Infection, Genetics and Evolution 12 (2012) 827–831

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Phylogeny of the genus *Mycobacterium*: Many doubts, few certainties

Enrico Tortoli *

Emerging Bacterial Pathogens Unit, San Raffaele Scientific Institute, Milan, Italy

ARTICLE INFO

Article history:
Available online 12 June 2011

Keywords:
Mycobacterium
Phylogeny
Genetic sequencing
Evolution
Multilocus phylogeny

ABSTRACT

The genus *Mycobacterium* is characterized by very limited interspecies genetic variability and this makes the definition of a robust phylogeny problematic. In this study a twofold phylogenetic approach was adopted. Phylogenetic trees were constructed using as targets the almost complete 16S rRNA gene sequences and the concatenated amino acid sequences coded by fragments of *hsp65* and *rpoB* genes. The comparison of the results made it possible to identify clusters of species sharing common phylogenetic pathway but for the majority of mycobacteria the definition of a robust phylogeny remained unreached.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Phylogenetic studies, initially based on phenotypic similarities (Bergey et al., 1923), made an enormous qualitative leap, from the 1960s onwards, thanks to the dramatic progress of genetic knowledge and, to a minor extent, to the development of mathematical models for the evolutionary analysis of sequences' similarities. Not less important was the contribution of another extraordinary achievement of the modern age, the global network, which makes genetic sequences available in real time on public domain databases.

The 16S rRNA gene has been the first and, for many years, the sole target of genetic sequencing in bacteria. Several unique characteristics make this gene the ideal candidate for mutations analysis. The 16S rRNA gene, devoted to the essential function of protein synthesis, has been present in all organisms since the beginning of evolution. It is furthermore characterized by an evolutionary rate high enough to produce interspecies variability but, at the same time, by a degree of conservation sufficient to minimize the intraspecies variability (Peix et al., 2009). Ribosomal genes are furthermore characterized by an often underestimated unique feature; differently from protein coding genes, whose silent mutations do not undergo the screen of the natural selection, in them every mutation has evolutionary relevance.

GenBank and the partners of the International Nucleotide Sequence Database Collaboration (http://www.insdc.org/) contain nowadays the 16S rDNA genetic sequences of thousands of mycobacterial strains; for many of them, including the type strains of every officially recognized species, the sequence of the almost complete gene is even available.

It is not surprising that almost all phylogenetic studies concerning the genus *Mycobacterium* (Rogall et al., 1990; Stahl and Urbance, 1990; Tortoli, 2003) are based on the 16S rDNA and in particular on the first 500 bp which include its major hypervariable regions. A nonnegligible limit of such phylogenetic reconstructions is however the poor robustness of the trees.

More recently the multilocus approach (Stackebrandt et al., 2002) has been used in two large studies (Devulder et al., 2005; Mignard and Flandrois, 2008) in which four and seven different nucleotide fragments, respectively, were concatenated.

The goal of present study, far from expecting to resolve the discrepancies present in the literature facing *Mycobacterium* phylogenesis, was to produce, targeting the amino acid residues composition instead of nucleotides, further data for future discussion. To increase its reliability two housekeeping genes were selected and concatenated. The idea of including other genes to improve the robustness of the tree, was set aside as the shortage, in the database, of the sequences related to genes other than the ones selected here would imply a dramatic cut of the number of species submitted to analysis. The results of such an approach were then compared, looking for similarities and diversities, with the ones that emerged from the analysis of the 16S rRNA gene.

2. Design and methods

The mycobacterial sequences used for the study were retrieved from GenBank.

The almost complete 16S rRNA gene sequences related to the type strains of different *Mycobacterium* species were collected; they were aligned using the Klustal W program (Thompson 1944) and were trimmed to the longest fragment available for all of them, which corresponded to a stretch of about 1400 bp spanning over the *Escherichia coli*-corresponding positions 53–1460.

* Address: Emerging Bacterial Pathogens Unit, San Raffaele Scientific Institute, San Michele Building, via Olgettina 60, 20132 Milan, Italy. Tel.: +39 02 26435684; fax: +39 02 26435183.
E-mail address: tortoli.enrico@hsr.it

1567-1348/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2011.05.025

The hypervariable fragment of *hsp65* (Telenti et al., 1993), a highly conserved housekeeping gene coding for a 65 kDa protein involved in the folding, assembly and transport of proteins, was chosen because of the completeness of the respective database. The sequences of the above mentioned fragment obtained from the type strains of different mycobacterial species were retrieved from GenBank; once aligned, they were trimmed to 399 nucleotides (starting from *Mycobacterium tuberculosis*-homologous position 443) and the corresponding 133 codons were translated to protein sequence.

Among the remaining nucleotide-databases related to mycobacterial genes, which are widely incomplete, the choice fell upon *rpoB*, probably the target most investigated in recent years. Of such a housekeeping gene, which codes for the β-subunit of RNA polymerase, two different regions have been proposed as target of genetic sequencing in mycobacteria (Adékambi et al., 2003; Kim et al., 1999), the largest and more variable, starting at *Mycobacterium smegmatis*-corresponding position 2593, was chosen here. Such sequences were available in GenBank for type strains belonging to 88 mycobacteria only; once aligned and trimmed to the largest fragment available, their length turned out to be variable, in different species, from 687 to 657 nucleotides corresponding to 229–219 codons; the respective amino acid sequences were then determined.

For each of the 88 species whose *rpoB* and *hsp65* sequences were both available in GenBank the translated amino acid strings were concatenated in a single stretch. The *hsp65*, and the 16S rDNA, sequences of the species other than the 88 above were no further investigated.

The phylogenetic analysis was conducted separately on the two sets of 88 aligned sequences: the nucleotide one, corresponding to 16S rDNA, and the protein one corresponding to concatenated *rpoB* and *hsp65* fragments. To each of them was added the corresponding sequence of *Nocardia farcinica* which was used as the outgroup. Trees were constructed using the neighbour-joining method under the total gap remotion and Kimura's two-parameter substitution model (Kimura, 1980) and visualized using the MEGA 5 software package (Tamura et al., 2011). The bootstrap analysis was conducted on 1000 resamplings.

3. Results

The phylogenetic tree inferred from the sequences of the almost complete 16S rDNA was in agreement with information obtained previously (Fig. 1) on the basis of the first third of such gene (Rogall et al., 1990; Stahl and Urbance, 1990). Rapid growers were clearly separated from slow growers and within the latter, two major clusters were evidently detectable. A group of slowly growing mycobacteria including *Mycobacterium simiae* presented the helix 18 of rRNA 12 nucleotides shorter than the majority of others while the group including the members of *Mycobacterium terrae* complex presented the helix 18 two nucleotides longer. A further well defined cluster was the one including the species of the *Mycobacterium avium* complex (MAC). Among rapid growers, two major clusters included the species related to *Mycobacterium fortuitum* and *Mycobacterium chelonae* respectively, with other mycobacteria scattered in a number of minor branches. A cluster including thermotolerant rapidly growing mycobacteria, a group tentatively proposed on the basis of an extra cytosine et *E. coli*-homologous position 184 (Kirschner et al., 1993; Springer et al., 1996) was hardly recognizable. The tree was characterized by poor robustness with less than a quarter of the nodes supported by bootstrap > 80%.

The position of three species only, among the 88 investigated, revealed somehow unexpected: the slowly growing *Mycobacterium tusciae* clustered with rapid growers, *Mycobacterium triviale* was

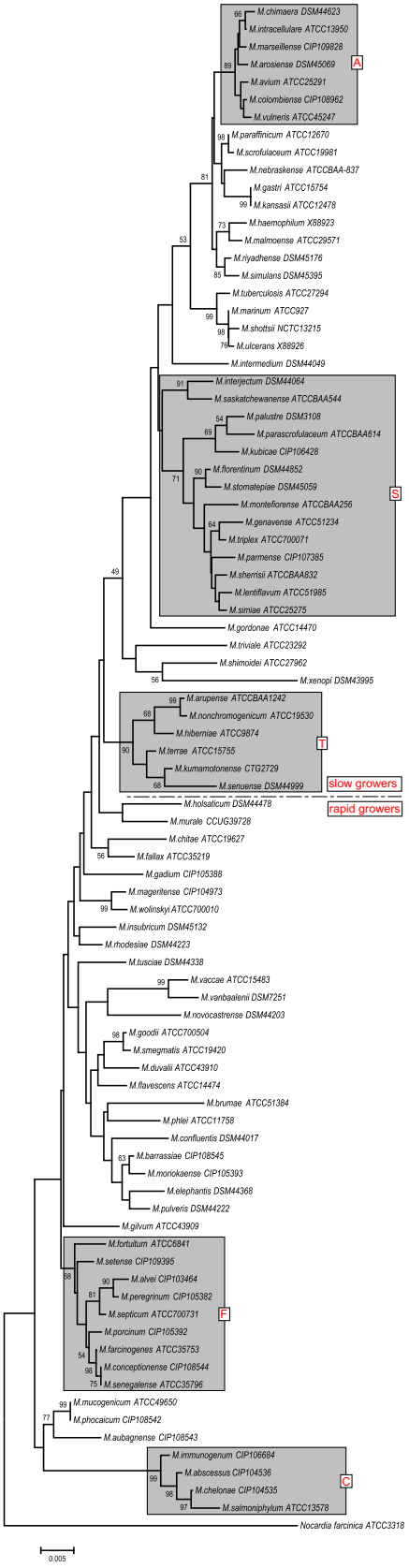

Fig. 1. Phylogenetic tree based on 16S rRNA gene almost complete sequences and constructed using the neighbour-joining method and the MEGA 5 package. The significance of branches (when >50) is indicated by bootstrap values calculated on 1000 replicates. Bar, 5 substitutions per 1000 nucleotides. Major clusters: (A) MAC; (C) *M. chelonae* group; (F) *M. fortuitum* group; (S) *M. simiae* group; and (T) *M. terrae* complex.

included among the slow grower despite the presence of a short helix 18, *Mycobacterium intermedium* although presenting the genetic signature (short helix 18) of the species included in the *M. simiae* group, was located in a separate branch.

The information provided by the evolutionary reconstruction based on the amino acid chain coded by the hypervariable fragment of *hsp65* gene appeared quite confusing because of the presence of a conspicuous number of slow growers (*Mycobacterium arupense*, *Mycobacterium gordonae*, *Mycobacterium nonchromogenicum*, *Mycobacterium hiberniae*, *M. triviale* and *M. tusciae*) scattered among rapid growers (Fig. 2). The MAC corresponded to a well defined cluster but included also the unrelated species *Mycobacterium kubicae*; other groups appeared fragmented: only 11 of the species related to *M. simiae* belonged to the same branch while the *M. terrae* complex split in three clearly distinct branches (including, the first: *Mycobacterium kumamotonense*, *Mycobacterium senuense* and *M. terrae*; the second: *M. nonchromogenicum* and *M. hiberniae*; and the third: *M. arupense* alone). The low resolution of the amino acid sequence coded by the *hsp65* gene was evident in the MAC branch, whose members shared an identical amino acid string, and within rapid growers characterized by several groups with an identical sequence, the largest of which included 12 species. Very poor was the robustness of the trees.

A clear distinction of slow- from rapid-growers was evident in the amino acid sequence coded by the hypervariable 700 bp fragment of *rpoB* gene (Fig. 3). The clusters of MAC and *M. terrae* complex were well recognizable. In contrast only seven of the slowly growing species characterized by a short helix 18 in 16S rDNA converged in the same branch, with the others being scattered. Among the rapid growers of note were the two clusters including the species related to *M. fortuitum* and *M. chelonae* respectively. The proportion of nodes supported by bootstrap >80% was close to 40%.

As expected, the evolutionary reconstruction based on the concatenation of the sequences of *rpoB* and *hsp65* produced a substantial increase in the robustness of the tree with bootstrap of about half of the nodes overcoming the 80% mark (Fig. 4).

From the comparison of the trees originating from the 16S rDNA nucleotide and the concatenated amino acid sequences significant similarities emerged.

The most important consensus concerned the clearly distinct evolutionary pathways of slow growing mycobacteria from the ones with rapid growth with the latter presenting as the ancestral feature. The close clustering of slow growers unambiguously supported their monophyletic origin in contrast with the polyphyletic evolution characterizing rapid growers.

The most consistent cluster, present in each of the reconstructions investigated here, was the one including the species belonging to the MAC. Equally unquestionable appeared the phylogenetic significance of the cluster including the species of *M. terrae* complex; this group, although missing in the tree based on *hsp65*, was in fact characterized by very high bootstrap in the trees obtained from 16S rDNA, *rpoB* and in the one resulting from concatenated amino acid sequences.

Controversial appeared the phylogenetic position of the species considered genetically related to *M. simiae*. All the species presenting a short helix 18 but one, *M. intermedium*, clustered together in the evolutionary reconstruction based on the 16S rDNA; they were, in contrast fragmented in a major and several minor groups in the ones based on *hsp65* and *rpoB*.

Among slow growers, minor consistent clusters, characterized by high bootstrap values, included: *Mycobacterium marinum*, *Mycobacterium ulcerans* and *Mycobacterium shottsii*, and *Mycobacterium kansasii* and *Mycobacterium gastri* respectively. Among rapid growers close clustering characterized the species *Mycobacterium goodii* and *M. smegmatis*, and *Mycobacterium barrassiae* and *Mycobacterium moriokaense* respectively.

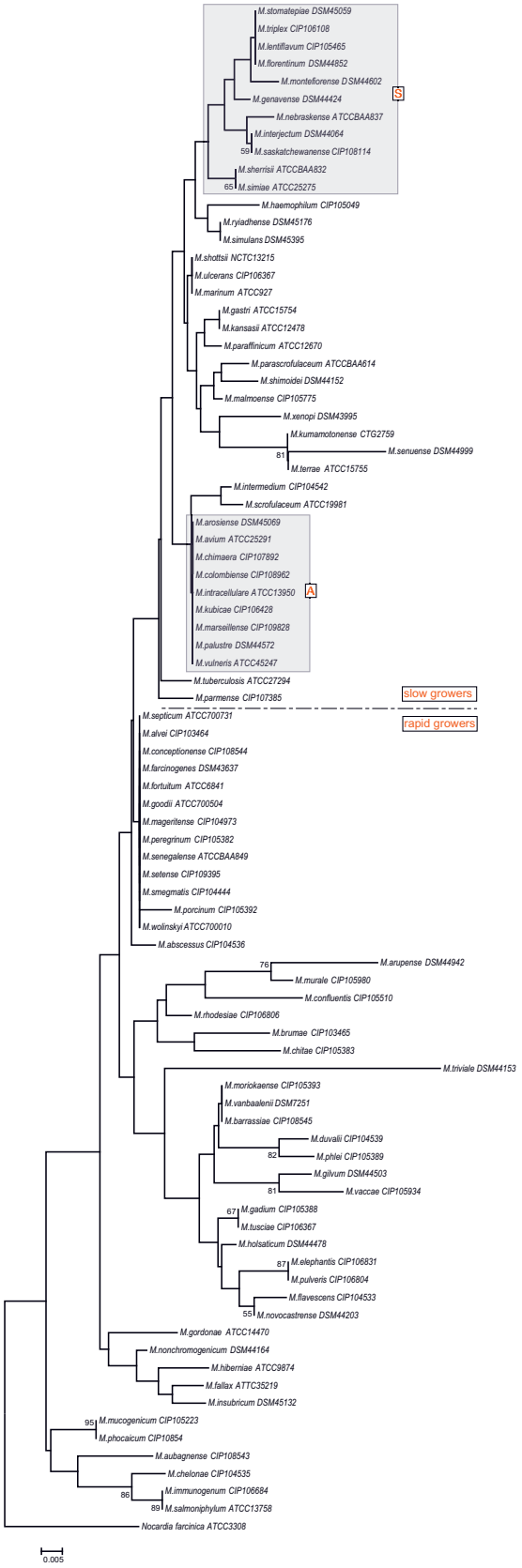

Fig. 2. Phylogenetic tree based on amino acid sequence coded by the hypervariable fragment of *hsp65* gene and constructed using the neighbour-joining method and the MEGA 5 package. The significance of branches (when >50) is indicated by bootstrap values calculated on 1000 replicates. Bar, 5 substitutions per 100 amino acid residues. Major clusters: (A) MAC and (S) *M. simiae* group.

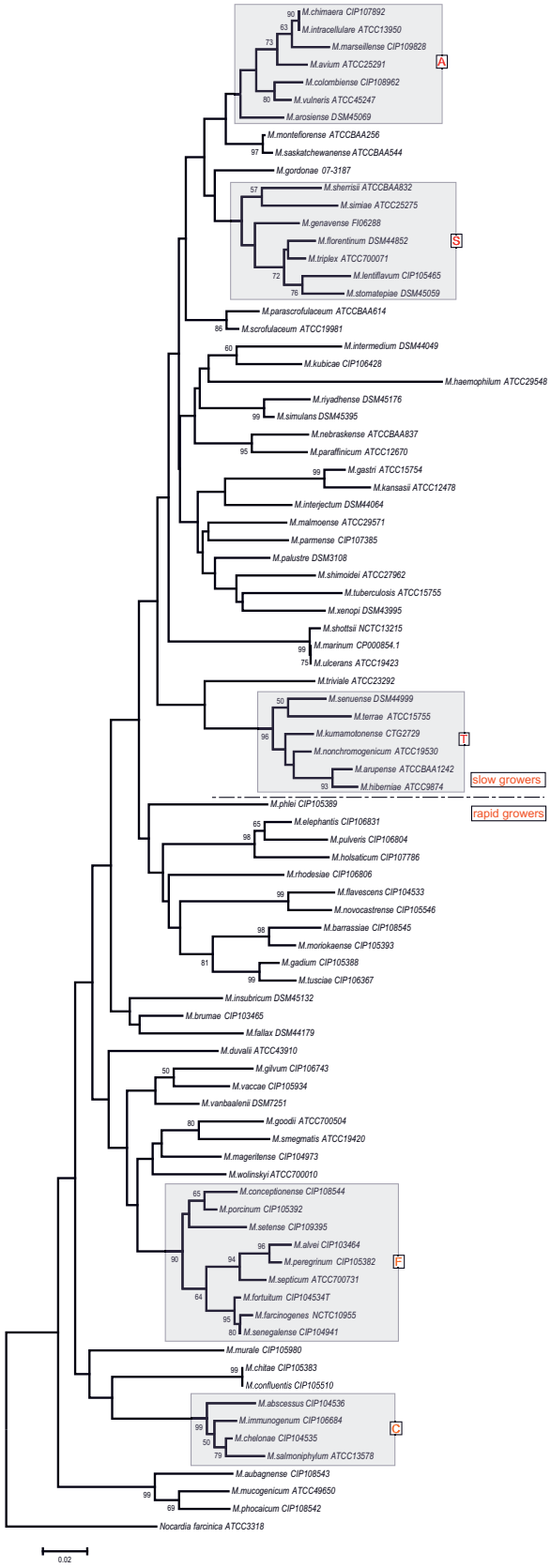

Fig. 3. Phylogenetic tree based on amino acid sequence coded by the hypervariable fragment of *rpoB* gene and constructed using the neighbour-joining method and the MEGA 5 package. The significance of branches (when >50) is indicated by bootstrap values calculated on 1000 replicates. Bar, 2 substitutions per 100 amino acid residues. Major clusters: (A) MAC; (C) *M. chelonae* group; (F) *M. fortuitum* group; (S) *M. simiae* group; and (T) *M. terrae* complex.

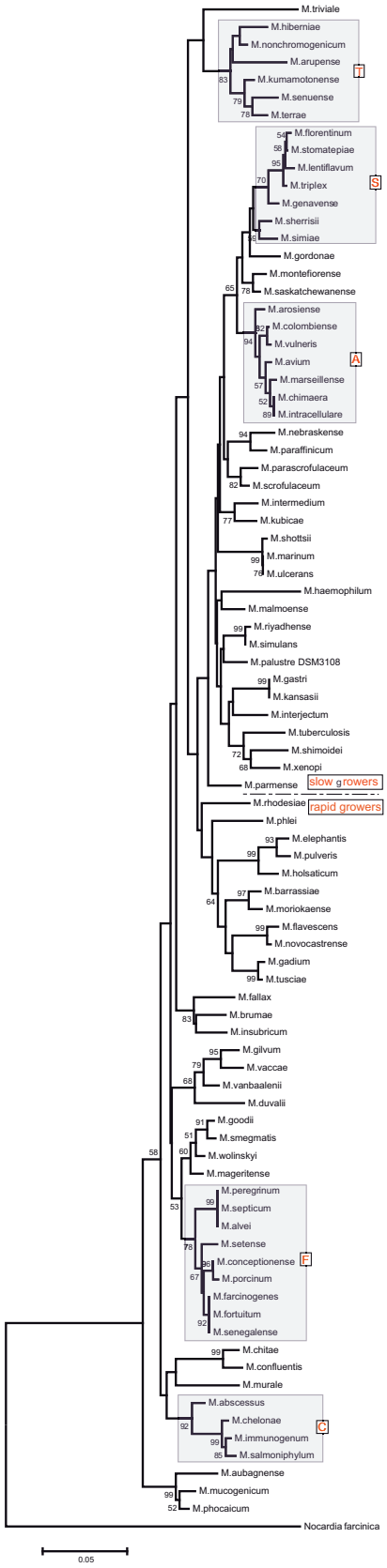

Fig. 4. Phylogenetic tree based on concatenated amino acid sequence coded by the hypervariable fragments of *rpoB* and *hsp65* genes and constructed using the neighbour-joining method and the MEGA 5 package. The significance of branches (when >50) is indicated by bootstrap values calculated on 1000 replicates. Bar, 5 substitutions per 100 amino acid residues. Major clusters: (A) MAC; (C) *M. chelonae* group; (F) *M. fortuitum* group; (S) *M. simiae* group; and (T) *M. terrae* complex.

The branches of the trees corresponding to rapidly growing mycobacteria made the interpretation of their evolution problematic. Quite consistent were only the clusters of the species related to *M. fortuitum* and *M. chelonae* respectively.

Notably the apparently improper positioning of *M. tusciae* and *M. triviale* present in the tree based on 16S was confirmed in the tree constructed with concatenated nucleotide sequences.

4. Discussion

The availability of computerized mathematical models often generates in users the conviction that phylogenetic analysis is an exact science. This belief, however, is often replaced by a feeling of frustration because of the difficulty of obtaining consistent results. The problem is even more complicated with mycobacteria which are characterized by interspecies similarity clearly higher than in other bacteria (Devulder et al., 2005).

The more popular targets of phylogenetic investigations remain at present the ribosomal operon and multiple concatenations of housekeeping genes (Stackebrandt et al., 2002).

In this study the 16S rDNA and two protein-coding housekeeping genes were investigated but, instead of combining the three sequences in a single stretch it was preferred to compare the phylogenetic trees based on 16S nucleotide sequence with the one emerging from the concatenation of amino acid sequences of the latter two genes; evolutionary credit was given only to clades characterized by consensus in the two approaches while the others were considered uncertain.

This study adds to the existing evidence that current mycobacterial species originated from an ancestral rapidly growing mycobacterium following various evolutionary branches only one of which leads to modern slow growers (Devulder et al., 2005; Mignard and Flandrois, 2008).

Within the phylum of slowly growing mycobacteria, a sharp-cut phylogenetic outline is evident for the clusters including the species belonging to MAC and the ones of the *M. terrae* complex respectively. Of note, two complexes overlapping the clusters above had been defined on the basis of the presence of shared phenotypic features, far before their genetic relatedness was acknowledged.

The monophyletic origin of MAC, already reported for *M. avium* and *Mycobacterium intracellulare* (Devulder et al., 2005; Mignard and Flandrois, 2008; Rogall et al., 1990) was confirmed here for all the seven species included in the complex.

Little was known about the evolution of *M. terrae* complex; previous studies, furthermore limited to three or four species, often split such group among separated branches (Devulder et al., 2005; Mignard and Flandrois, 2008) and even question about its position among slow growers. In the present analysis the monophyletic evolution within slow growers, furthermore supported by homogeneous phenotypic features, looks evident for all the seven species taking part in it.

The distribution of the slowly growing species presenting a short helix 18 in the 16S rRNA gene appears less clear: all but one cluster together in the tree based on such gene and only seven of them belong to the same branch, moreover very robust, in the tree inferred from concatenated amino acid sequences. Such arrangement is compatible with the analysis of Mignard and Flandrois (2008) but is not confirmed by others (Devulder et al., 2005). A common phylogeny can be reasonably assumed only for the seven species above while for the others the short helix 18 is probably an independent late acquisition. Two facts apparently support such hypothesis: the seven species clustered in the tree constructed on the basis of the amino acid sequence fully overlap a sub-cluster of the tree based on the 16S and, at phenotypic level, such species are the only mycobacteria characterized by a three-clustered HPLC profile of cell wall mycolic acids (Tortoli et al., 2011).

As far as the rapidly growing species are concerned, among the multiple branches present, a phylogenetic significance appears evident only for the ones including *M. fortuitum* and *M. chelonae* respectively. The clear cut separation of the latter two clusters is consistent with previous studies (Devulder et al., 2005; Mignard and Flandrois, 2008; Rogall et al., 1990) and, in one of them (Mignard and Flandrois, 2008), they are even characterized by a separation early in the nodes compatible with present data.

Apart from the points above in which several reliable branches are clearly defined, in particular for slow growers, many doubts remain about the phylogenetic position of most rapid growers as confirmed by the contrasting data of existing literature.

Nevertheless despite the important gap affecting present knowledge of mycobacterial phylogeny it seems likely that new horizons will be soon opened up thanks to the rapidly increasing availability of sequences compassing the whole bacterial genome.

Acknowledgement

The author is grateful to Mag Lafferty for the linguistic revision of the paper.

References

Adékambi, T., Colson, P., Drancourt, M., 2003. *RpoB*-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. J. Clin. Microbiol. 41, 5699–5708.

Bergey, D.H., Harrison, F.C., Breed, R.S., Hammer, B.W., Huntoon, F., 1923. Bergey's Manual of Determinative Bacteriology, first ed. The William & Willis Co., Baltimore.

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.

Kim, B.J., Lee, S.H., Lyu, M.A., Kim, S.J., Bai, G.H., Kim, S.J., Chae, G.T., Kim, E.C., Cha, C.Y., Kook, Y.H., 1999. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene *rpo*B. J. Clin. Microbiol. 37, 1714–1720.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.

Kirschner, P., Springer, B., Vogel, U., Meier, A., Wrede, A., Kiekenbeck, M., Bange, F.C., Böttger, E.C., 1993. Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory. J. Clin. Microbiol. 31, 2882–2889.

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.

Peix, A., Ramírez-Bahena, M.H., Velàsquez, E., 2009. Historical evolution and current status of the taxonomy of the genus *Pseudomonas*. Infect. Genet. Evol. 9, 1132–1147.

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.

Springer, B., Stockman, L., Teschner, K., Roberts, G.D., Böttger, E.C., 1996. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. J. Clin. Microbiol. 34, 296–303.

Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kampfer, P., Maiden, M.C.J., Nesme, X., Rossellò-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B., 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 52, 1043–1047.

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 [Epub ahead of print].

Tamura, K., Peterson, D., Peterson, N., Steker, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. [Epub ahead of print]

Telenti, A., Marchesi, F., Balz, M., Bally, F., Böttger, E.C., Bodmer, T., 1993. Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. J. Clin. Microbiol. 31, 175–178.

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

Tortoli, E., Böttger, E.C., Fabio, A., Falsen, E., Gitti, Z., Grottola, A., Klenk, H.P., Mannino, R., Mariottini, A., Messinò, M., Pecorari, M., Rumpianesi, F., 2011. *Mycobacterium europaeum* sp. nov., a scotochromogenic species related to *Mycobacterium simiae* complex. Int. J. Syst. Evol. Microbiol. 61, in press.