Mycolicibacterium stellerae sp. nov., a rapidly growing scotochromogenic strain isolated from Stellera chamaejasme

Imen Nouioui1*, Vartul Sangal2, Carlos Cortés1, Marlen Jando3, José Mariano Igual4, Hans-Peter Klenk1, Yu-Qin Zhang5, Michael Goodfellow1

1School of Natural and Environmental Sciences, Ridley Building 2, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom
2Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK
3Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Inhoffenstraße 7B, 38124 Braunschweig, Germany
4Instituto de Recursos Naturales y Agrobiología de Salamanca, Consejo Superior de Investigaciones Científicas (IRNASA-CSIC), c/Cordel de Merinas 40-52, 37008 Salamanca, Spain
5Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences & Peking Union Medical College, No. 1 of Tiantanxili Road, Beijing, Beijing 100050, China

Corresponding author: Imen Nouioui imen.nouioui@newcastle.ac.uk

Section: Actinobacteria

Keywords: Actinobacteria, polyphasic taxonomy, draft-whole genome sequence

Abbreviations: ANI, Average Nucleotide Identity; A2pm, diaminopimelic acid, BLAST, Basic Local Alignment Search Tool; CDP-x, CDP-alcohol phosphatidyltransferase family protein; dDDH, Digital DNA:DNA Hybridization; GGDC, Genome-to-Genome Distance Calculator; GTR, General Time-Reversible; GYM, Glucose-Yeast extract-Malt extract; LJ, Löwenstein Jensen; Mb, Megabases (millions of base pairs); MB7H10, Middlebrook 7H10; MIDI, Microbial IDentification Inc.; ML, Maximum-Likelihood; MP, Maximum-Parsimony; MUSCLE, MUltiple Sequence Comparison by Log-Expectation; PAUP, Phylogenetic Analysis Using Parsimony; PMG, Proteose peptone-Meat extract-Glycerol; RAxML, Randomized Axelerated Maximum Likelihood; MRE, Maximal-Relative-Error; TSA, Tryptic
Soy Agar; TNT, Tree analysis New Technology; PE, PhosphatidylEthanolamine; RAST, Rapid Annotation using Subsystem Technology.

The GenBank accession numbers of 16S rRNA gene and genome sequences for strain CECT 8783T are MH935827 and RARC00000000, respectively.
Abstract
A polyphasic study was undertaken to establish the taxonomic provenance of a rapidly growing Mycolicibacterium strain CECT 8783T, recovered from the plant Stellera chamaejasme L. in Yunnan Province, China. Phylogenetic analyses based upon 16S rRNA and whole-genome sequences showed that the strain formed a distinct branch within the evolutionary radiation of the genus Mycolicibacterium. The strain was most closely related to Mycolicibacterium moriokaense DSM 44221T with 98.4% 16S rRNA gene sequence similarity, but was distinguished readily from this taxon by a combination of chemotaxonomic and phenotypic features and by low average nucleotide identity and digital DNA:DNA hybridization values of 79.5% and 21.1%, respectively. Consequently, the strain is considered, to represent a novel species of Mycolicibacterium for which the name Mycolicibacterium stellerae sp. nov is proposed; the type strain is I10A-01893T (= CECT 8783T = KCTC 19843T = DSM 45590T).

Mycobacterium [1], the type genus of the family Mycobacteriaceae [2], can be distinguished from other genera classified in the order Corynebacteriales by using a combination of genotypic and phenotypic criteria [3]. The genus includes pathogenic and non-tuberculous mycobacteria that are common in the environment and can cause opportunistic infections in immunocompetent and immunosuppressed patients [4]. Mycobacteria can be divided into two groups based on their growth rates on solid media; slowly growing strains need seven or more days of incubation at optimal temperatures to form visible colonies from highly diluted inocula whereas colonies of rapidly growing strains are seen within seven days under comparable conditions [5]. Species can be assigned to these two groups using polyphasic procedures, as exemplified by the circumscription of strains previously lumped together within the Mycobacterium abscessus and Mycobacterium avium complexes [6, 7]. However, genomic based methods provide much better resolution between mycobacterial species compared to 16S rRNA generated phylogenies and also provide insights into their evolution [8-10]. The availability of whole genome sequences and associated bioinformatic tools has led to the unambiguous delineation of new mycobacterial species [11] and the recognition of heterotypic synonyms of Mycobacterium tuberculosis [10].

The genus Mycobacterium encompasses nearly 200 validly named species [12] representatives of which formed a monophyletic taxon in a genome-based classification of the phylum Actinobacteria [11]. Most of these species have been classified into five distinct monophyletic groups based on extensive phylogenomic and comparative genome analyses [13]. Gupta and his colleagues [13] recognised an emended genus Mycobacterium, which included all of the
major known pathogens as “Tuberculosis-Simiae” clade, and the novel genera *Mycolicibacillus*, *Mycolicibacterium*, *Mycobacteroides* which contain species assigned to clades designated as “Terrae”, “Triviale”, “Fortuitum-Vacciae”, and “Abscessus-Chelonae”, respectively [13]. The genus *Mycolicibacterium* encompasses 88 rapidly growing species, including *Mycolicibacterium fortuitum* the type species; members of this taxon have genomes that range in size from 3.95 – 8.0 Mbp and in G+C content 65.4 - 70.3 mol% [13].

In a continuation of our studies on mycobacterial diversity, strain CECT 8783$^T$ recovered from *Stellera chamaejasme* L. in Yunnan Province, China was the subject of a polyphasic study. The resultant datasets show that the strain represents a novel *Mycolicibacterium* species for which the name *Mycolicibacterium stellerae* sp. nov. is proposed.

A culture of CECT 8783$^T$ representing strain I10A-01893$^T$ was obtained from the Spanish Type Culture Collection (CECT), but was originally deposited in the German Collection of Microorganisms and Cell Cultures (DSMZ) and given the accession number DSM 45590$^T$. The strain was maintained together with its closest phylogenetic neighbour, *Mycolicibacterium moriokaense* CIP105393$^T$ [13, 14], on proteose peptone-meat extract-glycerol agar (PMG; DSMZ 250 medium) and as suspensions of cells in 35%, v/v glycerol at -80°C.

Cultural and morphological features of CECT 8783$^T$ were carried out under light and dark conditions on glucose-yeast extract-malt extract agar (GYM; DSMZ medium 65), Löwenstein-Jensen medium (LJ; [15]), Middlebrook 7H10 agar (MB7H10; [16]), nutrient agar (NA; [17]) and peptone-meat extract-glucose agar (PMG; DSMZ medium 250) and tryptic soy agar (TSA; [18]) after incubation for 14 days at 37°C. In addition the strain was examined for its ability to grow at 4°C, 10°C, 15°C, 25°C, 28°C, 37°C and 45°C as well as for its ability to grow anaerobically on PMG agar, using an anaerobic bag system (Sigma-Aldrich 68061). Strain CECT 8783$^T$ was examined for acid-alcohol-fastness [19]. The strain was acid-alcohol-fast, grew optimally on GYM, NA, PMG and TSA agar after 5 days at 37°C, but not under anaerobic conditions. Yellow pigmented colonies were produced on all of these media under both light and dark conditions.

Genomic DNA was extracted from a culture of CECT 8783$^T$ after Amaro et al. [20]. The genome of the strain was sequenced using an Illumina MiSeq instrument as described by Sangal et al. [21] and assembled into contigs using SPAdes 3.9.0 with a kmer length of 127 [22]. A complete 16S rRNA gene sequence was extracted from the draft genome (accession number MH935827). Corresponding 16S rRNA gene sequences of the type strains of closely related *Mycolicibacterium* species were retrieved from the EzBioCloud server [23]. Maximum-
likelihood (ML) and maximum-parsimony (MP) phylogenetic trees derived from the 16S rRNA were inferred using the genome-to-genome distance calculator (GGDC) web server [24] adapted to single gene inferences; the latter was also used to calculate pairwise sequence similarities [25, 26]. Multiple sequence alignments were generated using MUSCLE software [27] and a ML tree inferred from the alignment with RAxML [28] using rapid bootstrapping and the auto MRE criterion [29]. In turn, an MP tree was inferred from alignments with the Tree analysis New Technology (TNT) program [30] using 1000 bootstraps together with tree bisection and reconnection branch swapping and ten random sequence replicates. The X2 test implemented in PAUP* [31] was used to check for compositional bias in the sequences.

Blast analysis of the complete 16S rRNA gene sequence of strain CECT 8783T (1533 nucleotides) showed that it was most closely related to M. moriokaense CIP105393T and Mycolicibacterium goodii ATCC 700504T [13, 32], showing 16S rRNA gene sequence similarities with them of 98.4% and 98.6%, respectively; the corresponding pairwise 16S rRNA gene sequence similarities based on the GGDC platform were 98.3% and 98.4%. The strain formed a distinct branch in the 16S rRNA gene tree (Fig. 1), between Mycolicibacterium madagascarense ATCC 49865T branch [13, 33] and a poorly supported subclade that contained Mycolicibacterium celeriflavum AFPC-000207T [13, 34], M. moriokaense CIP105393T, Mycolicibacterium phlei DSM 43239T [13, 35], Mycolicibacterium pulveris DSM 44222T [13, 36]; strain CECT 8783T shared 16S rRNA gene sequence similarities with these organisms within the range 97.8%-98.4%. An adjacent well supported clade contained M. goodii ATCC 700504T and Mycolicibacterium smegmatis NCTC 8159T [13, 35, 37].

Due to poor bootstrap support for several nodes in the 16S rRNA gene tree, ML trees were constructed from core proteins and 400 universal protein sequences using PhyloPhlAn [38]. In brief, genome sequences were annotated using Prokka 1.11 [39] and the protein sequences were compared using BPGA 1.3 pipeline [40] to calculate the core genome. Poorly aligned regions and sites with the missing data was removed from the sequence alignment of the core proteins using Gblocks [41]. A ML phylogenetic tree was generated from the resulting alignment of 23,359 amino acids using IQ-Tree [42] with LG+F+I+G4 substitution model and 100,000 ultrafast bootstrap iterations and SH-like approximate likelihood ratio tests. Another ML tree was generated from the protein sequences using PhyloPhlAn which extracts subset of amino acids from 400 universal sequences and calculates ML phylogeny [38].
It can be seen from Figure 2 that strain CECT 8783T formed a well supported branch in the phylogenomic tree together with, *M. celeriflavum* DSM 46765T, *Mycolicibacterium flavescens* M6T [13, 43], *M. moriokaense* CIP 105393T and *M. phlei* CCUG21000T that was clearly separated from a corresponding clade contained *M. goodii* X7BT and *M. smegmatis* NCTC8159T (Fig 2a). The same topology was observed on the concatenated 400 universal amino acid sequences based tree (Fig. 2b).

The genome size of the strain was found to be ~6.9 Mb with 37.7X coverage, 88 contigs, 52 RNAs and 6836 coding sequences and an *in silico* G+C content of 65.8 mol%. Similarly, the type strain of *M. moriokaense* has a genome size of 6.2 Mb, 51 RNAs, 6114 coding sequences and an *in silico* G+C content of 66.0 mol%. The average nucleotide identity (ANI), estimated using blastANI algorithm [44], between the genomes of the two strains was 79.5%, a value well below the threshold of 95-96% used to delineate prokaryotic species [44-46]. Similarly, the digital DNA:DNA hybridization (dDDH) value between the two strains, calculated using GGDC based on formula 2 of Meier-Kolthoff et al. [25] (http://ggdc.dsmz.de/), was 21.1% [18.9 - 23.5%], this is well below the 70% cut-off recommended for assigning strains to the same species [47]. Using Artemis software [48], the molecular signatures specific to the genus *Mycolicibacterium* (LacI family transcriptional regulator (WP_036341761); CDP-x (WP_036344961); and CDP-diacylglycerol–serine (WP_066811333)), identified by Gupta et al. [13], were found in the genome of the strain CECT 8783T.

Biomass for most of the chemotaxonomic analyses on strain CECT 8783T was prepared in shake flasks (200 resolutions per minute) of PMG broth following incubation at 37°C for 5 days. Cells were washed three time in sodium chloride solution (0.9%, w/v), freeze dried and stored at room temperature. Standard chromatographic procedures were used to detect isomers of diaminopimelic acid (A2pm) [49]; mycolic acids [50], polar lipids [51] and cell wall sugars [52]. Strain CECT 8783T and *M. moriokaense* DSM 44221T produced whole organism hydrolysates rich in *meso*-A2pm, arabinose, galactose, glucose, mannose and ribose and had polar lipid profiles containing diphosphatidylglycerol (DPG), glycophospholipid (GPL), phosphatidylethanolamine (PE), phosphatidylinositol (PI), unidentified glycolipids (GL), unidentified lipids and phospholipid. In addition strain CECT 8783T contained an aminoglycolipid and *M. moriokaense* DSM 44221T an aminolipid (Fig. S1).

Wet biomass for the fatty acid was prepared in Middelbrook 7H10 broth following incubation for 5 days at 37°C. Cellular fatty acids were extracted and fatty acids methyl esters (FAMES) prepared after saponification and methylation using the procedure introduced by Miller.
The FAMES were analysed by gas chromatography (Agilent 6890N) instrument and identified based on myco6 database [54] and the standard Microbial Identification (MIDI) system, version 4.5.

Quantitative differences in the mixtures of saturated, unsaturated and 10-methyloctadecanoic (tuberclosteaeric) fatty acids were found between the two strains (Table S1). The major fatty acid of strain CECT 8783T (>25%) were C<sub>18:1</sub> ε9c and C<sub>16:0</sub> and that of M. moriokaense DSM 44221T was C<sub>16:0</sub> though the latter unlike the former, contained C<sub>17:1</sub> ε7c (Table S1). The mycolic acid profile of the strain CECT 8783T consisted of α, α'-mycolates, keto-mycolates, ω-carboxymycolate while M. moriokanese DSM 44221T lacked α'-mycolates (Table S1).

The two strains were examined, in duplicate, for a broad range of standard biochemical tests, namely: arylsulfatase after 3 and 20 days [55], reduction of potassium tellurite [56, 57], degradation of Tween 80 [58] and urea hydrolysis [59]. In addition, the strains were examined for their ability to metabolise a broad range of sole carbon and nitrogen sources, to grow on the presence of several concentrations of sodium chloride, at a range of pH values and in the presence of inhibitory compounds, using GENIII microplates and an Omnilog device (Biolog Inc., Hayward, USA). These tests were carried out in duplicate using freshly prepared inocula harvested from the mid-logarithmic growth phase of PMG agar plates incubated at 37°C for 7 days, as described by Nouioui et al. [60]. Opm package version 1.3.36 [61-62] was used to analyse the resultant data. Identical results were obtained for all of the duplicated tests.

Phenotypic features summarised in Table 1 clearly distinguish strain CECT 8783T from M. moriokaense DSM 44221T. Both strains found to be able to reduce potassium tellurite, but were unable to degrade Tween 80 or hydrolyse urea. However, only strain DSM 44221T produced arylsulfatase after 3 and 14 days.

It can be concluded from the chemotaxonomic, genomic and phenotypic data that strain CECT 8783T clearly forms a new centre of taxonomic variation within the genus Mycolicibacterium. Consequently, it is proposed that the strain be recognised to represent a new species, namely as Mycolicibacterium stellaiae sp. nov.

**Description of Mycolicibacterium stellaiae sp. nov.**

Mycolicibacterium stellaiae (stel’le.rae N.L. gen. n. stellaiae of Stellera, named referring to the host plant, Stellera chamaejasme, from which the strain was isolated).
Strictly aerobic Gram-stain positive, acid-alcohol fast, fast growing organism which produces orange coloured colonies on Middelbrook 7H10, nutrient agar, proteose peptone-meat extract-glycerol, glucose-yeast extract-malt extract and tryptic soy agar plates within 5 days at 37°C under dark and light conditions. Grows between 25°C and 37°C, optimally ~ 37°C, and at pH7. Arylsulfatase negative after 3 and 14 days at 37°C. Strain CECT 8783T reduced potassium tellurite, but was unable to degrade Tween 80 or hydrolyse urea. It was able to utilise L-arginine, L-aspartic acid and D-serine (amino acids); L-glutamic acid, L-lactic acid, D-malic acid, α-hydroxy-butyric acid, p-hydroxy-phenylacetic acid and L-pyroglutamic acid (organic acids); D-cellobiose, dextrin, glycerol, D-maltose, methyl pyruvate, sucrose, D-trehalose and turanose (sugars) and degraded pectin. Whole cell hydrolysates contained meso-diaminopimelic acid and arabinose, galactose, glucose, mannose and ribose; the polar lipid consists of diphosphatidylglycerol, a glycopospholipid, phosphatidylethanolamine, phosphatidylinositol, an aminoglycolipid, a phospholipid, three unidentified glycolipids and two unidentified lipids. The major fatty acids are C18:1ω9c and C16:0 with α, α′-mycolates, keto-mycolates, ω-carboxymycolate as mycolic acids. The genome size is 6.9 Mb with an in silico DNA G+C content of 65.8%.

The type strain I10A-01893T (= CECT 8783T = KCTC 19843T =DSM 45590T) was isolated from Stellera chamaejasme in Yunnan Province, China.

Acknowledgements

This project was supported by the School of Natural and Environmental Sciences at Newcastle University. IN is grateful to Newcastle University for a postdoctoral fellowship.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References:


36. Trevisan V. Generi e le Specie delle Batteriacee, Zanaboni and Gabuzi, Milano 1889.


<table>
<thead>
<tr>
<th></th>
<th>CECT 8783&lt;sup&gt;T&lt;/sup&gt;</th>
<th>M. moriokaense DSM 44221&lt;sup&gt;T&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utilisation of amino acids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Arginine, D-serine #2</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>L-histidine, L-serine</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Utilisation of organic acids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-Gluconic acid, glucuronamide</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>L-aspartic acid, L-glutamic acid, L-lactic acid, D-malic acid, α-hydroxy-butyric acid, p-hydroxy-phenylacetic acid, L-pyroglutamic acid</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Utilisation of sugars</strong> :</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-Cellobiose, dextrin, glycerol, D-maltose, methyl pyruvate, sucrose, D-trehalose turanose</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>D-arabitol, fructose, D-mannitol, D-sorbitol</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Resistance to</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fusidic acid, lincomycin, minocycline, rifamycin sv, 1% sodium lactate, tetrazolium blue, tetrazolium violet, vancomycin</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Degradation tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tween 40</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Pectin</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

+ Positive reaction; - negative reaction.
All of the strains metabolised acetic acid, acetoacetic acid, bromo-succinic acid, butyric acid, β-hydroxy-butryic acid, L-malic acid and propionic acid (organic acids), glycine-proline (amino acid); gelatin; grew in presence of aztreonam and nalidixic acid (antibiotics) and lithium acid, potassium tellurite, sodium bromate and sodium formate (salts). In contrast, none of the strains oxidised alanine, D-aspartic acid, γ-amino-n butyric acid, n-acetyl-D-galactosamine, n-acetyl-D-glucosamine, n-acetyl-β-d-mannosamine, n-acetyl-β-β-uraminic acid, citric acid, D-fructose-6-phosphate, D-fucose, L-galactonic acid-γ-lactone, D-galacturonic acid, D-glucose-6-phosphatae D-glucuronic acid, guanidine hydrochloride, α-keto-butyric acid, α-keto-glutaric acid, mucic acid, D-galactose, β-gentiobiose, inosine, D-lactic acid, α-D-lactose, D-lactic acid methyl ester, mannose, 3-o-methyl-D-glucose, β-methyl-D-glucoside, myo-inositol, melibiose, niaproof, quinic acid, D- raffinose, L-rhamnose, D-saccharic acid, D-salicin, D-serine, stachyose and troleandomycin.

Figures

**Fig. 1.** Maximum-likelihood phylogenetic tree based on almost complete 16S rRNA gene sequences inferred using the GTR+GAMMA model showing relationships between strain CECT 8783T and its closest phylogenetic neighbours. The numbers above the branches are bootstrap support values greater than 60% for ML (left) and MP (right). The scale bar is 0.007.
Fig. 2. Maximum-likelihood phylogenomic trees based on core genome sequences (a) and on concatenated amino acid sequences from 400 universal proteins (b) showing relationships between strain CECT 8783$^T$ and its nearest neighbours.