Formal description of *Mycobacterium neglectum* sp. nov. and *Mycobacterium palauense* sp. nov., rapidly growing actinobacteria

Imen Nouioui[1*], Lorena Carro[1], Vartul Sangal[2], Marlen Jando[3], José Mariano Igual[4], Michael Goodfellow[1], Hans-Peter Klenk[1]

[1]School of Natural and Environmental Sciences, Ridley Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom.

[2]Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK

[3]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany

[4]Instituto de Recursos Naturales y Agrobiología de Salamanca, Consejo Superior de Investigaciones Científicas (IRNASA-CSIC), c/Cordel de Merinas 40-52, 37008 Salamanca, Spain

To whom correspondence should be addressed: imen.nouioui@ncl.ac.uk

Abstract

The taxonomic positions of two fast growing mycobacteria (CECT 8778^T and CECT 8779^T) were established using a polyphasic approach. The strains were shown to have chemotaxonomic, cultural and morphological properties consistent with their classification in the genus *Mycobacterium.* Multi-locus sequence analyses (MLSA) show that strain CECT 8778^T forms a well-supported clade together with the type strains of *Mycobacterium aurum*, *Mycobacterium austroafricanum* and *Mycobacterium vanbaalenii* while strain CECT 8779^T presents as a distinct branch that is well separated from its nearest phylogenetic neighbours; it is also apparent from the MLSA genetic distances that these strains are most closely related to the type strains of *Mycobacterium mageritense* and *M. vanbaalenii*, respectively. Digital DNA:DNA hybridization and average nucleotide identity values between each of the strains and its close phylogenetic neighbour are below the 70 and 96% threshold values for definition of prokaryotic species; these results are underpinned by corresponding phenotypic data. Based upon the consensus of the phenotypic and phylogenetic analyses, it can be concluded that the two strains represent novel species within the genus *Mycobacterium* for which the following names are proposed: *Mycobacterium neglectum* sp. nov., with the type strain CECT 8778^T (BN 3150^T = DSM 44756^T) and *Mycobacterium palauense* sp. nov., with the type strain CECT 8779^T (= DSM 44914^T).

Keywords: Actinobacteria, phenotyping, phylogeny, polyphasic taxonomy

Introduction

The genus *Mycobacterium* (Lehmann and Neumann 1896), the sole representative of the family *Mycobacteriaceae* (Chester 1897) can be distinguished from all of the other genera classified in the order *Corynebacteriales* by using a selection of genotypic and phenotypic methods (Goodfellow and Jones 2012). The genus encompasses pathogenic and non-tuberculous mycobacteria (Magee and Ward 2012; Forbes 2017; Gcebe et al. 2017) which can be assigned to two groups based on growth rates. Slowly growing strains require 7 or more days of incubation at optimal temperature to produce visible colonies from highly diluted inocula whereas those of rapidly growing strains are evident in fewer than 7 days under comparable conditions (Wayne and Kubica 1986). Polyphasic taxonomic procedures are now used to detect novel mycobacterial species, as exemplified by the delineation of species previously aggregated within the *Mycobacterium abscessus* and *Mycobacterium avium*

complexes (Ben Salah et al. 2009; Tortoli et al. 2016). Developments such as these are needed to detect the causal agents of mycobacterial infections and to establish the primary reservoirs of individual mycobacterial species within natural habitats (Tran and Dahl 2016; Shahraki et al. 2017).

Environmental mycobacteria are common in aquatic and terrestrial ecosystems (Nishiuchi et al. 2017; Roguet et al. 2016), including biofilms of water distribution systems (September et al. 2004; Feazel et al. 2009; Gomez-Smith et al. 2015). This study was undertaken to establish the taxonomic status of two rapidly growing mycobacteria: strain CECT 8778 ^{T} (DSM 44756 ^{T}) was isolated from a biofilm of a water distribution system and strain CECT 8779^{T} from marine sediment. A 16S rRNA gene sequence of strain DSM 44756 (then coded BN 3150) was deposited in GenBank (accession number AJ580802) under the name "Mycobacterium neglectum". This code and species epithet have been used quite extensively in the literature (Thomas et al. 2008; Hussein et al. 2009; Jenkins et al. 2009; Salah et al. 2009; Loret and Creub 2010; Pontiroli et al. 2013; Nishiuchi et al. 2017). However, at no stage has a formal description been given for "M. neglectum" hence this name has no standing in nomenclature (Rule 29 of International Code of Nomenclature of Prokaryotes [2008 revision]; Parker et al. 2015). In the present polyphasic study, we provide the first formal description of *Mycobacterium neglectum* sp. nov., with the type strain CECT 8778^{T}, while a second novel species represented by strain CECT 8779^{T} is named *Mycobacterium palauense* sp. nov.

Materials and methods

Source, maintenance and cultivation of strains

Strain CECT 8778^{T}, isolated from a biofilm of an underground drinking water system in Duisburg, Germany in 1999, and strain CECT 8779^{T}, isolated from a marine sediment collected from the Republic of Palau in 2004, were obtained from the Spanish Type Culture Collection. The strains, together with *Mycobacterium aurum* DSM 43999^{T} (Tsukamura 1966), *Mycobacterium austroafricanum* DSM 44191^{T} (Tsukamura et al. 1983), *Mycobacterium mageritense* DSM 44476^{T} (Domenech et al. 1997) and *Mycobacterium vanbaalenii* DSM 7152^{T} (Khan et al. 2002) were maintained as suspensions in 35% (v/v) glycerol at -80°C. Biomass for

the chemotaxonomic and molecular systematic studies on the isolates was cultured in shake flasks (200 revolutions per minute) of proteose peptone-meat extract-glycerol agar medium (PMG; DSMZ medium 250); after incubation at 28°C for 5 days, cells were harvested and washed three times in sodium chloride solution (0.9%, w/v). Cells for the chemotaxonomic analyses were freeze dried and stored at room temperature; wet biomass for the fatty acid analyses was prepared under the same conditions.

Phylogeny

Genomic DNA was extracted from strains CECT 8778^T and CECT 8779^T using the procedure described by Amaro et al. (2008). The genomes of the strains were sequenced using an MiSeq instrument (Illumina), as described by Sangal et al. (2015) and assembled into contigs using SPAdes 3.9.0 with a kmer length of 127 (Bankevich et al. 2012). Annotation of the genomes was achieved using the RAST pipeline available on the RAST server (Aziz et al. 2008, 2012). Complete 16S rRNA gene sequences of strains CECT 8778^T and CECT 8779^T were extracted from the draft genomes (accession numbers NVQE00000000 and NVQF00000000, respectively) and deposited in GenBank under accession numbers MF769621 and MF769712. Corresponding 16S rRNA gene sequences of the type strains of closely related *Mycobacterium* spp. were retrieved from the EzBioCloud server (Yoon et al. 2017) and pairwise sequence similarities calculated using the Genome-to-Genome Distance Calculator (GGDC) web server (Meier-Kolthoff et al. 2013a, b). Phylogenies derived from the 16S rRNA gene sequences were inferred using the GGDC web server adapted to single genes (Meier-Kolthoff et al. 2014). Multiple sequence alignments were generated using MUSCLE software (Edgar 2004) and a maximum-likelihood (ML) tree inferred from the alignment with RAxML (Stamatakis 2014) using rapid bootstrapping together with the auto Maximal-Relative-Error (MRE) criterion (Pattengale et al. 2010). Similarly, a maximum-parsimony (MP) tree was inferred from the alignments with the 'Tree analysis New Technology' (TNT) program (Goloboff et al. 2008) using 1000 bootstraps together with tree bisection and reconnection branch swapping and ten random sequence replicates. The sequences were checked for computational bias using the X^2 test implemented in PAUP* (Phylogenetic Analysis Using Parsimony) (Swofford 2002).

Partial sequences of three housekeeping genes, *hsp*65 (heat shock protein), *rpo*B (RNA polymerase beta subunit) and *rec*A (recombination protein A) (McNabb et al. 2004; Ramaprasad et al. 2016), were drawn from the draft genomes of strains CECT 8778^T and CECT 8779^T and deposited in GenBank under the accession numbers MF774022, MF774023,

MF774024， MF77402， MF774026, MF774027, respectively. A multilocus sequence analysis (MLSA) tree was generated from 3203 nucleotides (nt) of concatenated sequences of the three housekeeping genes and corresponding 16S rRNA gene sequences and ML and MP trees inferred as described above. In addition, a neighbour-joining (NJ) tree (Saitou and Nei 1987) was generated from the MEGA 7 software package (Kumar et al. 2015). The alignment of the concatenated sequences and the corresponding evolutionary distances were carried out using CLUSTAL W software (Thompson et al. 1997) and the Kimura two parameter model (Kimura 1980), respectively.

The average nucleotide identity (ANI) between strain CECT 8778^T and *M. aurum* DSM 43999^T (genome accession number is NZ_CVQQ00000000), *M. austroafricanum* DSM 44191^T genome accession number is (NZ_HG964469) and *M. vanbaalenii* PYR1^T (genome accession number is CP000511) and between strain CECT 8779^T and *M. mageritense* DSM 44476^T (genome accession number is NZ_CCBF000000000), their respective near phylogenetic neighbours, were calculated according to Rodriguez and Konstantinidis (2014). Similarly, digital DNA-DNA hybridization (dDDH) similarities were determined between the two strains and their close phylogenetic neighbours using the GGDC server (Meier-Kolthoff et al. 2013a).

Chemotaxonomy

The chemotaxonomic profiles of strains CECT 8778^T and CECT 8779^T and their respective close phylogenetic neighbours were determined using standard thin-layer chromatographic procedures. To this end, the strains were examined for diaminopimelic acid isomers (A_2pm) (Staneck and Roberts 1974), predominant menaquinones (Collins 1985), mycolic acids (Mininkin et al. 1980), diagnostic sugars (Lechevalier and Lechevalier 1970) and polar lipids (Mininkin et al. 1984). In addition, cellular fatty acids were extracted from freeze dried biomass of the strains and fatty acid methyl esters (FAMES) prepared following saponification and methylation using the procedure introduced by Miller (1982), as modified by Kuykendall et al. (1988). The FAMES were analysed by gas chromatography (Agilent 6890 instrument) and the resultant peaks automatically integrated. The identities of the fatty acids were determined using the standard Microbial Identification (MIDI) System, version 4.5, and the Myco 6 database (Sasser 1990).

Growth and cultural properties

Strains CECT 8778^T and CECT 8779^T were examined for their ability to grow and form colonies and pigments on glucose-yeast extract-malt extract agar (GYM, DSMZ medium 65), Löwenstein-Jensen (LJ) medium (Jensen 1932), Middlebrook 7H10 agar supplemented with oleic acid, albumin dextrose and catalase (MB7H10, Lorian 1968), proteose peptone-meat extract-glycerol agar (PMG; DSMZ medium 250) and tryptic soy agar (TSA, MacFaddin 1985) for 14 days at 4, 10, 20, 25, 28 and 42°C under light and dark conditions. The strains were also examined for acid-alcohol-fastness using the Ziehl-Neelsen method (Runyon et al. 1980) and for their ability to grow on PMG agar under anaerobic conditions at 28°C using an anaerobic bag system (Sigma-Aldrich 68061**)**.

Phenotypic tests

The two strains and their close phylogenetic neighbours were examined for phenotypic tests found to be useful in mycobacterial systematics (Magee and Ward 2012, Nouioui et al. 2017). The strains were tested for their ability to use sole carbon and sole nitrogen sources, to grow in the presence of several concentrations of sodium chloride, at a range of pH values and in the presence of antibiotics using GENIII microplates and an Omnilog device (BIOLOG, Hayward, CA). The tests were carried out in duplicate using freshly prepared inocula (OD_{600}-0.3-0.6) harvested from the mid-logarithmic growth phase of PMG agar plates incubated at 28°C for 7 days. The resultant data were exported and analysed using the opm package version 1.3.36 (Vaas et al. 2012, 2013). The strains were also examined for their ability to produce arylsulfatase after 3 and 14 days (Tomioka et al. 1990), catalase (Palomino et al. 2007) and heat stable catalase (Sequeira de Latini and Barrera 2008) and for niacin accumulation (Kent and Kubica 1985), resistance to potassium tellurite (Kent and Kubica 1985; Kilburn et al. 1969), degradation of Tween 80 (Ribón 2012) and urea hydrolysis (Palomino et al. 2007) using the media and incubation conditions described in these references. All of these tests were carried out in duplicate using the standard inoculum.

Results and discussion

The chemotaxonomic, growth and staining properties of strains CECT 8778^T and CECT 8779^T were shown to be consistent with their classification in the genus *Mycobacterium* (Magee and Ward 2012). The organisms were found to be strictly aerobic, Gram-positive, acid-alcohol fast, rapid growing, rod-shaped bacteria which contain *meso*-diaminopimelic acid, arabinose, galactose, glucose, rhamnose and ribose in whole organism hydrolysates (wall chemotype IV *sensu* Lechevalier and Lechevalier 1970); mixtures of saturated, unsaturated and 10-methyl

octadecanoic (tuberculostearic) fatty acids; mycolic acids; dihydrogenated menaquinones with nine isoprene units (MK9(H_2)) as the predominant isoprenologue; and a polar lipid profile that includes diphosphatidylglycerol, phosphatidylethanolamine (diagnostic phospholipid), phosphatidylinositol, as well as a glycophospholipid and glycolipids (phospholipid type II; Lechevalier et al. 1977). Both strains were found to produce unpigmented colonies under both light and dark conditions on LJ, MB7H10, PMG and TSA plates after 5 days at 28°C; moderate growth was observed at 20°C, 25°C and 37°C; optimal growth was detected at 28°C on GYM, MB7H10 and PMG agar after 5 days. The strains were unable to grow on any of these media at 4°C, 15°C, 42°C, or 45°C or under anaerobic conditions at 28°C on PMG agar. Strain CECT 8778^T and its nearest phylogenetic neighbours, *M. aurum* DSM 43999^T, *M. austroafricanum* DSM 44191^T and *M. vanbaalenii* DSM 7152^T, share several features; they are all acid acid-alcohol fast, rapid growing bacteria that grew on MB7H10 and PMG media at 28°C though only the test strain formed non-pigmented colonies. Strain CECT 8779^T and *M. mageritense* DSM 44476^T were found to have very similar cultural and morphological traits though only the latter grew at 22, 30, 37 and 45°C on LJ media (Domenech, et al. 1997).

The pairwise 16S rRNA gene similarities between strain CECT 8778^T and *M. aurum* NCTC 10437^T, *M. austroafricanum* DSM 44191^T, *Mycobacterium pyrenivorans* DSM 44605^T (Derz et al. 2004), *Mycobacterium vaccae* ATCC 25954^T (Bönicke and Juhasz 1964) and *M. vanbaalenii* PYR-1 were found to be 99.2%, 99.3%. 98.6%, 98.9%, 99.2%, respectively. It can be seen from Figure 1 that the two strains were found to be well separated from the type strains of the remaining fast growing *Mycobacterium* species. Strain CECT 8778^T was shown to form a distinct branch at the periphery of a well-supported subclade that included all of the organisms cited above. In turn, the pairwise 16S rRNA gene similarities between strain CECT 8779^T and the type strains of *M. mageritense*, *Mycobacterium peregrinum* ATCC 14467^T (Kusunoki and Ezaki 1992) and *Mycobacterium wollinskyi* ATCC 700010^T (Brown et al. 1999) were found to be 98.8%, 98.8% and 98.9%, respectively; strain CECT 8779^T formed a well defined branch in a subclade that also contained these organisms, albeit one that was supported by a high bootstrap value only in the ML analysis. It can also be seen from Figure 1 that all of the rapidly growing strains were sharply separated from the type strain of *Mycobacterium Tuberculosis* (Zopf 1883; Lehmann and Neumann 1896), the type species of the genus.

The MLSA trees based on the sequences of the three housekeeping genes and the corresponding 16S rRNA sequences are shown in Figure 2; this tree was inferred from 3257 nt, 604 of which were variable and 324 of which were parsimony-informative. The average bootstrap supports for the ML and MP trees were found to be 90.0% and 92.3%, respectively.

It is evident from Figure 2 that strain CECT 8778^T forms a well supported subclade together with the type strains of *M. aurum, M. austroafricanum* and *M. vanbaalenii* while strain CECT 8779^T forms a branch distinct from all of the other mycobacteria, including the type strains of *M. mageritense, M. peregrinum* and *M. wollinskyi.* Moreover, it is clear from Table 1 that strains CECT 8778^T and CECT 8779^T share close genetic distances, namely 0.02 and 0.03, with the type strains of *M. vanbaalenii* and *M. mageritense*, respectively. However, in the corresponding NJ tree, strain CECT 8778^T was shown to be more closely related to the *M. austroafricanum* and *M. vanbaalenii* type strains than to the *M. aurum* strain though these relationships were not supported by high bootstrap values (Fig. S1).The phylogenetic trees based on the sequences of the individual housekeeping genes are shown in Figures S2-S4; the relationships between strain CECT 8778^T and the type strains of *M. aurum, M. austroafricanum* and *M. vanbaalenii* are evident in all of the trees though bootstrap values are low. In contrast, strain CECT 8779^T was found to form a distinct branch in all of these trees.

The genome sizes of strains CECT 8778^T (NVQE00000000) and CECT 8779^T (NVQF00000000) were both found to be ~6.2Mb with average *in silico* G+C contents of 65.3mol% and 69.4mol%, respectively. The draft genome of strain CECT 8778^T was generated from 123 contigs with N lengths of 116368 and was found to contain 6159 predicted protein coding sequences and 51 tRNA genes. Similarly, the draft genome of strain CECT 8779^T was compiled from 210 contigs with N lengths of 61684 and was shown to have 5802 predicted protein coding sequences and 71 tRNA genes. The coverage for strains CECT 8778^T and CECT 8779^T were 79X and 67X, respectively. Strains CECT 8778^T and CECT 8779^T contain, very similar subsystem gene functions as exemplified in Table S1.

Strain CECT 8778^T and *M. aurum* NCTC 10437^T, *M. austroafricanum* DSM 44191^T and *M. vanbaalenii* PYR-1^T, currently its closest phylogenetic neighbours, were found to share dDDH similarities of 20.1%, 21.1% and 21.1%, respectively, values well below the 70% cut off point recommended for the delineation of prokaryotic species (Wayne et al. 1987). The corresponding ANI similarities between strain CECT 8778^T and the three strains mentioned above were found to be 78.4%, 78.7% and 79.6%, values well below the 95-96% threshold used to distinguish between closely related species of prokaryotes (Goris et al. 2007; Richter and Rosselló-Móra 2009; Chun and Rainey 2014). Similarly, the dDDH and ANI values between strain CECT 8779^T and *M. mageritense* DSM 44476^T, its current closest phylogenetic neighbour, were 20.9% and 79.1%, respectively, values well below the species thresholds cited above.

Identical results were recorded for all of the phenotypic tests that were carried out in duplicate. It can be seen from Table 2 that strains CECT 8778^T and CECT 8779^T can be distinguished from one another and from their respective reference strains using a combination of phenotypic properties though it is apparent that all of these organisms share a common set of features. Strain CECT 8778^T, unlike the type strains of *M. aurum*, *M. austroafricanum* and *M. vanbaalenii*, was shown to grow in the presence of tetrazolium blue and tetrazolium violet and also differed from these strains by its inability to utilise L-alanine, butyric acid, α-*hydroxy*-butyric acid, α-*keto*-butyric acid, galactose, glycerol, L-lactic acid, pectin and sucrose or to grow in the presence of lithium chloride, nalidixic acid and sodium chloride (up to 8%, w/v). Similarly, strain CECT 8779^T, unlike *M. mageritense* DSM 44476^T, was shown to use D-arabitol, D-glucose-phosphate, D-maltose, D-mannose, methyl pyruvate, pectin, D-sorbitol, sucrose, D-trehalose and D-turanose as sole carbon sources and to grow in the presence of lithium chloride and sodium chloride (8%, w/v). In turn, strain CECT 8778^T, unlike strain CECT 8779^T, was seen to metabolise dextrin, *myo*-inositol, D-and L-malic acid, quinic acid, D- saccharic acid, D- salicin, L-serine and bromo-succinic acid. Conversely, strain CECT 8779^T was found to use L-alanine, L-aspartic acid, butyric acid, citric acid, D-glucose-6-phosphate, glycerol, α-*keto*-glutaric acid, L-histidine, L-lactic acid, D-maltose, N-acetyl-β-D-mannosamine, methyl pyruvate, pectin, L-rhamnose and sucrose and to grow in the presence of lithium chloride, nalidixic acid and sodium chloride.

The kind of mycolic acids synthesised by representatives of *Mycobacterium* species fall into several well established patterns of taxonomic value (Mininkin et al. 1985; Magee and Ward 2012). In the present study, the test strains were found to have different mycolic acid profiles: strain CECT 8778^T was shown to contain *α*- and *decarboxy*- mycolates and strain CECT 8779^T *α-, keto-* and *methoxy*- mycolates. This latter pattern serves to distinguish strain CECT 8779^T from *M. mageritense* DSM 44476^T which is characterised by the presence of α-, α'- and *epoxy*-mycolates (Domenech et al. 1997). Similarly, the mycolic acid profile distinguishes strain CECT 8778^T from *M. aurum* DSM 43999^T and *M. austroafricanum* DSM 44191^T as these strains have *α*- and *keto*-mycolates and wax esters (Mininkin et al. 1985; Magee and Ward 2012). Similarly, complex polar lipid pattern of strain CECT 8778^T serves to distinguish it from both strain CECT 8779^T and from the type strains of *M. aurum, M. austroafricanum* and *M. vanbaalenii;* all of these strains were found to contain diposphatidylglycerol, phosphatidylethanolamine (diagnostic lipid), phosphatidylinositol and

glycophospholipids. Strain CECT 8779^T, unlike its near phylogenetic neighbour, *M. mageritense* DSM 44476^T, was shown to have a lipid pattern that lacked phosphatidylglycerol.

All of the strains produced complex mixtures of straight-chain saturated, unsaturated and 10-methyl-octadecanoic (tuberculostearic) fatty acids, a profile typical of members of the genus *Mycobacterium* (Magee and Ward 2012). With few exceptions, strains CECT 8778^T, *M. aurum* DSM 43999^T, *M. austroafricanum* DSM 44191^T and *M. vanbaalenii* DSM 7152^T were found to have major proportions (>10% of total fatty acid) of $C_{16:0}$ (13.3-72.6%), $C_{18:1}$ ω9c (7.6-19.7.0%) and summed features 2 (12.6-43.2%) and 3 (12.3-16.8%) though the predominant component varied (Table 3). The fatty acid profiles of CECT 8778^T and the type strain of *M. vanbaalenii* were distinct; the test strain, for instance, produced higher proportions of summed feature 2 (43.2 against 27.6%). Even greater differences were found between the fatty acid profiles of CECT 8778^T and *M. aurum* DSM 43999^T; the latter, for instance, was especially rich in $C_{16:0}$ (25.4% against 13.3.0%). Marked differences were found between the fatty acid profiles of strain CECT 8779^T and *M. mageritense* DSM 44476^T as the former contained moderate proportions of $C_{16:1}$ ω9c and summed features 2 and a lower amount of $C_{16:0}$ (25.1 against 40.0%) (Table 3).

In summary, strain CECT 8778^T can be distinguished readily from *M. aurum* DSM 43999^T, *M. austroafricanum* DSM 44191^T and *M. vanbaalenii* DSM 7251^T, its close phylogenetic neighbours, in the 16S rRNA and MLSA gene trees, by low ANI and dDDH scores and by a range of chemotaxonomic and phenotypic markers. A similar wealth of taxonomic data separate strain CECT 8779^T from the type strain of *M. mageritense*, its close phylogenetic neighbour. These datasets clearly show that strains CECT 8778^T and CECT 8779^T represent new centres of taxonomic variation within the genus *Mycobacterium*; the names chosen for these species are *Mycobacterium neglectum* sp. nov. and *Mycobacterium palauense* sp. nov., respectively. The Digital Protologue database TaxoNumbers for strains CECT 8778^T and CECT 8779^T are TA00318 and TA000312, respectively.

Description of *Mycobacterium neglectum* sp. nov.

Mycobacterium neglectum (neg.lec'tum. L. adj. *neglectum*, neglected refecting the history of the strain)

Strict aerobic, Gram-stain positive, acid-alcohol fast, rapid growing organism which forms unpigmented colonies after growth on Middelbrook 7H10, proteose peptone-meat extract-glycerol and LJ media after incubation under the light and dark conditions after 5 days

at 28°C. Grows between 25°C and 28°C, optimally at 28°C and at pH 7. Produces arylsulfatase after 3 and 14 days, catalase, nitrate reductase, urease and grows in the presence of potassium tellurite. Additional phenotypic data are given in the text and in Table 1. Whole cell hydrolysates are rich in *meso*-diaminopimelic acid, arabinose, galactose, glucose, ribose and rhamnose; the polar lipid profile contains diphosphatidylglycerol, glycophospholipids (GPL1-2), phosphatidylethanolamine, phosphatidylinositol and a glycolipid; MK9 (H_2) is the predominant menaquinone and the major fatty acid is $C_{17:1}$ ω7c/18 alcohol. Contains *α*- and *decarboxy*-mycolic acids. The DNA G+C content determined from the draft genome of the type strain is 65.3 mol%.

The type strain, CECT 8778^T (= BN3150^T = DSM 44756^T) was isolated from a biofilm of an underground drinking water system in Germany. The Genbank accession number of the draft genome sequence of strain CECT 8778^T is NVQE00000000.

Description of *Mycobacterium palauense* sp. nov.

Mycobacterium palauense (pa.lau.en'se N.L. neut. adj. *palauense* referring to the Republic Palau, the source of the strain)

Strict aerobic, Gram-stain positive, acid-alcohol fast, rapid growing organism which forms unpigmented colonies after growth on Middelbrook 7H10, proteose peptone-meat extract-glycerol and LJ media after incubation under the light and dark conditions for 5 days at 28°C. Grows between 25°C and 28°C, optimally at 28°C and at pH 7and in the presence of up to 8% w/v NaCl. Produces arylsulfatase after 3 and 14 days, catalase, accumulates niacin, degrades Tween 80 and grows in presence of potassium tellurite. Additional phenotypic data are given in the text and in Table 1. Whole cell hydrolysates are rich in *meso*-diaminopimelic acid, arabinose, galactose, glucose, ribose and rhamnose; the polar lipid profile contains diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol, as well as glycophospholipids (GPL_{1-2}) and a glycolipid; MK9 (H_2) is the predominant menaquinone and the major fatty acids are $C_{16:0}$ and $C_{18:1}$ ω9c. Contains α, *keto- methoxy* mycolic acids. The DNA G+C content determined from the draft genome of the type strain is 69.4 mol %.

The type strain, CECT 8779^T (= DSM 44914^T) was isolated from marine sediment from the Republic of Palau. The Genbank accession number of the draft genome sequence of strain CECT 8779^T is NVQF00000000.

Acknowledgements This project was supported by the School of Natural and Environmental Sciences (Newcastle University). IN. and LC. are grateful to Newcastle University for postdoctoral fellowships.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical statement This article does not contain any studies inoculating human participants or animals.

References

Amaro A, Duarte E, Amado A, Ferronha H, Botelho A (2008) Comparison of three DNA extraction methods for *Mycobacterium bovis*, *Mycobacterium tuberculosis* and *Mycobacterium avium* subsp. *avium*. Lett Appl Microbiol 47:8-11

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K et al (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75

Aziz RK, Devoid S, Disz T, Edwards RA, Henry CS, Olsen GJ, Olson R et al (2012) SEED servers: high-performance access to the SEED genomes, annotations, and metabolic models. PLoS One 7:e48053

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-477

Ben Salah I, Cayrou C, Raoult D, Drancourt M. (2009) *Mycobacterium marseillense* sp. nov., *Mycobacterium timonense* sp. nov. and *Mycobacterium bouchedurhonense* sp. nov., members of the *Mycobacterium avium* complex. Int J Syst Evol Microbiol 59:2803-2808

Bönicke R, Juhasz SE (1964) Beschreibung der neuen Species *Mycobacterium vaccae* n. sp. Zentralb Bakteriol, Parasitenkd Infektionskr Hyg, Abt I Orig 192:133-135

Brown BA, Springer B, Steingrube VA, Wilson RW, Pfyffer GE, Garcia MJ, Menendez MC, et al (1999) *Mycobacterium wolinskyi* sp. nov. and *Mycobacterium goodii* sp. nov., two new rapidly growing species related to *Mycobacterium smegmatis* and associated with human wound infections: A cooperative study from the International Working Group on Mycobacterial Taxonomy. Int J Syst Bacteriol 49:1493-1511

Chester FD (1897) Report of the mycologist: bacteriological work. Del Agric Exp Sta Bull 9:38-145

Chun J, Rainey FA (2014) Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*. Int J Syst Evol Microbiol 64:316–324.

Collins MD (1985) Analysis of isoprenoid quinones. Meth Microbiol 18:329-366

Derz K, Klinner U, Schuphan I, Stackebrandt E, Kroppenstedt RM (2004) *Mycobacterium pyrenivoran*s sp. nov., a novel polycyclic-aromatic-hydrocarbon-degrading species. Int J Syst Evol Microbiol 54:2313-2317

386 Domenech P, Jimenez MS, Menendez MC, Bull TJ, Samper S, Manrique A, Garcia MJ (1997)
387 *Mycobacterium mageritense* sp. nov. Int J Syst Bacteriol 47:535-540
388 Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high
389 throughput. Nucleic Acids Res 32:1792-1797
390 Feazel LM, Baumgartner LK, Peterson KL, Frank DN, Harris JK, Pace NR (2009)
391 Opportunistic pathogens enriched in showerhead biofilms. PNAS 106: 16393-16399.

392 Forbes BA (2017) Mycobacterial taxonomy. J Clin Microbiol 55: 380-383.

393 Gcebe N, Rutten V, van Pittius NG, Naicker B, Michel A (2017) *Mycobacterium*
394 *malmesburyense* sp. nov., a non-tuberculous species of the genus *Mycobacterium*
395 revealed by multiple gene sequence characterization. Int J Syst Evol Microbiol 67: 832-
396 838
397 Goodfellow M, Jones A (2012) Order V. *Corynebacteriales* ord. nov. In Goodfellow M,
398 Kämper P, Busse HJ, Trujillo ME, Suzuki KI, Ludwig W, Whitman WB (eds) Bergey's
399 Manual of Systematic Bacteriology, 2nd edn, vol 5, The *Actinobacteria*, Part A. Springer,
400 New York, pp 232-243
401 Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis.
402 Cladistics 24:774-786
403 Gomez-Smith CK, . LaPara TM, Hozalski RM (2015) Sulfate reducing bacteria and
404 mycobacteria dominate the biofilm communities in a chloraminated drinking water
405 distribution system. Environ Sci Technol. 49: 8432-8440.
406 Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM (2007)
407 DNA–DNA hybridization values and their relationship to whole-genome sequence
408 similarities. Int J Syst Evol Microbiol 57:81–91.
409 Hussein Z, Landt O, Wirths B, Wellinghausen N (2009) Detection of non-tuberculous
410 mycobacteria in hospital water by culture and molecular methods. Int J Med Microbiol
411 299: 281–290.
412 Jensen KA (1932) Reinzüchtung und Typenbestimmung von Tuberkelbazillenstämmen.
413 Zentralbl Bakteriol I. Orig 125:222-239
414 Jenkins SN, Waite IS, Blackburn A, Husband R, Rushton SP, Manning DC, O'Donnell AG
415 (2009) Actinobacterial community dynamics in long term managed grasslands. Antonie
416 van Leeuwenhoek 95:319–334.

Kent PT, Kubica GPW (1985) Public Health Mycobacteriology A Guide for the Level III Laboratory. U.S. Departement. of Health and Human Services, Public Health Service, Centers for Disease Control, Atlanta

Khan AA, Kim SJ, Paine DD, Cerniglia CE (2002) Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as *Mycobacterium vanbaalenii* sp. nov. Int J Syst Evol Microbiol 52:1997-2002

Kilburn JO, Silcox VA, Kubica GP (1969) Differential identification of mycobacteria. V. The tellurite reduction test. Am Rev Resp Dis 99:94-100

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

Kumar S, Stecher G, Tamura K (2015) MEGA 7: Molecular Evolutionary Genetics Analysis version 7. 0 for bigger datasets. Mol Biol Evol 3:1870-1874

Kusunoki S, Ezaki T (1992) Proposal of *Mycobacterium peregrinum* sp. nov., nom. rev., and elevation of *Mycobacterium chelonae* subsp. *abscessus* (Kubica et al.) to species status: *Mycobacterium abscessus* comb. nov. Int J Syst Bacteriol 42:240-245

Kuykendall LD, Roy MA, O'Neill JJ, Devine TT (1988) Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum*. Int J Syst Evol Microbiol 38:358-361

Lechavalier MP, Lechevalier HA (1970) Composition of whole-cell hydrolysates as a criterion in the classification of aerobic actinomycetes. In: Prauser H (ed) The *Actinomycetales*, Gustav Fischer Verlag, Jena, pp 311-316

Lechevalier MP, De Biévre C, Lechevalier H (1977) Chemotaxonomy of aerobic actinomycetes: phospholipid composition. Biochem Syst Evol 5:249-260

Lehmann KB, Neumann R (1896) Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik. J. F. Lehmann, München

Loret JF, Greub G (2010) Free-living amoebae: Biological by-passes in water treatment. Int J Hyg Environ Heal 213: 167–175.

Lorian V (1968) Differentiation of *Mycobacterium tuberculosis* and Runyon Group 3 "V" strains on direct cord-reading agar. Am Rev Resp Dis 97:1133-1135

MacFaddin JF (1985) Media for Isolation - Cultivation - Identification - Maintenance of Medical Bacteria. vol 1, Williams & Wilkins, Baltimore

Magee JG, Ward AC (2012) Genus I. *Mycobacterium* Lehmann and Neumann 1896, 363^{AL}, In: Goodfellow M, Kämper P, Busse HJ, Trujillo ME, Suzuki KI, Ludwig W, Whitman

WB (eds) Bergey's Manual of Systematic Bacteriology, 2nd edn, vol 5. The *Actinobacteria*, Part A. Springer, New York, pp 312-375

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

Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M (2013a) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60

Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP (2013b) When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 195(6):413-418

Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C et al (2014) Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 9:2

Miller LT (1982) Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. J Clin Microbiol 16(3):584-586

Minnikin DE, Hutchinson IG, Caldicott AB, Goodfellow M (1980) Thin-layer chromatography of methanolysates of mycolic acid-containing bacteria. J ChromatogrA. 188: 221-233

Minnikin DE, O'Donnell AG, Goodfellow M, Alderson G, Athalye M, Schaal A, Parlett JH (1984) An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids. J Microbiol Meth 2(5):233-241

Minnikin DE, Minnikin SM, Parlett JH, Goodfellow M (1985) Mycolic acid patterns of some rapidly-growing species of *Mycobacterium*. Zentralbl Bakteriol Mikrobiol Hyg A. 259:446-60

Nishiuchi Y, Iwamoto T, Maruyama F (2017) Infection sources of a common non-tuberculous mycobacterial pathogen, *Mycobacterium avium* complex. Front Med. 4: 27.

Nouioui I, Carro L, Teramoto K, Igual JM, Jando M, Montero-Calasanz MC, Sutcliffe I et al (2017) *Mycobacterium eburneum* sp. nov., a non-chromogenic, fast-growing strain isolated from sputum. Int J Syst Evol Microbiol 67: 3174-3181.

Palomino JC, Leão SC, Ritacco V (2007) Tuberculosis 2007- From basic science to patient care. Institute of Tropical Medicine. Belgium, Brazil, Argentina

Parker CT, Tindall BJ, Garrity GM (2015) Inetrnational Code of Nomenclature of Prokaryotes (2008 Revision). Int J Syst Microbiol doi:10.1099/ijsem.0.000778

Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A (2010) How many bootstrap replicates are necessary? J Comput Biol 17(3):337-354

Pontiroli A, Khera TT, Oakley BB, Mason S, Dowd SE, Travis ER, Erenso G, Aseffa A, Courtenay O, Wellington EMH (2013) Prospecting environmental mycobacteria: combined molecular approaches reveal unprecedented diversity. PLOS One 8:e68648.

Ramaprasad EV, Rizvi A, Banerjee S, Sasikala C, Ramana CV (2016) *Mycobacterium oryzae* sp. nov., a scotochromogenic, rapidly growing species is able to infect human macrophage cell line. Int J Syst Evol Microbiol 66:4530-4536

Ribón W (2012) Biochemical isolation and identification of mycobacteria, In: Jimenez-Lopez JC (ed) Biochemical Testing. InTech. doi:10.5772/34309. https://www.intechopen.com/books/biochemical-testing/biochemical-isolation-and-identification-of-mycobacteria

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106(45):19126-19131

Rodriguez RLM, Konstantinidis KT (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1

Roguet A, Therial C, Saad M, Boudahmane L, Moulin L, Lucas F. S (2016) High mycobacterial diversity in recreational lakes. Antonie van Leeuwenhoek. 109: 619-631.

Runyon EH, Karlson AG, Kubica GP, Wayne LG (1980) *Mycobacterium*. American Society for Microbiology. Washington DC.

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

Salah IB, Ghigo E, Drancourt M (2009) Free-living amoebae, a training field for macrophage resistance of mycobacteria. Clin Microbiol Infect 15:894-905.

Sangal V, Jones AL, Goodfellow M, Hoskisson PA, Kämpfer P, Sutcliffe IC (2015) Genomic analyses confirm close relatedness between *Rhodococcus defluvii* and *Rhodococcus equi* (*Rhodococcus hoagii*). Arch Microbiol 197:113-116

Sasser MJ (1990) Identification of bacteria by gas chromatography of cellular fatty acids. Technical Note 101, Microbial ID USA: Inc, Newark, Del

September SM, Brözel VS, Venter SN (2004) Diversity of nontuberculoid *Mycobacterium* species in biofilms of urban and semi urban drinking water distribution systems. Appl Environ Microbiol 70: 7571–7573.

Sequeira de Latini MD, Barrera L (2008) Manual para el Diagnóstico Bacteriológico de la Tuberculosis: Normas y Guía Técnica, Parte I Baciloscopía. Organización Panamericana de la Salud

Shahraki AH, Trovato A, Mirsaeidi M, Borroni E, Heidarieh P, Hashemzadeh M, et al (2017) *Mycobacterium persicum* sp. nov., a novel species closely related to *Mycobacterium kansasii* and *Mycobacterium gastri.* Int J Syst Evol Microbiol 67:1766–1770

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313

Staneck JL, Roberts GD (1974) Simplified approach to identification of aerobic actinomycetes by thin layer chromatography. Appl Microbiol 28:226-231

Swofford DL (2002) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0. Sinauer Associates, Sunderland

Thomas V, Loret JF, Jousset M, Greub G (2008) Biodiversity of amoebae and amoebae-resisting bacteria in a drinking water treatment plant. Environ Microbiol 10: 2728–2745.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:876-4882

Tomioka H, Saito H, Sato K, Dawson DJ (1990) Arylsulfatase activity for differentiating *Mycobacterium avium* and *Mycobacterium intracellulare.* J Clin Microbiol 28:2104-2106

Tortoli E, Kohl TA, Brown-Elliott BA, Trovato A, Leão SC, Garcia MJ et al (2016) Emended description of *Mycobacterium abscessus*, *Mycobacterium abscessu*s subsp. *abscessus* and *Mycobacterium abscessus* subsp. *bolletii* and designation of *Mycobacterium abscessus* subsp. *massiliense* comb. nov. Int J Syst Evol Microbiol 66:4471-4479

Tran PM, Dahl JL (2016) *Mycobacterium sarraceniae* sp. nov. and *Mycobacterium helvum* sp. nov., isolated from the pitcher plant *Sarracenia purpurea*. Int J Syst Evol Microbiol 66:4480-4485

Tsukamura M (1966) Adansonian classification of mycobacteria. J Gen Microbiol 45:253-273

Tsukamura M, van der Meulen HJ, Grabow WOK (1983) Numerical taxonomy of rapidly growing, scotochromogenic mycobacteria of the *Mycobacterium parafortuitum* complex: *Mycobacterium austroafricanum* sp. nov. and *Mycobacterium diernhoferi* sp. nov., nom. rev. Int J Syst Bacteriol 33:460-469

550 Vaas LAI, Sikorski J, Michael V, Göker M, Klenk HP (2012) Visualization and curve-
551 parameter estimation strategies for efficient exploration of phenotype microarray
552 kinetics. PLoS ONE 7: e34846
553 Vaas LAI, Sikorski J, Hofner B, Fiebig A, Buddruhs N, Klenk HP, Göker M (2013) Opm: an
554 R package for analysing OmniLog (R) phenotype microarray data. Bioinformatics
555 29:1823–1824
556 Wayne LG, Kubica GP (1986) The mycobacteria, In: Sneath PHA, Mair S, Sharpe ME, Holt
557 JG (eds) Bergey's Manual of Systematic Bacteriology, vol 2, Williams & Wilkins,
558 Baltimore, pp 1435-1457
559 Wayne L, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, Moore LH
560 et al (1987) Report of the *ad hoc* committee on reconciliation of approaches to bacterial
561 systematics. Int J Syst Evol Bacteriol 37:463-464
562 Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J (2017) Introducing EzBioCloud: A
563 taxonomically united database of 16S rRNA and whole genome assemblies. Int J Syst
564 Evol Microbiol 67:1613-1617
565 Zopf W (1883) Die Spaltpilze. Edward Trewendt, Breslau, pp 1-100

566

567

568 **Table 1**. Genetic distances between strains CECT 8778^T and CECT 8779^T and between them and their close phylogenetic neighbours

		MLSA (Kimura 2-paramenter) distance													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Strain CECT 8779^T														
2	Strain CECT 8778^T	0,051													
3	*Mycobacterium peregrinum*	0,031	0,039												
4	*Mycobacterium mageritense*	0,030	0,039	0,009											
5	*Mycobacterium wolinskyi*	0,032	0,036	0,012	0,012										
6	*Mycobacterium chlorophenolicum*	0,036	0,034	0,031	0,034	0,034									
7	*Mycobacterium chubuense*	0,037	0,036	0,033	0,036	0,037	0,006								
8	*Mycobacterium psychrotolerans*	0,037	0,039	0,035	0,036	0,037	0,015	0,016							
9	*Mycobacterium austroafricanum*	0,048	0,028	0,046	0,048	0,049	0,029	0,032	0,036						
10	*Mycobacterium vanbaalenii*	0,047	0,026	0,044	0,046	0,046	0,026	0,029	0,035	0,003					
11	*Mycobacterium aurum*	0,050	0,029	0,038	0,041	0,041	0,032	0,030	0,037	0,028	0,026				
12	*Mycobacterium rufum*	0,040	0,036	0,034	0,037	0,036	0,009	0,012	0,018	0,035	0,034	0,036			
13	*Mycobacterium arcueilense*	0,054	0,044	0,030	0,035	0,036	0,049	0,051	0,046	0,054	0,052	0,046	0,051		
14	*Mycobacterium alvei*	0,056	0,043	0,032	0,036	0,039	0,049	0,052	0,047	0,059	0,057	0,046	0,052	0,008	
15	*Tsukamurella paurometabola*	0,071	0,077	0,074	0,076	0,073	0,073	0,073	0,073	0,083	0,081	0,087	0,077	0,084	0,086

569

570

571 **Table 2**. Phenotypic features that distinguish strains CECT 8778^T and CECT 8779^T from one another and from their near phylogenetic
572 neighbours. All data are from the present study.
573

	strain CECT 8778^T	*M. aurum* DSM 43999^T	*M. austroafricanum* DSM 44191^T	*M. vanbaalenii* DSM 7152^T	strain CECT 8779^T	*M. mageritense* DSM 44476^T
Biochemical tests:						
Arylsufatase 3 days	+	-	+	+	+	+
Heat stable catalase 68°C	+	-	-	-	-	-
Niacin, Tween 80	+	+	-	-	+	+
Urea hydrolysis	+	-	-	+	-	-
GEN III Biolog microplate tests						
Utilisation of sugars:						
D-Arabitol,	+	-	+	-	+	-
Dextrin	+	-	-	+	-	+
D-Galactose	-	+	+	-	-	-
N-acetyl-D-Glucosamine	+	-	+	-	+	+
3-*o*-methyl-D-Glucose	-	-	+	-	-	-
D-Glucose-6-phosphate	-	-	+	-	+	-
Glycerol	-	+	+	+	+	+
myo-Inositol	+	+	+	+	-	-
D-Maltose	-	-	+	-	+	-
N-acetyl-β-D-Mannosamine	-	-	+	-	+	+
D-Mannose, D-sorbitol, D-trehalose	+	+	+	+	+	-
L-Rhamnose	-	-	+	-	+	+
D-Salicin	+	-	+	-	-	-

Sucrose	-	+	+	+	+	-
D-Turanose	+	+	-	+	+	-
Utilisation of organic acids:						
Butyric acid	-	+	+	+	+	+
β-amino-*n*-Butyric acid	+	+	-	-	+	+
α-*hydroxy*-Butyric acid, α-*keto*-Butyric acid	-	+	+	+	-	+
Citric acid	-	-	-	-	+	+
D-Galacturonic acid	-	+	-	-	-	-
α-*keto*-Glutaric acid	-	+	-	+	+	+
L-Lactic acid	-	+	+	+	+	+
D-and L-Malic acid	+	+	+	+	-	+
Methyl pyruvate	-	+	+	+	+	-
Quinic acid	+	-	+	-	-	-
D-Saccharic acid	+	+	+	+	-	-
Bromo-Succinic acid	+	+	+	+	-	+
Utilisation of amino acids:						
L-Alanine	-	+	+	-	+	+
L-Aspartic acid	-	-	+	-	+	+
Glycyl- L-proline	+	-	+	+	+	+
L-Histidine	-	-	-	-	+	+
D-Serine #2, L-Pyroglutamic acid	-	+	-	-	-	+
L-Serine	+	-	+	-	-	+
Resistance to:						
Lincomycin	-	-	-	-	-	+
Lithium chloride	-	+	+	+	+	-
Nalidixic acid, sodium chloride (1% w/v)	-	+	+	+	+	+

Rifamycin SV, sodium bromate	+	+	-	+	+	+
Sodium chloride (4%w/v)	-	+	-	-	+	+
Sodium chloride (8%w/v)	-	+	-	-	+	-
Sodium formate	+	+	+	+	-	+
Tetrazolium violet	+	-	-	-	+	+
Tetrazolium blue	+	-	-	-	-	-
Troleandomycin	-	+	-	-	-	-
Vancomycin	-	+	-	+	-	-
Growth in presence of:						
Gelatin	-	+	-	-	-	+
Pectin	-	+	+	+	+	-
Tween 40	+	-	+	+	+	+
Chemotaxonomic traits						
Polar lipids	DPG, PE, PI, GL GPL1-2	DPG, PE, PI, GL GPL1-2	DPG, PE, PI, GL GPL1-2	DPG, PE, PI, GL GPL1-2	DPG, PE, PI, GL GPL1-2	DPG, PE, PI, PG, GL GPL1-2
Fatty acids (>20 %)	Summed feature 2	$C_{16:0}$	$C_{16:0}$	Summed feature 2	$C_{16:0}$, $C_{18:1}$ ω9c	$C_{16:0}$, $C_{18:1}$ ω9c
Mycolic acids	*α*- and *decarboxy*-mycolates	*α-keto*-mycolates and wax esters*	*α-keto*-mycolates and wax esters*	*α-keto*-mycolates and wax esters*	α, *keto-methoxy* mycolic acids	α-, α'- and *epoxy*-mycolates**
DNA G+C content (%)	65.3	67.5	67.4	67.8	69.4	66.6

574 + Positive reaction; - negative reaction.

575 Positive results recorded for all of the strains: arylsufatase (14 days) and catalase (biochemical tests); utilisation of acetic acid, acetoacetic acid, L-
576 arginine, β-*hydroxy*-butyric acid, D-glucose, D-frucose, D-gluconic acid, L-glutamic acid, D-mannitol and propionic acid; growth at pH 6 and in
577 presence of aztreonam, potassium tellurite and 1% sodium lactate. Negative results detected for all of the strains: utilisation of D-aspartic acid, D-

cellobiose, D-fructose-6-phosphate, D-and L-fucose, L-galactonic acid-γ-lactone, *N*-acetyl-D-galactosamine, β-gentiobiose, β-methyl-D-glucoside, glucuronamide, D-glucuronic acid, D-lactic acid methyl ester, α-D-lactose, D-melibiose, mucic acid, *N*-acetyl-neuraminic acid, *p-hydroxy*-phenylacetic acid, D-raffinose and stachyose; growth at pH 5 and in the presence of fusidic acid, guanidine hydrochloride, inosine, minocycline and niaproof. Abbreviation: DPG: diphosphatidylglycerol; GPL1-2: glycophospholipids; GL: glycolipid; PE: phosphatidylethanolamine; PI: phosphatidylinositol; PG: glycophospholipid; summed features 2, $C_{17:1}$ ω7c /18 alcohol; *data taken from Mininkin et al. 1985, Magee and Ward 2012; ** data taken from Domenech et al. 1997.

Table 3. Fatty acid profiles (%) of strain CECT 8778^T (1), *M. aurum* DSM 43999^T (2), *M. austroafricanum* DSM 44191^T (3), *M. vanbaalenii* DSM 7152^T (4), strain CECT 8779^T (5) and *M. mageritense* DSM 44476^T (6). All data are from the present study.

	1	2	3	4	5	6
$C_{14:0}$	3.6	3.4	2.6	2.4	5.2	7.3
$C_{16:1}$ ω9c	1.0	0.8	0.9	1.5	3.6	-
$C_{16:1}$ ω6c	4.2	6.9	10.7	6.8	6.5	10.3
$C_{16:1}$ ω7c	0.5	-	-	-	1.6	4.1
$C_{16:0}$	13.3	25.4	27.6	17.0	25.4	40.0
Summed feature 2	43.2	12.6	15.7	27.6	7.6	-
$C_{18:1}$ ω9c	7.6	19.7	8.9	13.8	26.2	28.3
$C_{18:0}$	-	1.6	2.7	1.6	-	-
10 Me-$C_{18:0}$	11.0	12.6	14.2	8.6	7.8	8.7
Summed feature 3	12.3	15.2	15.0	16.8	-	-
$C_{20:0}$	-	-	0.3	0.3	-	-

Summed features 2, $C_{17:1}$ ω7c /18 alcohol and summed feature 3, 20:0 ALC 18.838/ECL20:0 ALC.

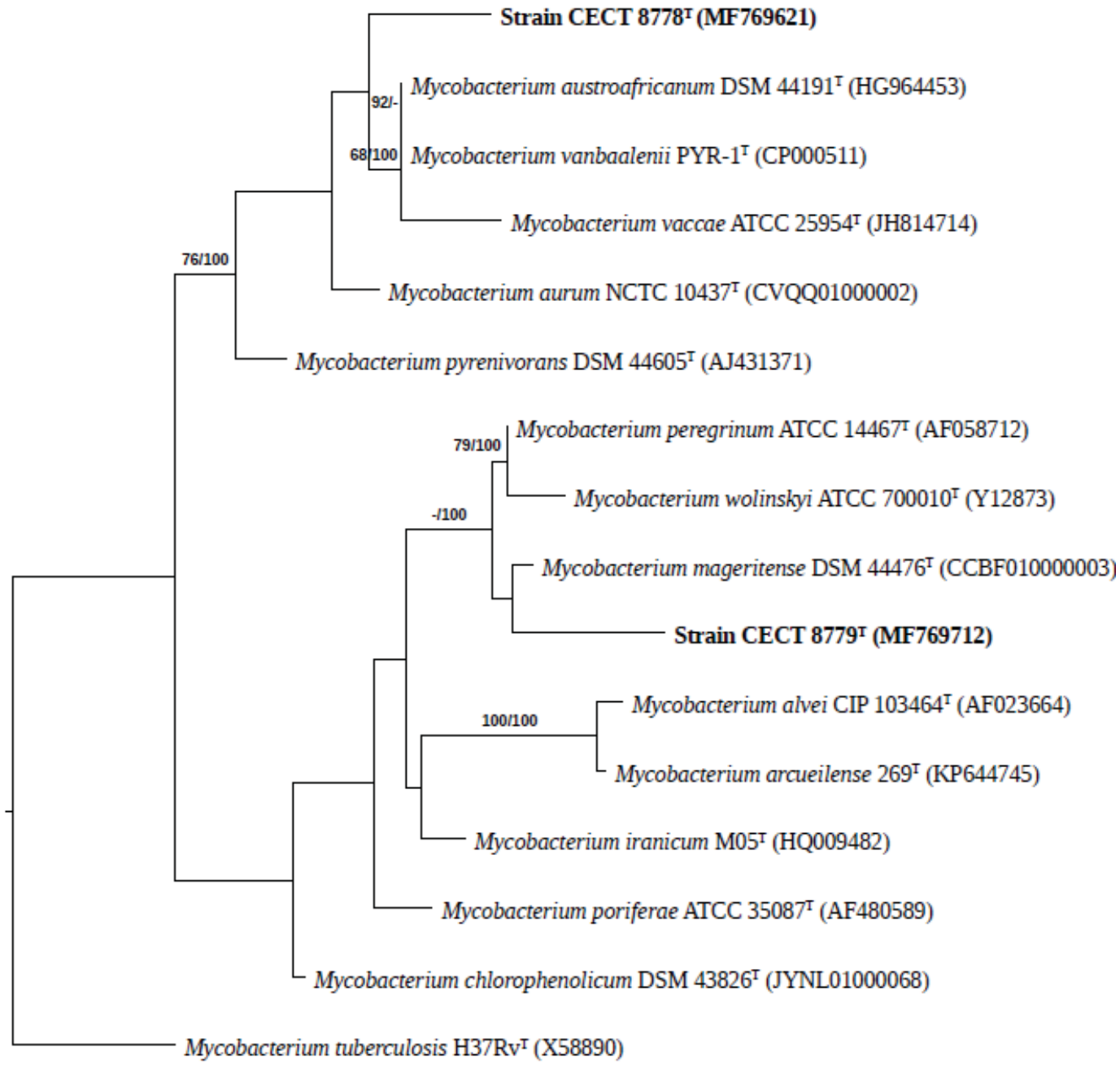
603 **Figure legends**

604 Fig. 1. Maximum-likelihood phylogenetic tree based on almost complete 16S rRNA gene
605 sequences generated using the GTR+GAMMA model and midpoint-rooting showing
606 relationships between strains CECT 8778^T and CECT 8779^T and between them and their close
607 phylogenetic neighbours. The numbers above the branches are bootstrap support values greater
608 than 60% for ML (left) and MP (right). The scale bar indicates 0.007 substitutions per site.

609 Fig. 2. Maximum-likelihood phylogenetic tree based on concatenated sequences of 16S rRNA,
610 *hsp*65, *rpo*B and *rec*A gene sequences (3257 nt) showing relationships between strains CECT
611 8778^T and CECT 8779^T and between them and their close phylogenetic neighbours. The tree
612 was inferred using the GTR+GAMMA model. The branches are scaled in terms of the expected
613 number of substitutions per site. The numbers above the branches are bootstrap support values
614 when larger than 60% from ML (left) and MP (right). The scale bar indicates 0.02 substitutions
615 per site. The accession numbers of the MLSA gene sequences are listed in Table S2.

616

617



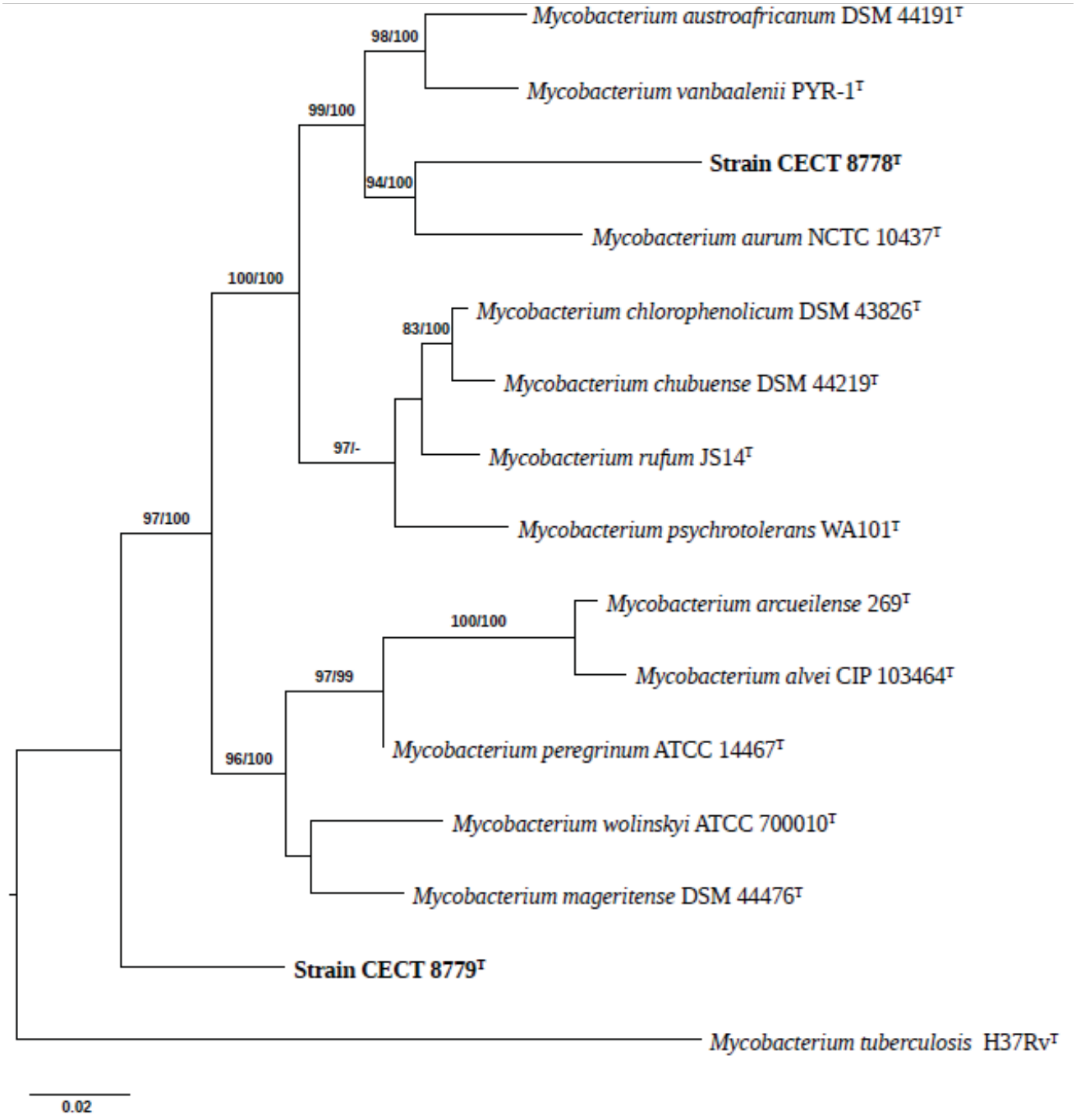
618

619

620

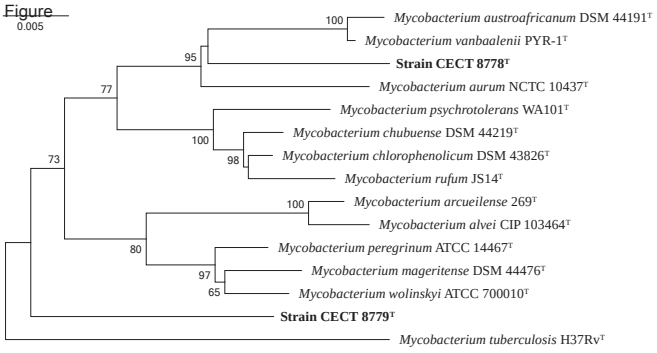
621

622


Mycobacterium austroafricanum DSM 44191T
98/100
Mycobacterium vanbaalenii PYR-1T
99/100
Strain CECT 8778T
94/100
Mycobacterium aurum NCTC 10437T
100/100
Mycobacterium chlorophenolicum DSM 43826T
83/100
Mycobacterium chubuense DSM 44219T
97/-
Mycobacterium rufum JS14T
97/100
Mycobacterium psychrotolerans WA101T
Mycobacterium arcueilense 269T
100/100
Mycobacterium alvei CIP 103464T
97/99
Mycobacterium peregrinum ATCC 14467T
96/100
Mycobacterium wolinskyi ATCC 700010T
Mycobacterium mageritense DSM 44476T
Strain CECT 8779T
Mycobacterium tuberculosis H37RvT
0.02

623

Figure



Click here to download Figure Fig. S1..pdf

Fig. S1. Neighbour-joining MLSA phylogenetic tree based on concatenated sequences of 16S rRNA, *hsp*65, *rpo*B and *rec*A genes showing relationships between strains CECT 8778^T and CECT 8779^T and between them and their close phylogenetic neighbours. The numbers above the branches are bootstrap support values greater than 60%. The scale bar indicates 0.005 substitutions per site.

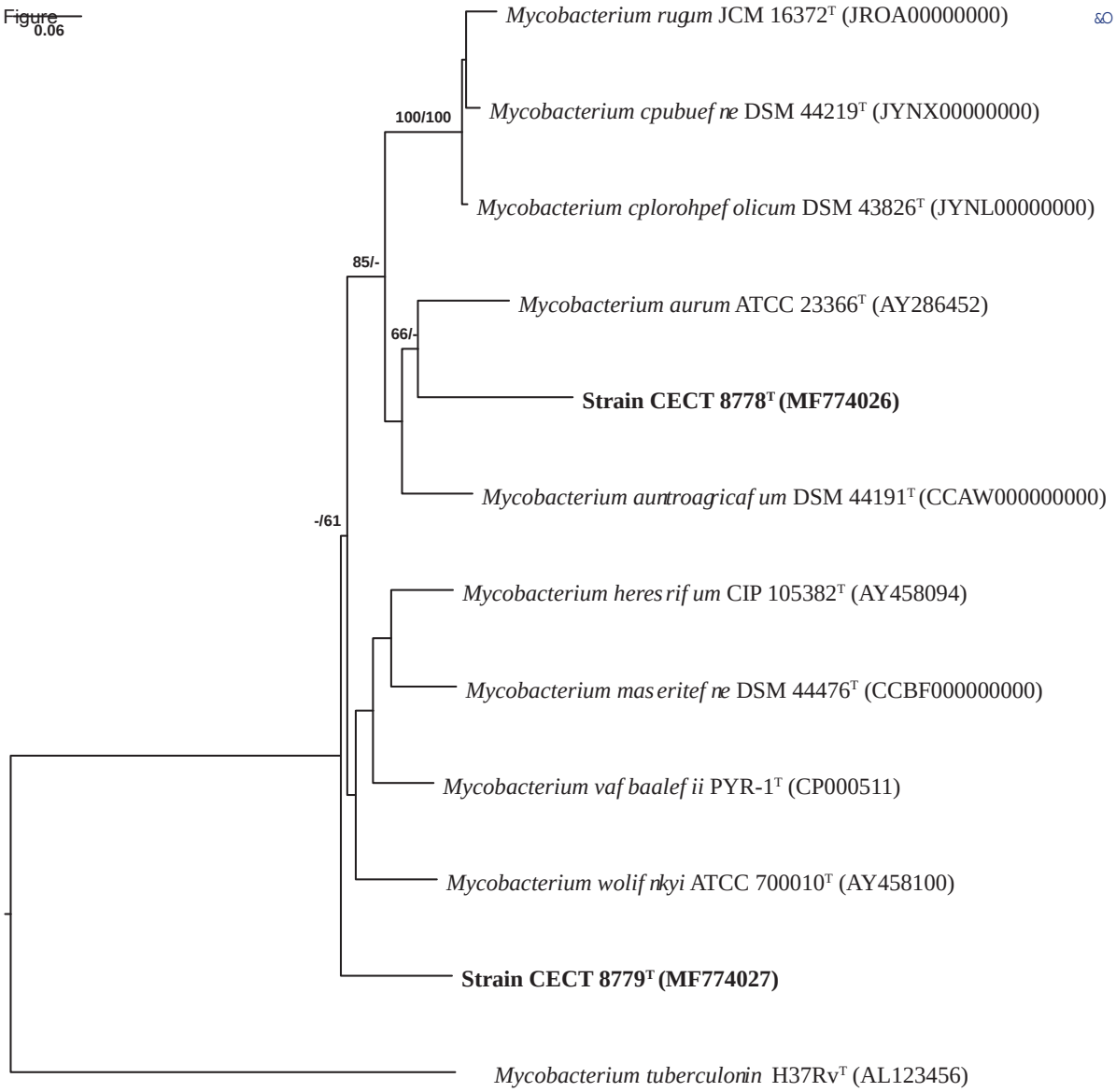


Fig. S2. Maximum-likelihood phylogenetic tree based on *rec*A partial gene sequences generated using the GTR+GAMMA model and midpoint-rooting showing relationships between strains CECT 8778^{T} and CECT 8779^{T} and between them and their close phylogenetic neighbours. The numbers above the branches are bootstrap support values greater than 60% for ML (left) and MP (right). The scale bar indicates 0.06 substitutions per site.

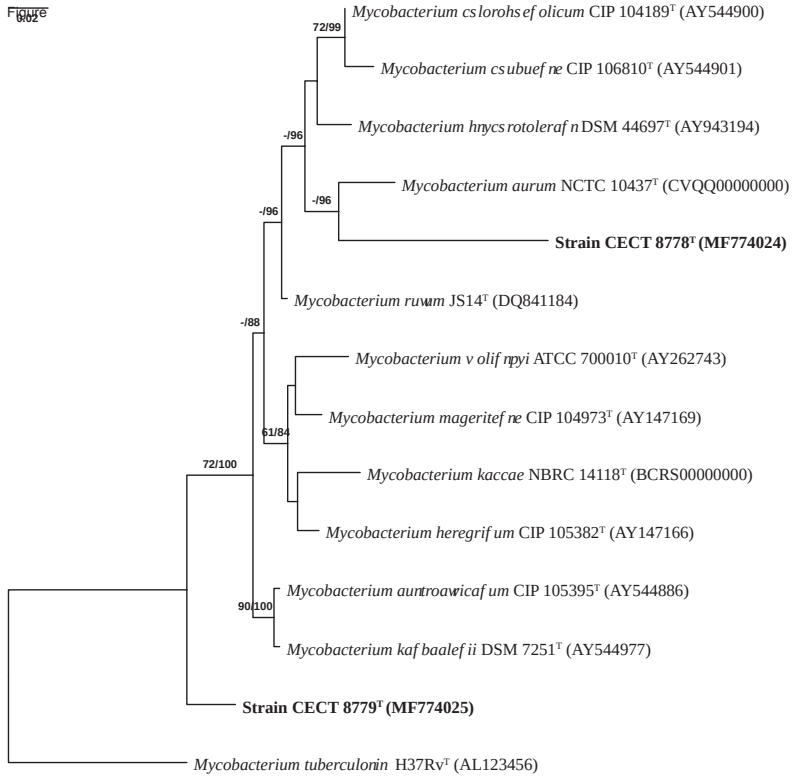


Fig. S3. Maximum-likelihood phylogenetic tree based on *rpo*B partial gene sequences generated using the GTR+GAMMA model and midpoint-rooting showing relationships between strains CECT 8778^{T} and CECT 8779^{T} and between them and their close phylogenetic neighbours. The numbers above the branches are bootstrap support values greater than 60% for ML (left) and MP (right). The scale bar indicates 0.02 substitutions per site.

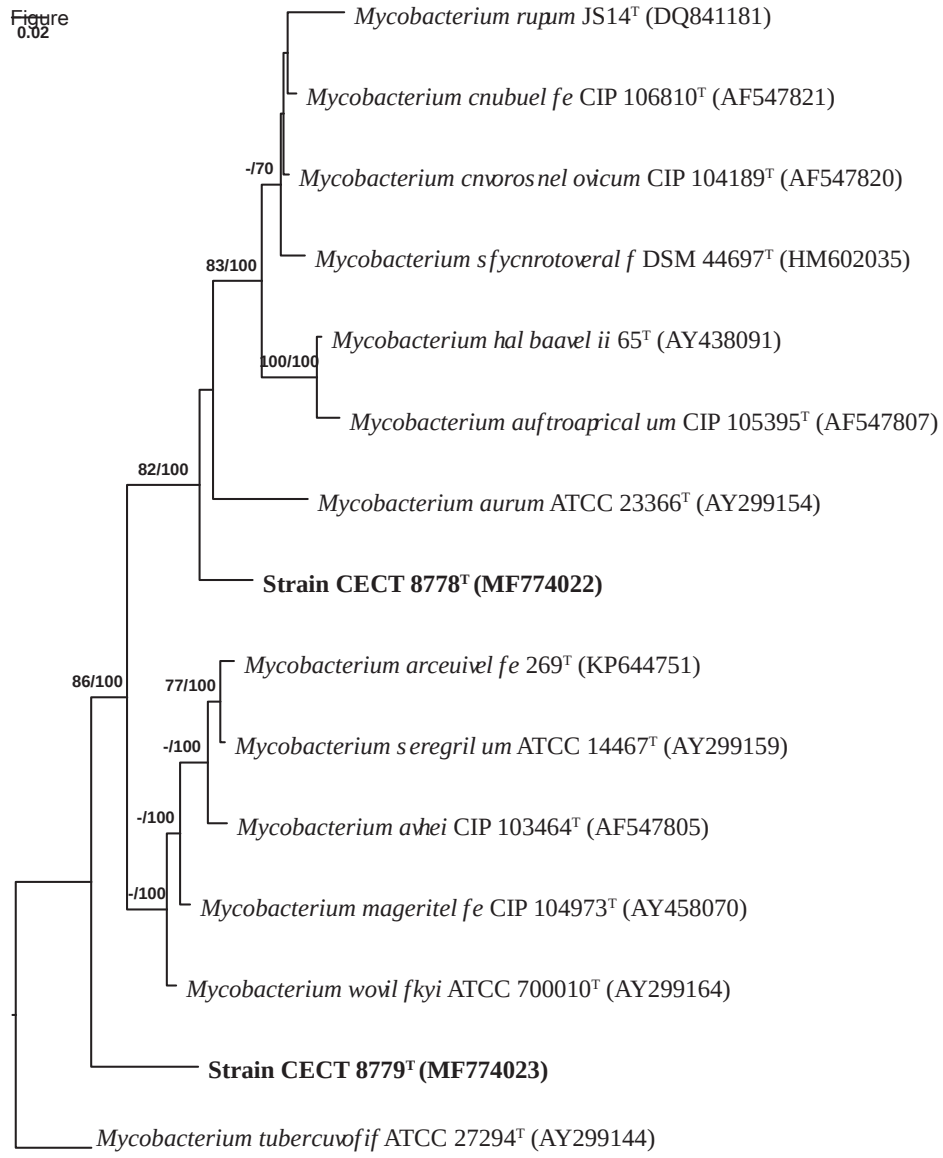


Fig. S4. Maximum-likelihood phylogenetic tree based on *hsp*65 partial gene sequences generated using the GTR+GAMMA model and midpoint-rooting showing relationships between strains CECT 8778^T and CECT 8779^T and between them and their close phylogenetic neighbours. The numbers above the branches are bootstrap support values greater than 60% for ML (left) and MP (right). The scale bar indicates 0.02 substitutions per site.