

# **Diagnostics of non-tuberculous mycobacteria**

Bruijnesteijn van Coppenraet, L.E.S.

# **Citation**

Bruijnesteijn van Coppenraet, L. E. S. (2009, March 5). *Diagnostics of non-tuberculous mycobacteria*. Retrieved from https://hdl.handle.net/1887/13665



**Note:** To cite this publication please use the final published version (if applicable).

# **Chapter 8.**

### **Comparison of gene targets for the differentiation of mycobacterial species.**

Lesla ES Bruijnesteijn van Coppenraet<sup>1</sup>, Jakko van Ingen<sup>4</sup>, Jessica L de Beer<sup>2</sup>, Petra EW de Haas<sup>4</sup>, D.M. Klomberg<sup>3</sup>, Corné HW Klaassen<sup>3</sup>, Edward J Kuijper<sup>1</sup>, Adri GM van der Zanden<sup>2</sup>, Dick van Soolingen<sup>4</sup>

<sup>1</sup>Department of Medical Microbiology, Center of infectious Diseases, Leiden University Medical Center, Leiden; <sup>2</sup>Medical Microbiology & Infection Prevention, Gelre Hospitals, Apeldoorn; <sup>3</sup>Departments of Medical Microbiology and Infectious Diseases, Canisius Wilhelmina Hospital, Nijmegen; <sup>4</sup>National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.

The work presented in this chapter is part of a more extensive analysis on the diagnostic value of gene targets for the identification of mycobacterial species, which is currently performed by the same group.

#### **Abstract.**

Sequencing of specific genomic targets like 16S and the 16S-23S spacer for the identification of *Mycobacterium* isolates is a widely used approach. However, it is not known whether the currently used targets are the most optimal for speciation of mycobacteria. Hence, the most widely used gene; the 16S rRNA, automatically remains the favourite target of choice for this purpose, while no independent study has been conducted to compare the utility of the other candidate genes.

In this study, the performance of Important for the functionality of a target is the quality of sequences (practicability), availability of sequences in public databases and inter- and intraspecies variation.

six gene targets (16S, ITS, *sodA*, *secA1*, *hsp65* and *rpoB*) were compared regarding the differentiation of 3 or 4 isolates of each of the five most frequently encountered *Mycobacterium* species; *M. avium, M. kansasii, M. gordonae, M. chelonae/abscessus, and M. malmoense* and 1 isolate of *M. tuberculosis.*

The results demonstrate that all targets can be used to adequately identify all mentioned species groups. Furthermore, 16S sequence information provided one intra-species distinction; between *M. gordonea* I and II. All five other targets revealed different levels of intra-species variation. Most different intra-species types among the 18 strains tested were provided by the *secA1* (17), followed by *sodA* (16), *hsp65* (13) and ITS (13). We conclude that apart from the currently used genomic targets for identification of mycobacteria, other loci contain genetic information to subdivide the currently recognised species. This should lead to a more detailed identification and eventually taxonomy, if there is a difference in the clinical relevance of isolation of bacteria of subgroupings.

#### **Introduction.**

To date, 128 species of the genus *Mycobacterium* have been registered in the approved list of bacterial species (DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, april 2006). In fact, the correct designation of Atypical mycobacteria is 'non-*Mycobacterium tuberculosis* complex' (NTM).

Infections by NTM have been increasingly recognized in the last decades, mainly due to the high prevalence of these infections in immune-compromised patients. However, the clinical significance of the isolation of these bacteria is often not clear. A part of the microbiological laboratories in The Netherlands are equipped to perform cultures and to identify of mycobacterial species by molecular tests. While several molecular assays *(1, 2, 3, 4)* have been designed for the direct detection of the genus *Mycobacterium* in clinical materials, culturing of the mycobacteria is still widely propagated to confirm the result of the molecular detection and to determine its susceptibility pattern.

Commercially available molecular assays for the identification of cultured mycobacteria are restricted to the Accuprobe hybridization assay (Gen-Probe Inc, San Diego, CA, USA) for the species *M. gordonae*, *M. avium-complex*, *M. kansasii* and *M. tuberculosis*, and and to reverse line hybridization assays: the GenoType Mycobacterium for a wide variety of *M. tuberculosis* complex and NTM species (Hain Lifesciences, Nehren, Germany), and the Inno-Lipa V2 Mycobacteria for NTM species (Innogenetics, Gent, Belgium). The latter two assays identify respectively 16 and 13 different species and both target the Internal Transcribed Spacer (ITS) region between the 16S and the 23S rRNA genes. For the identification of all other NTM species DNA sequencing is currently the golden standard and this is mostly based on the analysis of the 16S rRNA. However, in principle there are multiple semi-conserved genes suitable for identification procedures. The *hsp65* gene (encoding the 65 kDa heat shock protein) and the *rpoB* gene (encoding the ß subunit of RNA polymerase) are often used as sequence target for the identification of mycobacterial isolates. Other targets like the gyrase genes (*gyrA/B), dnaJ, sodA, secA1, recA* have also been reported to contain sufficient genetic variation for identification. Important for the utility of a target for identification at species level is the rate of DNA polymorphism among *Mycobacterium* species and between strains of the respective species. Suitable genes offer sufficient polymorphism to distinguish between species, but are so conserved in the evolution that strains of the same species have little to no variation in sequences.

Genes which reveal too much intra-species variation may not be suitable for the identification within the currently used taxonomy, but sub-division of groupings within species may be important in the light of studies on the clinical relevance of the isolation of certain NTM. Furthermore, the performance of the sequence analysis and the interpretability of the results is of importance.

The National Reference Laboratory for Mycobacteriology (RIVM) collects all *M. tuberculosis* complex, as well as the largest part of the clinical NTM isolates. In the last decade the most common mycobacterial species were *M. tuberculosis*-complex (n= 11777), *M. avium*-complex (n=1480), *M. gordonae* (n=468), *M. kansasii* (n=353), *M. malmoense* (n=185) and *M. chelonae/abscessus* (n=131).

We compared 6 gene targets (16S, ITS, *sodA*, *secA1*, *hsp65* and *rpoB*), all applied as gene targets for the identification of NTM isolates in diagnostic laboratories in the Netherlands, for their usefulness in the identification of isolates of the most commonly encountered species.

# **Strains & Methods.**

Of each NTM species, 3 or 4 strains were included in this study. The strains were all clinical isolates from the Netherlands collected by the RIVM in 2003: *M. chelonae/abscessus* (n=4), *M.avium*-complex (n=4), *M. malmoense* (n=3), *M. kansasii* (n=3), *M. gordonae* (n=3) and ATCC control strain *M. tuberculosis* H37RV. The isolates were randomly selected and no background information was available on clinical relevance of isolation of these mycobacteria or the clinical picture of the respective patients. Strains were identified at the RIVM by 16S sequencing (Table 1). This method is the current standard and has previously been described by Bottger et al. *(5)*.

DNA was extracted using the QiaAmp DNA extraction kit (Qiagen, Venlo, The Netherlands) and diluted to a final concentration of 250 ng/ $\mu$ l.

The DNA extract was divided into aliquots and send to all participating institutes which performed their method(s) of species identification as described below.

For the targets 16S, ITS, *sodA*, *hsp65* and *rpoB*, standard 3-step PCR protocols were used using the primers depicted in Table 1. The same primers were used in the PCR as for direct sequencing. For *secA1,* the protocol has been described by Zelazny et al *(6)*.

Both forward and reverse sequences were obtained and assembled. All sequences were sent to one institute for further analysing. All targets were subjected to a BLAST recognition in the NCBI database (http://www.ncbi.nlm.nih.gov/) and the 16S and ITS sequences in the RIDOM database (http://www.ridom.com/) as well.

A phylogenetic tree was assembled for each target to visualise the variability between the strains using Bionumerics software 3.5 with UPMGA applied as clustering method (Applied-Maths, Belgium). The lengths of all sequences were shortened to the shortest product per target to apply in phylogenetic tree calculations (table 1 / fig 1-6).

# **Results.**

The analysis of the *sodA*, 16S, ITS and *secA1* sequences resulted in readable sequences for all strains. For the target *hsp65* one *M. kansasii* strain repeatedly yielded no PCR product as was the case as well for the *rpoB* target with several *M. malmoense* and *M. gordonae* strains (table 2). Only a part of the *rpoB* sequences were successful for *M. malmoense* isolates. For both targets, alignments of known target-sequences revealed mismatches in the primer regions of several species.

Inter- and intraspecies diversity of the analysed NTM isolates per target is shown in figure 1.

ITS and *sodA* sequences show less than 90% sequence homology for strains in different species groups. *RpoB* sequences show less than 93% homology for strains of different species groups and 16S, *hsp65* and *secA1* sequences show less than 95% homology. All targets except for the 16S, yielded a clear difference between *M. abscessus* and *M. chelonae*.

Table 2 describes the amount of different sequences (subtypes) found within species groups and the amount of mismatches between strains of the same species group. Variability between isolates of the same species was significant for the ITS, SecA1, hsp65 and *sodA* targets. The *sodA* and *secA1* targets reveal the most subtypes over all species and the largest amount of mismatches within species. No variability between strains of the same species group or species is observed for the 16S target. However, a clear differentiation between species groups is observed.

As shown in table 3, the RIDOM database does not provide more information on the identity of the strains than given by the NCBi database. In fact, no ITS sequences of *M. malmoense* are present in the RIDOM database and are therefore not recognised. *M. massiliensis* and *M. bolletii* sequences are not present in the RIDOM database either. The 16S sequence of *M. malmoense* was present in both databases. The NCBI database provides identifications of all

strains of all targets to the species group level. However, discrepanties were seen between the strains belonging to the *M. avium*-complex. Both the *rpoB* and the *secA*1 targets identify *M. paratuberculosis* on multiple occasions while the other targets either identify *M. avium* or *M. avium*-complex. *M. gastri* is recognised twice by secA1 while all other targets yield *M. kansasii*. None of the discrepant results have a 100% concordance with the database sequences which indicates that the matching sequences are not available.

Fig 1. Phylogenetic trees per target.



#### 16S original identification



secA1







#### **Discussion.**

Molecular identification of mycobacterial species provides two primary advantages to phenotypic (conventional) identification: it is more rapid and accurate in identification. Although fast and easy to perform, "sequence based" methods, like species specific real-time PCR and PCR-RFLP have their limitations, since they only recognise the evaluated species. Most laboratories are equipped with a sequencer and are therefore able to perform sequence identifications. Six different sequence targets have been analysed for their ability to differentiate the most common clinical mycobacterial species in the Netherlands. Considerations for a preferable protocol are: practicability (high success rate of creating identifiable sequences and easy to use protocol) , availability of sequences in public databases and suitable levels of discrimination within the target. The level of resolution for routine diagnostic application of species differentiation is difficult to ascertain; clinical relevance of determining subspecies or only group levels is not known. Trial information is lacking on this subject.

The 16S rRNA gene has an advantage in its presence in all prokaryotic organisms, therefore a broadrange PCR would be able to detect sequences of all genera and differentiate based on sequence variation. Still, because this gene does not contain a high degree of variation compared to some of the other targets, the use of this target for the differentiation of mycobacteria is limited. Several species have identical 16S sequences, among which the species belonging to the *M. tuberculosis* complex, "*M. kansasii* and *M. gastri"*, and the subspecies of *M. avium*. Using only the 5'part of the gene, species as "*M. chelonae* and *M. abscessus"* and "*M. ulcerans* and *M. marinum"* cannot be differentiated either. In other words: 16S differentiation does not exceed the group level, as confirmed with the strains used here. An important advantage of the 16S gene is the abundance of sequences in the public databases. However, the reliability of sequence information can be questionable *(11).*

Both *rpoB* and *hsp65* contain a higher degree of divergence between species than the 16S target. The additional advantage of the *rpoB* target is that it contains species-specific sequences as well as polymorphisms responsible for rifamycin resistance and can therefore be used for the simultaneously detection of susceptibility and species identification. *rpoB* and *hsp65* protocols lack in reliability: sufficient amounts of sequences are available, but primer sequences are often not completely genus specific due to too much variation, which explains the unsuccessful PCR of several strains.

In 1993 the Internal Transcribed Spacer (ITS) region was added as a possible target in species differentiation and proven useful as such *(12, 13)* for the recognition of virtually all known mycobacterial species . It is used in the commercial assays InnoLipa V2 and GenoType Mycobacterium. The ITS-protocol has a similar performance but with more discriminative power than the 16S. However, while many sequences are available for the ITS target, differentiation between *M. kansasii* and *M. gastri* is not possible and no clear differentiation between species belonging to the *M. avium* group can be made. Also, many short and incomplete sequences are a pollution to the database and might cause false identifications when analysed incautiously.

In 1994 the *sodA* gene was proposed as possible target for differentiation *(14)* and since then sporadically applied *(10, 15) SecA1* is the most recent evaluated target gene used here for species identification *(16)*. Correct identifications can only be made with an extended database. S*odA* appears superior in species differentiation to *secA1*, probably due to the availability of more than 130 sequences (less species) while for *secA1* 42 sequences of 30 different species are available. Considering there are over 150 species known to date, a second sequence target would be necessary for the identification of uncommon (sub)species for both targets. Also, when the amount of interspecies variation within a target is too small, species differentiation will be compromised. However, if there is too much variation, the risk of mismatches in the primer sequences will increase as well as the occurrence of unidentifiable new sequences. In either way a second target will be necessary for the correct differentiation of uncommon species or difficult taxonomical groups.

None of the tested targets is capable of differentiation within the *M. tuberculosis*-complex, but all targets are 100% reliable in identifying the correct species group.

*M. gordonae* and *M. malmoense* are both clearly separated from other species by all targets and no complicated taxonomy intricate differentiation of these species.

*SecA1* is the only target to identify *M. gastri* in two cases while no other targets confirm these results. Both sequences are 99% similar to an ATCC strain of *M. gastri*, and therefore a 100% concordance with a *M. kansasii* would be optional if more sequences were available. However, ITS and 16S are not able to differentiate the two species, while *rpoB*, *hsp65* and *sodA* have no 100% assurance either for the identification of *M. kansasii*.

The *M. abscessus/chelonae* group is recently divided into more species. *M. massiliensis* and *M. bolletii* belong to this group. They are an example of the increasing complexity of mycobacterial taxonomy. Not for all targets the sequences of these species are known yet, but for *sodA* and the ITS the sequence variability within this species group is probably sufficient enough to harbour specific differences.

The *M. avium*-complex appears to be the most difficult group to differentiate. This species-group is known to harbour extensive sequence variation. Still, 16S and *hsp65* show no variability between the (sub)species of this complex. All strains yield discrepant results for the other targets. While one target shows two strains to have similar sequences, another target shows another combination of strains to be similar. Not one strain is identified consistently by different targets. While by two targets identified as such, the recognition of *M. paratuberculosis* is unlikely to be correct, due to the difficult growth kinetics of the species. The identification of this species group in particular would benefit by an extensive and quality controlled database.

Concluding: When is chosen for identification to the group level, then 16S sequencing is the most reliable. When identification to the (sub)species level is preferred, no target will be sufficient on its own and a combination of two targets will be necessary to give an accurate identification.



Table 1: Oligo's used in PCR and nucleotide sequence analysis. Table 1: Oligo's used in PCR and nucleotide sequence analysis.

\* M. malmoense is only 84 bp long for rpoB. \* M. malmoense is only 84 bp long for rpoB.

Table 2. Intraspecies difference in basepairs resulting in amount of subtypes per species. Table 2. Intraspecies difference in basepairs resulting in amount of subtypes per species.



Table 3: Comparison of sequences with the public databases. The 16S sequencing protocol used for the original identifications of the strains resulted in the sequences Table 3: Comparison of sequences with the public databases. The 16S sequencing protocol used for the original identifications of the strains resulted in the sequences described in the 16S column. described in the 16S column.

Analysed length of products can vary for some targets: only two-directional fragments compared to database. 16S = 153 bp, *secA1=* 459-761 bp, *rpoB* = 101-437 bp, *hsp65* = Analysed length of products can vary for some targets: only two-directional fragments compared to database. 16S = 153 bp, secA1= 459-761 bp, rpoB = 101-437 bp, hsp65 = 338-440 bp, ITS = 269-(approx)350 bp, sodA = 356-463 bp. 338-440 bp, ITS = 269-(approx)350 bp, *sodA* = 356-463 bp.



avium ptb = M. paratuberculosis of the M. avium-complex, MAC = M. avium-complex, aviu st 2 = M. avium serotype 2, tb-complex = M. tuberculosis-complex, absces = M. avium ptb = M. paratuberculosis of the M. avium-complex, MAC = M. avium-complex, aviu st 2 = M. avium serotype 2, tb-complex = M. tuberculosis-complex, absces = M.

abscessus, chel = M. chelonae, mass = M. massiliensis. abscessus, chel = M. chelonae, mass = M. massiliensis.

\* only a short sequence (84 bp) of the *M. malmoense* isolates was successful for the rpoB target. \* only a short sequence (84 bp) of the M. malmoense isolates was successful for the rpoB target.

#### **References: Chapter 8.**

- 1. Shrestha NK, Tuohy MJ, Hall GS, Reischl U, Gordon SM, Procop GW. Detection and differentiation of Mycobacterium tuberculosis and nontuberculous mycobacterial isolates by real-time PCR. J Clin Microbiol. 2003;41:5121-6.
- 2. Park H, Kim C, Park KH, Chang CL. Development and evaluation of triplex PCR for direct detection of mycobacteria in respiratory specimens. J Appl Microbiol. 2006;100:161-7.
- 3. Park H, Jang H, Song E, Chang CL, Lee M, Jeong S, Park J, Kang B, Kim C. Detection and genotyping of Mycobacterium species from clinical isolates and specimens by oligonucleotide array. J Clin Microbiol. 2005;43:1782-8.
- 4. Bruijnesteijn Van Coppenraet ES, Lindeboom JA, Prins JM, Peeters MF, Claas EC, Kuijper EJ. Realtime PCR assay using fine-needle aspirates and tissue biopsy specimens for rapid diagnosis of mycobacterial lymphadenitis in children. J Clin Microbiol. 2004;42:2644-50.
- 5. Bottger EC, Teske A, Kirschner P, Bost S, Chang HR, Beer V, Hirschel B. Disseminated "*Mycobacterium genavense*" infection in patients with AIDS. Lancet. 1992;340:76-80.
- 6. Zelazny AM, Calhoun LB, Li L, Shea YR, Fischer SH. Identification of Mycobacterium species by secA1 sequences. J Clin Microbiol. 2005;43:1051-8.
- 7. Van Der Zanden AG, Te Koppele-Vije EM, Vijaya Bhanu N, Van Soolingen D, Schouls LM. Use of DNA extracts from Ziehl-Neelsen-stained slides for molecular detection of rifampin resistance and spoligotyping of *Mycobacterium tuberculosis*. J Clin Microbiol. 2003;41:1101-8.
- 8. Telenti, A, Marchesi F, Balz M, Bally F, Böttger E, Bodmer T. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J. Clin. Microbiol. 1993;31:175-178.
- 9. Roth A, Reischl U, Streubel A, Naumann L, Kroppenstedt RM, Habicht M, Fischer M, Mauch H. Novel diagnostic algorithm for identification of mycobacteria using genus-specific amplification of the 16S-23S rRNA gene spacer and restriction endonucleases. J Clin Microbiol. 2000;38:1094-104.
- 10. Adekambi T, Drancourt M. Dissection of phylogenetic relationships among 19 rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *rpoB* gene sequencing. Int J Syst Evol Microbiol. 2004;54:2095-105.
- 11. Turenne CY, Tschetter L, Wolfe J, Kabani A. Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous *Mycobacterium* species. J Clin Microbiol. 2001;39:3637-48. Erratum in: J Clin Microbiol 2002;40:2316.
- 12. Roth A, Fischer M, Hamid ME, Michalke S, Ludwig W, Mauch H. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. J Clin Microbiol. 1998;36:139-47.
- 13. Tortoli E, Mariottini A, Mazzarelli G. Evaluation of INNO-LiPA MYCOBACTERIA v2: improved reverse hybridization multiple DNA probe assay for mycobacterial identification. J Clin Microbiol. 2003;41:4418- 20.
- 14. Zolg JW, Philippi-Schulz S. The superoxide dismutase gene, a target for detection and identification of mycobacteria by PCR. J Clin Microbiol. 1994;32:2801-12.
- 15. Domenech P, Jimenez MS, Menendez MC, Bull TJ, Samper S, Manrique A, Garcia MJ. *Mycobacterium mageritense* sp. nov. Int J Syst Bacteriol. 1997;47:535-40.
- 16. Zelazny AM, Calhoun LB, Li L, Shea YR, Fischer SH. Identification of *Mycobacterium* species by *secA1* sequences. J Clin Microbiol. 2005;43:1051-8.