Phylogenetic Analysis of Prevalent Tuberculosis and Non-\textit{Tuberculosis Mycobacteria} in Isfahan, Iran, Based on a 360 bp Sequence of the \textit{rpoB} Gene

Bahram Nasr Esfahani,¹ Sharareh Moghim,¹ Hajieh Ghasemian Safaei,¹ Mohsen Moghoofei,² Mansour Sedighi,¹ and Shima Hadifar¹,*

¹Department of Microbiology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, IR Iran
²Department of Virology, Faculty of Medicine, Iran University of Medical Sciences, Tehran, IR Iran

Corresponding author: Shima Hadifar, Department of Microbiology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, IR Iran. Tel: +98-3137922493, Fax: +98-313688597, E-mail: aseman.shima@yahoo.com

Received 2015 July 05; Revised 2016 January 18; Accepted 2016 February 01.

Abstract

\textbf{Background:} Taxonomic and phylogenetic studies of \textit{Mycobacterium} species have been based around the 16sRNA gene for many years. However, due to the high strain similarity between species in the \textit{Mycobacterium} genus (94.3\% - 100\%), defining a valid phylogenetic tree is difficult; consequently, its use in estimating the boundaries between species is limited. The sequence of the \textit{rpoB} gene makes it an appropriate gene for phylogenetic analysis, especially in bacteria with limited variation.

\textbf{Objectives:} In the present study, a 360bp sequence of \textit{rpoB} was used for precise classification of \textit{Mycobacterium} strains isolated in Isfahan, Iran.

\textbf{Materials and Methods:} From February to October 2013, 57 clinical and environmental isolates were collected, subcultured, and identified by phenotypic methods. After DNA extraction, a 360bp fragment was PCR-amplified and sequenced. The phylogenetic tree was constructed based on consensus sequence data, using MEGA5 software.

\textbf{Results:} Slow and fast-growing groups of the \textit{Mycobacterium} strains were clearly differentiated based on the constructed tree of 56 common \textit{Mycobacterium} isolates. Each species with a unique title in the tree was identified; in total, 13 nods with a bootstrap value of over 50\% were supported. Among the slow-growing group was \textit{Mycobacterium kansasii}, with \textit{M. tuberculosis} in a cluster with a bootstrap value of 98\% and \textit{M. gordonae} in another cluster with a bootstrap value of 90\%. In the fast-growing group, one cluster with a bootstrap value of 89\% was defined, including all fast-growing members present in this study.

\textbf{Conclusions:} The results suggest that only the application of the \textit{rpoB} gene sequence is sufficient for taxonomic categorization and definition of a new \textit{Mycobacterium} species, due to its high resolution power and proper variation in its sequence (85\% - 100\%); the resulting tree has high validity.

\textbf{Keywords:} Phylogeny, \textit{Mycobacterium}, \textit{rpoB} Gene

1. Background

More than 160 species of the \textit{Mycobacterium} genus have been identified, with a third of them having the potential to cause a wide variety of human (1). These medically important species are responsible for considerable human morbidity and mortality, and mostly belong to the \textit{Mycobacterium tuberculosis} complex (MTBC) (such as \textit{M. tuberculosis}, \textit{M. africanum}, \textit{M. canetti}, and \textit{M. bovis}) (2, 3) and the \textit{M. avium} intracellulare complex (MAC), which include atypical mycobacteria. Various phylogenetic studies have been performed to gain an understanding of mycobacterial taxonomic identity, as regards the nature of pathogenic \textit{Mycobacterium} species, the wide range of diseases they cause, and their high genetic similarity (4, 5). Phylogenetic studies have been primarily based on phenotypic characteristics and methods; however, significant progress in the genetic sciences and the development of statistical models for evolutionary analyses of nucleotide sequences have resulted in an improvement of such studies (6, 7).

For many years, phylogenetic and taxonomic studies of \textit{Mycobacterium} species have been based on 16sRNA analysis (8). The critical role of 16sRNA in protein synthesis and its high information content, conserved nature, and presence in all organisms have made this gene an appropriate candidate for phylogenetic analysis, as well as for identification (6, 9). The criterion for classification based on the 16sRNA gene is that bacterial strains differing in less than 10 - 15bp belong to the same species. However, interspecies genetic similarity in the members of the \textit{Mycobacterium} genus is as high as 94.3\% - 100\% and, in some species such as the fast-growing species \textit{Mycobacterium kansasii}, the similarity
is extremely high (> 99%) (10). The presence of more than one copy of 16S rRNA in this genus leads to complexity in interpretation of the sequence information and makes it difficult to draw a valid phylogenetic tree, thereby limiting the estimation boundary between species in phylogenetic analysis (10-12).

Accordingly, in recent decades, other genes such as recA (13), hsp65 (14, 15), and rpoB have been considered as candidates for phylogenetic studies and diagnosis (16, 17). From these, rpoB encoding β subunit of RNA polymerase enzyme is thought to be the most appropriate gene for such purposes, especially in bacteria with close relativity (18-20). The first attempt to apply this gene for the differentiation of Mycobacterial species was carried out by Kim et al. (16, 21). Consequently, different parts of this gene were used to sequence analysis (6, 10, 14), and further studies have acknowledged the gene’s adequacy for interference in phylogenetic relationships in bacterial groups with close links, such as Mycobacteria (22).

2. Objectives

More than 50 human pathogen species of the Mycobacterium genus have been identified, and interspecies similarity in the genus is high. It is important to determine the relationships between Mycobacterial species. In the present study, a 360 bp fragment of the rpoB gene, from Mycobacterium strains isolated from patients in Isfahan province, was used to obtain a precise taxonomic classification of these isolates.

3. Materials and Methods

3.1. Mycobacteria Strains and Molecular Methods

A total of 57 Mycobacterium isolates were obtained between February and October 2013 at the Mycobacterium collection of the department of microbiology, Isfahan University of Medical Sciences and Tuberculosis center of Isfahan. Of these 57 specimens, 41 were respiratory types, including sputum, bronchoalveolar lavage, and bronchial wash specimens, 8 were non-respiratory, and 8 were environmental samples such as water which was needed to break up the mucin; each respiratory specimen was treated with an equal volume of 3.5% NaOH in a 50 mL centrifuge tube, before being vortexed for 30 seconds. Specimens that were decontaminated after being in the tubes were incubated at room temperature for 15 minutes. A sterile phosphate buffer was added to stop the digestion decontamination process; the tubes were then mixed by inversion and centrifuged at 3,000 × g for 15 minutes. The supernatant was discarded and the resultant pellet re-suspended in 3.0 mL of sterile phosphate-buffered saline (Sigma, USA).

Non-respiratory specimens, after the homogenies, were processed in the same way, although sampling of water was done using the grab sampling method. Sodium thiosulfate, as an antichlor, and 0.04% cetylpyridinium chloride (Merck, Germany), as an antimicrobial agent, were added to 21 sterile Erlemeyer flasks. Samples amounting to 500 mL were passed from 0.45 μm filters. The filters were transferred directly onto 7H10 middle brook solid media (Merck, Germany), including 15% oleic acid, albumin, dextrose, and catalase.

3.2. Primary Identification of Isolates by Conventional Methods

The Mycobacterium species were subcultured on LJ media and middle brook 7H9 (Merck, Germany). The isolates were identified by primary conventional methods including acid-fast staining, colony characteristics, growth at 25°C, 37°C, and 42°C, pigment production, semi-quantitative catalase test, Tween 80 hydrolysis, arylsulfatase test (3 and 14 days), heat-stable catalase (pH 7, 68°C), pyrazinamidase (4 and 7 days), urease, nitrate reduction test, and colony morphology. The reference strains used in this study were Mycobacterium tuberculosis H37Rv and M. fortuitum (ATCC 49403) (23) (Table 1).

<table>
<thead>
<tr>
<th>Mycobacterium Species</th>
<th>No. and Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mycobacterium tuberculosis complex</td>
<td>18 clinical isolates, H37Rv</td>
</tr>
<tr>
<td>Mycobacterium fortuitum</td>
<td>ATCC 49403, 28 clinical and environmental isolates</td>
</tr>
<tr>
<td>Mycobacterium avium</td>
<td>1 clinical isolate</td>
</tr>
<tr>
<td>Mycobacterium kansasii</td>
<td>1 clinical isolate</td>
</tr>
<tr>
<td>Mycobacterium smegmatis</td>
<td>1 environmental isolate</td>
</tr>
<tr>
<td>Mycobacterium conceptionense</td>
<td>1 environmental isolate</td>
</tr>
<tr>
<td>Mycobacterium gordonae</td>
<td>5 clinical and environmental isolates</td>
</tr>
</tbody>
</table>

3.3. Preparation of Genomic DNA

Genomic DNA was extracted using the cetyltrimethyl ammonium bromide (CTAB) method (24) and subjected to a high pure PCR template preparation kit (Roche Applied Science, Germany). Purified DNA was stored at -70°C for subsequent experiments.

3.4. PCR Amplification of rpoB

The reaction was performed in a thermal cycler from PCR Hybrid (Omnigene). The partial
The rpoB gene (360bp) was amplified using primers (RPO5') 5'-TCAAGGAGAAGCGCTACGA-3' and (RPO3') 5'-GGATGTGTACGGGTCTGC3' (25). Each 25 µL PCR mixture contained 2 µL of DNA supernatant (5 ng genomic DNA), 1 µL of each primer (10 pmol/µL) (Eurofin MWG/Operon, Ebersberg, Germany), 1.25 µL of MgCl2 (1.5 mM) (Fermentas, Canada), 0.5 µL of dNTP (200 µM) (Fermentas Canada), 0.25 µL of Taq polymerase (500 U) (Cinnagen, Iran), and 2.5 µL of 10x buffer. The PCR program was performed in a thermo cycler (Eppendorf) including at 94°C for 5 minutes, 35 cycles of 94°C for 1 minute each, 60°C for 1 minute, 72°C for 1 minute, and a final step at 72°C for 7 minutes. The PCR products were run on 1.5% agarose gel, which was visualized by green viewer staining (Pars Tous, Iran) and gel documentation.

3.5. Sequencing and Analysis

PCR products from amplification of the rpoB gene were sent to the Pishgam Company (Iran) for purification and sequencing. Chromatogram sequences, using the Bio Edit software, were applied to investigate the outset and termination of gene sequences that were improperly read; they were removed, and a blast of each sequence, using NCBI blast, was performed. Mega software (version 5.2) (26) was used to construct the phylogenetic tree, and the alignment was performed using Clustal W format. The method used was maximum likelihood, and the phylogram stability was evaluated by parsimony bootstrapping with 500 simulations. A Nocardia asteroides sequence (accession number AB219431.1) was used as the out group.

4. Results

Of 57 isolates, 19 (33.33%) were identified as M. tuberculosis and 38 (66.66%) were identified as NTM, using the phenotypic method. Among the 38 NTM isolates, 30 (78.94%) were clinical and 8 (21.05%) were environmental. The clinical isolates included M. tuberculosis complex (MTBC), M. avium, M. fortuitum, M. gordonae, and M. kansasii, while the environmental isolates consisted of M. fortuitum, M. smegmatis, M. conceptionense, and M. gordonae.

The constructed phylogenetic tree of the 360 bp fragment of the rpoB gene sequence was more or less similar to the results obtained from other studies (10, 12, 22, 27). Based on the tree for 56 Mycobacterium isolates that are prevalent in Isfahan, fast and slow-growing groups could be completely separated so that each species was identified in the tree by a unique title. There were 13 nodes supported that had a bootstrap value of over 50%.

The slow-growing group included M. avium, M. kansasii, M. gordonae, and M. tuberculosis complex. Mycobacterium kansasii and M. tuberculosis were in a cluster with a bootstrap value of 98%, while M. gordonae was in another cluster with a bootstrap value of 90%. In addition, in the fast-growing group, one cluster with a bootstrap value of 89% was identified, which included all members of the fast-growing group that were studied in the current paper (M. fortuitum, M. smegmatis, and M. conceptionense). In this cluster, M. smegmatis was in one sub-cluster while the rest of the members were in another sub-cluster with a bootstrap value of 90%.

All of these results indicate the validity of drawing a tree based on the sequence of this gene (Figure 1). Evolutionary interval matrix results show that interspecies similarity in M. tuberculosis was 98% - 100%, while it was 99% - 100% in M. fortuitum species (in most of the M. fortuitum isolates, intraspecies similarity was 100%, and in a few of them it was 99%) and 98% - 100% in M. gordonae; these results illustrate high conservation in the rpoB gene.

Among the 57 investigated isolates, interspecies similarity was 83/73% between M. conceptionense and M. tuberculosis, 86/25% between Mycobacterium gordonae and M. fortuitum, and 90/80% between M. smegmatis and M. conceptionense (Figure 2).

5. Discussion

The 16S rRNA gene sequence has been used as a reference method for the detection and characterization of Mycobacteria and has helped to define over 45 new Mycobacterial species (19). The criteria used for the classification of Mycobacteria, based on the 16S rRNA gene, is that the bacterial strains would belong to the same species if they had only 10 - 15 base pair differences with other species. The high similarity between Mycobacterium species (about 94.3% - 100%) (99%) (10), the existence of 2 copies of 16S rRNA, and the complexity of interpretation of the resulting information made it difficult to define a valid phylogenetic tree and limited the use of it for estimating the boundaries between species in the phylogenetic analysis (10, 12, 22). Accordingly, in recent decades, other genes have been considered as candidates for phylogenetic studies, including hsp65 (14), recA (13), and rpoB (16). Among these, a single-copy rpoB gene encoding a β subunit of RNA polymerase enzyme has been identified as the most appropriate gene for phylogenetic analysis (18, 19, 22, 28). This gene has variable and conserved regions and different fragments of it are used for bacterial analysis (29, 30).

Many studies have used this gene for Mycobacterium genus analysis (10, 16, 18, 27). For the development and completion data resulting from 16S rRNA gene sequences to distinguish bacterial groups with close relationships, several studies have used the sequence of some housekeeping
Figure 1. A Phylogenetic Tree Based on rpoB Gene Sequences Shows the Relationship of the 57 Mycobacterium Isolates and 1 out-Group Strain.

This tree was constructed using the maximum likelihood method and the evolutionary history was inferred using the same method, based on the Jukes-Cantor model. Evolutionary analyses were conducted in MEGA5.

genes; the rpoB gene is one of those that is applied using the MLST method. This is supported by the resulting information of the rpoB gene sequence. Considering the results of tree drawing using other genes with a high bootstrap value, this gene has been deemed appropriate to infer the phylogenetic relationships of bacterial groups with close links, such as Mycobacteria (22, 29, 31). Accordingly, notice the frequency of Mycobacterium species, especially atypical Mycobacteria, in this geographic region, and that these bacteria are widely isolated from environmental, animal, and human resources.

Based on the above, it is necessary to determine interspecies diversity and conduct precise taxonomic classification in Isfahan. To this end, this study utilizes a portion of the gene sequence for classification of prevalent Mycobacterium isolates. According to the results of the phylogenetic tree in the present study, all the tested species were completely separated, so that slow and fast-growing atypical Mycobacteria and M. tuberculosis complex included separate clusters. These results are consistent with those of other studies; for example, in 2005, Devulder et al. (10) illustrated that the resulting phylogenetic tree of the rpoB gene can properly separate Mycobacterium genus members so that, based on a 396 bp sequence of this gene, fast-growing (such as M. smegmatis) and slow-growing (such as Mycobacterium tuberculosis) groups are also totally separated. Based on the results represented in this study, the bootstrap values of the phylogenetic tree for
Mycobacterium tuberculosis complex, M. kansasi-gastroi, and M. smegmatis-goodii were 100%, 84%, and over 50%, respectively. In the present study, the same results were obtained, with M. kansasi and M. tuberculosis included in the cluster with a bootstrap value of 98% and M. fortuitum in the cluster with a bootstrap value of 89%. In another survey from 2004, Adekambi et al. (22) demonstrated that the bootstrap value of the phylogenetic tree derived from the rpoB gene sequence for each cluster was more than 90%, and all tested species, according to this sequence, were completely separated. In addition, the tree resulting from this gene is more valid than those of the 16srRNA and recA genes.

In the present study, the slow-growing group included M. kansasi, M. tuberculosis complex, M. avium, and M. gordonae. M. kansasi and M. tuberculosis were in a cluster with a bootstrap value of 98% and M. gordonae in another cluster with a bootstrap value of 90%. Fast-growing group members in one cluster with a bootstrap value of 98% included all members of the fast-growing group in this study (M. conceptionense and M. smegmatis). Mycobacterium smegmatis was in one sub-cluster with a bootstrap value of 77%, while the cluster for M. fortuitum-farcinogenes-senegalense was 95%. Mycobacterium tuberculosis and M. avium, as members of the slow-growing group, were in their own distinct cluster.

In another study, Kim et al. (16) reported, based on this gene, that the studied fast and slow-growing groups were totally separated and that the bootstrap value of the drawn tree was 100% for the M. fortuitum cluster, 62% for the M. gordonae-intermedium cluster, and 100% for the M. avium-paratuberculosis cluster. This slight difference could be due to variations between geographical regions and the use of different parts of the gene. Like the inter-species and intra-species similarities shown in the present study, Kim et al. intra-species and inter-species similarity was 98% - 100% and 80.05% - 99%, respectively. Adekambi et al. (29), in 2003, reported similar results whereby inter- and intra-species similarity was 83.9% - 97% and 98.3% - 100%, respectively.
tively. Based on the rpoB gene sequence, in 2007 Simmon et al. (34) reported that inter-species similarity was 99.3% -100%, and also introduced the rpoB gene as a proper goal for study, with much more distinctive power than sodA and hsp65 in the study of Mycobacteria.

Our conclusion is that only the application of the rpoB gene sequence is sufficient for Mycobacterial phylogenetic study, due to its high resolution power and proper variation in its sequence (85% -100%) for taxonomic categorization and definition of new Mycobacterium species; the resulting tree has high validity.

Acknowledgments

The authors are grateful to Manuchehr Homae and Maryam Mokhtari and the staff of the Tuberculosis center, Isfahan provincial health office, Isfahan University of Medical Sciences, IR Iran.

Footnotes

Authors’ Contribution: All authors contributed to the research equally.

Financial Disclosure: The vice chancellor of research at Isfahan University of Medical Sciences financially supported this study.

Funding/Support: This work was financially supported by Isfahan University of Medical Sciences (grant 191123).

References


