Genomic molecular signatures determined characterization of *Mycolicibacterium gossypii* sp. nov., a fast-growing mycobacterial species isolated from cotton field soil

Rui-Rui Huang · Shen-Rong Yang · Cheng Zhen · Xian-Feng Ge · Xin-Kai Chen · Zhi-Qiang Wen · Ya-Nan Li · Wen-Zheng Liu

Received: 30 March 2021 / Accepted: 5 August 2021 / Published online: 15 August 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

**Abstract** A Gram-positive, acid-fast and rapidly growing rod, designated S2-37 T, that could form yellowish colonies was isolated from one soil sample collected from cotton cropping field located in the Xinjiang region of China. Genomic analyses indicated that strain S2-37 T harbored T7SS secretion system and was very likely able to produce mycolic acid, which were typical features of pathogenetic mycobacterial species. 16S rRNA-directed phylogenetic analysis referred that strain S2-37 T was closely related to bacterial species belonging to the genus *Mycolicibacterium*, which was further confirmed by pan-genome phylogenetic analysis. Digital DNA-DNA hybridization and the average nucleotide identity presented that strain S2-37 T displayed the highest values of 39.1% (35.7–42.6%) and 81.28% with *M. litorale* CGMCC 4.5724 T, respectively. And characterization of conserved molecular signatures further supported the taxonomic position of strain S2-37 T belonging to the genus *Mycolicibacterium*. The main fatty acids were identified as C16:0, C18:0, C20:3ω3 and C22:6ω3. In addition, polar lipids profile was mainly composed of diphosphatidylglycerol, phosphatidylethanolamine and phosphatidylinositol. Phylogenetic analyses, distinct fatty aids and antimicrobial resistance profiles indicated that strain S2-37 T represented genetically and phenotypically distinct from its closest phylogenetic neighbour, *M. litorale* CGMCC 4.5724 T. Here, we propose a novel species of the genus *Mycolicibacterium*: *Mycolicibacterium gossypii* sp. nov. with the type strain S2-37 T (= JCM 34327 T = CGMCC 1.18817 T).

**Keywords** *Mycolicibacterium* · Fast-growing · Cotton field soil · Molecular signatures

**Introduction**

The genus *Mycolicibacterium* presents the second largest population of Gram-positive, acid-fast and rod-shaped microbes in the family *Mycobacteriaceae* (Parte et al. 2020). At the time of writing, the genus *Mycolicibacterium* is composed of 92 recognized species with published names (https://lpsn.dsmz.de) (Parte et al. 2020), which primarily harbors rapid-growing species isolated from diverse range of environments, including river water, marine sediment and soil (Butler et al. 1993; Brown-Elliott et al. 2010;...
There are also relatively high proportion of members in the genus isolated from clinical specimens, indicating potential pathogenicity to humans and animals (Shojaei et al. 2000; Brown-Elliott and Wallace 2002).

It was well documented that mycobacterial species displayed relatively high diverses in their genomic features at species level (Fedrizzi et al. 2017), which enhanced the challenge to characterize the taxonomic position of microbial species belonging to this population. Much effort has been devoted to primarily delineate different evolved branches of this population using well established approaches based on analyses of the 16S rRNA gene, 16S-23S spacer, and house-keeping gene concatenated multilocus sequences (Roth et al. 1998; Mignard and Flandrois, 2008; Magee and Ward, 2012). However, the reliability of these methods used to well distinguish sub-groups (e.g., slow- and fast-growing species) of mycobacterial species remains of concern. Recently, Gupta et al. (2018) have developed a robust method consistently supporting the existence of five distinct monophyletic sub-groups of mycobacterial species, which are designated as the *Mycobacterium*, *Mycolicibacterium*, *Mycolicibacter*, *Mycolicibacillus* and *Mycobacteroides* genera. They have identified representative molecular markers in the form of conserved signature indels and proteins, which are uniquely shared by members of the five identified clades.

In this study, a putative novel species (strain S2-37$^\mathrm{T}$) belonging to mycobacterial species was isolated from a soil sample of cotton cropping field of Xinjiang in PR China. Phylogenetic analyses based on 16S rRNA gene and genomic sequences coherently agreed the closer relationship of strain S2-37$^\mathrm{T}$ with microbial species from the genus *Mycolicibacterium*. Finally, genome sequence of strain S2-37$^\mathrm{T}$ was mapped to previously documented molecular features specific for each genus (Gupta et al. 2018), ending up with the conclusion that the genome sequence of strain S2-37$^\mathrm{T}$ supported the classification into the genus *Mycolicibacterium* along with the polyphasic approaches.

### Materials and methods

**Isolation of the novel strain and cultivation**

Strain S2-37$^\mathrm{T}$ was isolated from a soil sample collected from a cotton cropping field (86°20’ N 44°62’ E, Xinjiang, China). The isolation procedures were described in the previous report with some modifications (Hopkins et al. 1991). Soil samples (5 g) were firstly naturally dried, then mixed with 100 ml of sterilized water in 250 ml flasks. Samples were placed on a shaker with vibration frequency of 160 r/min at 30 °C for 30 min, followed by homogenization treatment for 1 min using a sonicator (XO-3200DT, Nanjing Xianou laboratory equipment Co., Ltd) with a frequency of 40 kHz. Then, 100 μl serially diluted samples were plated on commercially available Gause’s synthetic No.1 agar (G1, 20 g soluble starch; 1 g KNO$_3$; 0.5 g K$_2$HPO$_4$; 0.5 g MgSO$_4$.7H$_2$O; 0.5 g NaCl; 0.01 g FeSO$_4$; 20 g Agar; 1000 ml distilled water; pH 7.2–7.4) for cultivation at 30 °C in the following 14 days. Bacterial colonies were sub-cultured on G1 agar for three times to achieve pure isolates. Strain S2-37$^\mathrm{T}$ was maintained aerobically on G1 agar at 30 °C, and stored at -80 °C in G1 broth supplemented with equal volume of 50% (v/v) sterilized glycerol for preservation (Prakash et al. 2013).

**Phylogenetic analysis based on 16S rRNA gene and genomic sequences**

Genomic DNA of strain S2-37$^\mathrm{T}$ was extracted and its 16S rRNA gene was achieved by PCR amplification using universal primers 27F and 1492R as described by Fan et al. (2008), followed by sequence alignment in order to preliminarily determine the taxonomic position of the strain S2-37$^\mathrm{T}$ using the BLAST function embedded in NCBI database (Federhen 2012). GGDC web server (available at http://ggdc.dsmz.de/) was employed for gene phylogenetic and similarity analysis between 16S rRNA gene sequences of S2-37$^\mathrm{T}$ and all published *Mycolicibacterium* type strains collected from the LPSN database (Parte et al. 2020). The phylogenetic trees were reconstructed based on 16S rRNA gene sequences of strains S2-37$^\mathrm{T}$ and its 22 closely phylogenetic relatives using the neighbor-joining (NJ) (Saitou and Nei 1987), maximum-parsimony (MP) (Sourdis and Nei 1988) and maximum-likelihood (ML) (Steel and Rodrigo 2008).
The phylogenetic trees were evaluated by performing a bootstrap analysis based on 1000 replications (Felsenstein 1985). Genomic DNA was extracted from strain S2-37 T using Bacterial Genomic DNA Rapid Extraction kit (Cat. No. B518225) supported by Sangon Biotech (Shanghai, China). Library construction, quality control and analysis were performed following the methods described by Zhu et al. (2001). The draft genome of S2-37 T was sequenced using Illumina HiSeq 4000 PE150 (Patnaik et al. 2016) at Beijing TSINGKE Bioinformatics Technology Co., Ltd, and assembled using SOAPdenovo (Luo et al. 2012), SPAdes (Bankevich et al. 2012) and Abyss software (Simpson et al. 2009), respectively. The final genome assembly was achieved by integrating three assemblies in CISA software (Lin and Liao 2013). The whole genomic sequences of strain S2-37 T and its 22 closely phylogenetic relatives were used to perform pan-genome analysis with a bacterial pan genome analysis pipeline (BPGA) (Chaudhari et al. 2016), in order to derive the exact phylogenetic affiliation. ONE CLICK MODE was performed in BPGA program and all the analyses were performed in a single step using default parameters (sequence identity cut-off = 50% and No. of iterations for pan-genome profile calculation = 20).

Genomic analysis

Functional categories of the strain S2-37 T genome were predicted using the Cluster of Orthologous Group of Proteins (COG) (Tatusov et al. 2003). AntiSMASH was employed to predict biosynthetic gene clusters of strain S2-37 T (Blin et al. 2019). Functional gene annotation and metabolic pathway prediction were performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2016). The virulence factor database (VFDB) was employed to predict virulence factors of strain S2-37 T (Liu et al. 2019). Digital DNA-DNA hybridization (dDDH) analysis was performed using an automated high-throughput platform for genome-based taxonomy called Type (Strain) Genome Server (TYGS; http://tygs.dsmz.de/) (Meier-Kolthoff et al. 2013), with the genome of strain S2-37 T as the only query sequence. Confidence intervals were calculated using the recommended settings of the GGDC 2.1 (Meier-Kolthoff et al. 2013). Additionally, average nucleotide identity (ANI) was calculated with the OrthoANIu algorithm using the server-based software (available at http://www.ezbiocloud.net/tools/ani) (Yoon et al. 2017).

Divergence analysis of conserved molecular signatures

Conserved molecular signatures, including indels (CSIs) and unique proteins (CSPs), specific for the genus Mycolicibacterium and Mycobacterium were directly collected from the work described by Gupta et al. (2018). Conserved protein sequences of different species with CSIs were achieved by BLASTp searches against the NCBI non-redundant (nr) database (Altschul et al. 2005), and the identified orthologous proteins with a minimum of 50% in sequence identity were used for divergence analyses. Phylogenetic analyses were performed using the software package MEGA 7.0 after multiple alignments of the sequence data with CLUSTAL W (Kumar et al. 2016). Phylogenetic trees were reconstructed using the neighbor-joining (NJ) algorithms (Saitou and Nei 1987). Partial sequence alignments of a conserved region of orthologous proteins from different species were displayed in same patterns as the previous work did (Gupta et al., 2018). Phylogenetically closely related species to strain S2-37 T (Mycolicibacterium litorale CGMCC 4.5724 T, Mycolicibacterium monacense DSM 44395 T, Mycobacterium neglectum CECT 8778 T and Mycobacterium lehmannii SN 1900 T) and randomly chosen species (Mycolicibacterium astroafricanum DSM 44191 T, Mycolicibacterium smegmatis NCTC 8159 T, Mycobacterium asiaticum DSM 44297 T, Mycobacterium avium 104 T and Mycobacterium bohemicum DSM 44277 T) were employed for the divergence analyses. The CSPs profile of strain S2-37 T was identified using BLASTp searches against the NCBI non-redundant (nr) database, in order to evaluate its phylogenetic relationship with genus Mycolicibacterium and Mycobacterium. Full names will always be used throughout the context for species belonging to the genus Mycobacterium, in order to avoid confusion when referring to species from the genus Mycolicibacterium.
Morphology, physiology, and biochemical analysis

Gram staining was performed using the methods described by Smibert and Krieg (1994). Acid fastness was conducted using methods described by Berd (1973). Cell morphology was determined when grown on Bennett’s agar (10 g glucose; 2 g N-Z amine; 1 g beef extract; 1 g yeast extract; 15 g agar; 1000 ml distilled water; pH 7.1–7.5) at 30 °C for 7 days using a light microscope (OLYMPUS BX43F; Olympus Corporation, Tokyo, Japan) and a cold field emission scanning electron microscope (SEM; Hitachi SU8010, Tokyo, Japan). SEM analysis was performed as described by Koon et al. (2019).

Growth with various NaCl concentrations (0–3% at 0.5% intervals and 3–7% at 1% intervals, w/v) and at different temperatures (10, 18, 25, 30, 37 and 40, 45 °C), was examined by growing the strains on Bennett’s medium as the basal medium. Growth at different pH values (4.0–10.0, at intervals of 1.0 pH unit) was examined on Bennett’s medium using the buffer system described by Xu et al. (2005). Catalase activity was determined using 3% H2O2, and gas production was identified as a positive reaction. Urease activity, nitrate reduction and tween 80 hydrolysis were determined using methods described by Kent and Kubica (1985). Resistance to antibiotics was determined using impregnated filter-paper discs (Goodfellow and Orchard 1974) containing cefalotin, cefoxitin, amikacin, ciprofloxacin, clarithromycin, doxycycline, tobramycin and sulfamethylisoxazole. Other biochemical properties of strain S2-37T were further tested using the API 20 NE and API ZYM systems (bioMérieux) according to the manufacturer’s instructions, and results were summarized in the species description.

Chemotaxonomic characterization

The polar lipids profile of the strain S2-37T was determined using standard thin-layer chromatographic procedures (Minnikin et al. 1984). In addition, cellular fatty acids were extracted from freeze dried biomass of the strain and were saponified and methylated to produce fatty acid methyl esters (FAMES) following the procedure described by Kuykendall et al. (1988). The FAMES were analyzed by gas chromatography (Agilent 6890 instrument) and the resultant peaks were automatically integrated (Lisec et al. 2006). The fatty acids profile was determined using the standard Microbial Identification (MIDI) System, version 4.5, and the Myco 6 database (Sasser 2001).

Results and discussion

Phylogenetic and genomic analyses

Phylogenetic analyses displayed that strain S2-37T showed the highest 16S rRNA gene sequence similarity with Mycolicibacterium pyrenivorans DSM 44605T (98.5%), Mycobacterium neglectum CECT 8778T (98.5%), followed by Mycolicibacterium austroafricanum DSM 44191T (98.3%) (Table S1). NJ algorithm-directed phylogenetic analysis showed that strain S2-37T was located in a clade adjacent to M. pyrenivorans DSM 44605T, Mycolicibacterium aurum NCTC 10437T and M. austroafricanum DSM 44191T, with a low bootstrap support (< 50%) (Fig. 1). The phylogenetic distribution pattern could also be reproduced by the MP and ML trees (Fig. S1, available in the online version of this article). Pan- and core-genome phylogenetic analyses presented distinct topologies of tested mycobacterial species as compared to that given by the tree reconstructed based on 16S rRNA genes, but coherently agreed that strain S2-37T showed closer relationship with species belonging to the genus Mycolicibacterium (Fig. S2).

The assembled genome of S2-37T was 5.9 Mbp with 12 contigs (all > 500 bp, with an N50 length of 495,170 bp) and the sequencing coverage was approximately × 100. The total length was 5,843,440 bp and the G + C content was 68.43 mol%. The results of functional annotation based on the COG database showed that genes involved in lipid transport and metabolism accounted for the largest proportion, except for genes with general function prediction only (Fig. S3). The antiSMASH biosynthetic gene clusters of S2-37T were shown in Supplementary table 2, which could provide ideas for guiding the screening of active secondary metabolites. According to the KEGG analysis, the majority of genes were involved in cell metabolism (Fig. S4), and some genes were classified into pathways associated with antimicrobial resistance and bacterial infection (Table S3). According to the VFDB analysis, strain S2-37T was predicted to be...
able to produce mycolic acid (see details in Table S4), which was considered as a virulence factor shared by large number of mycobacteria (Tortoli 2003). In addition, type VII secretion system (T7SS) was also identified, which has been proved to be involved in the secretion of virulence-associated proteins, the interaction between pathogens and hosts, the balance of zinc/iron in microbes as well (Cao et al. 2016). Particularly, PE/PPE proteins classified as members of T7SS were detected in the genome of strain S2-37 T. These proteins have been identified to localize at the cell surface and/or be secreted, inducing strong immune responses in the host and playing crucial roles in the virulence and pathogenesis of Mycobacterium tuberculosis (Choudhary et al. 2003; Sampson 2011). All these results suggested the potential pathogenicity of strain S2-37 T to human being, the feature of which was shared by many mycobacteria pathogens.

We then determined that strain S2-37 T presented highest dDDH values (average and confidence interval sequences are shown in parentheses. Bacillus subtilis subsp. subtilis JCM 1465 T was employed as the outgroup. Filled circles indicate that the corresponding nodes are also recovered in trees reconstructed by the ML and MP algorithms. Bar, 0.02 substitutions per nucleotide position in parentheses) of 39.1% (35.7–42.6%) and 34.9% (35.0–38.4%) with M. litorale CGMCC 4.5724 T and M. monacense DSM 44395 T. This corroborated the result achieved from the phylogenetic trees constructed based on pan- and core-genomes, suggesting closely phylogenetic relationships of strain S2-37 T with M. litorale CGMCC 4.5724 T and M. monacense DSM 44395 T. The ANI values between strain S2-37 T and M. litorale CGMCC 4.5724 T and M. monacense DSM 44395 T were 81.28% and 81.09%, respectively. Overall, all calculated values were below the suggested threshold for the delineation of a novel species (Chun et al. 2018).

Divergence analysis of conserved molecular signatures

Four (LacI family transcriptional regulator, cyclase, CDP-alcohol phosphatidyltransferase and phosphatidylserine synthase) and two (UPF0182 family protein and 23S rRNA (guanosine(2251)-2'-O)-
methyltransferase RlmB) conserved proteins with CSIs specific for most members of the genus *Mycolicibacterium* and *Mycobacterium* were employed for the divergence analysis, respectively. Phylogenetic analyses based on whole sequences of conserved proteins indicated that strain S2-37 T was more closely related to members from the genus *Mycolicibacterium* as compared to those from the genus *Mycobacterium*. And in the partial sequence alignments of conserved proteins, the amino acid insertion patterns of strain S2-37 T were consistent with the molecular signatures (Fig. S5a, c, e, f), which were specific for most members within the genus *Mycolicibacterium* identified by Gupta et al. (2018). However, conserved regions of two protein sequences did not give a clear clue where strain S2-37 T could be taxonomically positioned (Fig. S5b, d). The CSPs profile of strain S2-37 T was summarized in Table 1. It was clearly shown that strain S2-37 T possessed some identified CSPs specific for members of the genus *Mycolicibacterium*, but none for members of the genus *Mycobacterium* and other slow growers of mycobacterial species. Thus, we concluded that strain S2-37 T represented a novel bacterial species in the genus *Mycolicibacterium*.

Intriguingly, we observed that *Mycobacterium neglectum* CECT 8778 T and *Mycobacterium lehmannii* SN 1900 T displayed similar CSIs patterns with members of the genus *Mycolicibacterium*. It has been reported that these two species were capable of forming colonies within 7 days (Nouioui et al. 2017, 2018), and their most phylogenetically closely related species (e.g., *Mycobacterium aurum*, *Mycobacterium mageritense* and *Mycobacterium vanbaalenii*) have been reclassified into the genus *Mycolicibacterium* (Gupta et al. 2018). Therefore, we suggested that *Mycobacterium neglectum* and *Mycobacterium lehmannii* should be modified to *Mycolicibacterium neglectum* and *Mycolicibacterium lehmannii*, respectively.

**Table 1** Similarity of conserved signature proteins (CSPs) between strain S2-37 T and members of the genus *Mycolicibacterium*, slow growers of mycobacterial species and the genus *Mycobacterium*

| Gene or protein | Accession number of reference protein | Sequence coverage (%) | Sequence identity (%) | Specificity |
|----------------|--------------------------------------|-----------------------|-----------------------|-------------|
| Hypothetical protein | WP_048630777.1 | 96 | 71.59 | *Mycolicibacterium* |
| Hypothetical protein | WP_048632025.1 | – | – | |
| Hypothetical protein | WP_048632497.1 | 56 | 57.78 | |
| Hypothetical protein | WP_048634851.1 | – | – | |
| Hypothetical protein | WP_048633467.1 | 49 | 47.47 | |
| Hypothetical protein | WP_048633322.1 | 96 | 40.94 | |
| Hypothetical protein | WP_048631132.1 | 48 | 29.07 | |
| Hypothetical protein | WP_048634509.1 | 70 | 46.03 | |
| Hypothetical protein | WP_048630657.1 | 41 | 47.62 | |
| Hypothetical protein | WP_048632441.1 | 60 | 48.77 | |
| PPE Family protein | YP_177721.1 | 6 | 38.24 | Slow-Growers of mycobacterial species |
| PE Family protein PE36 | YP_178025.1 | – | – | |
| PE Family protein | WP_011725130.1 | – | – | |
| MAP_RS07685 | WP_003874405.1 | – | – | |
| Histone-like protein HNS | NP_218639.1 | – | – | *Mycobacterium* |
| Rv4010 | YP_004837050.1 | – | – | |
| Membrane protein | NP_217322.1 | – | – | |

*Previously presented by Gupta et al. (2018)*

The symbol: –, not detected
Morphology, physiology, and biochemical analysis

Strain S2-37\textsuperscript{T} showed more robust growth on Bennett’s agar as compared to G1 and R2A agar. It could form round, yellowish colonies stably on Bennett’s agar. Strain S2-37\textsuperscript{T} displayed Gram-indefinite in Gram staining, but positive in acid-fast staining (Fig. S6), which was frequently observed in mycobacterial species (Nakamura et al. 1991). SEM observation revealed that strain S2-37\textsuperscript{T} is rod-shaped, approximately 1.0–1.5 \textmu m in length and 0.4 \textmu m in diameter (Fig. 2). Positive for catalase activity and negative for Tween 80 hydrolysis, other results achieved from strains S2-37\textsuperscript{T}, M. litorale CGMCC 4.5724\textsuperscript{T} and M. monacense DSM 44395\textsuperscript{T}, were summarized in Table 2. All of these tests were carried out in duplicate using the standard inoculum.

Chemotaxonomic characterization

The major polar lipids of strain S2-37\textsuperscript{T} were identified as diphosphatidylglycerol, phosphatidylethanolamine and phosphatidylinositol (Fig. S7). The major fatty acids (> 10\%) of strain S2-37\textsuperscript{T} were composed of C\textsubscript{16:0}, C\textsubscript{18:0}, C\textsubscript{20:3o3} and C\textsubscript{22:6o3}, of which the last two types of fatty acids were not detected or only traced in small amounts (< 1\%) in M. litorale CGMCC 4.5724\textsuperscript{T} and M. monacense DSM 44395\textsuperscript{T} (Reischl et al. 2006; Zhang et al. 2012).

In conclusion, the phenotypic, chemotaxonomic and phylogenetic data supports the delineation of strain S2-37\textsuperscript{T} as a novel species of the genus Mycolicibacterium. We propose the name Mycolicibacterium gossypii sp. nov. for the species.

Description of Mycolicibacterium gossypii sp.nov.

Mycolicibacterium gossypii (gos.sy’pi.i, N.L. gen. n. gossypii of the cotton genus Gossypium).

Cells are aerobic, Gram-positive, acid-fast, non-spore-forming and short rods. It could form smooth, yellowish colonies on Bennett’s agar within 7 days at 30 °C. Cells are approximately 0.4 \textmu m in diameter and 1.0–1.5 \textmu m long. Positive for catalase activity and negative for Tween 80 hydrolysis. In API 20NE tests, arginine dihydrolase activity is positive. Negative for nitrate reduction, indole production, \beta-galactosidase, \beta-glucosidase, gelatinase and urease activities, as well as fermentation of D-glucose. In addition, results are negative for assimilation of D-glucose, L-arabinose, D-mannitol, maltose, gluconate, N-acetyl-glucosamine, adipic acid, malic acid, capric acid and phenylacetic acid, but positive for citrate acid. In API ZYM tests, alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cystine arylamidase, acid phosphatase, naphthol-AS-BI-phosphohydrolase, \alpha-glucosidase activities are positive. Negative for lipase (C14), trypsin, \alpha-chymotrypsin, \alpha-galactosidase, \beta-galactosidase, \beta-glucuronidase, N-acetyl-\beta-glucosaminidase, \alpha-mannosidase and \alpha-rucosidase activities. Susceptible to amikacin, ciprofloxacin, clarithromycin, doxycycline, tobramycin and sulfamethoxazole, but resistant to cefalotin and cefoxitin. The major fatty acids (> 10\%) are C\textsubscript{16:0}, C\textsubscript{18:0}, C\textsubscript{20:3o3} and C\textsubscript{22:6o3}.

The polar lipids profile is mainly composed of diphosphatidylglycerol, phosphatidylethanolamine and phosphatidylinositol. The genome size of strain S2-37\textsuperscript{T} is 5.9 Mbp and the DNA G + C content of the type strain is 68.4 mol\%. The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA sequence is MW295419. This Whole Genome Shotgun project has been deposited at GenBank/EMBL/DDBJ under the accession number JAFEVR000000000. The version described in this paper is version JAFEVR010000000.
The type strain, S2-37 $^T$ (= JCM 34327 $^T$ = CGMCC 1.18817 $^T$), was isolated from cotton field soil in Xinjiang, PR China.

Acknowledgements The authors thank Dr. Zhi-Dong Zhang from Xinjiang Academy of Agricultural Sciences of China for his assistance with sampling.

Authors’ contributions RRH performed the experiments and finished the draft of the manuscript. RRH, SRY, CZ, XFG and XKC performed strain isolation and phenotypic analyses. RRH and SRY performed genomic analyses. ZQW and YNL assisted to improve the manuscript. WZL designed all the experiments and supervised the manuscript.

Funding This work was supported by grants from the Natural Science Foundation of Jiangsu Province (Grant No. BK20190703) and The Natural Science Research of the Jiangsu Higher Education Institutions of China (Grant No. 19KJB530010).

Declarations

Conflict of interest The authors declare no conflicts of interest.

Availability of data and material The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of the strain S2-37 $^T$ is MW295419. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAFEVR000000000. The version described in this paper is version JAFEVR010000000.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent All authors agree to publish this work.

References

Altschul S, Wootton J, Gertz E, Agarwala R, Morgulis A, Schäffer A, Yu Y (2005) Protein database searches using compositionally adjusted substitution matrices. FEBS J 272:5101–5109. https://doi.org/10.1111/j.1742-4658.2005.04945.x

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov A, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Alekseyev M, Pevzner P (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

Berd D (1973) Laboratory identification of clinically important aerobic actinomycetes. Appl Microbiol 25(4):665–681. https://doi.org/10.1128/aem.25.4.665-681.1973

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee S, Medema M, Weber T (2019) AntiSMASH 5.0: updates to
the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:81–87. [https://doi.org/10.1093/nar/gkz210]

Brown-Elliott BA, Wallace RJ (2002) Clinical and taxonomic status of pathogenic nonpigmented or late-pigmenting rapidly growing mycobacteria. Clin Microbiol Rev 15:716–746. [https://doi.org/10.1128/CMR.15.4.716-746.2002]

Brown-Elliott BA, Wallace RJ, Petti CA, Mann LB, McGlasson Choudhary RK, Mukhopadhyay S, Chakhaiyar P, Sharma N, Cao Z, Casabona MG, Kneuper H, Chalmers JD, Palmer T Fedrizzi T, Meehan C, Grottola A, Giacobazzi E, Fregni Serpini Federhen S (2012) The NCBI taxonomy database. Nucleic Acids Res 40:68–72. [https://doi.org/10.1093/nar/gkr1178]

Fedrizzi T, Meehan C, Grottola A, Giacobazzi E, Fregni Serpini G, Tagliazucchi S, Fabio A, Bettua C, Bertorelli R, De Sanctis V, Rumpanesi F, Pecorari M, Jousson O, Tortoli E, Segata N (2017) Genomic characterization of Nontuberculous Mycobacteria. Sci Rep 7:45258. [https://doi.org/10.1038/srep45258]

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4):783–791. [https://doi.org/10.2307/2408678]

Goodfellow M, Orchard VA (1974) Antibiotic sensitivity of some nocardioform bacteria and its value as a criterion for taxonomy. J Gen Microbiol 83(2):375–387. [https://doi.org/10.1099/00221287-83-2-375]

Gupta RS, Lo B, Son J (2018) Phylogenomics and comparative genomic studies robustly support division of the genus Mycobacterium into an emended genus Mycobacterium and four novel genera. Front Microbiol. [https://doi.org/10.3389/fmicb.2018.00067]

Hopkins DW, Macnaughton SJ, O’Donnell AG (1991) A dispersion and differential centrifugation technique for representatively sampling microorganisms from soil. Soil Biol Biochem 23(3):217–225. [https://doi.org/10.1016/0038-0717(91)90055-O]

Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428(4):726–731. [https://doi.org/10.1016/j.jmb.2015.11.006]

Koon MA, Almohammed Ali K, Speaker RM, McGrath JP, Linton EW, Steinhilb M (2019) Preparation of prokaryotic and eukaryotic organisms using chemical drying for morphological analysis in scanning electron microscopy (SEM). J Vis Exp 143:e58761. [https://doi.org/10.3791/58761]

Kent PT, Kubica GP (1985) Public health mycobacteriology: a guide for the level III laboratory. Centers for Disease Control, U.S. Department of Health and Human Services, Atlanta, GA, Washington, D.C

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolution analyses using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 33(7):1870–1874. [https://doi.org/10.1093/molbev/msw054]

Kuykendall LD, Roy MA, O’Neill JJ, Devine TE (1988) Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of Bradyrhizobium japonicum. Int J Syst Bacteriol 38(4):358–361. [https://doi.org/10.1099/00207713-38-4-358]

Lin SH, Liao YC (2013) CISA: Contig integrator for sequence assembly of bacterial genomes. PLoS ONE 8(3):e60843. [https://doi.org/10.1371/journal.pone.0060843]

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) A genome-scale map of the metabolic and regulatory network in Arabidopsis thaliana. Science 313(5797):1551–1557. [https://doi.org/10.1126/science.1127781]

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung D, Yiu S, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang J, Wang J, Lam T, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1(1):18. [https://doi.org/10.1186/2047-217X-1-18]

Magee GM, Ward AC (2012) Genus I. Mycobacterium. Lehmann and Neumann. In: Goodfellow M, Kampfer P, Busse HJ, Trujillo ME, Suzuki K, Ludwig W, Whitman W (eds) Bergey’s manual of systematic bacteriology, vol 5. Springer, New York, pp 312–375

Meier-Kolthoff JP, Aich AF, Klefnk HP, Goker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinform 14:60. [https://doi.org/10.1186/1471-2105-14-60]

Mignard S, Flandrois J (2008) A seven-genе, multilocus, genusswide approach to the phylogeny of mycobacteria using
supertrees. Int J Syst Evol Microbiol 58(6):1432–1441. https://doi.org/10.1099/ijsem.0.65658-0

Minnikin DE, O’Donnell AG, Goodfellow M, Alderson G, Athalye M, Schaal A, Parlett J (1984) An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids. J Microbiol Methods 2(5):233–241. https://doi.org/10.1016/0167-7012(84)90018-6

Nakamura M, Harano Y, Koga T (1991) Effect of heat-staining procedure on the Gram staining properties of mycobacteria. Nippon Sakingaku Zasshi 46(2):533–539. https://doi.org/10.3412/jsb.46.533

Nouiou I, Sangal V, Carro L, Teramoto K, Jando M, Montero-Minnikin DE, O’Donnell AG, Goodfellow M, Alderson G, Parte A, Carbasse J, Meier-Kolthoff J, Reimer L, Göker M, Saitou N, Nei M (1987) The neighbor-joining method: a new phylogeny reconstruction procedure on the Gram staining properties of mycobacteria: Mycobacterium lehmannii sp. nov. and Mycobacterium neumannii sp. nov. Int J Syst Evol Microbiol 67(12):4948–4955

Nouiou I, Sangal V, Carro L, Teramoto K, Jando M, Montero-Calasanz MDC, Igual JM, Sutcliffe I, Goodfellow M, Klenk HP (2017) Two novel species of rapidly growing mycobacteria: Mycobacterium lehmannii sp. nov. and Mycobacterium neumannii sp. nov. Int J Syst Evol Microbiol 67(12):4948–4955

Parte A, Carbasse J, Meier-Kolthoff J, Reimer L, Göker M (2020) List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol 70(11):5670–5612. https://doi.org/10.1099/ijsem.0.04332

Patnaik BB, Park SY, Kang SW, Hwang HJ, Wang TH, Park E, Chung J, Song D, Kim C, Kim S, Lee J, Jeong O, Park H, Han Y, Lee Y (2016) Transcriptome profile of the Asian Giant Hornet (Vespa mandarinia) using Illumina HiSeq 4000 sequencing: De novo assembly, functional annotation, and discovery of SSR markers. Int J Genomics 2016:4169587. https://doi.org/10.1155/2016/4169587

Prakash O, Nimonkar Y, Shouche YS (2013) Practice and prospects of microbial preservation. FEMS Microbiol Lett 339:1–9. https://doi.org/10.1111/1574-6968.12034

Reischl U, Melz H, Kroppendstedt RM, Mietlhe T, Naumann L, Mariotti G, Azzarelli G, Tortoli E (2006) Mycobacterium monacense sp. nov. Int J Syst Evol Microbiol 56:2575–2578. https://doi.org/10.1099/ijsem.0.04527-0

Roth A, Fischer M, Hamid M, Michalke S, Ludwig W, Mauch H (1998) Differentiation of phylogenetically related slowly growing mycobacteria based on 16S–23S rRNA gene internal transcribed spacer sequences. J Clin Microbiol 36(1):139–147. https://doi.org/10.1128/JCM.36.1.139-147.1998

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I (2009) Abyss: a parallel assembler for short read sequence data. Genome Res 19(6):1117–123. https://doi.org/10.1101/gr.089532.108

Smibert RM, Krieg NR (1994) Phenotypic characterization. In: Gerhardt P, Murray RGE, Wood WA, Krieg NR (eds) Methods for general and molecular bacteriology. American Society for Microbiology, Washington, pp 607–654

Souidis J, Nei M (1988) Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. Mol Biol Evol 11(2):261–277. https://doi.org/10.1093/oxfordjournals.molbev.a040497

Steel M, Rodrigo A (2008) Maximum likelihood supertrees. Syst Biol 57(2):243–250. https://doi.org/10.1080/1063515080203014

Tatusov R, Fedorova N, Jackson J, Jacobs A, Kiryutin B, Koonin E, Krylov D, Mazumder R, Smirnov S, Nikolayskaya A, Rao B, Mekhedov S, Sverlov A, Vasudevan S, Wolf Y, Yin J, Natale D (2003) The COG database: an updated version includes eukaryotes. BMC Bioinform 4(4):1–14. https://doi.org/10.1186/1471-2105-4-41

Tortoli E (2003) Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. Clin Microbiol Rev 16:319–354. https://doi.org/10.1128/CMR.16.2.319-354.2003

Xu P, Li W, Tang S, Zhang Y, Chen G, Chen H, Xu L, Jiang C (2015) Naxibacter alkalitolerans gen. nov., sp. nov., a novel member of the family ‘Oxalobacteraceae’ isolated from China. Int J Syst Evol Microbiol 55:1149–1153. https://doi.org/10.1099/ijsem.0.04307-0

Yoon S, Ha S, Lim J, Kwon S, Chun J (2017) A large-scale evaluation of algorithms to calculate average nucleotide identity. Anton Leeuw Int J G 110(10):1281–1286. https://doi.org/10.1007/s10482-017-0844-4

Zhang Y, Zhang J, Fang C, Pang H, Fan J (2012) Mycobacterium litorale sp. nov., a rapidly growing mycobacterium from soil. Int J Syst Evol Microbiol 62:1204–1207. https://doi.org/10.1099/ijsem.0.033449-0

Zhang D, Chen X, Zhang X, Zhi X, Yao J, Jiang Y, Xiong Z, Li W, Tang S, Zhang Y, Zhang J, Fang C, Pang H, Fan J (2012) Mycobacterium sediminis sp. nov. and Mycobacterium arabiense sp. nov., two rapidly growing members of the genus Mycobacterium. Int J Syst Evol Microbiol 63:4081–4086. https://doi.org/10.1099/ijsem.0.050567-0

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: A SMART TM approach for full-length cDNA library construction. Biotechniques 30:892–897. https://doi.org/10.2144/00001817

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.