probe signals by target. Thus, 16s probes would have one color, and gyrB probes would have a second color. This assay design would divide the discrimination of NTMs from MTBC, with the 16s signal used to differentiate NTM species from one another, and the gyrB signal used to differentiate MTBC species. Both mycobacterial gene products could be detected in a single tube, in two colors.
Automation of Data Acquisition

In a clinical context, to identify pathogenic mycobacterial species in a human sample, diagnosis and data handling should be rapid and objective. In order to detect mycobacterial species using the LATE-PCR Lights-On/Lights-Off technology, data must be handled and objectively interpreted to classify species based on fluorescence profiles. To transfer the data from the output of the PCR machine to a graphical display, where the normalization scheme could be carried out, data must be imported into a practical data infrastructure. Before the automation in MatLab, this procedure was laborious and quantitative analysis was sometimes impossible.

Using MatLab to consolidate information, mycobacterial fluorescence data was analyzed, manipulated and classified with ease. For each replicate, a temperature vector and all channel fluorescence vectors were grouped with the nominal features of the sample into a standard data object. Replicates of the same species were added to one species standard object. This grouping allowed the simultaneous access to species name, reaction data, replicate number, and the matrix of fluorescence-temperature vectors. Fluorescence data could be normalized, differentiated, visualized and classified in an automated procession. This automated algorithm of data handling allows the species classification to be accessible to point-of-care devices. Future programming efforts will produce a graphical user interface that allows the user to specify the diagnostic context and the standards used for classification.
Visualization of Fluorescence Profiles

Principal component analysis of second derivative mycobacterial fluorescence profiles demonstrates two different scenarios for classification potential in the separate channels, but does not reflect the complete variety of species signatures. In PCA space, the FAM channel fluorescence distinguishes members of the MTBC from the rest of mycobacterial isolates tested. In a different space, the CO560 channel fluorescence creates variation that separates NTMs more distinctly than members of the MTBC.

Dimensionality reduction of the CO560 fluorescence curves confirms separation between many NTM species. PCA demonstrates that pathogenic species such as *M. xenopi, M. intracellulare, M. szulgai, M. kansasii* replicates segregate into their respective species groupings. The members of the MTBC are defined and distinguished in the FAM channel PCA.

Principal component analysis of FAM fluorescence successfully distinguishes members of the *Mycobacterium tuberculosis* complex from other mycobacteria tested. The visualization separates the *M. tuberculosis* cluster from the clusters of *M. africanum, M. bovis, and M. microti*. Given the lack of need for discrimination between *M. africanum* and *M. microti*, the fluorescence assay constructed separately identifies all members of the MTBC that could cause human disease, as confirmed in the visualization of the FAM channel PCA data.

Assessment of Fluorescent Data Variation between PCR Reactions
Run-to-run variation within the *Mycobacterium tuberculosis* complex was assessed, and *M. tuberculosis* showed the greatest variation between runs. Each member of the complex was tested in the LATE-PCR fluorescence assay in three different reactions with 5 replicates each reaction. Thus, a total of 15 fluorescent signatures were procured for each member of the complex.

The difference in the two types of *M. tuberculosis* profiles observed in the second derivative reflected results from different reaction mixtures. Rotation of the FAM PCA space also reveals two distinct populations of *M. tuberculosis* curves. Within one set of *M. tuberculosis* raw fluorescence curves, there was a difference in amplitudes that was not reconciled by normalization methods. These *M. tuberculosis* fluorescence discrepancies indicate a nonhomogenous reaction mixture. With this conclusion, the aberrant *M. tuberculosis* profiles were removed from mathematical analysis and classification. In order to completely assess the amount of variation possible in this diagnostic assay, many more mycobacterial replicates should be generated for analysis. Clearly, an ideal diagnostic assay would have minimal run-to-run variation between the same species' fluorescence data.

Even in 3-D PCA space, the variance between replicates is not completely presented. For each temperature dimension, there is a principal component with a certain direction of fluorescence variation. Figures 7a,b only show the first three directions of variance, leaving out certain discriminating features of the fluorescence. For example, *M. branderi* and *M. simiae* cluster at the origin with the members of the MTBC in the C0560 channel. The fluorescent profiles of these
species in the CalOrange channel are different from one another, but this variance is not captured in the chosen principal components.

**Objective Classification Using Euclidean Distance**

The final piece of the diagnostic assay was developing a method that would objectively classify the unknown PCR fluorescent curve as the most likely mycobacterial out of a set of known mycobacterial fluorescent profiles. Artificial replicates, each representing a hypothetical unknown sample, were generated to test the assay’s power of identification. For classification, the Euclidean distance was used to categorize an unknown curve as a known species fluorescence curve. To verify the function of this diagnostic assay, artificial fluorescence curves replaced actual mycobacterial fluorescence curves, due to experimental constraints.

In uniquely identifying mycobacterial species, both MTBC and NTM species had high rates of classification. For classifying *M. tuberculosis*, *M. africanum/M. microti* and *M. bovis*, the classifier demonstrated a 99-100% probability of correctly classifying the pathogenic species. Additionally, pathogenic NTM species such as *M. fortuitum*, *M. kansasii*, and *M. szulgai* were classified correctly 100% of the time. All other species were successfully classified with true positive probabilities of 100%, except for *M. avium* which demonstrated a 99.9% probability of true classification.

In one instance, this classifier cross-classified species as one another due to a high similarity between their fluorescent profiles. *M. intracellulare* and *M. avium* were classified as one another, slightly increasing their probability of false positive and negative classification. In a clinical context, it is much more likely that an
immunocompromised patient with a lung disease would have *M. avium* infection over an *M. intracellulare* infection. Successful classification of simulated mycobacterial curves by species was achieved. Future experimentation will encompass testing a wide variety and larger number of mycobacterial isolates.

**Conclusion**

This thesis provides a proof-of-principle demonstration that LATE-PCR amplification and Lights-On/Lights-Off probe analysis can generate highly informative melt curves that can be objectively classified in a manner that will eventually serve as output of an automated clinical system. Quantitative species identification has been accomplished, given the sophisticated fluorescence data produced by LATE-PCR and Lights-On/Lights-Off technologies. The data set analyzed here was derived from 19 species of mycobacteria, but the potential applications of the system described here extend to virtually any target or set of targets amplified and analyzed using these chemistries. That said, experiments described here also have revealed the need for future refinements and improvements that should be implemented both for analysis of mycobacteria specifically and for analysis of this type of fluorescent data generally. These recommendations include:

1. The main limiting factor is the amount and variety of mycobacterial DNA available for testing. More genomic DNA would allow assessment of true species identification power within the genus of mycobacteria. Further, one could
accurately assess the reproducibility of the fluorescence profiles, given variability in reaction mixtures and DNA concentration.

2. Addressing the assay construction, the probe design could be altered to create separate dimensions of fluorescence data that respectively identify non-tuberculous mycobacteria and the *Mycobacterium tuberculosis* complex. Partitioning probe sets in this manner would allow more discriminate quantitative analysis.

3. Analysis of this complex fluorescence data should be available to inexperienced clinician, and complete automation of this analysis could be accomplished by creating a graphical user interface (GUI). This GUI would allow the user to specify diagnostic conditions, the catalog of species fluorescence profiles to be used in classification, and other such details imperative to accurate diagnosis.
Chapter 5: Literature Cited

34. Bicmen, C., et al., *Molecular detection and identification of mycobacterium tuberculosis complex and four clinically important nontuberculous
mycobacterial species in smear-negative clinical samples by the genotype
35. Omar, S.V., et al., Analytical performance of the Roche LightCycler(R)
Mycobacterium Detection Kit for the diagnosis of clinically important
37. Pierce, K.E., et al., Linear-After-The-Exponential (LATE)-PCR: primer design
criteria for high yields of specific single-stranded DNA and improved real-time
method of asymmetric PCR and its uses in quantitative real-time analysis. Proc
39. Wangh, L.R., J.; Reis, A.; Rice, L.; Carver-Brown, R.; Tetrault, S., Fluorescent
Signatures for Variable DNA Sequences. Classification: Biological Sciences:
Biochemistry, 2012.
40. Marley, L., Discrimination, classification, identification of microorganisms
hybridization pattern recognition analysis of generic Mycobacterium DNA
42. Rajalahl, T. and O.M. Kvalheim, Multivariate data analysis in pharmaceutics: a
43. Boas, M.L., Mathematical methods in the physical sciences. 3rd ed2006,
44. Jain, A.M., M.; Flynn, P., Data Clustering: A Review. ACM Computing Surveys,
1999. 31(3).
45. Palais, R. and C.T. Wittwer, Mathematical algorithms for high-resolution DNA
46. van der Stoep, N., et al., Diagnostic guidelines for high-resolution melting curve
(HRM) analysis: an interlaboratory validation of BRCA1 mutation scanning
Chapter 6: Appendix
Brandeis Biosafety Protocol #09013

TO: Lawrence Wangh, Department of Biology, MS 008
FROM: KC Hayes, Acting Chair, Brandeis Institutional Biosafety Committee
CC: IBC Protocol File
DATE: March 3, 2009
RE: Protocol #09013 PCR Assay Development Using DNA Sequences from Tuberculosis bacillus

The Brandeis Institutional Biosafety Committee (IBC) has reviewed the above referenced protocol renewal at its February 26, 2009 meeting and approved it under NIH Category 5 to be conducted under Biosafety Containment Level (BSL) 1.

This approval is valid for the period: February 26, 2009 – February 25, 2010.

If your research will continue beyond the approval expiration date listed above, please submit an IBC Biosafety Approval Form – Protocol Renewal by the submission deadline posted on the IBC webpage in time to receive a new approval date for the upcoming work. Forms, procedures, and deadlines are available at http://www.brandeis.edu/csp/ibcintro.html.

Always notify the IBC if you plan to make changes to the scope, personnel, location or continuance of the research.

Please contact the IBC Administrator (ibc@brandeis.edu or x 6 8136) should you require certification of this approval to an outside funding agency.

Additional information requested:
None.