Chapter 1

The Taxonomy of the Genus *Mycobacterium*

Enrico Tortoli
Emerging Bacterial Pathogens Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

The genus *Mycobacterium* includes more than 190 species and belongs to the family of *Mycobacteriaceae*, class *Corynobacteriales*, type *Actinobacteria*, and kingdom *Bacteria*. It was first proposed in 1896 (Lehmann and Neuman, 1896) to host organisms considered at that time to be halfway between fungi and bacteria.

The Swedish botanist Carl Linnaeus is considered the founder of taxonomy: a hierarchical classification of plants (Linnaeus, 1735), subsequently extended to animals, based on similarity of phenotypic characters. A similarly important innovation introduced by Linnaeus is the binomial nomenclature, still in use for naming species.

Following the discovery of DNA (Watson and Crick, 1953), genetic characters started to be investigated and to be used for taxonomic purposes. The determination of DNA base composition (guanine + cytosine%) represented the first step (Barbu et al., 1956). A few years later Wayne et al. proposed to measure the DNA relatedness among strains in a paper aiming to reconcile the competing taxonomic approaches: phenotypic and genotypic (Wayne et al., 1987). He suggested that in members of different species this parameter, measured by DNA−DNA hybridization (DDH), should be lower than 70%. The 70% threshold is still considered a gold standard for species circumscription, the DDH test is, however, hardly performed in modern laboratories (Chan et al., 2012).

The genetic sequencing introduced by Sanger (Sanger et al., 1977), a milestone in biological sciences, had an enormous impact on taxonomy and represented significant progress toward a phylogeny-based classification. The rRNA, which is highly conserved because of the essential role of ribosomes in the protein synthesis, soon became the primary target. Among its three subunits, the 16S has been by far the most frequently investigated, and at

Nontuberculous Mycobacteria (NTM). DOI: https://doi.org/10.1016/B978-0-12-814692-7.00001-2

present the sequence of this locus is available, for every known species, in public databases.

Comparative studies between 16S rRNA sequence and DDH lead to identify a 16S similarity of 97% as the threshold corresponding to 70% DDH (Gevers et al., 2005). Despite the cutoff being subsequently revised and raised to 98.8%–99% (Stackebrandt and Ebers, 2006) it still remains unsuitable to appraise the divergence between several species within the genus *Mycobacterium*.

The genus *Mycobacterium* is characterized, at phenotypic level, by unique characteristics. In the cell wall, extremely rich in lipids, the mycolic acids play a major role. Although they are present in a few other *Actinobacteria*-related genera, only the genus *Mycobacterium* has chains as long as 60–90 C (Brennan and Nikaido, 1995). On the basis of growth rate, the mycobacteria are conventionally divided into two groups: the slow growers require more than 1 week to develop visible colonies on solid media, while the rapid growers may require 3–7 days, thus growing slower in comparison with other cultivable bacteria (Tortoli, 2003).

Genetically, the mycobacteria differ from the large majority of bacteria for the high G + C content (ranging from 62% to 70%). The number of copies of the ribosomal operon is low: two copies in the rapid growers and only one in the slow growers; there are very few exceptions (Böddinghaus et al., 1990).

The variability of 16S rRNA is moderate among mycobacteria with a number of species presenting 100% identical sequence (Tortoli, 2003). The 16S rRNA is about 1500 bp long and includes mostly highly conserved regions with few interposed variable traits. The two major variable strings are located in the first third of the gene; they are known as hypervariable regions A (*E. coli* positions 130–210) and B (*E. coli* positions 430–500) with the latter including the helix 18. Minor variable regions are located in the remaining two thirds of the gene (Stahl and Urbance, 1990). The hypervariable region A hosts most of the species-specific polymorphisms. The hypervariable region B is characterized by three mayor formats; it may host, in the helix 18, a 14-nucleotide insertion, a 12-nucleotide insertion, or no insertion at all.

Since the first taxonomic analyses of the genus *Mycobacterium* based on the sequence of 16S rRNA, a number of features have emerged. Rapid and slow growers appeared clearly separated in the phylogenetic tree. All the rapidly growing species had a short helix 18 (no insertion in the hypervariable region B). The majority of slow growing species had a long helix 18: 12-nucleotide insertion in the hypervariable region B. The slow growers *M. terrae* and *M. nonchromogenicum* had a 14-nucleotide insertion and were located in a separate clade interposed between slow and rapid growers. *M. simiae* did not have insertion in the hypervariable region B, but clustered

with slow growers according to its phenotype (Rogall et al., 1990; Stahl and Urbance, 1990) (Fig. 1.1).

In subsequent years, the inclusion in the analysis of further species did not produce substantial changes in the topology of the phylogenetic tree. Among the newly added species, the new rapid growers clustered with the previous ones and confirmed the short helix 18; the species which added to the clade of *M. terrae* and *M. nonchromogenicum* had the 14-nucleotide insertion; most of the slow growers confirmed the presence of a 12-nucleotide insertion. The few slow growers with short helix 18 grouped with *M. simiae* except two, *M. doricum* and *M. tusciae*, which fell among the rapid growers (Figs. 1.2 and 1.3).

On the basis of such results, the presence of signatures within the 16S rRNA was theorized:

- The short helix 18 is a marker of rapid growers.
- A small group of slow growers related to *M. simiae* (*M. simiae* complex), recognizable for sharing the same sequence in the hypervariable region B, have the short helix 18 as well.
- The long helix 18 (12-nucleotide insertion) is a marker of slow growers.

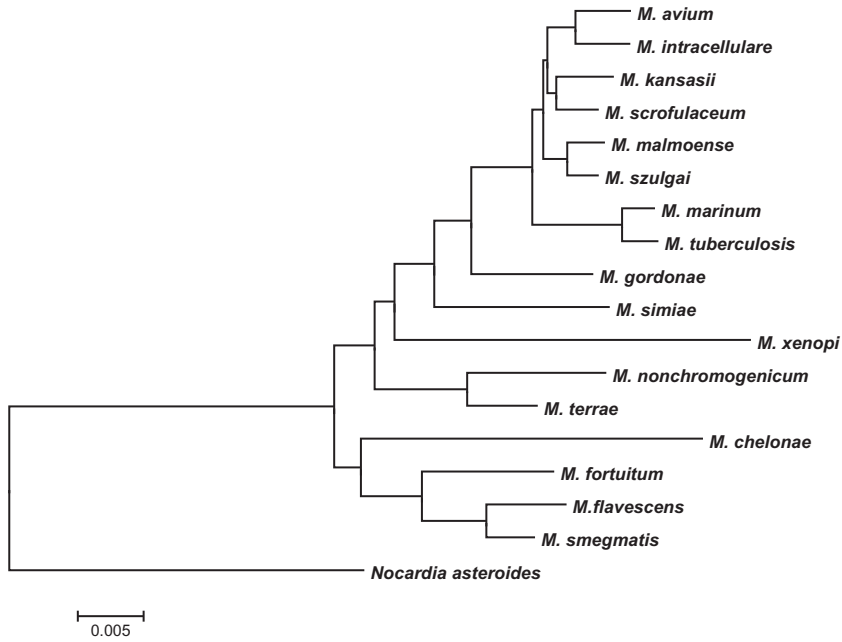

FIGURE 1.1 Phylogenetic tree inferred with the neighbor-joining algorithm from the sequence dataset investigated by Rogall et al. (1990). The tree was rooted with *Nocardia asteroids* as outgroup. Bar, 0.005 substitutions per nucleotide position.

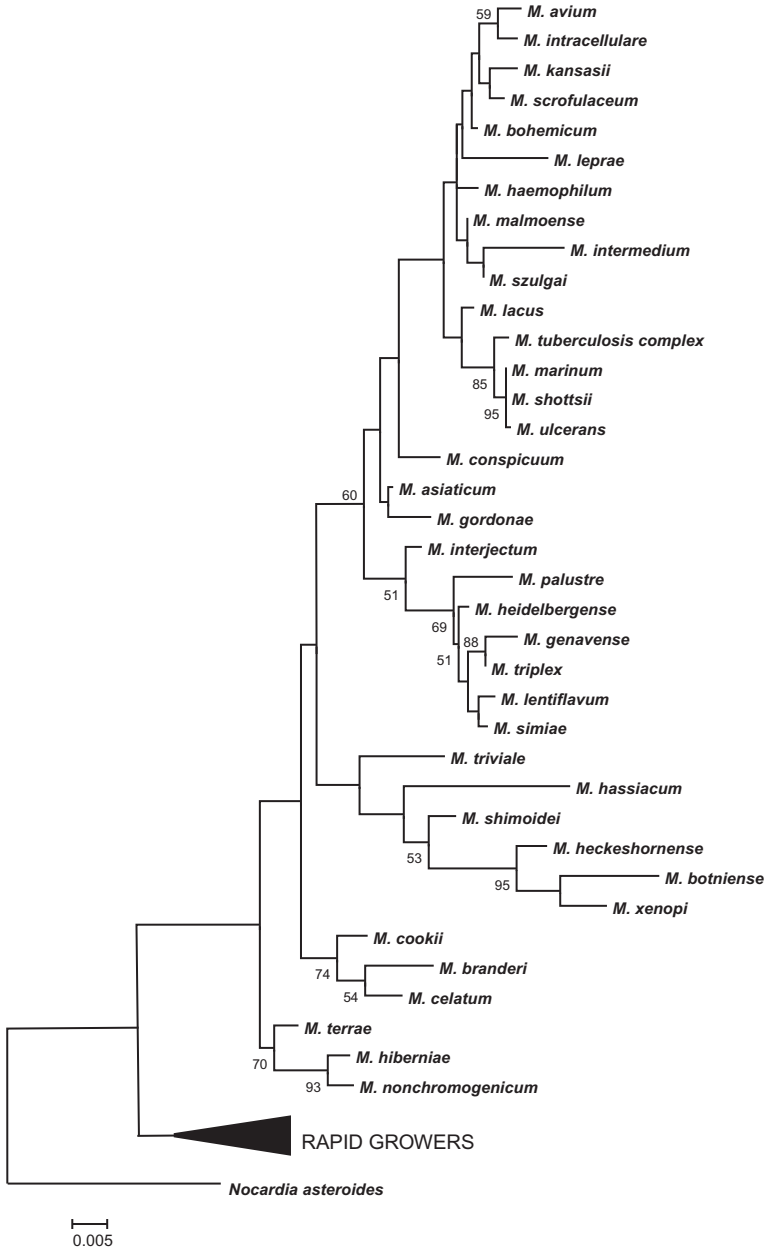

FIGURE 1.2 Phylogenetic tree of slow growers inferred with the neighbor-joining algorithm from the sequence dataset investigated by Tortoli (2012), bootstrapped 500 times. The tree was rooted with *Nocardia asteroides* as outgroup. Bar, 0.005 substitutions per nucleotide position.

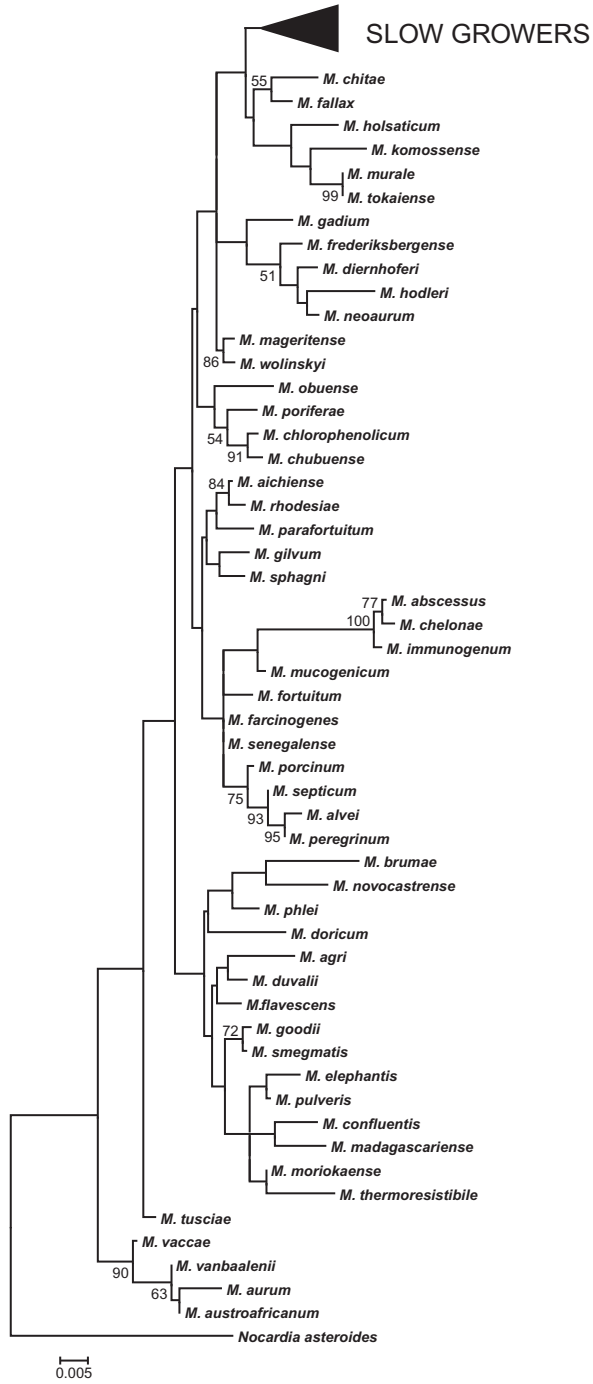

FIGURE 1.3 Phylogenetic tree of rapid growers inferred with the neighbor-joining algorithm from the sequence dataset investigated by Tortoli (2012), bootstrapped 500 times. The tree was rooted with *Nocardia asteroides* as outgroup. Bar, 0.005 substitutions per nucleotide position.

- A small group of slow growers related to *M. terrae* (*M. terrae* complex) have an even longer helix 18 (14-nucleotide insertion).

M. doricum and *M. tusciae* where regarded as the exceptions to the theorem above.

In the first decade of the 21st century, with the objective of improving the robustness of the phylogenetic trees, concatenated sequences of a few housekeeping genes were used for inference (Devulder et al., 2005; Mignard and Flandrois, 2008; Tortoli, 2012). The reduction of the weight of 16S rRNA, due to the presence of other genes in the concatenated sequence, was expected to minimize or even to sponge out the theory of the signatures. This was not the case, the 16S signatures continued to predict the topology of the phylogenetic supertree. *M. doricum* and *M. tusciae* were confirmed as the only exceptions.

The ultimate taxonomy of the genus *Mycobacterium* is being written right now. So far, two studies based on whole genomes have dealt with the phylogeny of the genus *Mycobacterium* (Fedrizzi et al., 2017; Tortoli et al., 2017). The number of genomes were different, as well as the approaches. Trees were reconstructed starting from the core genomes (concatenated sequences of the genes shared by all strains), the gene presence/absence clustering (Fedrizzi et al., 2017), and the average nucleotide identity (ANI) (Tortoli et al., 2017). The first tree was inferred from concatenated sequences using the maximum likelihood (ML) algorithm; the second was obtained, again with the ML algorithm, after converting the clusters to a binary matrix encoding the presence of each strain in the clusters; the third was constructed from the distance matrix of pairwise ANI-divergences using the ML, the neighbor-joining and the unweighted pair group method using arithmetic averages (UPGMA) algorithms. The similarity between topologies of different trees is striking.

The most ancestral branch leads to the species of the *M. abscessus-chelonae* complex. The other rapid growers emerge from a more recent branch giving rise to a major clade, accommodating the species of the *M. fortuitum-smegmatis* group, and a large number of dispersed species. Other progressively more recent branches lead to the *M. terrae* complex and to slow growers. Among slow growers several clusters of species are present. Two minor close groupings include the species related to *M. xenopi* and *M. celatum*. A large cluster includes the major pathogens (*M. tuberculosis*, *M. leprae*, *M. ulcerans*) and a number of species frequently involved in human diseases. Other clusters include the species of the *M. simiae* complex and the *M. avium* complex. Most of the 16S rRNA signatures are confirmed, only the one related to the *M. simiae* complex emerges poorly specific as part of the species presenting the signature do not cluster with *M. simiae*, but are dispersed among other slow growers not part of the complex. Interestingly, *M. doricum* and *M. tusciae* cluster again with rapid growers (Figs. 1.4 and 1.5).

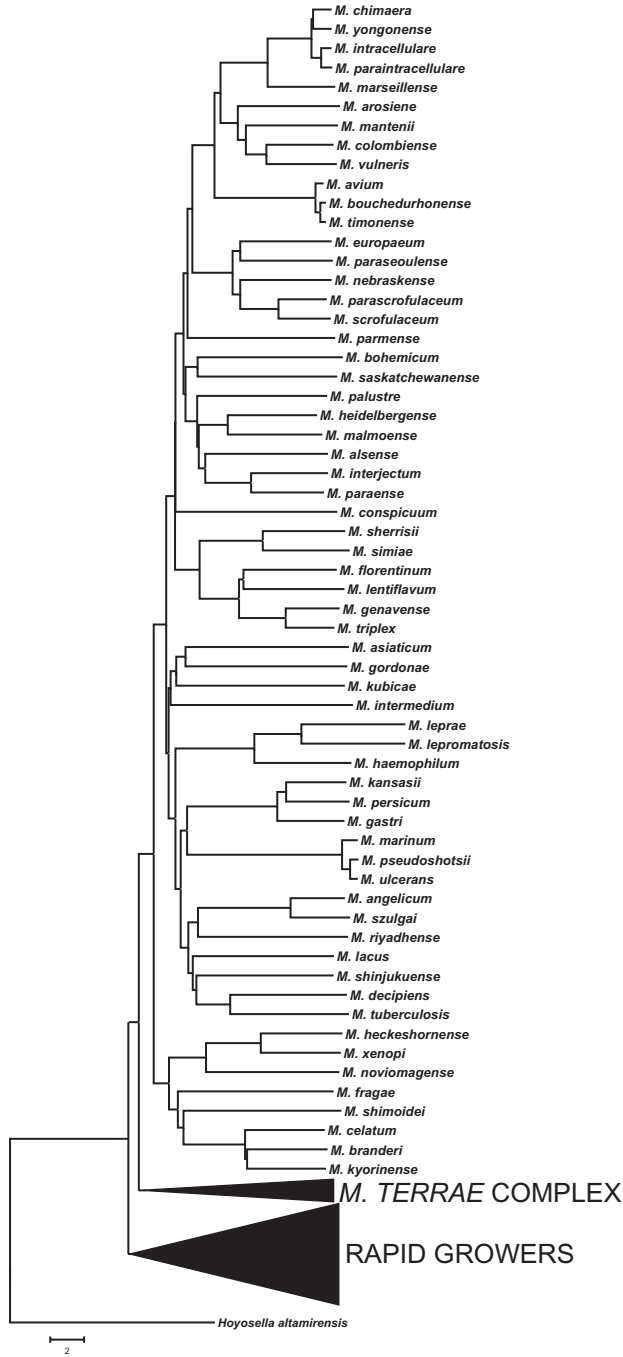

FIGURE 1.4 Phylogenetic tree of slow growers (excluding *M. terrae* complex) inferred with the maximum likelihood algorithm on 10,878 ANI-divergence scores (Tortoli, 2017), bootstrapped 500 times. The tree was rooted with *Hoyosella altamirensis* as outgroup. Bar, 2 units difference of ANI-divergence values.

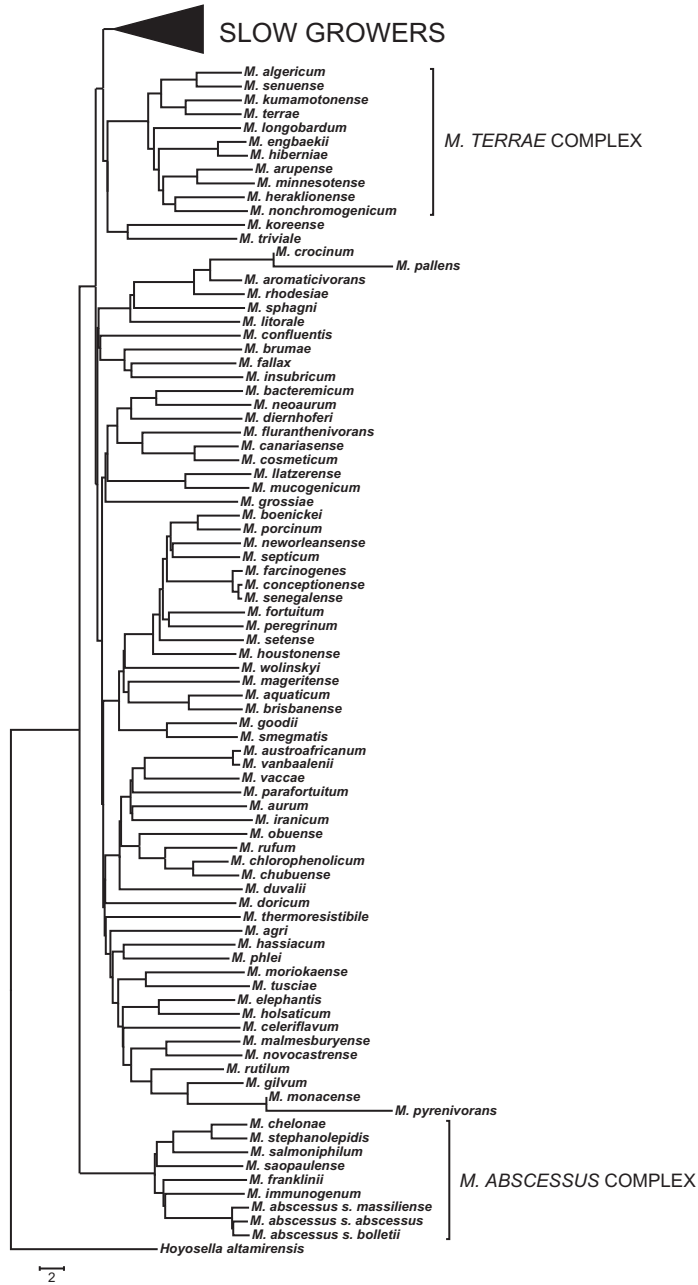

FIGURE 1.5 Phylogenetic tree of *M. terrae* complex and rapid growers inferred with the maximum likelihood algorithm on 10,878 ANI-divergence scores (Tortoli, 2017), bootstrapped 500 times. The tree was rooted with *Hoyosella altamirensis* as outgroup. Bar, 2 units difference of ANI-divergence values.

Several major traits characterize the phylogenetic scenario emerging from whole genome data. Primordial mycobacteria were rapid growers more closely related to the *M. abscessus-chelonae* complex. The evolution toward slow growers, likely associated with the acquisition of a 14-nucleotide insertion in the helix 18, leads to the species related to the present *M. terrae* complex. The deletion of two nucleotides in the helix 18 made the 12-nucleotide insertion the marker of the most successful slow growers. In contrast with previous knowledge, the lost of the whole 12-nucleotide insertion is not a specific marker of *M. simiae* complex. It is actually shared by a number of paraphyletic species.

Interestingly, the present pathogenic species and most of those frequently responsible of opportunistic diseases in humans appear to have evolved from a common ancestor. The latter hypothesis is not new, it was in fact hypothesized on the basis of the presence, in many such species, of genes related with virulence (van Ingen et al., 2012).

A recent paper (Gupta et al., 2018), published after the completion of present review, redraws in deep the classification of mycobacteria. The Authors split the genus *Mycobacterium* in five genera basing on molecular markers consisting, either in insertions/deletions of amino acids, or in proteins exclusively found in evolutionarily-related groups of species. Far from distorting the aforesaid phylogenetic reconstruction, the new genera exactly overlap the major clades of classical taxonomy: *M. abscessus* complex, other rapid growers, *M. terrae* complex and *M. triviale-related species*.

REFERENCES

Barbu, E., Lee, K.Y., Wahl, R., 1956. Contenu en bases puriqniue et pyrimidiniques des acides deoxyribonucleiques bacteriques. Ann. Inst. Pasteur. (Paris) 91, 212–224.

Böddinghaus, B., Rogall, T., Flohr, T., Blöcker, H., Böttger, E.C., 1990. Detection and identification of mycobacteria by amplification of rRNA. J. Clin. Microbiol. 28, 1751–1759.

Brennan, P.J., Nikaido, H., 1995. The envelope of mycobacteria. Annu. Rev. Biochem. 64, 29–63.

Chan, J.Z., Halachev, M.R., Loman, N.J., Constantinidou, C., Pallen, M.J., 2012. Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. BMC. Microbiol. 12, 302.

Devulder, G., Perouse, D.M., Flandrois, J.P., 2005. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. Int. J. Syst. Evol. Microbiol. 55, 293–302.

Fedrizzi, T., Meehan, C.J., Grottola, A., Giacobazzi, E., Fregni Serpini, G., Tagliazucchi, S., et al., 2017. Genomic characterization of nontuberculous mycobacteria. Sci. Rep. 7, 45258.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., et al., 2005. Opinion: re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3, 733–739.

Gupta, R.S., Lo, B., Son, J., 2018. Phylogenomics and comparative genomic studies robustly support division of the genus Mycobacterium into an emended genus Mycobacterium and four novel genera. Front. Microbiol. 9, 67.

Lehmann, K.B., Neuman, R., 1896. Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Disagnostik. Lehmann, J.F., Munich.

Linnaeus, C., 1735. Systema naturae, sive regna tria naturae systematice proposita per classes, ordines, genera et species. Apud Theodorum Haak, Leiden.

Mignard, S., Flandrois, J.P., 2008. A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. Int. J. Syst. Evol. Microbiol. 58, 1432–1441.

Rogall, T., Wolters, J., Flohr, T., Böttger, E.C., 1990. Towards a phylogeny and definition of species at the molecular level within the genus *Mycobacterium*. Int. J. Syst. Bacteriol. 40, 323–330.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Stackebrandt, E., Ebers, J., 2006. Taxonomic parameters revised: tarnished gold standards. Microbiol. Today 33, 152–155.

Stahl, D.A., Urbance, J.W., 1990. The division between fast- and slow-growing species corresponds to natural relationship among the mycobacteria. J. Bacteriol. 172, 116–124.

Tortoli, E., 2003. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin. Microbiol. Rev. 16, 319–354.

Tortoli, E., 2012. Phylogeny of the genus *Mycobacterium:* many doubts, few certainties. Infect., Genet. Evol. 12, 827–831.

Tortoli, E., Fedrizzi, T., Meehan, C.J., Trovato, A., Grottola, A., Giacobazzi, E., et al., 2017. The new phylogeny of the genus *Mycobacterium:* the old and the news. Infect. Genet. Evol. 56, 19–25.

van Ingen, J., Boeree, M.J., van Soolingen, D., Iseman, M.D., Heifets, L.B., Daley, C.L., 2012. Are phylogenetic position, virulence, drug susceptibility and in vivo response to treatment in mycobacteria interrelated? Infect. Genet. Evol. 12, 832–837.

Watson, J.D., Crick, F.H., 1953. Molecular structure of nucleic acid: a structure for deoxyribose nucleic acid. Nature 171, 737–738.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., et al., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.