Characterization of 17 strains belonging to the *Mycobacterium simiae* complex and description of *Mycobacterium paraense* sp. nov.

Ana R. Fusco da Costa,[1] Tarcisio Fedrizzi,[2] Maria L. Lopes,[1] Monica Pecorari,[3] Wana L. Oliveira da Costa,[1] Elisabetta Giacobazzi,[2] Jeann R. da Costa Bahia,[1] Veronica De Sanctis,[2] Karla V. Batista Lima,[1] Roberto Bertorelli,[2] Antonella Grottola,[3] Anna Fabio,[3] Alessandro Mariottini,[4] Pamela Ferretti,[2] Francesca Di Leva,[2] Giulia Fregni Serpini,[3] Sara Tagliazucchi,[3] Fabio Rumpianesi,[3] Olivier Jousson,[2] Nicola Segata[2] and Enrico Tortoli[5]

Correspondence
Enrico Tortoli
tortoli.enrico@hsr.it

[1]Bacteriology and Mycology Section, Evandro Chagas Institute, Pará, Brazil

[2]NGS Facility, Laboratory of Biomolecular Sequence and Structure Analysis for Health, Centre for Integrative Biology, University of Trento, Italy

[3]Microbiology and Virology Unit, Policlinico University Hospital, Modena, Italy

[4]Genetic Unit, Careggi University Hospital, Florence, Italy

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

Fourteen mycobacterial strains isolated from pulmonary samples of independent patients in the state of Pará (Brazil), and three strains isolated in Italy, were characterized using a polyphasic approach. Thorough genetic investigation, including whole-genome sequencing, demonstrated that the strains belong to the *M. simiae* complex, being most closely related to *Mycobacterium interjectum*. For 14 of the strains, evidence emerged supporting their inclusion in a previously unreported species of the genus *Mycobacterium*, for which the name *Mycobacterium paraense* sp. nov. is proposed (type strain, IEC26^T=DSM 46749^T=CCUG 66121^T). The novel species is characterized by slow growth, unpigmented or pale yellow scotochromogenic colonies, and a HPLC mycolic acid profile different from other known mycobacteria. In different genetic regions, high sequence microheterogeneity was detected.

The mycobacteria related to *Mycobacterium simiae* constitute, at the time of writing, the largest group or complex within the genus *Mycobacterium*. The group actually includes 17 officially recognized species (Tortoli *et al.*, 2011).

Seventeen strains sharing phenotypic and genotypic characteristics consistent with this complex were isolated, in many cases repeatedly (39 isolations in total), from 16 patients with pulmonary symptoms. Most of such mycobacteria were found during the reidentification (by means of DNA sequencing) of clinical isolates, which, with PCR Restriction Analysis (PRA), had either been previously identified as *Mycobacterium asiaticum*, or had presented unknown PRA-patterns (da Costa *et al.*, 2010). The genetic sequences of these strains showed a high degree of similarity, with *Mycobacterium interjectum* (Springer

Abbreviations: ITS, internal transcribed spacer; SNPs, single nucleotide polymorphisms.

The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA gene sequences of strains FI-10043, FI-07156, FI-13041, IEC23–IEC34, IEC1808 and IEC1883 are HM770870, KJ920347, KJ920348 and KJ948993–KJ949006, respectively; those for the *hsp65* sequence of the same strains are HM775985, EU370529, KJ957808, KJ949035, KJ949036, HM056136–HM056138, KJ949037, HM056140–HM056143, KJ949038, HM056145, KJ949039 and KJ949040, respectively; and those for the *rpoB* sequence of the same strains are KJ586587, KJ957807, KJ920346 and KJ949007–KJ949020, respectively. The GenBank/EMBL/DDBJ accession numbers for the ITS1 sequence of strains FI-70043, FI-07156, FI-13041, IEC23–IEC34, IEC1883 and *Mycobacterium interjectum* DSM 44064^T are KJ586582, KJ920347, KJ920348, KJ949021–KJ949033 and KJ586583, respectively.

Three supplementary figures are available with the online Supplementary Material.

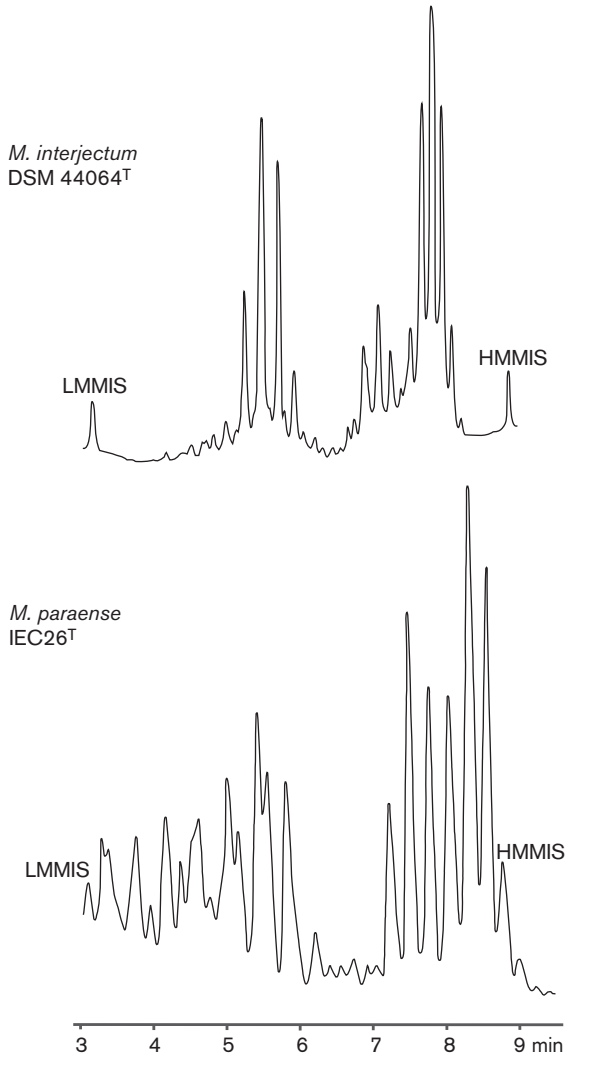

Fig. 1. Representative mycolic acid pattern of *M. interjectum* DSM 44064^T and *Mycobacterium paraense* sp. nov. IEC26^T. LMMIS, low molecular mass internal standard; HMMIS, high molecular mass internal standard.

et al., 1993) being the most closely related species. The polyphasic characterization of such strains led us to uphold the assignment of 14 of the strains to a novel species of the genus *Mycobacterium*.

Thirty-six mycobacteria were isolated between 1999 and 2010 at the Evandro Chagas Institute, from sputum samples of 14 patients resident in the state of Pará, Brazil. All the patients presented respiratory symptomatology, such as chronic cough and chest pain, persisting for two years or more. Some of the patients presented bronchiectasis that was occasionally associated with haemoptysis. Three strains were isolated in Italy, two of them (FI-07156 and FI-10043) were obtained from the same patient at a three year interval.

Major biochemical tests recommended for the speciation of mycobacteria were performed as described previously (Kent & Kubica, 1985), including tests for niacin accumulation, nitrate reduction, Tween 80 hydrolysis (10 days), urease, β-glucosidase, tellurite reduction, and catalase. All strains presented negative results for the majority of the tests performed; only catalase (thermostable and semi-quantitative) tests were uniformly positive, while discordant results were obtained for Tween 80 hydrolysis. The majority of the strains grew smooth, pale yellow, scotochromogenic colonies on Löwenstein–Jensen in 2 weeks or more at 37 °C. Several strains however remained unpigmented, both in the light and in the dark. Growth was slower at 30 °C and inhibited at 42 °C. Colonies were grown on media supplemented with p-nitrobenzoate (500 μg ml^{-1}) and thiacetazone (10 μg ml^{-1}), not on those supplemented with hydroxylamine (500 μg ml^{-1}) or isoniazid (1 μg ml^{-1}).

Cell-wall mycolic acids were analysed as reported previously (Rhodes *et al.*, 2005). In short, colonies grown for 15 days at 37 °C on Middlebrook 7H11 agar were saponified with KOH (25 % in H_2O), extracted with chloroform, derivatized according to the manufacturer's

Table 1. Minimal inhibitory concentrations (μg ml^{-1}) of antimycobacterial drugs suitable for slowly growing mycobacteria

Drug	MIC indicating resistance	Strain										
		IEC26^T	IEC30	IEC34	IEC24	IEC31	IEC27	IEC32	IEC28	IEC33	IEC29	FI-07156
Amikacin	>32	8	16	8	8	8	16	8	16	8	8	8
Ciprofloxacin	>2	8	16	16	16	16	16	8	8	16	8	>16
Clarithromycin	>16	2	2	2	4	16	>64	2	4	2	1	2
Doxycycline	>4	>16	>16	>16	>16	>16	>16	>16	>16	>16	>16	>16
Ethambutol	>4	16	>16	>16	>16	16	>16	16	>16	16	8	>16
Linezolid	>16	16	16	16	16	16	32	16	16	16	16	16
Moxifloxacin	>2	2	2	2	2	2	2	2	2	4	0.5	8
Rifabutin	>2	0.25	1	1	0.25	0.5	1	1	>8	0.25	0.25	0.5
Rifampicin	>1	4	>8	>8	1	>8	>8	4	>8	8	>8	8
Streptomycin	NE*	32	64	>64	16	64	>64	8	>64	16	4	16
Sulfamethoxazole	>38	>152	19	76	152	76	38	9.5	9.5	>152	4.75	>152

*NE, Not established.

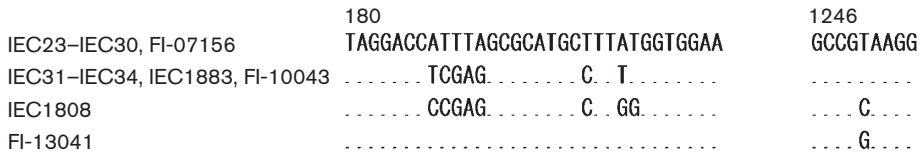

Fig. 2. Sequence heterogeneity presented in the 16S rRNA gene by different test strains. Numbers indicate *E. coli* corresponding positions; dots indicate identities.

instructions (MIDI) and loaded onto an Agilent ChemStation HPLC (Agilent Technologies). Mycolic acids were separated with a gradient of methanol and 2-propanol (starting ratio 75/25 %, end ratio 95/5 %) and analysed using the software Sherlock Version Myco 1.0 and the database MICAG1 1.02. All strains, with the exception of IEC1808 whose pattern was unique, were characterized by a late emerging major cluster of peaks (between 7 and 9 min) (Fig. 1); first came a series of peaks presenting very variable heights among the strains. Such mycolic acid pattern was clearly different from the one of *M. interjectum* (Tortoli *et al.*, 1996). The system identified such profiles as *Mycobacterium simiae*, with a low similarity index (<40 %).

To test antimicrobial susceptibility, minimal inhibitory concentrations were determined, according to CLSI recommendations (CLSI, 2011), using a commercially available microdilution method (SLOMYCOI, Sensititer) including the drugs with potential activity against slowly growing nontuberculous mycobacteria. The majority of the strains were susceptible to amikacin, clarithromycin, linezolid and rifabutin, and resistant to ciprofloxacin, doxycycline, ethambutol, rifampicin and streptomycin. Variable susceptibility was observed for sulfamethoxazole (Table 1).

For gene sequencing, double-strand sequences from all the strains included in the study were determined using BigDye Terminator chemistry on an AB3130 DNA sequencer (Applied Biosystems) following the standard protocol of the supplier. The regions investigated included the genes encoding 16S rRNA (Kirschner *et al.*, 1993), 65 kDa heat-shock protein (*hsp65*) (McNabb *et al.*, 2004), RNA polymerase β-subunit (*rpoB*) (Adékambi *et al.*, 2003) and the internal transcribed spacer (ITS1) interposed between the 16S and 23S rRNA genes (Roth *et al.*, 1998).

In the almost-complete 16S rRNA gene sequence (1448 bp), four sequence variants (sequevars) were detected which differed for a variable number of nucleotide substitutions, from 1 to 9, at *Escherichia coli* corresponding positions 187–91, 200, 203, 204 and 1250 (Fig. 2). The first sequevar was shared by nine strains: IEC23–IEC30 and FI-07156; the second was shared by IEC31–IEC34, IEC1883 and FI-10043; the third and the fourth sequevars were presented by IEC1808 and FI-13041, respectively. *M. interjectum* was the most closely related species to all of them, differing 13 bp from the first (similarity 99.1 %), 7 bp from the second and the third (99.5 %) and 12 bp from the fourth (99.2 %).

In the 399 bp of the hypervariable region of the *hsp65* sequence, seven sequevars were present, shared by six (IEC26^T, IEC27, IEC29, IEC32, FI-07156 and FI-10043), four (IEC24, IEC25, IEC28, IEC31), two (IEC33, IEC34), two (IEC1808, IEC1883) and, the other three, by one strain each. The highest similarity (97.2–97.5 %) was (with the exception of strains IEC23, IEC1808, IEC1883 and FI-13041) with *Mycobacterium parmense* (Fanti *et al.*, 2004).

In the 710 bp trait of the *rpoB* sequence, 12 sequevars were present, one shared by six strains (IEC24–IEC29) and the others presented by one strain each; the closest species was, for all of them, *M. interjectum* with similarity ranging from 96.2 to 97.7 %.

In the complete ITS1 sequence (223 bp), seven sequevars were detected, presented by ten (IEC24–IEC30, IEC32, IEC34, IEC1883), two (IEC33, FI-07156) and, the remaining five, by one strain each. Except for IEC1808 and FI-13041, *M. interjectum* presented the highest resemblance (similarity range: 92.5–96.6 %).

The distribution of different sequevars in various strains is summarized in Table 2.

Table 2. Combination of different sequevars in various strains

Strain	16S rRNA gene	*hsp65*	*rpoB*	ITS1
IEC23*	A	E	B	C
IEC24	A	B	A	A
IEC25	A	B	A	A
IEC26^T	A	A	A	A
IEC27	A	A	A	A
IEC28	A	B	A	A
IEC29	A	A	A	A
IEC30	A	F	C	A
IEC31	B	B	D	D
IEC32	B	A	E	A
IEC33	B	C	F	B
IEC34	B	C	G	A
IEC1808*	C	D	H	E
IEC1883	B	D	I	A
FI-07156	A	A	J	B
FI-10043	B	A	K	F
FI-13041*	D	G	L	G

*Strains that were not considered included in the novel species.

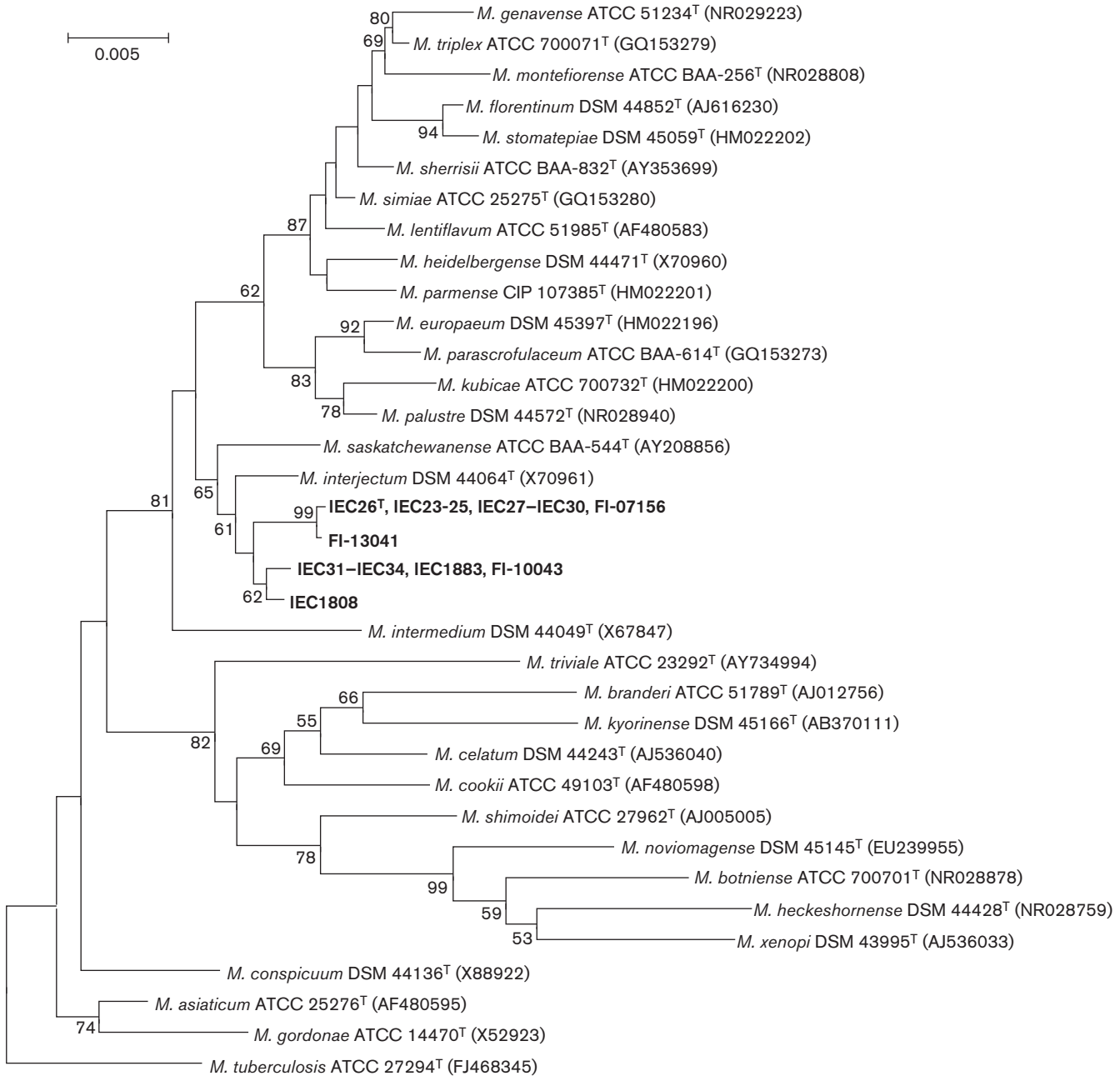

Fig. 3. Phylogenetic tree based on 16S rRNA gene sequences, reconstructed using the neighbour-joining method bootstrapped 1000 times. Bootstrap values are given at nodes. Bar, 0.005 substitutions per nucleotide position.

For phylogenetic analyses, sequences of the type strains of species most closely related to the test strains were retrieved from the GenBank database, aligned using CLUSTAL W software (Thompson *et al.*, 1994), and trimmed to start and finish at the same position. Phylogenetic analysis was conducted for each investigated genetic region and *Mycobacterium tuberculosis* ATCC 27294^T was used as an outgroup. The neighbour-joining method (Saitou & Nei, 1987), supported by MEGA 6.06 software (Tamura *et al.*, 2013), was used for the reconstruction of phylogenetic trees; 1000 bootstrap replications were implemented.

In the phylogenetic tree reconstructed using the almost-complete sequences of the 16S rRNA gene, the 17 strains characterized in this study belonged to the same cluster and were most closely related to *M. interjectum* DSM 44064^T (Fig. 3). In the tree based on partial *hsp65* sequences, 13 strains were included in the same robust cluster and were

most closely related to *Mycobacterium parmense* CIP 107385^T, while strains IEC23, IEC1808, IEC1883 and FI-13041 were more distant (Fig. S1, available in the online Supplementary Material). In the dendrogram reconstructed on the partial *rpoB* sequences, all strains were most closely related to *M. interjectum* DSM 44064^T, but strains IEC23, IEC1808 and FI-13041 clustered in a separate group with respect to the other strains (Fig. S2). In the tree inferred by complete ITS1 sequences, all strains presented the closest similarity with *Mycobacterium palustre* DSM 44572^T (Torkko *et al.*, 2002), however strains IEC1808 and FI-13041 were not included in the same cluster as the other strains (Fig. S3). Using concatenated sequences of the almost-complete 16S rRNA gene, complete ITS1 and partial *hsp65* and *rpoB* (about 2900 bp in total), a phylogenetic tree was reconstructed. The 17 test strains were confirmed as closely related to *M. interjectum* DSM 44064^T, but again strains IEC23, IEC1808 and FI-13041 belonged to separate branches (Fig. 4).

In conclusion, from the phylogenetic analysis, in three out of four investigated genetic regions, three strains (IEC23, IEC1808 and FI-13041) branched outside of the cluster including the other 14 strains; only in the 16S rRNA gene sequence analysis did the 17 test strains group together.

To further support the phylogenetic position of different strains, five strains were selected and subjected to whole-genome sequencing. The selection included the proposed

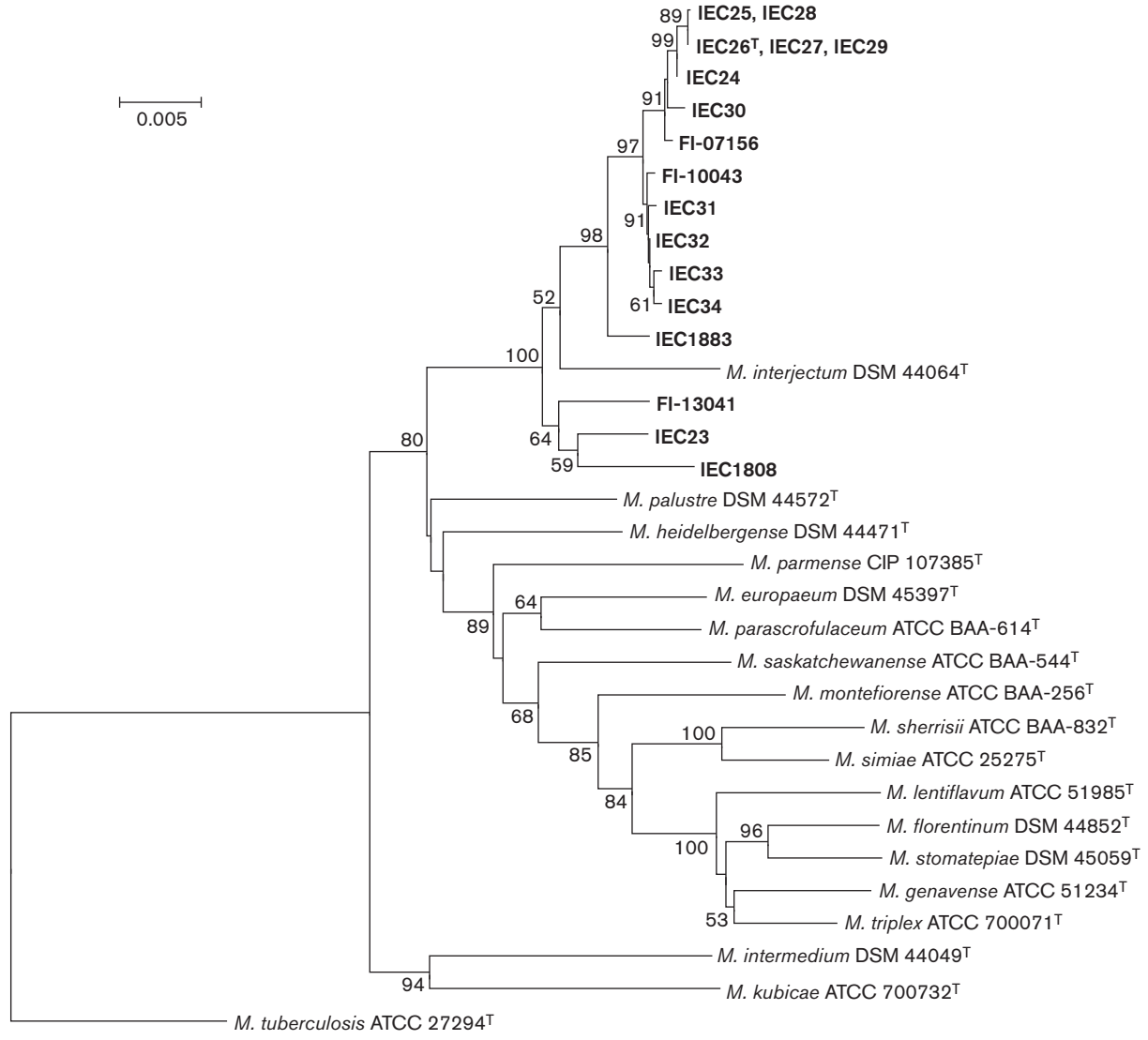

Fig. 4. Phylogenetic tree based on concatenated sequences of 16S rRNA gene, ITS1, *hsp65* and *rpoB*, reconstructed using the neighbour-joining method bootstrapped 1000 times. Bootstrap values are given at nodes. Bar, 0.005 substitutions per nucleotide position.

type strain of the novel species (IEC26^T), one of the three strains presenting as more divergent from the others (IEC1808), one strain located in a sub-cluster not harbouring the type strain in different genetic regions investigated (IEC33), and two strains isolated from the same patient at three years interval (FI-07156 and FI-10043).

Paired-end libraries were prepared from 1 ng of total bacterial DNA using Nextera XT DNA Sample Preparation kit and Nextera XT Index kit (Illumina) according to manufacturer's protocol. Libraries were then normalized to 2 nM, pooled for multiplexing in equal volumes, and sequenced at 10 pM on the Illumina HiSeq 2000 platform with 100 nt paired end reads to achieve a coverage >100 × per base. Assembly was performed with SPAdes 3.0.0 software (Bankevich *et al.*, 2012) and whole-genome sequence comparison with Mauve software (Darling *et al.*, 2010). The analysis showed that two isolates (FI-07156 and FI-10043) presented overlapping almost all single nucleotide polymorphisms (SNPs) (>99.9 %), a level of similarity indicative of representing a single strain (Goris *et al.*, 2007) (Fig. 5). The genome-wide similarities of strains IEC26^T, IEC33 and FI-07156/FI-10043 were high enough (>99 %) to support their belonging to a single species as confirmed by the high sequence conservation and the very limited difference in the gene content (Fig. 5). Strain IEC1808 was, in contrast, more divergent and did not belong to the same species as the other strains (Fig. 5).

In conclusion, whole-genome sequencing fully confirmed the results of phylogenetic analysis based on 16S rRNA gene, *hsp65*, *rpoB* and ITS1 sequences, and revealed that the two strains isolated from the same patient represented, despite several nucleotide diversities in the genetic regions above, a single strain. The 14 strains which cluster together in the phylogenetic analysis are concluded to represent a novel species of the genus *Mycobacterium*, for which the name *Mycobacterium paraense* sp. nov. is proposed.

Description of *Mycobacterium paraense* sp. nov. *Mycobacterium paraense* (pa.ra.en′se. N.L. neut. adj. *paraense* of, or belonging to, Pará, isolated in the state of Pará, Brazil).

Cells are Gram-stain-positive, non-motile, non-spore-forming, acid–alcohol-fast bacilli. Unpigmented or light yellow colonies develop in two weeks or more at temperatures ranging between 25 and 37 °C, both in the light and the

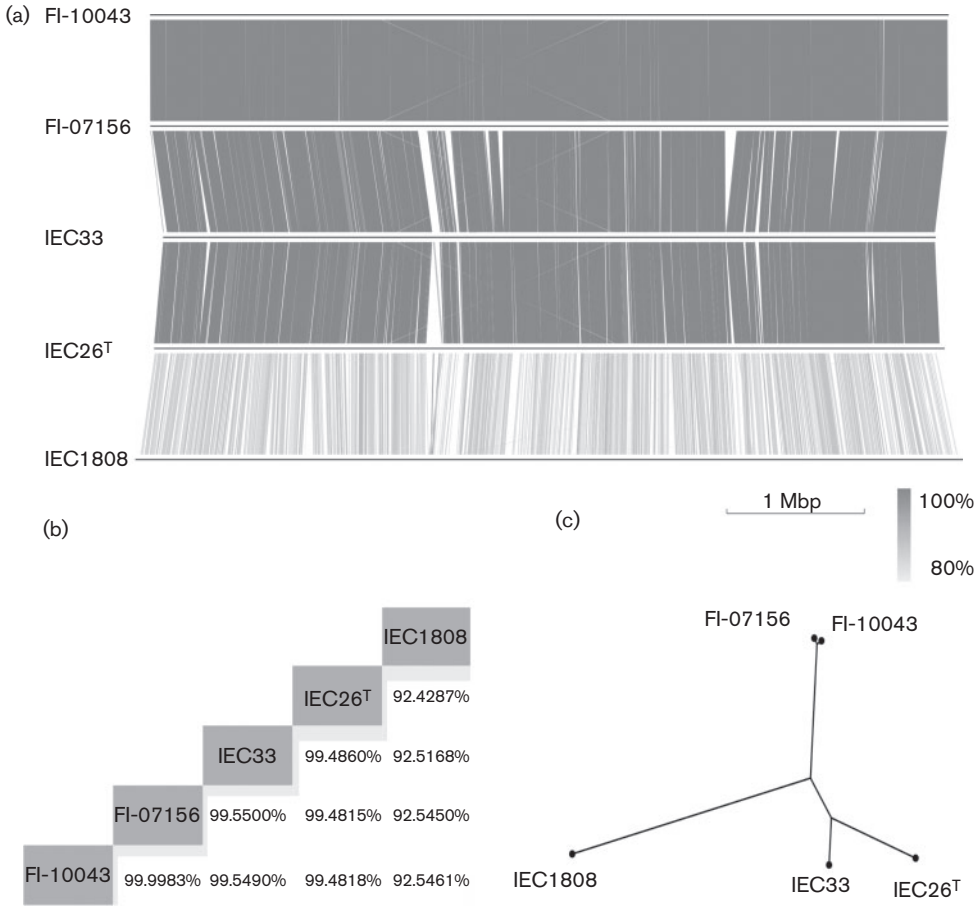

Fig. 5. Whole-genome comparison of five sequenced and assembled strains showing: (a), genome alignment (Sullivan *et al.*, 2011); (b), pairwise SNPs concordance rates; (c), SNPs-based phylogenetic tree on the whole-genome alignments.

dark. Among biochemical features, tests for niacin accumulation, nitrate reduction, urease, β-glucosidase and tellurite reduction were negative; catalase activity was the only uniformly positive test. These biochemical features are therefore not suitable to differentiate the novel species from other slowly growing scotochromogenic species. The HPLC profile of mycolic acids is characterized by one late cluster of peaks. Characterized by large microheterogeneity in 16S rRNA gene, *hsp65*, *rpoB* and ITS1 sequences, and is most closely related to *M. interjectum*, and, to a lesser extent, to other species of the *M. simiae* complex.

The type strain is IEC26^T (=DSM 46749^T=CCUG 66121^T), and was isolated from the sputum of a patient from Parauapebas, Pará, Brazil.

References

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

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S. & other authors (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477.

CLSI (2011). *Susceptibility Testing of Mycobacteria, Nocardiae and Other Aerobic Actinomycetes*; Approved Standard, 2nd edn, M24-A2. Wayne, PA: Clinical and Laboratory Standards Institute.

da Costa, A. R., Lopes, M. L., Furlaneto, I. P., de Sousa, M. S. & Lima, K. V. (2010). Molecular identification of nontuberculous mycobacteria isolates in a Brazilian mycobacteria reference laboratory. *Diagn Microbiol Infect Dis* **68**, 390–394.

Darling, A. E., Mau, B. & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147.

Fanti, F., Tortoli, E., Hall, L., Roberts, G. D., Kroppenstedt, R. M., Dodi, I., Conti, S., Polonelli, L. & Chezzi, C. (2004). *Mycobacterium parmense* sp. nov. *Int J Syst Evol Microbiol* **54**, 1123–1127.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81–91.

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

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

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

Rhodes, M. W., Kator, H., McNabb, A., Deshayes, C., Reyrat, J. M., Brown-Elliott, B. A., Wallace, R., Jr, Trott, K. A., Parker, J. M. & other authors (2005). *Mycobacterium pseudoshottsii* sp. nov., a slowly growing chromogenic species isolated from Chesapeake Bay striped bass (*Morone saxatilis*). *Int J Syst Evol Microbiol* **55**, 1139–1147.

Roth, A., Fischer, M., Hamid, M. E., 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**, 139–147.

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

Springer, B., Kirschner, P., Rost-Meyer, G., Schröder, K. H., Kroppenstedt, R. M. & Böttger, E. C. (1993). *Mycobacterium interjectum*, a new species isolated from a patient with chronic lymphadenitis. *J Clin Microbiol* **31**, 3083–3089.

Sullivan, M. J., Petty, N. K. & Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Torkko, P., Suomalainen, S., Iivanainen, E., Tortoli, E., Suutari, M., Seppänen, J., Paulin, L. & Katila, M. L. (2002). *Mycobacterium palustre* sp. nov., a potentially pathogenic, slowly growing mycobacterium isolated from clinical and veterinary specimens, and from Finnish stream waters. *Int J Syst Evol Microbiol* **52**, 1519–1525.

Tortoli, E., Bartoloni, A., Burrini, C., Colombrita, D., Mantella, A., Pinsi, G., Simonetti, M. T., Swierczynski, G. & Böttger, E. C. (1996). Characterization of an isolate of the newly described species *Mycobacterium interjectum*. *Zentralbl Bakteriol* **283**, 286–294.

Tortoli, E., Böttger, E. C., Fabio, A., Falsen, E., Gitti, Z., Grottola, A., Klenk, H. P., Mannino, R., Mariottini, A. & other authors (2011). *Mycobacterium europaeum* sp. nov., a scotochromogenic species related to the *Mycobacterium simiae* complex. *Int J Syst Evol Microbiol* **61**, 1606–1611.