Survey of 150 strains belonging to the *Mycobacterium terrae* complex and description of *Mycobacterium engbaekii* sp. nov., *Mycobacterium heraklionense* sp. nov. and *Mycobacterium longobardum* sp. nov.

Enrico Tortoli,[1] Zoe Gitti,[2] Hans-Peter Klenk,[3] Stefania Lauria,[4] Roberta Mannino,[4] Paola Mantegani,[1] Alessandro Mariottini[5] and Ioannis Neonakis[2]

Correspondence
Enrico Tortoli
e.tortoli@libero.it

[1]Emerging Bacterial Pathogens Unit, San Raffaele Scientific Institute, Milan, Italy

[2]Department of Clinical Bacteriology, Parasitology, Zoonoses and Geographical Medicine, University Hospital of Heraklion, Heraklion, Greece

[3]DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

[4]Microbiology and Virology Laboratory, Careggi University Hospital, Florence, Italy

[5]Cytogenetics and Genetics Unit, Careggi University Hospital, Florence, Italy

A thorough phenotypic and genotypic analysis of 150 strains belonging to the *Mycobacterium terrae* complex resulted in the identification of a number of previously unreported sequevars (sqvs) within the species known to belong to the complex. For the species *Mycobacterium arupense*, three sqvs were detected in the 16S rRNA gene, six sqvs in the *hsp65* gene and 15 sqvs in the *rpoB* gene; in *Mycobacterium senuense* two sqvs were present in each of the three genetic regions; in *Mycobacterium kumamotonense* four, two and nine sqvs were found, respectively, and in *M. terrae* three, four and six sqvs were found, respectively. The inappropriate inclusion of *Mycobacterium triviale* within the *M. terrae* complex was confirmed. The limited utility of biochemical tests and of mycolic acid analyses for the differentiation of the members of *M. terrae* complex was also confirmed. The survey allowed the recognition of three previously undescribed species that were characterized by unique sequences in the 16S rRNA, *hsp65* and *rpoB* genes. *Mycobacterium engbaekii* sp. nov. (proposed previously 40 years ago but never validly published) was characterized by pink photochromogenic pigmentation and rapid growth; phylogenetically it was related to *Mycobacterium hiberniae*. The type strain of this species, of which eight strains were investigated, is ATCC 27353^T (=DSM 45694^T). A cluster of 24 strains was the basis for the description of *Mycobacterium heraklionense* sp. nov., which has an intermediate growth rate and is unpigmented; nitrate reductase activity is typically strong. Closely related to *M. arupense* with respect to the 16S rRNA gene, *M. heraklionense* sp. nov. could be clearly differentiated from the latter species in the other genetic regions investigated. The type strain is NCTC 13432^T (=LMG 24735^T=CECT 7509^T). *Mycobacterium longobardum* sp. nov., represented in the study by seven strains, was characterized by a unique phylogenetic location within the *M. terrae* complex, clearly divergent from any other species. The type strain is DSM 45394^T (=CCUG 58460^T).

Abbreviations: MTC, *Mycobacterium terrae* complex; PRA, PCR restriction analysis; sqvs, sequevars.

The GenBank accession numbers of the almost complete 16S rRNA sequences for the type strains of *M. engbaekii* sp. nov., *M. heraklionense* sp. nov. and *M. longobardum* sp. nov. are AF480577, GU084182 and JN571166, respectively. Full details of the GenBank accession numbers determined for all the strains in this study are available as a supplementary table.

Two supplementary figures and six supplementary tables are available with the online version of this paper.

The *Mycobacterium terrae* complex (MTC) is a group within the genus *Mycobacterium* that was created in the 1970s to gather *Mycobacterium nonchromogenicum*, *Mycobacterium terrae* and *Mycobacterium triviale*, three species that could not be differentiated using the biochemical and cultural methods available at that time. *M. nonchromogenicum* described in 1965 (Tsukamura, 1965), *M. terrae* described in 1966 (Wayne, 1966) and *M. triviale* described in 1970 (Kubica *et al.*, 1970) share

important cultural features including intermediate growth rate (from 5 to 15 days are required for the development of clearly visible colonies from diluted inocula on solid media) and lack of pigmentation.

In the early 1990s, the detection of a unique genetic signature: the presence of a two-nucleotide insertion in helix 18 of the 16S rRNA gene (Kirschner *et al.*, 1993; Springer *et al.*, 1996), in comparison with other slow-growing mycobacteria, confirmed the consistence of the MTC. This signature still remains the most reliable marker for the attribution of mycobacteria to the MTC. At the same time, the presence, in *M. triviale*, as also seen in rapid growers, of a helix 18 that is 14 nt shorter, unquestionably demonstrated the unrelatedness of this species to the complex.

A gap of more than 20 years separates the recognition of the original members of the MTC from the description of a novel species related to this group, *Mycobacterium hiberniae* (Kazda *et al.*, 1993). This novel mycobacterium is characterized by a unique phenotypic feature, the pink pigmentation of the colonies, but the major role in its differentiation was played by the genetic analysis that was, at that time, beginning to emerge. The boom years of the MTC start however in 2006 with the description of three novel species *Mycobacterium kumamotonense*, *Mycobacterium senuense* and *Mycobacterium arupense* (Cloud *et al.*, 2006; Masaki *et al.*, 2006; Mun *et al.*, 2008).

Identification at species level within the members of the MTC still remains problematic As well as biochemical and cultural tests (Wayne & Kubica, 1986), the analysis of cell-wall lipids is poorly discriminative. More recently, DNA probes specific for the species of the MTC have not been introduced by any of the commercial hybridization kits; probably as a consequence of the limited interest aroused by organisms that have been grossly labelled as non-pathogenic. Unexpectedly, even their identification by means of genetic sequencing remains elusive as hundreds of sequences related to members of the MTC are stored in public domain databases without any, or with unreliable, species allocations.

The aim of this study was to investigate the phylogenetic and taxonomic structure of the MTC with a large number of isolates and to make a panel of species-specific genetic sequences characterized by unambiguous labels available in GenBank.

METHODS

Strains. For this study, all the strains assigned to the MTC on the basis of routine identifications performed in our laboratory from 1996 onwards were investigated. All such strains (a total of 156) had been grown from clinical specimens and stored at −80 °C. While the large majority of them were isolated either in Careggi Hospital laboratory or in other Italian hospitals, 27 strains were obtained from laboratories of other countries. After thawing, each strain was grown on Middlebrook 7H11 medium at 37 °C. The reference strain of '*M. engbaekii*' (ATCC 27353) and type strains of *M. hiberniae* (ATCC 49874^T), *Mycobacterium kumamotonense* (DSM 45093^T), *Mycobacterium senuense* (DSM 44999^T) and *M. terrae* (CIP 104321^T) were also included in the study; other reference strains investigated were *M. nonchromogenicum* (PI140330001) and *M. triviale* (PI141030004).

Genetic sequencing. Three different regions were chosen for genetic characterization: the genes coding for the 16S rRNA, for the 65 kDa heat shock protein (*hsp65*) and for the β-subunit of the RNA polymerase (*rpoB*). For the 16S rRNA gene, a region spanning 479 bp was sequenced starting from the position corresponding to *Escherichia coli* position 28, according to a previously reported procedure (Reischl *et al.*, 1998). The almost complete 16S rRNA gene sequences were determined for the strains which are proposed as representing novel species in this study. A stretch of 399 bp was sequenced in the hypervariable region of the *hsp65* gene (McNabb *et al.*, 2004) starting from the position corresponding to *Mycobacterium tuberculosis* position 443. For the *rpoB* gene, the stretch recently proposed for the differentiation of rapidly growing mycobacteria (Adékambi *et al.*, 2003) was investigated; the length of the nucleotide sequence (starting at the position corresponding to *Mycobacterium smegmatis* position 2554) ranged from 711 to 726 bp in different strains. In all the regions above, both the forward and reverse strands were determined using Big Dye terminator chemistry and an AB3730 DNA sequencer (Applied Biosystems).

For the designation of sequevars (sqvs) for which assignation to a species was possible, the first three letters (capitalized) of the species name were used followed, for the 16S rRNA gene, by a small letter (a, b, . . .), for *hsp65*, by a number (1, 2, . . .) and, for *rpoB*, by a Roman numeral (i, ii, . . .).

Sequences of protein-coding genes (*hsp65*, 133 codons; *rpoB*, 237–242 codons) were also translated to the amino acid residue compositions to distinguish silent mutations from ones affecting protein structure.

Phylogenetic analysis. The phylogenetic analysis was conducted according to the neighbour-joining method (Saitou & Nei, 1987) under the total gap removal and Kimura's two-parameter substitution model (Kimura, 1980), and was evaluated by bootstrap analysis based on 1000 replicates using MEGA software version 5 (Tamura *et al.*, 2011). The trees were rooted using *M. tuberculosis* as the outgroup. Sequences for *M. tuberculosis* retrieved from GenBank had been added previously to various alignments.

The phylogenetic reconstruction based on 16S rRNA gene sequences included 22 different sqvs detected among the strains investigated in this study. The sequences, downloaded from GenBank, of the most closely related slowly and rapidly growing mycobacteria were also added.

To improve the robustness of the tree (Devulder *et al.*, 2005; Mignard & Flandrois, 2008; Stackebrandt *et al.*, 2002), the sequences of the three genetic regions were concatenated in a single filament including a number of nucleotides ranging, in different strains, from 1589 to 1604 (16S rRNA, 479; *hsp65*, 399; *rpoB*, 711–726). In the investigation using the concatenated sequences, all the combinations (75 in total) of 16S rRNA, *hsp65* and *rpoB* sqvs detected in our strains were included.

Lipid investigations. HPLC of cell-wall mycolic acids was carried out on all the strains after esterification to bromophenacyl esters as described previously (CDC, 1996).

Biochemical and cultural tests. For all the strains, nitrate reduction, growth rate and pigmentation of colonies were investigated. For the strains considered to represent novel species, a number of randomly selected strains (10 of *M. heraklionense* sp. nov., five of *M. engbaekii* sp. nov. and two of *M. longobardum* sp. nov.) were investigated with a wider panel of tests according to standard procedures (Kent & Kubica, 1985).

Susceptibility testing. Susceptibility testing was performed on randomly selected strains belonging to the novel species proposed here (four of *M. heraklionense* sp. nov., four of *M. engbaekii* sp. nov. and two of *M. longobardum* sp. nov.). The MICs of drugs selected for their activity on slowly growing mycobacteria were determined using commercially available microdilution plates (SLOMYCO, VersaTREK) following the Clinical and Laboratory Standards Institute (CLSI) recommendations (CLSI, 2011).

RESULTS

Genetic sequencing

The alignment of the 22 sqvs detected in the 16S rRNA gene of the 156 strains investigated added to the existing evidence that *M. triviale* does not belong to the MTC. Helix 18 of the 16S rRNA gene was in fact, 14 nt shorter in the six strains of *M. triviale* present in our panel than in the remaining strains. Such a feature not only excludes *M. triviale* from the MTC, but even places this species within the group of rapid growers. These six strains were therefore excluded from the study.

Six of the 21 remaining 16S rRNA gene sqvs turned out to overlap the sequences of the type strains of the six species known to belong to the MTC and this allowed them to be assigned to *M. arupense*, *M. nonchromogenicum*, *M. hiberniae*, *M. senuense*, *M. terrae* and *M. kumamotonense*. Another sequevar was 100 % identical to the sequence of '*Mycobacterium engbaekii*'. Only nine of the remaining 16 sqvs were found in the GenBank database, five of them had been previously deposited by one of the authors of this study, while the species assignation of two others was either lacking or incorrect. The pairwise matrix of distances (Table S1 available at IJSEM online) allowed the identification of four clusters of sqvs which were assigned to the species *M. arupense*, *M. senuense*, *M. kumamotonense* and *M. terrae* on the basis of the inclusion of the sequences of the respective type strains. Within such clusters the intraspecies variability was <1 %. One sequevar shared by two strains (FI-07105/FI-11038) differed by only one nucleotide from '*Mycobacterium paraterrae*' (Lee *et al.*, 2010). The three remaining sqvs (NEW1, GN-9188, FI-09379), differed from all type strains and did not fit any cluster.

A total of 30 sqvs were found in the *hsp65* gene. With the similarity matrix (Table S2), eight clusters were recognized in this region, which were characterized by pairwise distances <3 % (with three exceptions with values up to 3.36 %). Five clusters were immediately attributed to the species *M. arupense*, *M. senuense*, *M. kumamotonense*, *M. terrae* and '*M. engbaekii*', because of the inclusion of the sequences of the respective type strains; they all were detected in strains assigned to the same species on the basis of the 16S rRNA gene sequence. One of the remaining clusters (NEW2, including two sqvs), although clearly separated from the cluster of *M. arupense*, had been detected in strains assigned, on the basis of 16S rRNA gene sequences, to the latter species. The other two clusters (NEW3 and NEW4) included two and three orphan (unassigned to any known species) sqvs, respectively. Interestingly two (FI-07105 and FI-11038) of the three sqvs included in NEW4 belonged to the strains showing close similarity with '*M. paraterrae*' in 16S rRNA gene sequences, but clearly differed from the latter in the *hsp65* gene region. In the species *M. nonchromogenicum* and *M. hiberniae*, one single sequevar was present. Three sqvs (NEW1, GN-9188 and FI-05196) differed from all type strains and did not fit any of the clusters above. Interestingly, the sequevar NEW1 was detected in the strains classified as NEW1 also on the basis of the 16S rRNA gene sequence.

As expected, large variability was detected in the *rpoB* gene fragment, with the presence of 58 sqvs. The pairwise matrix of distances (data not shown) included ten clusters with, in the large majority of cases, internal variability below the limit (3 %) proposed for the *rpoB* region (Adékambi & Drancourt, 2004) (there were five exceptions with values up to 5.3 %). For this gene, the presence in GenBank of only one MTC sequence overlapping to ours (and furthermore not assigned to any known species), did not allow us to attribute any sequevar to a species; the attribution was therefore inferred from that achieved on the basis of the 16S rRNA and *hsp65* sqvs for each strain. Following this approach, it was possible to classify one cluster within each of the species: *M. arupense* (10 sqvs), *M. nonchromogenicum* (five sqvs), '*M. engbaekii*' (four sqvs), *M. senuense* (two sqvs), *M. kumamotonense* (nine sqvs) and *M. terrae* (six sqvs). Of the remaining four clusters, one (NEW1, with two sqvs) was detected in strains classified as NEW1 on the basis of 16S rRNA and *hsp65* gene sequences; and one (NEW2, with seven sqvs) was detected in strains classified as NEW2 on the basis of *hsp65* gene sequences). One cluster (NEW5, with three sqvs) was detected in strains assigned to the species *M. arupense* on the basis of 16S rRNA and *hsp65* gene sequences, it was however very distant from the *rpoB* sqvs of the strains presenting in the 16S rRNA and *hsp65*, the typical sqvs of this species. The last cluster (NEW6, with two sqvs) remained orphan. The sqvs FI-06258, FI-07105/FI-09015/FI-09399, FI-09379, FI-05396/FI-06246/FI-05196 and FI-11038 did not fit any cluster. Interestingly the five cases in which the intraspecies variability exceeded 3 % were detected within the cluster of *M. terrae* where the sequevar obtained from the type strain (CIP 104321^T) clearly differed from all the others.

Of the sqvs detected, apart from the ones of the type strains and the ones previously deposited by some of the authors, only a limited number were already present in GenBank. As regards the 16S rRNA gene, two of the five sequevars present were not assigned to any species while the label of the remaining three was correct in two cases and incorrect in one. Concerning *hsp65* gene, there were three sqvs correctly assigned, two lacking any species attribution and three that were mislabelled; furthermore, in this region, our sequence of the type strain of *M. senuense* presented one mismatch in comparison with the one present in the database. For *rpoB* sqvs, only two, both assigned to the

species *M. terrae*, were present in GenBank although not identified. In this region, one type strain (*M. terrae*) presented one nucleotide discordance.

The combinations of the 16S rRNA, *hsp65* and *rpoB* sqvs detected in the strains investigated here are reported in Table S3.

Phylogenetic analysis

The tree constructed using the sequences of the 16S rRNA genes revealed a clear separation of the strains included in the MTC from both the slow- and the rapid-growers (Fig. 1). The strains investigated here were distributed among two major branches including in the first *M. senuense*, *M. terrae* and *M. kumamotonense* and in the second *M. arupense*, *M. nonchromogenicum*, *M. hiberniae*, '*M. engbaekii*' and the clusters NEW1 and NEW2. The branch including *M. triviale* appeared very distant from the species of the MTC.

The phylogenetic tree inferred from *hsp65* gene sequences was fully in agreement with the one obtained from 16S rRNA gene sequences (Fig. S1), while a polyphyletic distribution of different species emerged (Fig. S2) in the tree constructed based on the *rpoB* gene sequences.

The phylogenetic analysis conducted on the concatenated sequences of the 75 different combinations detected in the 137 MTC strains that it was possible to assign to a species produced a quite robust dendrogram characterized by a high percentage of nodes, including the early ones, with very high bootstrap values (Fig. 2). Different species-specific clusters were clearly separated although belonging to two major groupings: the first including *M. arupense*, NEW2, *M. nonchromogenicum*, *M. hiberniae*, '*M. engbaekii*' and NEW1; the second comprising *M. senuense*, *M. terrae* and *M. kumamotonense*.

The analysis of translated nucleotide sequences revealed a large number of synonymous mutations with, in the *hsp65* region, only 12 different amino acid sequences being coded by 31 different sqvs and, in the *rpoB* region, only 27 different amino acid sequences being coded by 59 sqvs.

HPLC of cell-wall mycolic acids

The HPLC of cell-wall mycolic acids performed on the whole initial group of 156 strains revealed two different types of pattern (Fig. 3). The motif shared by the large majority of the strains was characterized by the presence of two, clearly distinct, clusters of peaks, with the first one, including three major peaks, starting to elute after 3 min and the second, including three minor peaks, eluting about 3 min later. The other pattern was presented only by the six strains of *M. triviale* subsequently excluded from the study; in this species only a late emerging cluster of peaks was present (Fig. 3d).

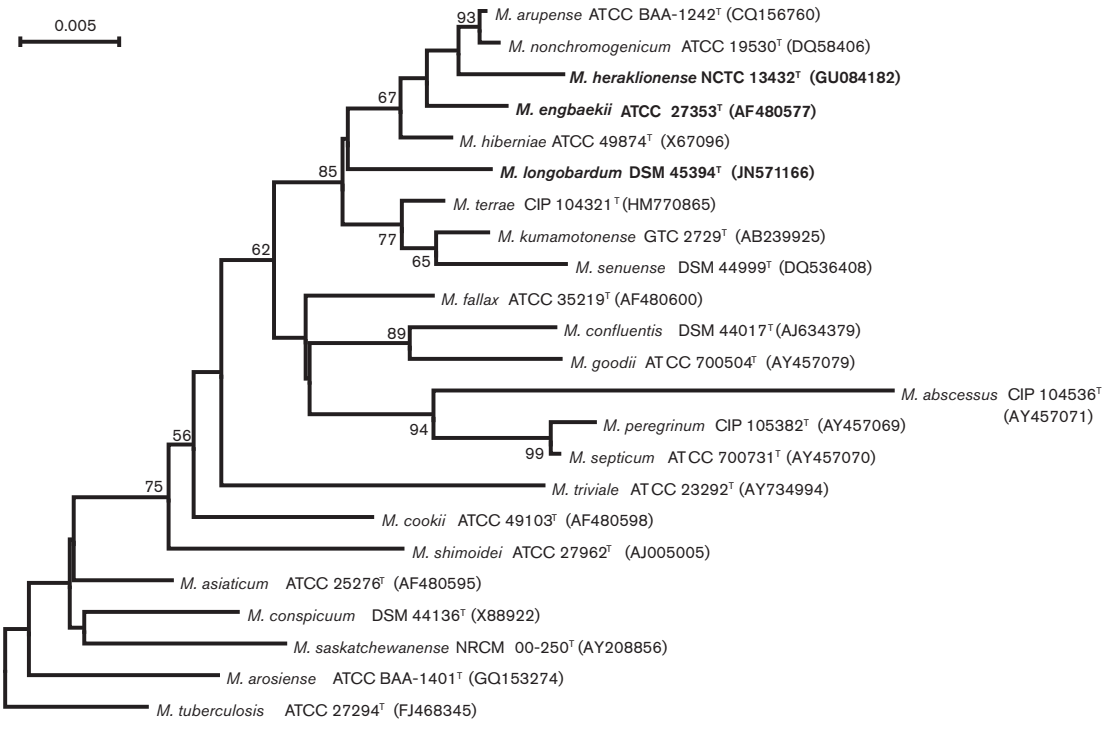

Fig. 1. Phylogenetic tree based on 16S rRNA sequences constructed using the neighbour-joining method (bootstrapped 1000 times). Bootstrap values >50 % are given at nodes. Bar, 0.005 substitutions per nucleotide position.

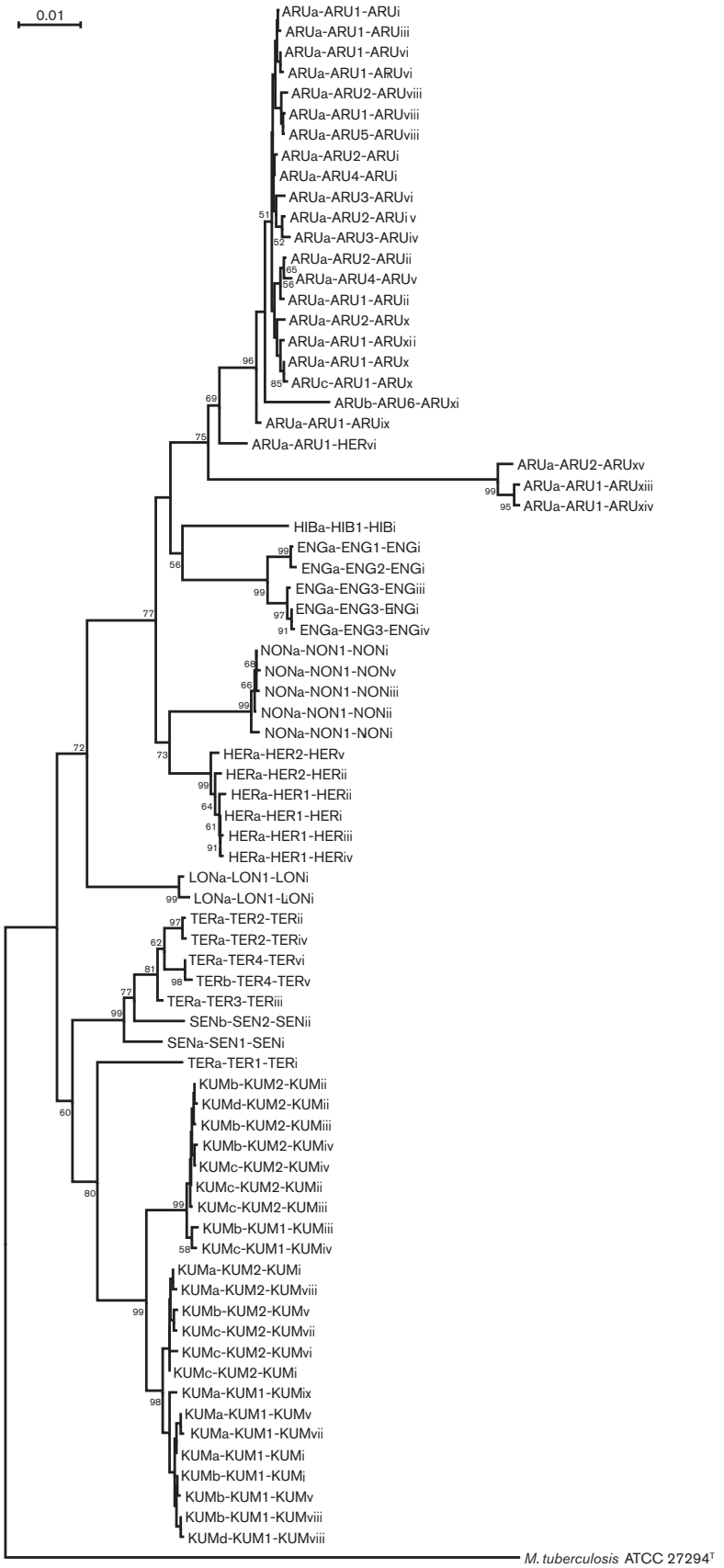

Fig. 2. Phylogenetic tree based on concatenated 16S rRNA, *hsp65* and *rpoB* gene sequences constructed using the neighbour-joining method. Every combination of sequevars detected in the strains investigated was included. The bootstrap was replicated 1000 times; only values >50 % are given at nodes. Bar, 0.01 substitutions per nucleotide position.

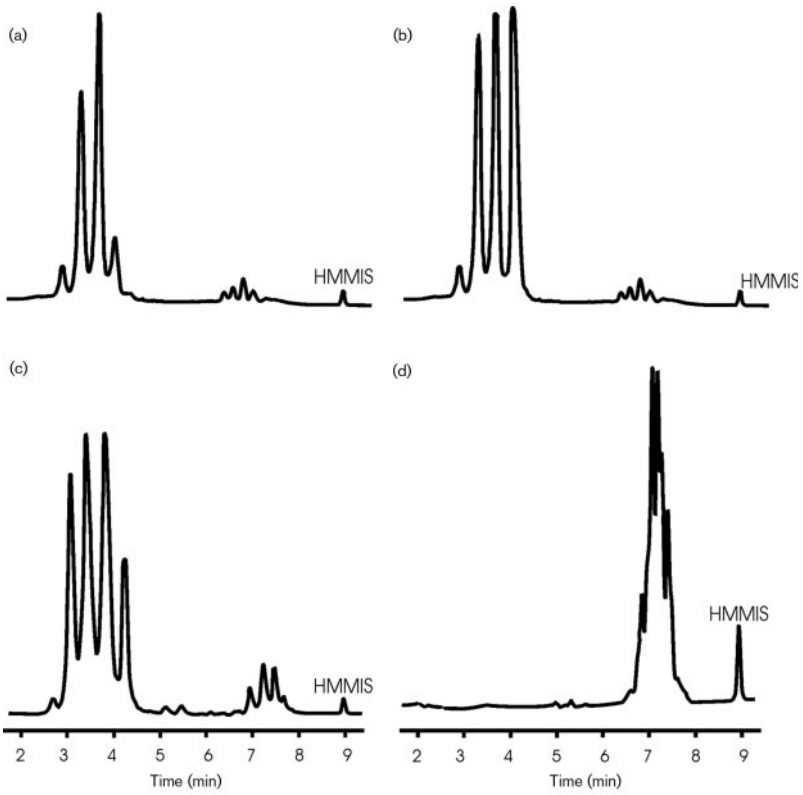

Fig. 3. Most frequently detected HPLC profiles of mycolic acids. (a), pattern presented by *M. nonchromogenicum*, *M. longobardum* sp. nov. and most strains of *M. arupense* (70 %) but occasionally found also within other species of the *M. terrae* complex; (b), pattern prevalent in *M. kumamotonense*; (c), most frequent profile of the species *M. terrae*; (d), pattern of *M. triviale*. HMMIS, high molecular mass internal standard.

Because of the common motif shared by all the members of the MTC with only a limited variability in the relative height in the major peaks of the first cluster, we tried to correlate these profiles with the different species. If the lower peaks were not taken into account, almost all the strains presented three major peaks in the first clusters. The most common motif was characterized by a highest central peak followed by the first and the third in decreasing order (Fig. 3a). This pattern was presented by the strains belonging to the species *M. nonchromogenicum* and to the NEW1 group. It was also shared by 70 % of the strains of *M. arupense* and was also represented in about half of the strains of '*M. engbaekii*' and of the NEW2 group. Less frequent, but scattered among various groupings, was the motif in which the third peak was higher than the first, with the second being the highest. Unique to the species *M. kumamotonense*, although not shown by all such strains, was a pattern characterized by three almost equal peaks with the height slightly rising from the first to the third (Fig. 3b). Equally unique to the species *M. terrae* was the profile presenting four major peaks instead of three in the first cluster (Fig. 3c).

Biochemical and cultural tests

The nitrate test was selected as the ability to reduce nitrate to nitrite is considered the sole classical biochemical feature suitable for discrimination within the otherwise phenotypically homogeneous species of the MTC (Wayne & Kubica, 1986). Nitrate reductase activity was shown by strains of *M. kumamotonense*; it was absent in strains of *M. arupense* and *M. terrae* (Table S4).

Colonies of the strains investigated were buff coloured and predominantly smooth; exceptions were shown by *M. hiberniae* and '*M. engbaekii*' which featured pink pigmentation (Table S4).

At 37 °C, growth on solid media from standardized inocula became distinctly visible, on average, after 7–14 days; faster growth was occasionally observed, while *M. senuense* grew typically slowly (Table S4).

DISCUSSION

Several obvious conclusions seem to emerge from the results of the genotypic and phenotypic investigations carried on a large number of strains belonging to the MTC.

(i) The legitimacy of the currently officially recognized species of the MTC is clearly confirmed.

(ii) The group we temporarily named NEW1 includes strains that are clearly different from any other species of the MTC in all the three genetic regions investigated. This group (of which seven strains were characterized in this study), is suggested as representing a novel species for which the name *Mycobacterium longobardum* sp. nov. is proposed.

(iii) The strains included so far in the NEW2 group, which had been initially assigned to the species *M. arupense* on the basis of the analysis of the 16S rRNA sequence, are

actually clearly distant from the latter in the *hsp65* and *rpoB* regions and represent a previously unreported species for which the name *Mycobacterium heraklionense* sp. nov. is proposed. Twenty three strains of this novel species were characterized in this study.

(iv) The seven strains investigated here and assigned to '*M. engbaekii*' support the formal proposal of *M. engbaekii* sp. nov. The name '*M. engbaekii*' was proposed in 1972 for 15 rapidly growing strains showing pink pigmentation (Korsak & Boisvert, 1972); although a reference strain of '*M. engbaekii*' was deposited in the American Type Culture Collection, no formal species description followed and the name was not validly published.

(v) Within the species *M. arupense*, two variants can be clearly distinguished on the basis of the *rpoB* gene sequence. Sufficient evidence has not emerged so far to justify the elevation of this new variant (indicated in this study as NEW6) to species rank, despite the divergence in the *rpoB* gene region that was clearly above proposed cut-off values.

(vi) Surprisingly, the recently described species *M. arupense* and *M. kumamotonense* are by far the most frequently isolated members of the MTC. One of the reference strains investigated here (*M. nonchromogenicum* PI140330001) turned out to be a member of *M. arupense*, which supports the hypothesis that the latter species and *M. kumamotonense* have been identified in the past (and also in the present) as *M. nonchromogenicum* and *M. terrae*, respectively. In our study, the isolations of *M. nonchromogenicum* and *M. terrae* were not as frequent as generally acknowledged. This brings up for discussion the widespread conviction that *M. nonchromogenicum* is the only member of the MTC potentially responsible for disease (Tsukamura *et al.*, 1983).

(vii) A number of strains have been detected that show, within 16S rRNA, *hsp65* and *rpoB* gene sequences, either conflicting sqvs of MTC species (three cases) or unreported sqvs that are not related to other, official or proposed, MTC species (10 cases; data not shown). Such strains (Table S5) need further characterization and have not been included in these phylogenetic investigations.

(viii) Questions are raised by the type strain of *M. terrae*. While the 16S rRNA gene sequence clusters with those of a number of other strains (Table S1), it clearly diverges from the latter in the *hsp65* gene (Table S2) and, even more evidently, in the *rpoB* gene. A question remains thus far unanswered. Is the type strain of *M. terrae* a rare organism that is poorly representative of the species? Are the other strains assigned here to *M. terrae* actually members of an as yet undescribed novel species? The latter option would imply that *M. terrae* is so rare that not a single isolate of this species was detected in our long-term study.

(ix) A hoary, unresolved, problem is that of the advisability of describing a novel species based on a single isolate. It is not our aim to deal with this topic here; nevertheless in this study, a dozen unique strains emerged as potentially exploitable for the description of the same number of novel species!

Bacterial strains

Seven strains of *M. longobardum* sp. nov. were isolated from two Italian hospitals between 2006 and 2009 from the sputum samples of one Lebanese and six Italian patients. Microscopy was negative and growth was obtained on solid media only (despite in four cases liquid media having also been inoculated). In no case was the strain considered responsible for disease.

A total of 23 strains of *M. heraklionense* sp. nov. were isolated from Greece (10 strains), Italy (7) and India (6). The Greek strains were grown from one outpatient and nine hospitalized patients on the island of Crete between 2002 and 2003. The Italian strains were isolated between 2005 and 2011 in five different hospitals. No information is available for the Indian strains except that they were isolated before 2005.

Of the seven strains of *M. engbaekii* sp. nov., six had been isolated in Italy, in three different hospitals between 1998 and 2011, and one had been isolated in Guadeloupe.

Clinical and epidemiological information available for some of the strains mentioned above are reported in Table 1.

Genetic sequences

In the 16S rRNA gene sequences, *M. arupense* was the recognized species most closely related to *M. longobardum* sp. nov.; however 25 bp out of 1488 bp differed (similarity 98.3 %). *M. longobardum* sp. nov. showed the closest resemblance in the *hsp65* gene (11 mismatches in 399 bp, similarity 97.2 %) with *M. kumamotonense* sp. nov.

As far as *M. heraklionense* sp. nov. is concerned, it was characterized by high similarity with *M. arupense* both in the 16S rRNA (only seven discrepancies out of 1427 bp, similarity 99.5 %) and in the *hsp65* genes (99.2 % similarity, three mismatches out of 399 bp).

For the three proposed novel mycobacteria, the species most closely related on the basis of the *rpoB* gene sequence still showed very low similarity (<94 %); this value is however affected by the limited coverage of the GenBank database for this genetic region. The comparison of their sequences with those determined in this study for the species belonging to the MTC revealed, for each of them, a clear divergence from the most closely related species; *M. heraklionense* sp. nov. differed by 2.9 % from *M. nonchromogenicum*; *M. longobardum* sp. nov. differed by 3.6 % from *M. heraklionense* sp. nov. and *M. engbaekii* sp. nov. differed by 4.2 % from *M. heraklionense* sp. nov.

PCR restriction analysis (PRA)

PRA patterns (Telenti *et al.*, 1993) inferred on the basis of restriction sites present in the *hsp65* gene sequences were different for each of the three potential novel species (Table 2); two biotypes were detected for *M. engbaekii* sp. nov. The PRA patterns were unique and suitable for the differentiation

Table 1. Epidemiological, microbiological and clinical characteristics of six strains of *M. engbaekii* sp. nov., 16 strains of *M. heraklionense* sp. nov. and seven strains of *M. longobardum* sp. nov. –, Negative; NA, no data available.

Strain	Age	Sex	Specimen type	Microscopy	Culture (positive/done)	Disease	Site*	Year	Sequevars (16S rRNA/*hsp65*/*rpoB*)
***M. engbaekii* sp. nov.**									
FI-04007	74	M	Gastric washing	–	NA	NA	GL	1998	a/2/iii
FI-98002	NA	M	NA	NA	NA	NA	MI	1998	a/1/ii
FI-06007	62	M	Urine	–	1/3	NA	VI	2006	a/3/iii
FI-98058	NA	M	NA	NA	NA	NA	TR	1998	a/1/i
FI-98001	NA	M	NA	NA	NA	NA	MI	1998	a/1/iv
HSR-11012	75	M	Bronchial aspirate	–	1/1	NA	VI	2011	a/1/ii
***M. heraklionense* sp. nov**									
GN01^T	74	M	NA	–	1/1	Renal failure, heart failure	H	2002	a/2/ii
GN02	76	M	NA	–	1/1	Myelodysplastic syndrome	H	2003	a/2/ii
GN04	83	M	NA	–	1/1	COPD†	H	2003	a/2/ii
GN05	42	M	NA	–	1/1	Tuberculosis	H	2003	a/2/ii
GN06	70	M	NA	–	1/1	Lung cancer	H	2003	a/2/ii
GN08	59	M	NA	–	1/1	Non-Hodgkin lymphoma	H	2003	a/2/ii
GN09	35	M	NA	–	1/1	Rheumatoid arthritis	H	2003	a/2/ii
GN10	82	M	NA	–	1/1	Pulmonary fibrosis	H	2003	a/2/ii
GN12	77	M	NA	–	1/1	Lung cancer	H	2003	a/2/ii
FI-10248	61	M	NA	NA	NA	NA	AN	2010	a/1/iii
HRS-11013	82	M	Bronchial lavage	–	1/2	NA	NO	2011	a/1/iii
FI-06009	62	M	Sputum	–	1/3	NA	NO	2006	a/1/i
FI-06255	NA	F	Sputum	NA	NA	NA	CO	2007	a/2/v
FI-08098	NA	M	NA	–	1/1	NA	MI	2008	a/1/iii
FI-08101	NA	M	NA	NA	NA	NA	MI2	2008	a/1/ii
FI-05158	74	F	Sputum	–	1/2	NA	NO	2005	a/1/iii
***M. longobardum* sp. nov.**									
FI-09110	71	M	NA	–	1/2‡	Tuberculosis	VA	2009	a/1/ii
FI-07054	65	M	NA	–	1/3‡	COPD†	VA	2006	a/1/ii
FI-06254	78	M	NA	–	1/1‡	Pneumonia	VA	2006	a/1/i
FI-09059	51	M	NA	–	1/9‡	Bronchitis	VA	2008	a/1/i
FI-07034^T	72	F	NA	–	1/4	Broncho-pneumonitis	VA1	2006	a/1/i
FI-07020	76	M	NA	–	1/2	Lung cancer	VA1	2006	a/1/ii
FI-07089	39	F	NA	–	2/3	Suspected tuberculosis in LES§ patient	VA1	2006	a/1/i

*H, Greece (Heraklion); GL, Guadeloupe; other acronyms refer to different Italian cities.
†Chronic obstructive pulmonary disease.
‡Positive in liquid medium, but negative in solid medium.
§Systemic lupus erythematosus.

of *M. engbaekii* sp. nov., *M. heraklionense* sp. nov. and *M. longobardum* sp. nov. from any other known species of the genus *Mycobacterium* (http://app.chuv.ch/prasite/).

Phylogenetic analysis

In the phylogenetic reconstructions based on 16S rRNA, *hsp65* and *rpoB* gene sequences, the three novel species proposed belonged to a sub-branch of the MTC that also included *M. arupense*, *M. nonchromogenicum* and *M. hiberniae*; with *M. kumamotonense*, *M. senuense* and *M. terrae* located on a different branch. The tree emerging from the concatenated sequences clearly showed *M. longobardum* sp. nov. separated from the other species; *M. heraklionense* sp. nov. closer to *M. nonchromogenicum* and *M. engbaekii* sp. nov. closer to *M. hiberniae* (Fig. 2).

HPLC of cell-wall mycolic acids

The low discriminating power, within the MTC, of the HPLC of cell-wall mycolic acids was also confirmed by the strains belonging to the species *M. engbaekii* sp. nov., *M.*

Table 2. Patterns detected by PCR restriction analysis with enzymes *BstE*II and *Hae*III for the strains for which novel species status is proposed

Species	Sequevars	Restriction patterns	
		***BstE*II**	***Hae*III**
M. engbaekii sp. nov.	ENG1	289-96	124-58-54
	ENG2-3-4	304-96	118-87-58
M. heraklionense sp. nov.	HER1-2	210-96-94	118-87-77
M. longobardum sp. nov.	LON1	210-190	118-112-69

heraklionense sp. nov. and *M. longobardum* sp. nov. which presented the typical profile that was characterized by an early major and a late minor clusters of peaks (Fig. 3).

Biochemical and cultural tests and susceptibility testing

The major differences, revealed by biochemical and cultural tests (Table 3) between the three novel species, concerned the morphology and pigmentation of the colonies (rough, pink, scotochromogenic in *M. engbaekii* sp. nov., rough and unpigmented in *M. longobardum* sp. nov. and smooth and unpigmented in *M. heraklionense* sp. nov.), the hydrolysis of Tween 80 (negative for *M. longobardum* sp. nov. only), nitrate reduction (negative in *M. engbaekii* sp. nov. only), arylsulfatase activity at three days (positive in *M. longobardum* sp. nov. only) and the β-glucosidase activity (positive in *M. heraklionense* sp. nov. only). With respect to other known members of the MTC, the discriminative power of biochemical and cultural tests was, as expected, very limited (data not shown) and confirmed the perception which lead to the proposal about 40 years ago that they should be included in a complex.

Antibiotic susceptibility testing revealed that the three novel species were susceptible to clarithromycin and resistant to quinolones with the activity of other molecules being variable (Table 4).

Description of *Mycobacterium engbaekii* sp. nov.

Mycobacterium engbaekii sp. nov. (eng.ba.e′ki.i. N.L. gen. masc. n. *engbaekii* of Engbaek, to honour of the Danish mycobacteriologist H. C. Engbaek).

Cells are typically acid–fast, rod-shaped, with some coccoid forms, but without branches or aerial hyphae; spores are not produced. Mature growth is obtained in solid media at temperatures between 25 and 37 °C in less than 10 days. Colonies grow as rough and unpigmented in the dark but develop a pink pigmentation after exposure to light. Gives a positive result for Tween 80 hydrolysis and tellurite reduction, and is negative in tests for niacin accumulation, arylsulfatase at 3 days, nitrate reduction, and urease and β-glucosidase activities. Catalase at 68 °C is present and more than 45 mm foam is produced in the semiquantitative test.

Table 3. Results of biochemical and cultural tests for the three proposed novel species

Taxa; 1, *M. engbaekii* sp. nov.; 2, *M. heraklionense* sp. nov. 3, *M. longobardum* sp. nov. +, Positive, −, negative; v, variable. All taxa give a negative result in tests for niacin accumulation and urea hydrolysis. All taxa give a positive result for growth at 25 °C, catalase at 68 °C, catalase reaction >45 mm, and tolerance of p-nitrobenzoic acid, thiophene carboxylic acid, tiacetazone and isoniazid.

Characteristic	1	2	3
Growth rate	Rapid	Intermediate	Intermediate
Growth at 45 °C	v	−	−
Pigmentation	Pink*	Absent	Absent
Colony morphology	Rough	Smooth	Rough
Nitrate reduction	−	+	+
Tween 80 hydrolysis	+	+	−
Tellurite reduction	+	−	−
Arylsulfatase 3 day	−	−	+
β-Glucosidase	−	+	−
Tolerance of			
MacConkey medium	−	−	v
Hydroxylamine	+	v	+
Oleate	+	−	+

*Photochromogenic.

Isolates are susceptible to amikacin, clarithromycin, ethambutol, linezolid and rifabutin and resistant to doxycycline and sulfamethoxazole. Mycolic acids produce an early major and a late minor cluster of peaks. The sequences are unique in the 16S rRNA gene, in the *hsp65* gene (where four sqvs are present) and in the *rpoB* (four sqvs) gene. Phylogenetically the species is included in the MTC and is most closely related to *M. hiberniae.*

The type strain is ATCC 27353^T=DSM 45694^T.

Description of *Mycobacterium heraklionense* sp. nov.

Mycobacterium heraklionense sp. nov. (he.ra.kli.on.en′se. N.L. neut. adj. *heraklionense* of or belonging to Heraklion, the city in Crete where many strains were isolated).

Cells are typically acid–fast, rod-shaped, with some coccoid forms, but without branches or aerial hyphae; spores are not produced. Mature growth is obtained in solid media at temperatures between 25° and 37 °C in 5–12 days. Colonies grow as smooth and unpigmented both in the dark and after light exposure. Positive result in tests for nitrate reductase, Tween 80 hydrolysis and β-glucosidase activity. Negative result in tests for niacin accumulation, arylsulfatase activity at 3 days, tellurite reduction and urease activity. Catalase is present at 68 °C and more than 45 mm foam is produced in the semiquantitative test. Isolates are susceptible to clarithromycin and resistant to quinolones, rifampicin, sulfamethoxazole and doxycycline.

Table 4. Minimal inhibitory concentrations (MIC) to antimycobacterial drugs for several strains of *M. engbaekii* sp. nov., *M. heraklionense* sp. nov. and *M. longobardum* sp. nov.

I, Intermediate; R, resistant, S, susceptible.

Drug	*M. engbaekii* sp. nov.		*M. heraklionense* sp. nov.		*M. longobardum* sp. nov.	
	MIC	Interpretation	MIC	Interpretation	MIC	Interpretation
Ciprofloxacin	2	I	⩾16	R	16	R
Moxifloxacin	2–4	R	>8	R	⩾8	R
Linezolid	⩽1	S	4–16	S/I	64	R
Rifampicin	4	I	⩾8	R	4–8	I/R
Rifabutin	⩽0.25	S	⩽0.25	S	0.5	I
Sulfamethoxazole	⩾76	R	>152	R	5–9.5	S
Doxycycline	16	R	⩾16	R	4	I
Ethambutol	⩽0.5	S	2	I	4	I
Clarithromycin	1	S	1	S	2	S
Amikacin	2–8	S	1–32	S/I	32	I
Streptomycin	2–8	S/I	2– >64	S/I/R	32	R

Mycolic acids produce an early major and a late minor cluster of peaks. The sequences are unique in the 16S rRNA gene, where two sqvs closely related to *M. arupense* are present, in the *hsp65* gene (three sqvs), where it most closely resembles *M. nonchromogenicum*, and in the *rpoB* gene (seven sqvs), equally divergent from *M. arupense* and *M. nonchromogenicum*. Phylogenetically the species is included in the MTC and is most closely related to *M. nonchromogenicum*.

The type strain is GN-1^T (=NCTC 13432^T=LMG 24735^T=CECT 7509^T).

Description of *Mycobacterium longobardum* sp. nov.

Mycobacterium longobardum sp. nov. (lon.go′bar.dum. N.L. neut. adj. of or pertaining to Lombardy, the region where the strains were isolated).

Cells are typically acid–fast, rod-shaped, with some coccoid forms, but without branches or aerial hyphae; spores are not produced. Mature growth is obtained in solid media at temperatures between 25° and 37 °C in 7–14 days. Colonies grow rough and unpigmented both in the dark and after light exposure. Positive in tests for nitrate reductase and arylsulfatase activity at 3 days and negative result for niacin accumulation, Tween 80 hydrolysis, tellurite reduction and urease and β-glucosidase activities. Catalase is present at 68 °C and more than 45 mm foam is produced in the semiquantitative test. Isolates are susceptible to sulfamethoxazole and clarithromycin and resistant to quinolones, linezolid and streptomycin. Mycolic acids produce an early major and a late minor cluster of peaks. The sequences are unique in the 16S rRNA gene, in the *hsp65* gene and in the *rpoB* gene (where two sqvs are present). Phylogenetically the species is included in the MTC and is clearly separated from other species.

The type strain is FI-07034^T (=DSM 45394^T=CCUG 58460^T).

ACKNOWLEDGEMENTS

We thank Antonella Grottola (Laboratory of Microbiology and Virology, Modena University Hospital, Modena, Italy) for her valuable comments on the manuscript. We are grateful to Meg Lafferty for reviewing the language of the manuscript.

REFERENCES

Adékambi, T. & Drancourt, M. (2004). 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* **54**, 2095–2105.

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

CDC (1996). *Standardized Method for HPLC Identification of Mycobacteria*. Atlanta: U.S. Department of Health and Human Services, Public Health Service.

Cloud, J. L., Meyer, J. J., Pounder, J. I., Jost, K. C., Jr, Sweeney, A., Carroll, K. C. & Woods, G. L. (2006). *Mycobacterium arupense* sp. nov., a non-chromogenic bacterium isolated from clinical specimens. *Int J Syst Evol Microbiol* **56**, 1413–1418.

CLSI (2011). *Susceptibility Testing of Mycobacteria, Nocardiae and Other Aerobic Actinomycetes; Approved Standard*, 2nd edn. Wayne, PA: Clinical and Laboratory Standards Institute.

Devulder, G., Pérouse de Montclos, 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.

Kazda, J., Cooney, R., Monaghan, M., Quinn, P. J., Stackebrandt, E., Dorsch, M., Daffé, M., Müller, K., Cook, B. R. & Tarnok, Z. S. (1993). *Mycobacterium hiberniae* sp. nov. *Int J Syst Bacteriol* **43**, 352–357.

Kent, P. T. & Kubica, G. P. (1985). *Public Health Mycobacteriology. A Guide for the Level III Laboratory*. Atlanta: U.S. Department of Health and Human Services.

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.

Korsak, T. & Boisvert, H. (1972). Mycobactéries à pigment rose. *Ann Inst Pasteur* (*Paris*) **122**, 31–41 (in French).

Kubica, G. P., Silcox, V. A., Kilburn, J. O., Smithwick, R. W., Beam, R. E., Jones, W. D. & Stottmeier, K. D. (1970). Differential identification of mycobacteria. VI. *Mycobacterium triviale* Kubica sp. nov. *Int J Syst Bacteriol* **20**, 161–174.

Lee, H., Lee, S. A., Lee, I. K., Yu, H. K., Park, Y. G., Jeong, J., Lee, S. H., Kim, S. R., Hyun, J. W. & other authors (2010). *Mycobacterium paraterrae* sp. nov. recovered from a clinical specimen: novel chromogenic slow growing mycobacteria related to *Mycobacterium terrae* complex. *Microbiol Immunol* **54**, 46–53.

Masaki, T., Ohkusu, K., Hata, H., Fujiwara, N., Iihara, H., Yamada-Noda, M., Nhung, P. H. H. M., Hayashi, M., Asano, Y. & other authors (2006). *Mycobacterium kumamotonense* sp. nov. recovered from clinical specimen and the first isolation report of *Mycobacterium arupense* in Japan: novel slowly growing, nonchromogenic clinical isolates related to *Mycobacterium terrae* complex. *Microbiol Immunol* **50**, 889–897.

McNabb, A., Eisler, D., Adie, K., Amos, M., Rodrigues, M., Stephens, G., Black, W. A. & Isaac-Renton, J. (2004). Assessment of partial sequencing of the 65-kilodalton heat shock protein gene (*hsp65*) for routine identification of *Mycobacterium* species isolated from clinical sources. *J Clin Microbiol* **42**, 3000–3011.

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.

Mun, H. S., Park, J. H., Kim, H., Yu, H. K., Park, Y. G., Cha, C. Y., Kook, Y. H. & Kim, B. J. (2008). *Mycobacterium senuense* sp. nov., a slowly growing, non-chromogenic species closely related to the *Mycobacterium terrae* complex. *Int J Syst Evol Microbiol* **58**, 641–646.

Reischl, U., Emler, S., Horak, Z., Kaustova, J., Kroppenstedt, R. M., Lehn, N. & Naumann, L. (1998). *Mycobacterium bohemicum* sp. nov., a new slow-growing scotochromogenic mycobacterium. *Int J Syst Bacteriol* **48**, 1349–1355.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

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., Kämpfer, P., Maiden, M. C. J., Nesme, X., Rosselló-Mora, R., Swings, J. & other authors (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.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739.

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.

Tsukamura, M. (1965). A group of mycobacteria from soil sources resembling nonphotochromogens (group 3). A description of *Mycobacterium nonchromogenicum*. *Igaku To Seibutsugaku* **71**, 110–113 (in Japanese).

Tsukamura, M., Kita, N., Otsuka, W. & Shimoide, H. (1983). A study of the taxonomy of the *Mycobacterium nonchromogenicum* complex and report of six cases of lung infection due to *Mycobacterium nonchromogenicum*. *Microbiol Immunol* **27**, 219–236.

Wayne, L. G. (1966). Classification and identification of mycobacteria. 3. Species within group 3. *Am Rev Respir Dis* **93**, 919–928.

Wayne, L. G. & Kubica, G. P. (1986). Family *Mycobacteriaceae* CHESTER 1897, 63^{AL}. In *Bergey's Manual of Systematic Bacteriology*, pp. 1435–1457. Edited by P. H. A. Sneath, N. S. Mair, M. E. Sharpe & J. G. Holt. Baltimore: The Williams & Wilkins Co.

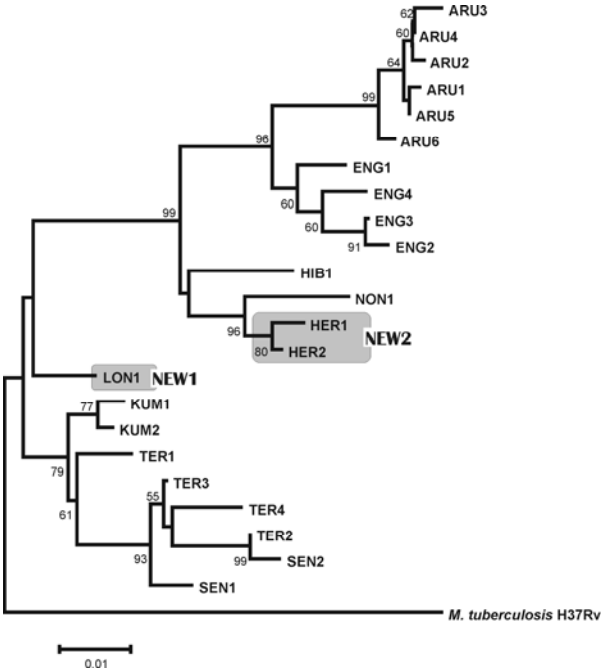

Fig. S1. Phylogenetic tree based on *hsp65* sequevars constructed using the neighbour-joining method (bootstrapped 1000 times). Bootstrap values >50% are given at nodes. Bar, 0.01 substitutions per nucleotide position.

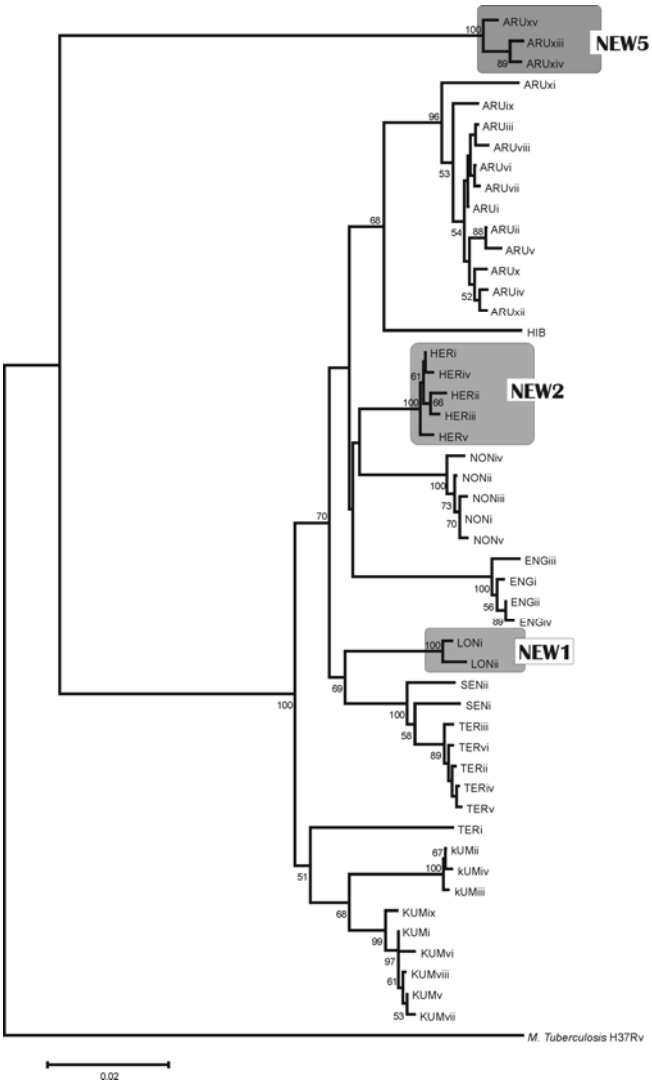

Fig. S2. Phylogenetic tree based on *rpoB* sequevars constructed using the neighbour-joining method (bootstrapped 1000 times). Bootstrap values >50% are given at nodes. Bar, 0.02 substitutions per nucleotide position

.

E. Tortoli, Z. Gitti, H.-P. Klenk, S. Lauria, R. Mannino, P. Mantegani, A. Mariottini and I. Neonakis (2013). Survey of 150 strains belonging to the *Mycobacterium terrae* complex and description of *Mycobacterium engbaekii* sp. nov., *Mycobacterium heraklionense* sp. nov. and *Mycobacterium longobardum* sp. nov. *Int J Syst Evol Microbiol* **63**, 401–411.

Table S1. Pairwise distances among the sequevars detected in the 16S rRNA gene

ARU, *M. arupense*; NON, *M. nonchromogenicum*; ENG, '*M. engbaekii*'; HIB, *M. hiberniae*; SEN, *M. senuense*; KUM, *M. kumamotonense*; TER, *M. terrae*; PAR, '*M. paraterrae*'; TRI, *M. triviale*; HER, *M. heraklionense*; LON, *M. longobardum*; NEW, temporary attribution.

Clusters	Sequevars	ARUa	ARUb	ARUc	HERa	HERb	NON	ENG	HIB	LON	SENa	SENb	KUMa	KUMb	KUMc	KUMd	TERa	TERb	TERc	GN-9188	FI-07105	FI-09379	TRI
ARU	ARUa																						
ARU	ARUb	0.000																					
ARU	ARUc	0.002	0.002																				
ARU	HERa	0.002	0.002	0.004																			
ARU	FI-101	0.002	0.002	0.004	0.004																		
NON	NON	0.011	0.011	0.013	0.013	0.013																	
ENG	ENG	0.009	0.009	0.011	0.011	0.011	0.002																
HIB	HIB	0.020	0.020	0.022	0.017	0.017	0.017	0.015															
NEW1	LON	0.026	0.026	0.028	0.024	0.024	0.022	0.020	0.017														
SEN	SENa	0.026	0.026	0.028	0.024	0.024	0.022	0.020	0.017	0.000													
SEN	SENb	0.033	0.033	0.031	0.031	0.031	0.028	0.026	0.024	0.006	0.006												
KUM	KUMa	0.033	0.033	0.035	0.031	0.031	0.028	0.026	0.024	0.006	0.006	0.004											
KUM	KUMb	0.031	0.031	0.033	0.028	0.028	0.026	0.024	0.022	0.004	0.004	0.002	0.002										
KUM	KUMc	0.035	0.035	0.033	0.033	0.033	0.031	0.028	0.026	0.009	0.009	0.002	0.002	0.004									
KUM	KUMd	0.026	0.026	0.028	0.024	0.024	0.022	0.020	0.017	0.000	0.000	0.006	0.006	0.004	0.009								
TER	TERa	0.028	0.028	0.031	0.026	0.026	0.024	0.022	0.020	0.002	0.002	0.009	0.009	0.006	0.011	0.002							
TER	TERb	0.028	0.028	0.026	0.031	0.026	0.024	0.022	0.024	0.006	0.006	0.009	0.013	0.011	0.011	0.006	0.009						
TER	TERc	0.017	0.017	0.020	0.015	0.020	0.020	0.017	0.020	0.024	0.024	0.031	0.031	0.028	0.033	0.024	0.026	0.031					
GN-9188	GN-9188	0.006	0.006	0.009	0.009	0.009	0.009	0.006	0.013	0.022	0.022	0.028	0.028	0.026	0.031	0.022	0.024	0.024	0.011				
FI-07105/110	PAR	0.009	0.009	0.011	0.006	0.011	0.011	0.009	0.011	0.020	0.020	0.026	0.026	0.024	0.028	0.020	0.022	0.026	0.009	0.002			
FI-09379	FI-09379	0.006	0.006	0.004	0.004	0.009	0.013	0.011	0.017	0.024	0.024	0.026	0.031	0.028	0.028	0.024	0.026	0.026	0.020	0.013	0.011		
TRI	TRI	0.037	0.037	0.035	0.039	0.035	0.044	0.042	0.042	0.046	0.046	0.044	0.048	0.046	0.046	0.046	0.048	0.044	0.046	0.039	0.042	0.035	

Table S2. Pairwise distances detected among the sequevars detected in the *hsp65* gene

ARU, *M. arupense*; NON, *M. nonchromogenicum*; ENG, '*M. engbaekii*'; HIB, *M. hiberniae*; SEN, *M. senuense*; KUM, *M. kumamotonense*; TER, *M. terrae*; HER, *M. heraklionense*; LON, *M. longobardum*; NEW, temporary attribution.

Clusters	Sequevars	ARU1	ARU2	ARU3	ARU4	ARU5	ARU6	HER1	HER2	NON1	ENG1	ENG2	ENG3	ENG4	HIB1	LON1	SEN1	SEN2	KUM1	KUM2	TER1	TER2	TER3	TER4	GN-9188	FI-06258	FI-009379	FI-07105	FI-10193	FI-11038	FI-05196
ARU	ARU1																														
ARU	ARU2	0.008																													
ARU	ARU3	0.010	0.008																												
ARU	ARU4	0.005	0.003	0.005																											
ARU	ARU5	0.003	0.005	0.008	0.003																										
ARU	ARU6	0.013	0.010	0.013	0.008	0.010																									
NEW2	HER1	0.054	0.062	0.065	0.059	0.057	0.057																								
NEW2	HER2	0.057	0.065	0.067	0.062	0.059	0.059	0.008																							
NON	NON1	0.065	0.073	0.075	0.070	0.067	0.073	0.023	0.031																						
ENG	ENG1	0.033	0.041	0.044	0.038	0.036	0.031	0.054	0.057	0.051																					
ENG	ENG2	0.044	0.046	0.054	0.049	0.046	0.041	0.049	0.054	0.062	0.023																				
ENG	ENG3	0.038	0.046	0.049	0.044	0.041	0.036	0.044	0.049	0.057	0.018	0.005																			
ENG	ENG4	0.038	0.041	0.049	0.044	0.041	0.036	0.044	0.049	0.062	0.023	0.018	0.018																		
HIB	HIB1	0.054	0.057	0.065	0.059	0.057	0.057	0.033	0.038	0.046	0.057	0.049	0.054	0.046																	
NEW1	LON1	0.073	0.075	0.084	0.078	0.075	0.075	0.051	0.054	0.065	0.067	0.062	0.062	0.059	0.059																
SEN	SEN1	0.089	0.097	0.097	0.095	0.092	0.092	0.067	0.070	0.081	0.086	0.084	0.078	0.089	0.075	0.044															
SEN	SEN2	0.111	0.108	0.117	0.111	0.114	0.108	0.081	0.078	0.095	0.106	0.097	0.097	0.097	0.095	0.057	0.033														
KUM	KUM1	0.084	0.081	0.089	0.084	0.086	0.081	0.057	0.059	0.070	0.081	0.073	0.073	0.073	0.067	0.028	0.036	0.044													
KUM	KUM2	0.081	0.078	0.086	0.081	0.084	0.078	0.054	0.057	0.067	0.078	0.070	0.070	0.070	0.065	0.025	0.033	0.041	0.008												
TER	TER1	0.081	0.084	0.092	0.086	0.084	0.084	0.057	0.059	0.073	0.084	0.073	0.073	0.073	0.067	0.031	0.033	0.038	0.018	0.020											
TER	TER2	0.106	0.103	0.111	0.106	0.108	0.103	0.075	0.073	0.089	0.100	0.092	0.092	0.092	0.089	0.051	0.028	0.005	0.038	0.036	0.033										
TER	TER3	0.089	0.092	0.100	0.095	0.092	0.092	0.062	0.065	0.075	0.086	0.078	0.078	0.078	0.070	0.033	0.010	0.023	0.025	0.023	0.023	0.018									
TER	TER4	0.103	0.106	0.108	0.108	0.106	0.106	0.075	0.078	0.089	0.100	0.092	0.092	0.092	0.084	0.046	0.020	0.031	0.038	0.036	0.036	0.025	0.013								
GN-9188	GN-9188	0.092	0.095	0.103	0.097	0.095	0.095	0.062	0.059	0.081	0.092	0.078	0.078	0.078	0.075	0.038	0.033	0.041	0.031	0.028	0.025	0.036	0.028	0.041							
NEW3	FI-06258/07088/09100	0.089	0.086	0.092	0.089	0.092	0.086	0.062	0.065	0.070	0.086	0.078	0.078	0.078	0.070	0.038	0.044	0.051	0.020	0.018	0.031	0.046	0.033	0.046	0.038						
NEW3	FI-009379	0.075	0.073	0.081	0.075	0.078	0.073	0.054	0.057	0.067	0.073	0.065	0.065	0.070	0.057	0.031	0.031	0.044	0.018	0.015	0.020	0.038	0.025	0.038	0.031	0.023					
NEW4	FI-07105	0.036	0.044	0.046	0.041	0.038	0.038	0.044	0.049	0.057	0.033	0.028	0.023	0.023	0.044	0.059	0.081	0.100	0.070	0.067	0.073	0.095	0.075	0.089	0.081	0.075	0.067				
NEW4	FI-10193	0.033	0.041	0.044	0.038	0.036	0.031	0.044	0.041	0.062	0.020	0.020	0.015	0.020	0.057	0.065	0.081	0.097	0.075	0.073	0.073	0.092	0.081	0.095	0.078	0.081	0.067	0.028			
NEW4	FI-11038	0.033	0.041	0.044	0.038	0.036	0.036	0.044	0.046	0.057	0.031	0.031	0.025	0.023	0.046	0.057	0.078	0.097	0.067	0.065	0.070	0.092	0.073	0.086	0.078	0.073	0.065	0.003	0.025		
FI-05196	FI-05196	0.062	0.070	0.073	0.067	0.065	0.065	0.015	0.015	0.033	0.062	0.054	0.049	0.046	0.033	0.059	0.075	0.084	0.065	0.062	0.065	0.078	0.070	0.081	0.065	0.070	0.062	0.044	0.049	0.046	

Table S3. Combinations of 16S rRNA, *hsp65* and *rpoB* gene sqvs detected in the strains investigated

ARU, *M. arupense;* NON, *M. nonchromogenicum;* ENG, '*M. engbaekii*'; HIB, *M. hiberniae*; SEN, *M. senuense*; KUM, *M. kumamotonense*; TER, *M. terrae;* HER, *M. heraklionense*; LON, *M. longobardum*; NEW, temporary attribution.

Genetic region			Combinations number	Species
16S rRNA	***hsp65***	***rpoB***		
ARUa	ARU1	ARUi		
ARUa	ARU1	ARUii	9	ARU
ARUa	ARU1	ARUiii	4	ARU
ARUa	ARU1	ARUix	1	ARU
ARUa	ARU1	ARUvi	2	ARU
ARUa	ARU1	ARUvii	1	ARU
ARUa	ARU1	ARUviii	4	ARU
ARUa	ARU1	ARUx	3	ARU
ARUa	ARU1	ARUxii	1	ARU
ARUa	ARU1	ARUxiii	2	ARU
ARUa	ARU1	ARUxiv	1	ARU
ARUa	ARU2	ARUi	2	ARU
ARUa	ARU2	ARUii	2	ARU
ARUa	ARU2	ARUiv	1	ARU
ARUa	ARU2	ARUviii	2	ARU
ARUa	ARU2	ARUx	1	ARU
ARUa	ARU2	ARUxv	1	ARU
ARUa	ARU3	ARUiv	1	ARU
ARUa	ARU3	ARUvi	1	ARU
ARUa	ARU4	ARUi	1	ARU
ARUa	ARU4	ARUv	1	ARU
ARUa	ARU5	ARUviii	1	ARU
ARUb	ARU6	ARUxi	1	ARU
ARUc	ARU1	ARUx	1	ARU
ENGa	ENG1	ENGi	2	ENG
ENGa	ENG1	ENGii	2	ENG
ENGa	ENG1	ENGiv	1	ENG
ENGa	ENG2	ENGiii	1	ENG
ENGa	ENG3	ENGiii	1	ENG
HERa	HER1	HERi	6	HER
HERa	HER1	HERii	1	HER
HERa	HER1	HERiii	4	HER
HERa	HER1	HERiv	1	HER
HERa	HER2	HERii	11	HER
HERa	HER2	HERv	1	HER
HIBa	HIB1	HIBi	1	HIB
KUMa	KUM1	KUMi	1	KUM
KUMa	KUM1	KUMv	1	KUM
KUMa	KUM1	KUMvii	5	KUM
KUMa	KUM1	KUMx	1	KUM
KUMa	KUM2	KUMi	1	KUM
KUMa	KUM2	KUMviii	1	KUM
KUMb	KUM1	KUMi	3	KUM
KUMb	KUM1	KUMv	1	KUM
KUMb	KUM1	KUMviii	2	KUM
KUMb	KUM1	kUMiii	1	KUM
KUMb	KUM2	KUMii	3	KUM
KUMb	KUM2	kUMiii	1	KUM
KUMb	KUM2	kUMiv	1	KUM
KUMb	KUM2	KUMv	1	KUM

E. Tortoli, Z. Gitti, H.-P. Klenk, S. Lauria, R. Mannino, P. Mantegani, A. Mariottini and I. Neonakis (2013). Survey of 150 strains belonging to the *Mycobacterium terrae* complex and description of *Mycobacterium engbaekii* sp. nov., *Mycobacterium heraklionense* sp. nov. and *Mycobacterium longobardum* sp. nov. *Int J Syst Evol Microbiol* **63**, 401–411.

KUMc	KUM1	kUMix	1	KUM
KUMc	KUM2	KUMi	1	KUM
KUMc	KUM2	kUMii	3	KUM
KUMc	KUM2	kUMiii	1	KUM
KUMc	KUM2	kUMiv	1	KUM
KUMc	KUM2	KUMvi	1	KUM
KUMc	KUM2	KUMvii	1	KUM
KUMd	KUM1	KUMviii	1	KUM
KUMd	KUM2	KUMii	1	KUM
LONa	LON1	LONi	4	LON
LONa	LON1	LONii	3	LON
NONa	NON1	NONi	3	NON
NONa	NON1	NONii	3	NON
NONa	NON1	NONiii	1	NON
NONa	NON1	NONiv	1	NON
NONa	NON1	NONv	1	NON
SENa	SEN1	SENi	1	SEN
SENb	SEN2	sENii	1	SEN
TERa	TER1	TERi	1	TER
TERa	TER2	TERii	1	TER
TERa	TER2	TERiv	1	TER
TERa	TER3	TERiii	1	TER
TERa	TER4	TERvi	1	TER
TERb	TER4	TERv	1	TER

Table S4. Phenotypic features detected in different species of *M. terrae* complex

Species	**Number of strains**	**Nitrate reductase (% positive)**	**Pigmentation (% positive)**	**Growth rate**
M. arupense	51	Negative (4%)	None (0%)	Intermediate
'*M. engbaekii*'	7	Positive (43%)	Pink (85%)	Rapid
M. hiberniae	1	Negative (0%)	Pink (100%)	Intermediate
M. kumamotonense	34	Positive (55%)	None (6%)	Intermediate
M. nonchromogenicum	9	Variable (55%)	None (0%)	Intermediate
M. senuense	2	Variable (50%)	None (0%)	Slow
M. terrae	6	Negative (0%)	None (0%)	Intermediate

Table S5. Strains presenting conflicting or new sqvs, not assigned to any of the species of the *M. terrae* complex

Genetic region			**Combinations number**
16S rRNA	***hsp65***	***rpoB***	
FI-07105	FI-07105	FI-07105	1
FI-07105	FI-11038	FI-11038	1
FI-09379	FI-09379	FI-09379	1
FI-09015	FI-09015	ARUx	1
ARUa	ARU1	FI-07105	2
ARUa	ARU1	HERvi	1
HERa	FI-05196	FI-05196	1
HERa	ENG4	FI-05196	2
FI-101	FI-10193	FI-10193	1
KUMa	FI-06258	FI-09100	1
KUMa	KUM1	HERvii	1
TERc	FI-06258	FI-06258	1
TERc	FI-06258	FI-09100	1

Table S6. GenBank accession numbers determined for the strains characterized in this study

Species	**16S rRNA**	***Hsp65***	***rpoB***
M. engbaekii	AF480577, JN571172	JN571194–7	JN571242–5
M. heraklionense	GU084182	JN571191–2	JN571230–6
M. longobardum	JN571166	JN571199	JN571247–8
M. arupense	JN571167–9	FJ263631, JN571186–90	JN571215–29
M. hibernae	JN571173	JN571198	JN571246
M. kumamotonense	JN571176–9	JN571202–3	JN571251–58, JN571260
M. nonchromogenicum	JN571171	JN571193	JN571237–41
M. senuense	JN571174–5	JN571200–1	JN571249–50
M. terrae	JN571180–2	JN571204–7	JN571261–9
Unassigned strains of *M. terrae* complex	JN571183–5	JN571208–14	JN571267–73

E. Tortoli, Z. Gitti, H.-P. Klenk, S. Lauria, R. Mannino, P. Mantegani, A. Mariottini and I. Neonakis (2013). Survey of 150 strains belonging to the *Mycobacterium terrae* complex and description of *Mycobacterium engbaekii* sp. nov., *Mycobacterium heraklionense* sp. nov. and *Mycobacterium longobardum* sp. nov. *Int J Syst Evol Microbiol* **63**, 401–411.