Careful use of 16S rRNA gene sequence similarity values for the identification of Mycobacterium species

M. Beye1, N. Fahsi1, D. Raoult1,2 and P.-E. Fournier1
1) URMITE, UM63, CNRS 7278, IRD 198, Institut Hospitalo-Universitaire Méditerranée-infection, Aix-Marseille Université, Faculté de Médecine, Marseille, France and 2) Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia

Abstract

In order to evaluate the suitability of 16S rRNA nucleotide sequence similarity for the classification of new Mycobacterium isolates at the species level, we systematically studied the pairwise identity values of this gene for 131 Mycobacterium species with standing in nomenclature. Only one of the studied species, M. poriferae (0.76%), strictly respected the 95% and 98.65% threshold values currently recommended to determine the affiliation of bacterial isolates to an existing or new genus or species, respectively. All other species exhibited at least an identity value >98.65% and/or <95% with another Mycobacterium species. Therefore, we suggest that interpretation of interspecies 16S rRNA identity values should be made cautiously when classifying a new mycobacterial isolate at the species level.

© 2017 The Authors. Published by Elsevier Ltd.

Keywords: 16S rRNA, Mycobacterium, sequence similarity, taxonomy, threshold value

Original Submission: 20 October 2017; Accepted: 20 December 2017

Article published online: 29 December 2017

Corresponding author: P.-E. Fournier, URMITE, UM63, CNRS 7278, IRD 198, Institut Hospitalo-Universitaire Méditerranée-infection, 19-21 Bd Jean Moulin, 13005 Marseille, France
E-mail: pierre-edouard.fournier@univ-amu.fr

Introduction

Taxonomy provides scientists with essential information, enabling them to understand the relationships between living organisms and their different ecosystems [1]. For prokaryotes, taxonomy allows the reliable identification of microbial strains from clinical or environmental samples [2]. Bacterial taxonomy was initiated in the late 19th century, when phenotypic characteristics were incorporated into bacterial description, including motility, growth requirements, morphology, staining properties, colony size and colour, and chemical reactions [3]. Between the mid-1950s and the 1980s, new parameters were progressively added, notably chemotaxonomy [4], numerical taxonomy, genomic DNA-DNA hybridization and G+C content [5]. In the 1980s, the advent of DNA amplification and sequencing techniques, in particular of the 16S rRNA gene, constituted a major step forwards by facilitating bacterial classification [6,7]. The 16S rRNA gene is a highly conserved gene that is made of nine hypervariable domains separated by more preserved fragments in which universal primers can be designed. More than three million 16S rRNA gene sequences are currently available in public databases [8]. In 1996, Vandamme et al. [9] suggested that polyphasic taxonomy, which takes into account all available phenotypic and genotypic data and integrates them into a consensus classification, should include 16S rRNA gene sequence identity. In 2010, Tindall et al. [10], in a reevaluation of the various available methods, proposed a combination of phenotypic and genotypic criteria within which 16S rRNA gene sequence similarity, and phylogeny was included as a first-line tool.

In 1994, scientists considered two strains as belonging to different species if they shared 16S rRNA gene sequence similarity values <97% and to a distinct genus if this value was <95% [11]. The cutoff value at the species level was later reevaluated at 98.7% [12] and then 98.65% [13]. However, several authors have shown that these thresholds, originally designed to standardize the use of sequences of 16S rRNA genes in taxonomy, are not applicable to multiple genera. In 2015, we demonstrated that many of the current bacterial species with validly published names do not respect the 95% and 98.7% thresholds [14].
In 2000, Woo et al. [15] proposed that 16S rRNA gene sequencing was the reference standard for the identification of *Mycobacterium* species. Genotypic investigations based on the sequencing of the 16S rRNA gene have played a significant role in the taxonomic classification of members of the genus *Mycobacterium* [16]. However, to date, no systematic study of the degree of 16S rRNA divergence among *Mycobacterium* species has been conducted.

Here we evaluate the value of current 16S rRNA cutoff values at the species and genus levels by systematically calculating the pairwise degree of 16S rRNA similarity between all *Mycobacterium* species with standing in nomenclature.

### Methods

### Collection of 16S rRNA gene sequences from members of the genus *Mycobacterium*

Within the List of Prokaryotic Names with Standing in Nomenclature website (http://www.bacterio.net/mycobacterium.html),

© 2017 The Authors. Published by Elsevier Ltd, NMNI, 22, 24–29

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
we selected all *Mycobacterium* species with a validly published name as of 25 March 2016, and we collected the 16S rRNA gene accession numbers from type strains. As a result of the wide heterogeneity in length and quality of the 16S rRNA gene sequences of type strains, we did not use sequences shorter than 1320 nt. We created a FASTA format file containing all selected sequences.

16S rRNA gene sequence analysis: calculation of pairwise 16S rRNA gene sequence similarities

Sequences were aligned using Muscle software with default settings [17]. In this study, pairwise 16S rRNA gene sequence similarities between all species of the genus *Mycobacterium* were first estimated by MEGA 5 phylogeny software [18]. Then the highest and lowest values computed by this software were

---

**FIG. 1.** Phylogenetic distribution of *Mycobacterium* species used in present study based on comparison of 16S rRNA sequences. Sequences were aligned by MUSCLE [14], and phylogenetic inferences were obtained by maximum likelihood method and Kimura two-parameter model in MEGA software. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis 1000 times to generate majority consensus tree. *Pseudonocardia acaciae* (EU921261) was used as outgroup.
TABLE 2. Species that do not respect pairwise similarity thresholds of <95% and >98.65%

| Species (accession no.) | No. for <95% threshold | No. for >98.65% threshold |
|-------------------------|------------------------|---------------------------|
| Mycobacterium abscessus subsp. bolletii (AY859681) | 41 | 2 |
| Mycobacterium africanum (AF480605) | 23 | 7 |
| Mycobacterium agar (AJ92895) | 32 | 0 |
| Mycobacterium amnicum (X55598) | 18 | 11 |
| Mycobacterium alei (AF023664) | 17 | 13 |
| Mycobacterium anyangense (KJ850563) | 3 | 5 |
| Mycobacterium arquense (EF004881) | 16 | 11 |
| Mycobacterium arquense (DQ157760) | 12 | 0 |
| Mycobacterium atrautopticum (AF480595) | 6 | 4 |
| Mycobacterium ambiguous (AY859683) | 6 | 3 |
| Mycobacterium aureum (X55595) | 23 | 0 |
| Mycobacterium australis (X93182) | 17 | 1 |
| Mycobacterium avium subsp. avium (A536037) | 30 | 9 |
| Mycobacterium avium subsp. silvaticum (EF521891) | 23 | 13 |
| Mycobacterium baetilodurense (EF590105) | 20 | 11 |
| Mycobacterium brandeni (AF480574) | 66 | 0 |
| Mycobacterium britannense (AJ012577) | 1 | 1 |
| Mycobacterium bromae (AF480676) | 32 | 0 |
| Mycobacterium canadensis (AY255478) | 20 | 4 |
| Mycobacterium capri (AJ131120) | 17 | 7 |
| Mycobacterium caudatum (L08169) | 23 | 0 |
| Mycobacterium clementii (KJ607136) | 0 | 1 |
| Mycobacterium chimaera (AY157072) | 47 | 2 |
| Mycobacterium chimaera (X56603) | 17 | 0 |
| Mycobacterium chlorophanocaulum (X79292) | 0 | 3 |
| Mycobacterium chylubum (AF480597) | 0 | 3 |
| Mycobacterium congoense (AY859685) | 2 | 17 |
| Mycobacterium cuniculorum (AY859684) | 8 | 3 |
| Mycobacterium cunettii (AY449728) | 1 | 0 |
| Mycobacterium cursorium (DQ534008) | 5 | 6 |
| Mycobacterium dierdorfi (AF480599) | 23 | 5 |
| Mycobacterium dolacum (AF264700) | 42 | 0 |
| Mycobacterium duvalii (UP4745) | 3 | 2 |
| Mycobacterium elephantis (A701074) | 6 | 0 |
| Mycobacterium flexneri (AF264800) | 6 | 2 |
| Mycobacterium favae (AY67904) | 36 | 0 |
| Mycobacterium flavescens (A701074) | 5 | 10 |
| Mycobacterium flavoneuri (AE220684) | 10 | 5 |
| Mycobacterium fortuitum subsp. fortuitum (AY427066) | 6 | 16 |
| Mycobacterium fortuitum subsp. oceani (F4273722) | 6 | 16 |
| Mycobacterium fraenulum (X998451) | 17 | 0 |
| Mycobacterium frenkleri (JQ76274) | 11 | 3 |
| Mycobacterium gudu (X55594) | 6 | 8 |
| Mycobacterium gudini (AF480602) | 26 | 7 |
| Mycobacterium genavense (X66070) | 11 | 4 |
| Mycobacterium gervaisi (Y28372) | 11 | 10 |
| Mycobacterium geron (K52923) | 27 | 1 |
| Mycobacterium hassiacum (UP4745) | 52 | 0 |
| Mycobacterium hedingeri (AB282084) | 6 | 5 |
| Mycobacterium hadenii (X93184) | 5 | 0 |
| Mycobacterium halocarbonum (A310467) | 1 | 0 |
| Mycobacterium haustoriiense (AY457067) | 2 | 21 |
| Mycobacterium interjectum (AY037998) | 1 | 1 |
| Mycobacterium intermedium (X67847) | 20 | 0 |
| Mycobacterium incarcerum (A536036) | 29 | 12 |
| Mycobacterium iranica (HCO39482) | 13 | 0 |
| Mycobacterium kansaii (AZ36053) | 26 | 7 |
| Mycobacterium kansaii (X55591) | 10 | 0 |
| Mycobacterium koreense (JF721826) | 6 | 1 |
| Mycobacterium kudzuki (AF43790) | 3 | 1 |
| Mycobacterium kyushuense (AB707011) | 70 | 0 |
| Mycobacterium lacus (AF467683) | 17 | 14 |
| Mycobacterium lentiflavum (AF480583) | 32 | 6 |
| Mycobacterium leporis (GU979640) | 20 | 0 |
| Mycobacterium lipoferum (A701074) | 14 | 1 |
| Mycobacterium madagascarense (AB537170) | 0 | 3 |
| Mycobacterium magnetense (AE69399) | 0 | 1 |
| Mycobacterium malmoense (X53930) | 57 | 5 |
| Mycobacterium manitensis (FH28997) | 5 | 7 |
| Mycobacterium mannuin (AF456240) | 12 | 7 |
| Mycobacterium marsellei (EU266632) | 22 | 14 |
| Mycobacterium macro (AF480584) | 23 | 6 |
| Mycobacterium macro (AF480594) | 23 | 6 |
| Mycobacterium macro (AF480591) | 35 | 1 |
| Mycobacterium macro (X79292) | 0 | 3 |
| Mycobacterium macro (K522204) | 4 | 11 |
| Mycobacterium madera (AB537171) | 1 | 0 |
| Mycobacterium nebraskense (AS36456) | 20 | 12 |
| Mycobacterium neohydrophovora (AF480593) | 5 | 4 |
| Mycobacterium neovenetense (AY457068) | 2 | 17 |
| Mycobacterium novamagense (EU399553) | 48 | 0 |
studied *Mycobacterium* species is presented in Fig. 1. Among the 131 studied species, the pairwise 16S rRNA gene sequence similarity values ranged from 93.00% between *M. chelonae* and *M. kyriinense* to 100% between *M. fortuitum* subsp. *acet-amidolyticum* and *M. fortuitum* subsp. *fortuitum*, *M. africanum* and *M. caprae*, *M. farcinogenes*, *M. houstonense* and *M. senegalense*, *M. gastri* and *M. kansaisi*, *M. mucogenium* and *M. phocaicum*, *M. murale* and *M. takaiense*, and *M. paraseoulineu* and *M. seoulense*, respectively (Supplementary Table S1).

Of the 131 studied *Mycobacterium* species, 90 (68.7%) exhibited at least one 16S rRNA gene sequence similarity value greater than 98.65% with another species in this genus (Table 2, Supplementary Table S1). Among 131 studied species, 123 (93.9%) exhibited at least one 16S rRNA gene sequence similarity value lower than 95% with another species in the genus (Table 2, Supplementary Table S1). Only one (0.76%) of the 131 studied species, i.e. *M. poriferae*, exhibited only expected values (Table 2, Supplementary Table S1). At the intraspecies level, only expected values were observed.

**Discussion**

Over the past decade, several authors suggested that the inter- and intraspecies discriminatory power of 16S rRNA gene sequences was insufficient for some bacterial genera [19,20]. As examples, *Streptococcus pneumoniae* and *S. mitis* exhibit only a 3 nt difference (99.7% identity), which would classify them in the same species. In contrast, major interspecies differences may be observed, as is the case in the genus *Clostridium*, with *C. tetani* and *C. innocuum* exhibiting a 104 nt divergence (93.7% identity). The strict application of the 95% threshold would justify their classification in distinct genera [19]. In addition, in 2010, Pei et al. [21] identified an intragenomic sequence divergence greater than 1.3% among 16S rRNA genes copies in 11 bacterial species. Among these, *Borreli aafzelii*, an agent of Lyme disease in humans, exhibits a similarity of only 79.62% between its two 16S rRNA gene copies [21]. Thus, a strict application of the 98.65% threshold would classify these bacteria in different species depending on the 16S rRNA gene copy analysed [21,22]. According to Rossi-Tamisier et al. [14], among 158 studied bacterial genera, only members of 17 genera strictly respected the 95% and 98.65% thresholds. Among other studied genera, the percentage of species that respected strictly both thresholds varied from 0 (Brucella) to 93.9% (*Nocardia*) [14].

In the present report, we observed that the currently used 16S rRNA gene sequence similarity thresholds for delineating bacterial species are valid for only 0.76% of 131 studied *Mycobacterium* species with standing in nomenclature. Because our study covers 71.97% of the currently validly published *Mycobacterium* species names, we believe that the 95% and 98.65% thresholds are not suitable for this genus and should at the maximum be used as indicators, not as a reference standard, for classifying new *Mycobacterium* species.

**Acknowledgement**

This study was funded by the Fondation Méditerranéenne Infection.

**Conflict of interest**

None declared.

**Appendix A. Supplementary data**

Supplementary data related to this article can be found at https://doi.org/10.1016/j.nmni.2017.12.009.

**References**

[1] Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, et al. Opinion: Re-evaluating prokaryotic species. Nat Rev Microbiol 2005;3:733–9.

[2] Moore ER, Mihaylova SA, Vandamme P, Krichevsky MI, Dijkshoorn L. Microbial systematics and taxonomy: relevance for a microbial commons. Res Microbiol 2010;161:430–8.

[3] Lehmann KB, Neumann R. Atlas und Gundriss des Bakteriologie und Mikrobiologie. 1896. Munich.

[4] Minnikin DE, Almahaomy L, Goodfellow M. Differentiation of *Mycobacterium*, *Nocardia*, and related taxa by thin-layer chromatographic analysis of whole-organism methylamines. J Gen Microbiol 1975;88:200–4.

[5] Brenner DJ, Fanning GR, Johnson KE, Citarella RV, Falkow S. Polynucleotide sequence relationships among members of *Enterobacteriaceae*. J Bacteriol 1969;98:637–50.

[6] Coenye T, Vandamme P. Use of the genomic signature in bacterial classification and identification. Syst Appl Microbiol 2004;27:175–85.

[7] Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 2007;10:304–9.

[8] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41:D590–6.

[9] Vandamme P, Pot B, Gillis M, De Vos P, Kersters K, Swings J. Polysacharide taxonomic, a consensus approach to bacterial systematics. Microbiol Rev 1996;60:407–38.

[10] Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010;60:249–66.

[11] Stackebrandt E, Goebel BM. Taxonomic note: a place for DNA-DNA reassocation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bact 1994;44:846–9.
[12] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.

[13] Kim M, Oh HS, Park SC, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol 2014;64:346–51.

[14] Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE. Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. Int J Syst Evol Microbiol 2015;65:1929–34.

[15] Woo PC, Tsoi HW, Leung KW, Lum PN, Leung AS, Ma CH, et al. Identification of Mycobacterium neoaurum isolated from a neutropenic patient with catheter-related bacteremia by 16S rRNA sequencing. J Clin Microbiol 2000;38:3515–7.

[16] Tortoli E. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin Microbiol Rev 2003;16:319–54.

[17] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 2004;5:113.

[18] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–9.

[19] Bosshard PP, Zbinden R, Abels S, Boddinghaus B, Altwegg M, Bottger EC. 16S rRNA gene sequencing versus the API 20 NE system and the VITEK 2 ID-GNB card for identification of nonfermenting Gram-negative bacteria in the clinical laboratory. J Clin Microbiol 2006;44:1359–66.

[20] Mignard S, Flandrois JP. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. J Microbiol Methods 2006;67:574–81.

[21] Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, et al. Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl Environ Microbiol 2010;76:3886–97.

[22] Acinas SG, Marcelino LA, Klepac-Ceraj V, Polt MF. Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol 2004;186:2629–35.