

# Gene and whole genome analyses reveal that the mycobacterial strain JS623 is not a member of the species *Mycobacterium smegmatis*

Maria J. Garcia<sup>1\*</sup> and Susanne Gola<sup>2\*\*</sup>

<sup>1</sup>Universidad Autonoma de Madrid, Madrid, Spain.

<sup>2</sup>Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain.

## Summary

Unexpected differences were found between the genome of strain JS623, used in bioremediation studies, and the genome of strain mc<sup>2</sup>155, a model organism for investigating basic biology of mycobacteria. Both strains are currently assigned in the databases to the species *Mycobacterium smegmatis* and, consequently, the environmental isolate JS623 is increasingly included as a representative of that species in comparative genome-based approaches aiming at identifying distinctive traits of the different members of the genus *Mycobacterium*. We applied traditional molecular taxonomic procedures – inference of single and concatenated gene trees – to re-evaluate the membership of both strains to the same species, adopting the latest accepted cut-off values for species delimitation. Additionally, modern whole genome-based *in silico* methods were performed in a comprehensive molecular phylogenetic analysis of JS623 and other members of the genus *Mycobacterium*. These analyses showed that all relevant genome parameters of JS623 clearly separate this strain from *M. smegmatis*. The strain JS623 should be corrected as *Mycobacterium* sp. not only in the literature but, even more importantly, in the database entries, as inclusion of the genome wrongly attributed to the *M. smegmatis* species in comparative studies will result in misleading conclusions.

## Introduction

*Mycobacterium smegmatis* (Trevisan) Lehmann and Neumann 1899 is a typical environmental rapid-growing

Received 23 July, 2015; revised 14 October, 2015; accepted 17 October, 2015. For correspondence. \*E-mail mariaj.garcia@uam.es; Tel. +34 914972753; Fax +34 914975353. \*\*E-mail sgola@cnb.csic.es; Tel. +34 915854686; Fax +34 915854506.

*Microbial Biotechnology* (2016) 9(2), 269–274  
doi:10.1111/1751-7915.12336

mycobacterium (Kazda *et al.*, 2009) that occasionally causes infections in animals and humans (Vonmoos *et al.*, 1986; Bercovier and Vincent, 2001; Brown-Elliott and Wallace, 2002). It gained importance in the tuberculosis research field as a non-pathogenic model system for certain aspects of mycobacterial biology (Reyrat and Kahn, 2001; Shiloh and Champion, 2010) and as a working-horse for the development of new molecular genetic tools (e.g. Ehrh *et al.*, 2005; Boldrin *et al.*, 2010). The subsequent application of these tools in *Mycobacterium tuberculosis* greatly enhanced our knowledge on the pathogen in the recent past. Besides, *M. smegmatis* is also an interesting model on its own right. Its investigation stimulates (myco) bacterial research with basic discoveries, as the identification of a novel conjugation mechanism (Mizuguchi and Tokunaga, 1971; Gray *et al.*, 2013), and impacts on state-of-the-art technologies, as illustrated by the utilization of the *M. smegmatis* porin A in nanopore sequencing procedures (Derrington *et al.*, 2010). The use of *M. smegmatis* as referential non-pathogenic mycobacterium has been boosted by the arrival of whole genome sequence (WGS) technologies.

The type strain, *M. smegmatis* ATCC 19420, is an offspring of the ancient isolate of Alvarez and Tavel, who recovered the microorganism from smegma (<http://www.bacterio.net/mycobacterium.html>). In spite of its role as type strain (which means that it is the reference to which all candidate strains have to be compared with in order to determine whether they belong to the *M. smegmatis* species), the genome of ATCC 19420 has not been completed yet. On the contrary, the genome of several other *M. smegmatis* strains, such as mc<sup>2</sup>155 and JS623, has been sequenced recently. The former is the most widely used strain for genetic approaches within the genus because it is easily transformable with replicating plasmids, the latter is an environmental isolate from soil, used in bioremediation (Coleman and Spain, 2003; Jin and Mattes, 2008; Jin *et al.*, 2010). Strain JS623 became popular after publication of its complete genome sequence. It can be foreseen that these strains will be increasingly included in comparative genome-based studies in the mycobacterial field, both under genetic and environmental aspects.

In the context of a general, taxonomically motivated, approach to the genus, a preliminary study on the use of

**Table 1.** Databases that include JS623 labelled as *M. smegmatis*.

Database*	Web-site link to JS623 information
JGI	<a href="http://genome.jgi-psf.org/mycsm/mycsm.info.html">http://genome.jgi-psf.org/mycsm/mycsm.info.html</a>
KEGG	<a href="http://www.genome.jp/kegg-bin/show_organism?org=msa">http://www.genome.jp/kegg-bin/show_organism?org=msa</a>
NCBI	<a href="http://www.ncbi.nlm.nih.gov/genome/genomes/1026">http://www.ncbi.nlm.nih.gov/genome/genomes/1026</a>
PATRIC	<a href="https://www.patricbrc.org/portal/portal/patric/Genome?cType=genome&amp;cld=710686.3">https://www.patricbrc.org/portal/portal/patric/Genome?cType=genome&amp;cld=710686.3</a>
UniProt	<a href="http://www.uniprot.org/taxonomy/710686">http://www.uniprot.org/taxonomy/710686</a>

\* Joint Genome Institute (JGI), Kyoto Encyclopedia of Genes and Genomes (KEGG), National Center for Biotechnology Information (NCBI), Pathosystems Resource Integration Center (PATRIC), Universal Protein Resource (UniProt).

WGS data for the classification of mycobacteria was undertaken. In those studies, we stumbled upon discrepancies when comparing strain JS623 with the strain mc<sup>2</sup>155. That finding was unexpected, because both strains appear labelled as *M. smegmatis* in main molecular databases (Table 1).

At least two recent other studies, which followed the rationale to cover a broad range of mycobacterial genomes for comprehensive comparative analyses, included the strain JS623 and detected differences, compared with *M. smegmatis*, for single traits: the screening of mycobacterial genomes for prophages, identified an prophage-like element specific in JS623 (Fan *et al.*, 2014), and Osório and colleagues, interested in the distribution of L-lactate dehydrogenase genes among mycobacteria, found different numbers of *lldD* paralogues in JS623 compared with other *M. smegmatis* strains (N.S. Osório, pers. comm.).

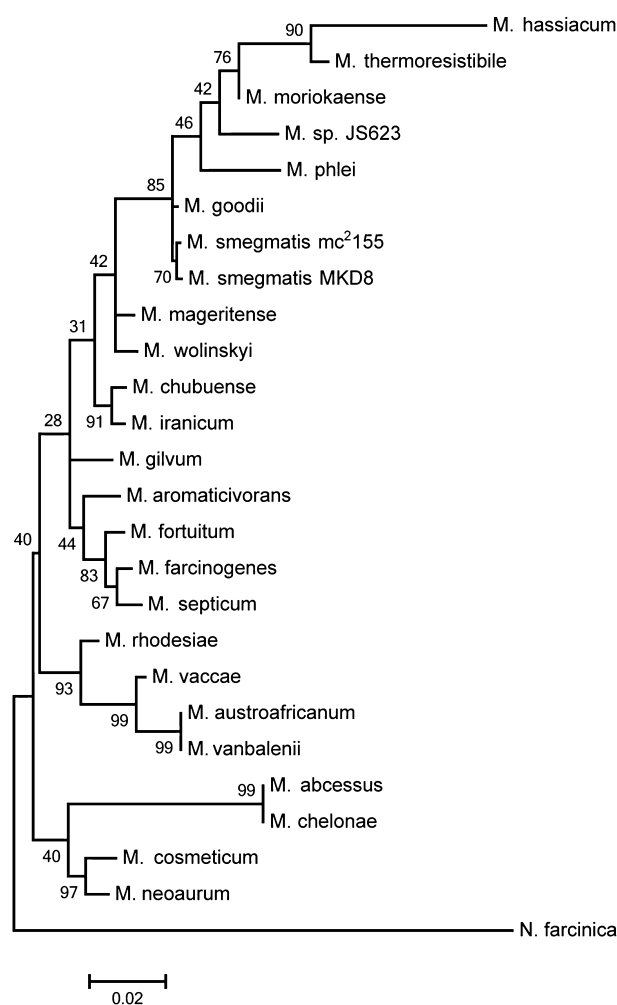
To find out whether the origins of the *M. smegmatis* strains with available WGS data could offer an explanation for the detected differences, the history of the different isolates was traced back (see Supporting Information). Yet, historical data did not allow a clear allocation of the strain JS623 as a member of the species *M. smegmatis*.

To clarify if the strain JS623 was or was not a member of the species *M. smegmatis*, phylogenetic relationships were inferred from genotypes by traditional approaches (16S rRNA gene and concatenated housekeeping genes similarities) and three WGS based methods.

## Results and discussion

### Gene-scale comparisons

The availability of genome sequences allows now also for mycobacteria to infer phylogenetic relationships based on the 16S rRNA gene that are not only sustained by the similarity of the hypervariable regions but the whole coding sequence, leading to more precise results (Tortoli, 2012; Yarza *et al.*, 2014).

**Fig. 1.** Molecular phylogenetic analysis based on the 16S rRNA coding sequence.

Numbers are the percentage of trees in which the associated taxa clustered together in the bootstrap analysis. The tree is drawn to scale and the bar length indicates 0.02 substitutions per site. See Supporting Information Table S1 for accession number of the 16S rRNA genes and text and Experimental Procedures in Supporting Information for details.

In order to cover a broad range of fast-growing *Mycobacterium* species, analysing at the same time the longest 16S rRNA coding region possible, the sequences from 22 representative rapid growers with approved names (<http://www.bacterio.net/mycobacterium.html>) were obtained from the NCBI genome database during October 2014. In cases when the genome has not been annotated yet, a 16S rRNA gene sequence was sought manually in the scaffolds or contigs (Supporting Information Table S1). Additionally, the 16S rRNA gene sequences from *M. goodii*, *M. moriokaense* and *M. wolinskyi* were obtained from GenBank (Supporting Information Table S2).

Figure 1 shows the relationship of strain JS623 to other rapid-growing mycobacteria, based on the analysis of the

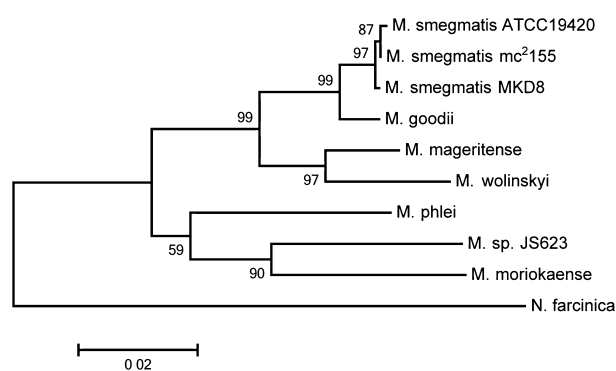
nearly full-length 16S rRNA coding sequence. Strain JS623 clusters within the group of the so-called thermo-tolerant rapid growers (Tortoli, 2003), but is clearly distant from the two other *M. smegmatis* strains included in the tree, as well as from their closest known species *M. goodii* (Brown *et al.*, 1999). Rather, that tree suggested that strain JS623 was related to *M. moriokaense*. This coincides with the observation made by Coleman and Spain, 2003, that strain JS623 is, based on 16S rRNA gene identity, closely related to another environmental isolate, strain JS619, which in turn was previously classified as being probably *M. moriokaense* (Coleman *et al.*, 2002).

The sequence of the 16S rRNA gene is frequently used for species identification in mycobacteria (Tortoli, 2012). However, employing it as the sole criterion for classification is insufficient as the same 16S rRNA gene sequence has been detected in different species (i.e. *M. gastri* and *M. kansasii*; Brown *et al.*, 1999) and, on the other hand, more than one sequence has been detected in the same genome of a single species (i.e. *M. celatum*; Menendez *et al.*, 2002). Thus, it is considered that the discriminative power of the 16S rRNA gene alone is limited, and more robust phylogenies can be obtained by building a tree with concatenated sequences of several housekeeping genes (Gadagkar *et al.*, 2005; Tortoli, 2012). We used this strategy to zoom into the clade comprising the thermo-tolerant non-pigmented species plus the thermo-tolerant pigmented *M. phlei*, as the last also seems to be related to strain JS623 on the basis of the 16S rRNA gene sequences (Fig. 1). To build the tree of concatenated genes, only partial sequences could be compiled of the 16S rDNA, *hsp65*, *rpoB* and *tuf* sequences from *M. smegmatis* candidate strains, including the type strain and selected thermo-tolerant rapid growers (Supporting Information Table S2). The *M. smegmatis* type strain could be included in this genotype analysis, the only method with sufficient available data for that strain.

The concatenated tree substantiates the phylogenetic distance of JS623 to *M. smegmatis* and, again, locates this strain closest to *M. moriokaense* and *M. phlei* (Fig. 2). Contrary to JS623, the other two strains labelled as *M. smegmatis* clustered as expected with the *M. smegmatis* type strain (Fig. 2). The species clustering derived from our concatenated tree agreed with other previously published concatenated trees of the genus *Mycobacterium* (Devulder *et al.*, 2005).

#### Genome-scale comparisons

Determining the relatedness of whole genomes should result in a better resolution of the phylogenetic relationship than single and multiple gene identities (Richter and Roselló-Mora, 2009). This principle has always been taken into account when determining whole genomic DNA



**Fig. 2.** Molecular phylogenetic analysis based on concatenation of partial 16S rRNA, *hsp65*, *rpoB* and *tuf* gene sequences. Numbers are the percentage of trees in which the associated taxa clustered together in the bootstrap analysis. The tree is drawn to scale and the bar length indicates 0.02 substitutions per site. See Supporting Information Table S2 for accession number of the genes and Experimental Procedures in Supporting Information for details.

relatedness by the traditional wet-lab DNA–DNA hybridization (DDH), which is still considered the gold-standard for species delimitation of bacteria, including the genus *Mycobacterium* (Goris *et al.*, 2007; Ramasamy *et al.*, 2014). Nowadays, WGS offer different and global approaches to support species delineation.

We have compared the sequences of three independent isolates that are listed in the databases as *M. smegmatis* and for which a WGS is available, i.e. strains JS623, mc<sup>2</sup>155 and MKD8. Three different procedures that deliver species identification-relevant parameters from WGS data were applied: Genome to Genome Distance Calculator (GGDC), Average Nucleotide Identity (ANI), and gene synteny. The genome of *M. phlei* was additionally included in the GGDC and ANI analyses.

GGDC (<http://ggdc.dsmz.de>) is a web-based procedure that was established to replace the so far gold standard for species delimitation, DDH, by *in silico* genome-to-genome comparison (dddH). Two genomes are considered to belong to the same species when their dDDH value is higher than 70%, which corresponded to the genome distance value (GD) lower than 0.0258 (Auch *et al.*, 2010; Meier-Kolthoff *et al.*, 2013).

The ANI procedure compares the nucleotide sequences of conserved shared genes between two genomes (Konstantinidis and Tiedje, 2005; <http://enve-omics.ce.gatech.edu/ani/>). The currently accepted cut-off value indicating that genomes derive from the same species is about 95% ANI value, corresponding to 70% DDH (Goris *et al.*, 2007).

Gene synteny analysis determines the order of genes in the genome. It is believed that changes in the gene order are accumulated during genome evolution and species delineation. Gene synteny analysis was performed with

**Table 2.** WGS comparison between *Mycobacterium* strain JS623 and other strains.

Strain	JS623			mc <sup>2</sup> 155		MKD8
	MKD8	mc <sup>2</sup> 155	Mph	MKD8	Mph	Mph
% dDDH – GD*	20.3% – 0.21	20.3% – 0.22	21.5% – 0.21	82.5% – 0.021	21.8% – 0.20	21.9% – 0.20
% ANI	81.8%	81.9%	80.1%	98.4%	80.4%	80.8 %
% GR	95%	94%	nd	0.24%	nd	nd

\* Percentage of digital DNA–DNA hybridization (dDDH) and correspondent genome distance (GD), percentage of Average Nucleotide Identity (ANI) and percentage of gene rearrangements (GR) between genomes of strain JS623, *M. smegmatis* mc<sup>2</sup>155, *M. smegmatis* MKD8 and *M. phlei* RIVM601174.

Mph, *Mycobacterium phlei*; nd, not done.

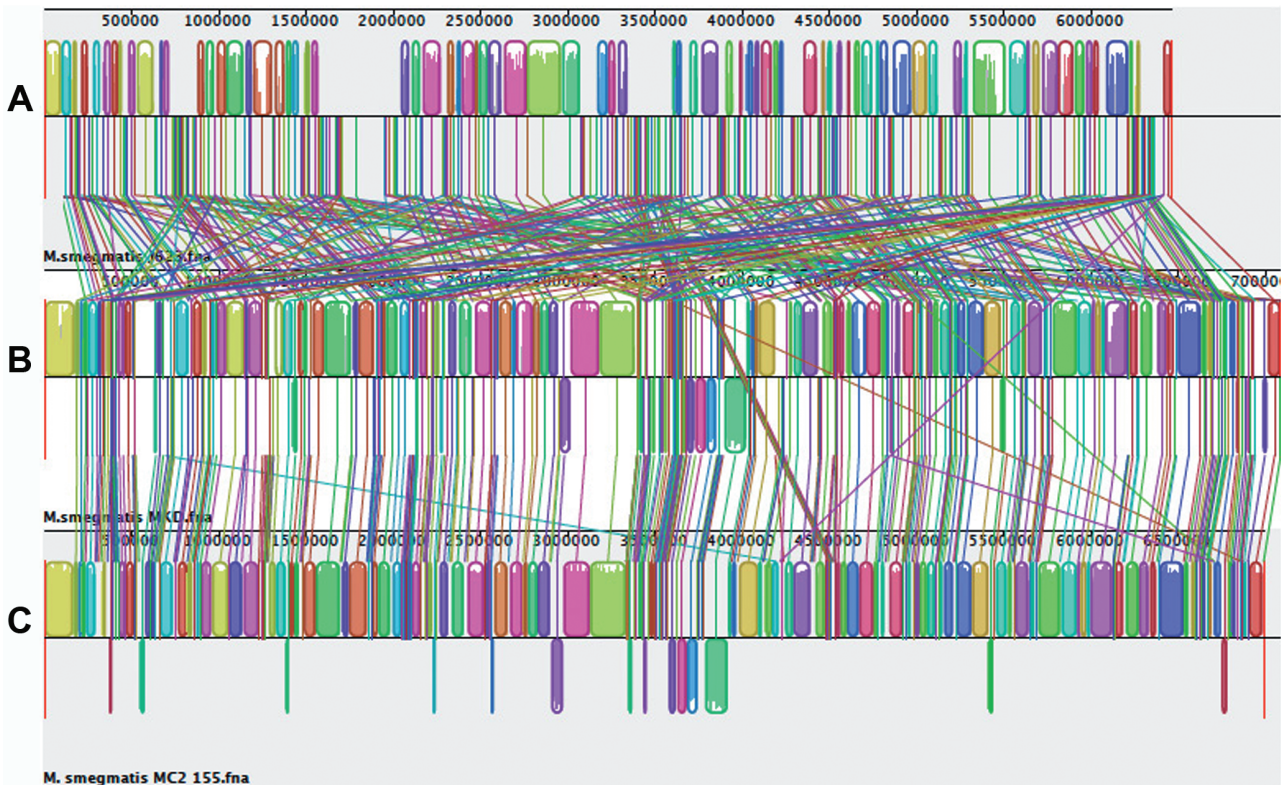
See Supporting Information Table S2 for accession numbers of the genomes and text and Experimental Procedures in Supporting Information for further details.

the open-source software MAUVE (Darling *et al.*, 2004) (<http://darlinglab.org/mauve/>). A value of more than 70% of gene rearrangement is considered to indicate that two genomes belong to different mycobacterial species (Garcia-Betancur *et al.*, 2012).

The three independent methods used for genome comparison showed a clear separation of JS623 from the other two *M. smegmatis* genomes (Table 2 and Fig. 3). According to the cut-off values for species delimitation, the strains MKD8 and mc<sup>2</sup>155 belong to the same

species, and are at the same time both members of a different species than strain JS623 (Table 2). A differentiation at the species level was also found after comparison with the genome of *M. phlei* (Table 2).

Strains of *M. smegmatis*, namely mc<sup>2</sup>155 and MKD8, showed some differences in their gene order (Table 2 and Fig. 3), concordant with their independent origin (see Supporting Information). Those differences do not include wide gene rearrangements (Table 2 and Fig. 3), which are usually found when comparing genomes from separate

**Fig. 3.** Gene alignment of genomes of the mycobacterial strains using MAUVE multiple alignments.

Homologous regions that align to segments of another genome are designated by coloured blocks connected by lines. Genomes of strains from top to bottom: (A) JS623; (B) MDK8; (C) mc<sup>2</sup>155.

See Supporting Information Table S2 for accession numbers of the genomes. See text and Experimental Procedures in Supporting Information for details.

mycobacterial species (Garcia-Betancur *et al.*, 2012). On the contrary, extensive segment shuffling was observed when genomes of these two strains were compared with the genome of JS623 (Fig. 3), therefore suggesting that the latter belongs to a different species than *M. smegmatis*.

## Conclusions

The first description of *Mycobacterium* strain JS623 included its preliminary classification as being most similar to *M. smegmatis*, based on the 96.7% identity value of a 421 bp 16S rRNA gene alignment (Coleman and Spain, 2003), and, consequently, the strain was considered in the literature and in several databases (Table 1) as a member of the species *M. smegmatis*. Upon description, the percentage was within the range of species delimitation (95–97%), however, according to the established and currently accepted values afterwards (Ramasamy *et al.*, 2014), that percentage is below the species limit of differentiation (98.7%). We reckoned, therefore, that JS623 was probably not a member of *M. smegmatis*. Our subsequently undertaken methodical gene and genome analyses showed that strain JS623 is a mycobacterium more related to *M. moriokaense* than to *M. smegmatis*, and indicate that it is not a member of this last species, as was previously believed. Consequently, and in order to avoid future confusions, the database entries should be corrected accordingly, and JS623 should be referred to as *Mycobacterium* sp. awaiting a formal species description.

The data presented in this work, derived from a multiple approach, underline the strength of genome comparisons to identify erroneously classified genome entries in the databases. Recently, Baez-Hidalgo *et al.* (2015) detected misclassification of deposited genomes within the genus *Aeromonas* by using a similar combined strategy.

In conclusion, systematic application of the novel *in silico* tools helps to avoid misleading conclusions, as, for example, those derived from comparative genome-based analyses of mycobacteria, when a genome wrongly attributed to a certain species was included in the analysis.

## Acknowledgements

This work has been partially supported by funds from the International Cooperation UAM-Banco Santander and Latin America (CEAL-UAM) and the Spanish Ministry of Health (AES, PI13-01218) (MJG) as well as BFU2014-52070-C2-1 from the Spanish Government (SG). We wish to thank M. Pavelka for helpful comments on the history of the MKD8 strain and N.S. Osório for the communication of unpublished results on conservation of *ltdD*.

## References

- Auch, A.F., Klenk, H.P., and Goker, M. (2010) Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* **2**: 142–148.
- Baez-Hidalgo, R., Hossain, M.J., Liles, M.R., and Figueras, M.-J. (2015) Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for *Aeromonas* genomes in the GenBank database. *PLoS ONE* **10**: e0115813.
- Bercovier, H., and Vincent, V. (2001) Mycobacterial infections in domestic and wild animals due to *Mycobacterium marinum*, *M. fortuitum*, *M. chelonae*, *M. porcinum*, *M. farcinogenes*, *M. smegmatis*, *M. scrofulaceum*, *M. xenopi*, *M. kansasii*, *M. simiae* and *M. genavense*. *Rev Sci Tech* **20**: 265–290.
- Boldrin, F., Casonato, S., Dainese, E., Sala, C., Dhar, N., Palu, G., *et al.* (2010) Development of a repressible mycobacterial promoter system based on two transcriptional repressors. *Nucleic Acids Res* **38**: e134.
- Brown, B.A., Springer, B., Steingrube, V.A., Wilson, R.W., Phyffer, G.E., Garcia, M.J., *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.
- Brown-Elliott, B.A., and Wallace, R.J., Jr (2002) Clinical and taxonomic status of pathogenic nonpigmented or late-pigmenting rapidly growing mycobacteria. *Clin Microbiol Rev* **15**: 716–746.
- Coleman, N.V., and Spain, J.C. (2003) Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading *Mycobacterium* strains. *Appl Environ Microbiol* **69**: 6041–6046.
- Coleman, N.V., Mattes, T.E., Gossett, J.M., and Spain, J.C. (2002) Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. *Appl Environ Microbiol* **68**: 6162–6171.
- Darling, A.C., Mau, B., Blattner, F.R., and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–1403.
- Derrington, I.M., Butler, T.Z., Collins, M.D., Manrao, E., Pavlenok, M., Niederweis, M., and Gundlach, J.H. (2010) Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci USA* **107**: 16060–16065.
- Devulder, G., Perouse de Montclos, M., and Flandrois, J.P. (2005) A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int J Syst Evol Microbiol* **55**: 293–302.
- Ehrt, S., Guo, X.V., Hickey, C.M., Ryou, M., Monteleone, M., Riley, L.W., and Schnappinger, D. (2005) Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res* **33**: e21.
- Fan, X., Xie, L., Li, W., and Xie, J. (2014) Prophage-like elements present in *Mycobacterium* genomes. *BMC Genomics* **15**: 243.
- Gadagkar, S.R., Rosenberg, M.S., and Kumar, S. (2005) Inferring species phylogenies from multiple genes:

- concatenated sequence tree versus consensus gene tree. *J Exp Zool B Mol Dev Evol* **304**: 64–74.
- Garcia-Betancur, J.C., Menendez, M.C., Del Portillo, P., and Garcia, M.J. (2012) Alignment of multiple complete genomes suggests that gene rearrangements may contribute towards the speciation of *Mycobacteria*. *Infect Genet Evol* **12**: 819–826.
- Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.
- Gray, T.A., Krywy, J.A., Harold, J., Palumbo, M.J., and Derbyshire, K.M. (2013) Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. *PLoS Biol* **11**: e1001602.
- Jin, Y.O., and Mattes, T.E. (2008) Adaptation of aerobic, ethene-assimilating *Mycobacterium* strains to vinyl chloride as a growth substrate. *Environ Sci Technol* **42**: 4784–4789.
- Jin, Y.O., Cheung, S., Coleman, N.V., and Mattes, T.E. (2010) Association of missense mutations in epoxyalkane coenzyme M transferase with adaptation of *Mycobacterium* sp. strain JS623 to growth on vinyl chloride. *Appl Environ Microbiol* **76**: 3413–3419.
- Kazda, J., Pavlik, I., Falkinham, J.O., III, and Hruska, K. (2009) *The Ecology of Mycobacteria: Impact on Animal's and Human's Health*, 2nd edn. Dordrecht, Heidelberg, London, New York: Springer.
- Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102**: 2567–2572.
- Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., and Goker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**: 60.
- Menendez, M.C., Garcia, M.J., Navarro, M.C., Gonzalez-y-Merchand, J.A., Rivera-Gutierrez, S., Garcia-Sanchez, L., and Cox, R.A. (2002) Characterization of an rRNA operon (*rrnB*) of *Mycobacterium fortuitum* and other mycobacterial species; implications for the classification of mycobacteria. *J Bacteriol* **184**: 1078–1088.
- Mizuguchi, Y., and Tokunaga, T. (1971) Recombination between *Mycobacterium smegmatis* strains Jucho and Lacticola. *Jpn J Microbiol* **15**: 359–366.
- Ramasamy, D., Mishra, A.K., Lagier, J.C., Padhmanabhan, R., Rossi, M., Sentausa, E., *et al.* (2014) A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. *Int J Syst Evol Microbiol* **64**: 384–391.
- Reyrat, J.M., and Kahn, D. (2001) *Mycobacterium smegmatis*: an absurd model for tuberculosis? *Trends Microbiol* **9**: 472–474.
- Richter, M., and Roselló-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.
- Shiloh, M.U., and Champion, P.A. (2010) To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? *Curr Opin Microbiol* **13**: 86–92.
- Tortoli, E. (2003) Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin Microbiol Rev* **16**: 319–354.
- Tortoli, E. (2012) Phylogeny of the genus *Mycobacterium*: many doubts, few certainties. *Infect Genet Evol* **12**: 827–831.
- Vonmoos, S., Leuenberger, P., Beer, V., and de Haller, R. (1986) [Pleuropulmonary infection caused by *Mycobacterium smegmatis*. Case description and literature review]. *Schweiz Med Wochenschr* **116**: 1852–1856.
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., *et al.* (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**: 635–645.

### Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Table S1.** Sources of 16S rRNA gene sequences.

**Table S2.** Sources of gene sequences used in the concatenated tree and genome sequences used in WGS comparisons.

**Appendix S1.** Historical approach to the *M. smegmatis* strains with available WGS data and Experimental Procedures.