INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

RESEARCH ARTICLE
Riojas *et al.*, *Int J Syst Evol Microbiol*
DOI 10.1099/ijsem.0.002507

MICROBIOLOGY SOCIETY

Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium microti* and *Mycobacterium pinnipedii* are later heterotypic synonyms of *Mycobacterium tuberculosis*

Marco A. Riojas,[1,*] Katya J. McGough,[1,2] Cristin J. Rider-Riojas,[3] Nalin Rastogi[4] and Manzour Hernando Hazbón[1]

Abstract

The species within the *Mycobacterium tuberculosis* Complex (MTBC) have undergone numerous taxonomic and nomenclatural changes, leaving the true structure of the MTBC in doubt. We used next-generation sequencing (NGS), digital DNA–DNA hybridization (dDDH), and average nucleotide identity (ANI) to investigate the relationship between these species. The type strains of *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium microti* and *Mycobacterium pinnipedii* were sequenced via NGS. Pairwise dDDH and ANI comparisons between these, previously sequenced MTBC type strain genomes (including '*Mycobacterium canettii*', '*Mycobacterium mungi*' and '*Mycobacterium orygis*') and *M. tuberculosis* H37Rv^T were performed. Further, all available genome sequences in GenBank for species in or putatively in the MTBC were compared to H37Rv^T. Pairwise results indicated that all of the type strains of the species are extremely closely related to each other (dDDH: 91.2–99.2 %, ANI: 99.21–99.92 %), greatly exceeding the respective species delineation thresholds, thus indicating that they belong to the same species. Results from the GenBank genomes indicate that all the strains examined are within the circumscription of H37Rv^T (dDDH: 83.5–100 %). We, therefore, formally propose a union of the species of the MTBC as *M. tuberculosis*. *M. africanum*, *M. bovis*, *M. caprae*, *M. microti* and *M. pinnipedii* are reclassified as later heterotypic synonyms of *M. tuberculosis*. '*M. canettii*', '*M. mungi*', and '*M. orygis*' are classified as strains of the species *M. tuberculosis*. We further recommend use of the infrasubspecific term 'variant' ('var.') and infrasubspecific designations that generally retain the historical nomenclature associated with the groups or otherwise convey such characteristics, e.g. *M. tuberculosis* var. bovis.

INTRODUCTION

The species within the *Mycobacterium tuberculosis* Complex (MTBC) have undergone numerous taxonomic and nomenclatural changes, leaving the true structure of the MTBC in doubt.

At the time of writing, the species within the MTBC with validly published names are *Mycobacterium tuberculosis* (also the type species of the genus), *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium caprae*, *Mycobacterium microti* and *Mycobacterium pinnipedii*, which are all very closely related [1, 2]. Even these species have undergone some taxonomic and nomenclatural changes. For example, the species most recently known as *M. caprae* was first proposed as *M. tuberculosis* subsp. *caprae* [3], then was later renamed as *M. bovis* subsp. *caprae* [4], and finally elevated to the rank of species [5]. These nomenclatural changes have resulted in equally valid basonyms for the same organism. Numerous other 'species' with similar properties have been identified but have not been

Author affiliations: [1]ATCC®, Manassas, VA 20110, USA; [2]George Mason University, Fairfax, VA 22030, USA; [3]Unaffiliated; [4]Institut Pasteur de la Guadeloupe, Les Abymes 97139, Guadeloupe, France.
***Correspondence:** Marco A. Riojas, mriojas@atcc.org
Keywords: *Mycobacterium tuberculosis* complex; *Mycobacterium tuberculosis*; digital DNA–DNA hybridization; average nucleotide identity; species reclassification.
Abbreviations: ANI, average nucleotide identity; dDDH, digital DNA–DNA hybridization; GGD, genome-to-genome distance; ICNP, International Code of Nomenclature of Prokaryotes; MTBC, *Mycobacterium tuberculosis* Complex; NGS, next-generation sequencing; RD, regions of difference.
The GenBank accession numbers for the whole genome draft sequences are as follows: *Mycobacterium tuberculosis* var. africanum ATCC® 25420™ - MWXF01.1, *Mycobacterium tuberculosis* var. bovis ATCC® 19210™ - MWXE01.1, *Mycobacterium tuberculosis* var. caprae ATCC® BAA-824™ - MWXD01.1, *Mycobacterium tuberculosis* var. microti ATCC® 19422™ - MWXC01.1, *Mycobacterium tuberculosis* var. pinnipedii ATCC® BAA-688™ - MWXB01.1.
Eight supplementary tables are available with the online version of this article.

officially accepted into bacterial nomenclature: '*Mycobacterium canettii*' [6], '*Mycobacterium mungi*' [7], and '*Mycobacterium orygis*' [8].

M. tuberculosis was first identified in 1882 [9] and publications describing species of the MTBC date as far back as 1957 for *M. microti* [10], 1969 for *M. africanum* [11], and 1970 for *M. bovis* [12]. While these species were no doubt characterized according to the best available methods of their respective generations, technology has clearly advanced considerably and now allows much greater analytical resolution, enabling the identification and delineation of species with greater accuracy. In particular, next-generation sequencing (NGS) and powerful bioinformatics tools allow the classification of species based upon the entirety of their genomes, rather than just a few potentially misleading phenotypic observations or even a small number of genomic loci (16S, *hsp65*, *rpoB*, etc.).

MTBC phylogeny is typically based on regions of difference (RDs) and SNPs [13]. Molecular strain typing methods currently used for the identification of species of MTBC include IS*6110*-RFLP, spoligotyping, mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU-VNTR), repetitive-sequence-based PCR (rep-PCR) and whole-genome sequencing [14]. By characterizing species/strains of the MTBC using molecular typing techniques, the species/strains have been categorized into various lineages [13, 15–17]. Although the development of such phylogenetic lineages may have important clinical and epidemiological applications, conflating what are likely host-adapted ecotypes [18, 19] with actual species may have a negative and confounding effect on mycobacterial systematics, particularly taxonomy and nomenclature.

It was suggested as early as 1982 that the typically accepted species within the MTBC may actually represent a single species [20]. Recent research comparing the genomes of species of the MTBC determined that whole-genome similarities support this idea. For example, the work of Garcia-Betancur, *et al.* compared *M. tuberculosis* H37Rv^T to nine other MTBC genomes and found the strains to be closely enough related to be considered as a single species. While their results are sound, their work compared non-type strains of members of the MTBC. Thus, their conclusion that 'mycobacterial scientists should agree an accord that designates MTBC as a single species in the official taxonomic rules of nomenclature' [21] is overstated. The applicability of their findings is inherently limited by the scope of the strains examined: their results indicate that those nine strains fall within the circumscription of *M. tuberculosis*. However, to suggest that the species *M. africanum*, *M. bovis*, and '*M. canettii*' – and, by extension, those specific epithets – should be consolidated into *M. tuberculosis* is premature. Such a conclusion could only be reached by comparing the type strains of the species considered. To our knowledge, such an examination of the type strains of the MTBC has not been reported prior to the current work.

The use of DNA–DNA hybridization (DDH) has been considered the gold standard for the genomic circumscription of bacterial species, with 70 % relatedness generally considered the threshold for species delineation [22]. However, the technique is labour-intensive, error-prone and poorly reproducible [23]. A recently developed *in silico* adaptation of DDH, digital DDH (dDDH), that allows the pairwise calculation of similarly-scaled genome-to-genome distances (GGDs) from whole-genome DNA sequences [23, 24] has been shown to be useful in species-level identification of bacterial strains, the identification of subspecies [25], and the development of detailed phylogenies for difficult taxa such as *Escherichia* [25] the *Bacillus cereus* Group [26] and *Aeromonas* [27].

The current work uses NGS and phylogenomic analysis based on dDDH and average nucleotide identity (ANI) to investigate the genomic coherence among type strains of the MTBC.

METHODS

Bacterial strains and DNA extraction

We obtained the following type strains of currently recognized species of the MTBC from ATCC: *M. africanum* ATCC 25420^T, *M. bovis* ATCC 19210^T, *M. caprae* ATCC BAA-824^T, *M. microti* ATCC 19422^T and *M. pinnipedii* ATCC BAA-688^T. DNA was extracted in BSL-3 conditions using BMBL-recommended safety precautions [28]. After verifying that the DNA extracts contained no viable organisms ($\geq$28 days), subsequent work was performed at BSL-2 conditions. Previously sequenced genomes from species in or putatively in the MTBC include *M. tuberculosis* H37Rv^T, '*M. canettii*' CIPT 140010059, '*M. mungi*' BM22813, and '*M. orygis*' 112400015. Because they are not validly published, no officially accepted type strains of '*M. canettii*', '*M. mungi*', and '*M. orygis*' exist. The aforementioned strains are the earliest identified strains or the earliest published strains for which whole-genome sequencing data was available. For the purposes of this work, these strains were treated as type strains. *M. pseudoshottsii* L15^T was used as an intermediate phylogenetic outlier, i.e. within *Mycobacterium* but outside the MTBC. *Nocardia asteroides* NBRC 15531^T was used as an extrageneric phylogenetic outlier. The list of accession numbers for the genomes used is shown in Table 1.

Next-generation sequencing

DNA was prepared using a KAPA Biosystems Hyper Prep Kit and sequenced using Illumina HiSeq 2500 (2$\times$100 bp). Sequencing reads were quality assessed with FastQC, data were filtered using Sickle, and *de novo* contigs were assembled using Velvet 1.2.10.

Calculation of genomic distance

MTBC type strain genomes sequenced during this work were combined with the previously sequenced genomes to compose the main dataset (Table 1). We calculated GGDs using dDDH via the Genome-to-Genome Distance

Table 1. Genomes of type (or treated as type) strains of the MTBC

Current organism name	Strain	Genome	Genome source
M. tuberculosis	H37Rv^T	NC_000962.3	GenBank
M. africanum	ATCC 25420^T	MWXF01.1	This Work
M. bovis	ATCC 19210^T	MWXE01.1	This Work
M. caprae	ATCC BAA-824^T	MWXD01.1	This Work
M. microti	ATCC 19422^T	MWXC01.1	This Work
M. pinnipedii	ATCC BAA-688^T	MWXB01.1	This Work
'*M. canettii*'	CIPT 140010059	NC_015848.1	GenBank
'*M. mungi*'	BM22813	LXTB01.1	GenBank
'*M. orygis*'	112400015	APKD01.1	GenBank
M. pseudoshottsii	L15^T	BCND01.1	GenBank
Nocardia asteroides	NBRC 15531^T	BAFO02.1	GenBank

Calculator (GGDC) v2.1 using the recommended Formula 2 [23, 24]. GGDs between *M. tuberculosis* H37Rv^T and the genomes published in GenBank were calculated using the same methods. The complete list of genomes analyzed is available together with the results in Tables S1–S6 (available in the online version of this article). In order to corroborate independently the dDDH results, we calculated GGDs via ANI between the type strains using OrthoANI v0.93 [29]. The species delineation thresholds used were 80 % for dDDH [25] and 96 % for ANI [30].

Phylogenomic analysis

For a better phylogenetic context, the whole genomes of the sequenced type strains were compared to each other and to the whole genome FASTAs from other species (n=12). After the GGD results from this larger pairwise whole-genome comparison were transformed into PHYLIP format, a phylogeny was inferred with FastME 2.0 using the BioNJ tree-building algorithm [31]. The resulting phylogenomic Newick tree was visualized using iTOL 3.5.3 and rooted with *Nocardia asteroides* NBRC 15531^T. The accession numbers for the genomes used in this analysis are provided in Fig. 1.

In silico spoligotyping and clade analysis

We calculated spoligotypes *in silico* for the available genomes of the strains of MTBC (limited to 10 strains of the species *M. tuberculosis*) using SpoTyping v2.0 [32]. The resulting spoligotypes were analyzed using the SITVITWEB database [33].

RESULTS

Genomic distances between the genomes of type strains

GGDs obtained from the analysis of the MTBC genomes are shown in Table 2. For the nine MTBC type (or treated as type) strains, the results from both pairwise comparison methods demonstrated that each strain is closely related to each of the other strains (dDDH: 91.2–98.9 %, ANI: 99.21–99.92 %), whereas genomic distances to the outgroups were far lower (dDDH: 18.8–22.3 %, ANI: 70.75–79.37 %). In all cases, the results from comparisons to *M. tuberculosis* H37Rv^T greatly exceed the respective dDDH or ANI species delineation thresholds, demonstrating that these all belong to the same species [22, 24, 25, 30]. Both sets of results show that the most distantly related member of the MTBC is '*M. canettii*', which is consistent with evidence suggesting that it is the likely progenitor from which the remaining MTBC members diverged [15].

Phylogenomic analysis

The phylogenetic tree inferred from the pairwise whole-genome GGDs is shown in Fig. 1. The results clearly show that the members of the MTBC form an extremely tight clade that is very distant from all of the other species, further supporting the hypothesis that they represent a single species.

Genomes in GenBank

Table 3 summarizes the dDDH GGDs calculated from GenBank genomes, illustrating that all the strains showed a high similarity (83.5–100 %) to *M. tuberculosis* H37Rv^T. (The complete data are shown in Tables S1–S6.) A single outlier represents a strain that has most likely been misclassified and is not included in Table 3; this strain was analyzed separately (see below).

In silico spoligotyping and clade analysis

The spoligotypes calculated for the strains of the MTBC and their SITVITWEB clade classification are shown in Table S7.

Outlier strain

When compared to *M. tuberculosis* H37Rv^T, *M. tuberculosis* TKK-01–0051 (JLXW01.1) provided GGDs indicating it is not a strain of *M. tuberculosis* (dDDH: 22.2 %, ANI: 79.39 %). We, therefore, compared this genome to a reference genome set composed of all the available type strains of species/subspecies of the genus *Mycobacterium* (n=81). These results demonstrate that strain TKK-01-0051 represents a strain of *M. colombiense* (dDDH: 81.3 %, ANI: 97.73 %; see Table S8).

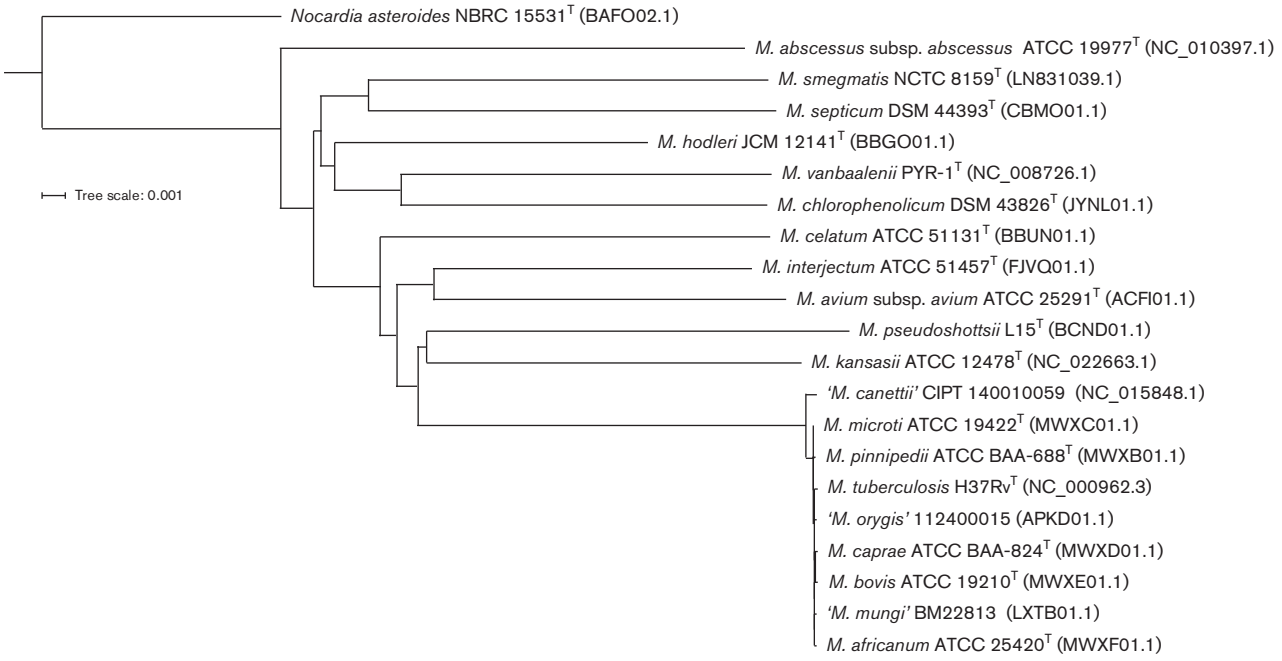

Fig. 1. Phylogenomic tree showing the relationship between the whole genomes of type strains of species of MTBC and the type strains of various other species of the genus *Mycobacterium* with *Nocardia asteroides* as an outgroup. Bar, 0.001 substitutions per site.

DISCUSSION

Pairwise GGDs for the nine type strains of the species in or putatively in the MTBC fall well within both the dDDH and ANI thresholds for delineation of bacterial species. It is also clear that (with the exception of a single misclassified strain) all genomes in GenBank deposited as *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. caprae*, *M. microti* and '*M. canettii*' are genomically within the circumscription of *M. tuberculosis*.

Thus, our analysis of all the 'species' of MTBC demonstrates clearly that taxonomically they actually represent a single species. While this is not the first work to challenge the MTBC species concept [20, 21], our work provides sufficient justification in the form of comprehensive whole-genome comparisons showing extremely high similarities between the type strains of the members. NGS-based phylogenomic analysis supports reclassification of all the species of the MTBC as a single species. Were such a reclassification to occur, Rule 42 of the International Code of Nomenclature of Prokaryotes (ICNP) [34] requires that the oldest validly published specific epithet be retained, in this case *M. tuberculosis*.

In their review of past and future definitions of species of bacteria, Roselló-Móra and Amann have recommended a species concept that would allow 'unequivocal identification' into categories that circumscribe 'monophyletic, and genomically and phenotypically coherent populations of individuals that can be clearly discriminated from other such entities by means of standardized parameters' [35]. Their analysis of genomes in the NCBI database shows some specific epithets applied to genomes are incorrect. Their results show a trimodal distribution: between the inter-species/intra-genus range and the intra-species range, the central mode represents a 'fuzzy zone' wherein organisms may represent either genomovars of the same species or different species based on the stability of differential phenotypes [35]. (Although their research used ANI as the measure of genomic similarity, a similar trend would be expected from data generated using dDDH.) Such phenotypes may be at odds with the genomic similarities observed. In this 'fuzzy zone', strains that should be considered a single 'genomospecies' (e.g. $\geq$80 % GGD) may still display stable differential phenotypes, essentially constituting different 'phenospecies'. One such example from the MTBC is '*M. canettii*', which phenotypically retains a smooth colony morphology unlike any of the other species of the MTBC, but is genomically well within the circumscription of *M. tuberculosis*.

There are defined characteristics (certain morphologies, SNPs, spoligotypes, etc.) that can differentiate the MTBC into various coherent lineages. At the same time, we observe that the overall genomic similarity within the MTBC is quite high – in most cases far higher than the 79–80 % dDDH GGD threshold that would delineate bacterial subspecies [25]. Thus, considering the currently accepted species of the MTBC as genomovars or host-adapted ecotypes of a single species represents something of a compromise between the

Table 2. Genomic Distance Analysis of type strains of members of the MTBC

Pairwise genome-to-genome distances (GGDs) for the type strains of the species within the *Mycobacterium tuberculosis* complex ('MTBC') and two outgroups ('OUT'). Results from digital DNA–DNA hybridization (dDDH) are shown above the self-comparison diagonal, and average nucleotide identity (ANI) results are shown below the diagonal. Species delineation thresholds used were 80 % for dDDH and 96 % for ANI.

	Species/Strain (ANI \ dDDH)	MTBC1	MTBC2	MTBC3	MTBC4	MTBC5	MTBC6	MTBC7	MTBC8	MTBC9	OUT1	OUT2
MTBC1	*M. tuberculosis* H37Rv^T	100	97.7	97.5	97.9	98.7	97.3	91.2	97.9	97.6	22.1	18.8
MTBC2	*M. africanum* ATCC® 25420^T	99.88	100	98	98.5	98.9	98.2	92.2	98.9	98.4	22.2	18.8
MTBC3	*M. bovis* ATCC® 19210^T	99.78	99.82	100	98.2	98.2	97.8	91.6	98.1	97.8	22.3	19
MTBC4	*M. caprae* ATCC® BAA-824^T	99.86	99.87	99.83	100	98.8	97.7	91.7	98.6	98.2	22.1	18.8
MTBC5	*M. microti* ATCC® 19422^T	99.89	99.89	99.79	99.91	100	99.2	93.3	99	98.9	22.3	18.9
MTBC6	*M. pinnipedii* ATCC® BAA-688^T	99.84	99.86	99.78	99.84	99.92	100	91.6	98.3	97.6	22.1	18.8
MTBC7	"*M. canettii*" CIPT 140010059	99.25	99.28	99.21	99.26	99.36	99.26	100	92.5	91.8	22.2	19
MTBC8	"*M. mungi*" BM22813	99.88	99.91	99.81	99.88	99.92	99.89	99.31	100	98.6	22.2	18.8
MTBC9	"*M. orygis*" 112400015	99.85	99.87	99.80	99.87	99.91	99.86	99.28	99.88	100	22.1	18.8
OUT1	*M. pseudoshottsii* L15^T	79.33	79.27	79.21	79.34	79.34	79.25	79.29	79.25	79.37	100	19.2
OUT2	*Nocardia asteroides* NBRC 15531^T	70.92	70.76	70.96	70.94	70.88	71.02	70.99	71.00	70.84	70.75	100

Color Key	
ANI	dDDH
Self-Comparison	Self-Comparison
96 – 100	80 – 100
< 96	< 80

classical/phenotypic species definitions and the reality of their global genomic similarities.

We, therefore, propose that the currently recognized species of the MTBC: *M. africanum* Castets *et al.* 1969 (Approved Lists 1980), *M. bovis* Karlson and Lessel 1970 (Approved Lists 1980), *M. caprae* (Aranaz *et al.* 1999) Aranaz *et al.* 2003, *M. microti* Reed 1957 (Approved Lists 1980), *M. pinnipedii* Cousins *et al.* 2003, and *M. tuberculosis* (Zopf 1883) Lehmann and Neumann 1896 (Approved Lists 1980) should be united as *M. tuberculosis*. Further, we propose that the taxa '*M. canettii*', '*M. mungi*', and '*M. orygis*', which have not yet been validly published, be similarly considered as later heterotypic synonyms of *M. tuberculosis*.

However, due to the stability of certain differential phenotypic and/or genomic characteristics within lineages, we recommend that the MTBC lineages be considered infrasubspecific subdivisions (i.e. variants) of *M. tuberculosis*. Although infrasubspecific subdivisions are not governed by the ICNP, it is useful to discuss some of the associated issues. With respect to the appropriate infrasubspecific term, the nature of the differential characteristics between the variants should inform its selection. Some variants, e.g. '*M. canettii*' and the 'smooth tuberculosis bacilli' (STB), may be identified according to unique morphologies [17] and would best be considered morphovars. Others may be considered biovars according to differential biochemical or physiological properties. In many cases, lineages are separated primarily based upon RDs, SNPs, and/or spoligotyping; these would best be considered genomovars. As a single such term does not appear to be appropriate between all the lineages, we recommend the general infrasubspecific term 'variant' ('var.'). With respect to the appropriate infrasubspecific designations, we recommend the use of designations that generally retain the structure of the current lineages, the historical nomenclature associated with the groups, or otherwise convey such characteristics (Table 4). Use of the recommended variant designations, e.g. *M. tuberculosis* var. bovis, will be useful in minimizing the confusion that might otherwise arise. As infrasubspecific subdivisions are not governed by the ICNP, no formal proposals are made in the current work regarding their usage, but we believe that the

Table 3. dDDH Analysis of type strains of members of the MTBC

dDDH GGD values are between the GenBank genomes identified as respective species and the type strain of *M. tuberculosis* (H37Rv^T).

GenBank organism identifier	*n*	dDDH GGD to *M. tuberculosis*, H37Rv^T (%)		
		Minimum	Mean	Maximum
M. africanum	30	96.7	97.5	98.3
M. bovis	69	95.7	97.8	99.1
'*M. canettii*'	9	80.1	89.8	94.1
M. caprae	2	97.4	97.7	97.9
M. microti	1	97.1	97.1	97.1
M. tuberculosis	3631*	83.5	98.6	100.0

*Excludes one single outlier which was examined separately (see text and Table S8).

Table 4. Recommended infrasubspecific designations and reference strains

Current Name	Recommended Name	Reference Strain
M. tuberculosis	*M. tuberculosis* var. tuberculosis	H37Rv^T (ATCC 27294^T) (type strain of species)
M. africanum	*M. tuberculosis* var. africanum	ATCC 25420
M. bovis	*M. tuberculosis* var. bovis	ATCC 19210
M. bovis BCG	*M. tuberculosis* var. BCG	–
M. caprae	*M. tuberculosis* var. caprae	ATCC BAA-824
M. microti	*M. tuberculosis* var. microti	ATCC 19422
M. pinnipedii	*M. tuberculosis* var. pinnipedii	ATCC BAA-688
'*M. canettii*'	*M. tuberculosis* var. canettii	CIPT 140010059
'*M. mungi*'	*M. tuberculosis* var. mungi	BM22813
'*M. orygis*'	*M. tuberculosis* var. orygis	112400015
'*M. suricattae*'*	*M. tuberculosis* var. suricattae	–
Dassie bacillus*	*M. tuberculosis* var. dassie	–
Chimpanzee bacillus*	*M. tuberculosis* var. chimpanzee	–

*These organisms were not included in this study, but recommended infrasubspecific designations (based upon the typically used nomenclature for these organisms) are included for completeness. Note that the strains that were previously listed as the species type strains are now considered reference strains of the respective varieties.

use of infrasubspecific subdivisions as described above would be ideal.

From a nomenclatural standpoint, it can be expected that the proposal to reclassify the species of the MTBC as *M. tuberculosis* might initially seem unnecessarily disruptive; however, it is not without precedent. Among the key principles of the International Code of Nomenclature of Prokaryotes is that the governing nomenclatural system should aim to bring stability to the names used in prokaryotic microbiology [34]. Accordingly, systematic changes as potentially major as those proposed in the current work have been infrequent, though not unheard of. One of the most significant such changes – which bears important parallels to the current work – occurred to the genus *Salmonella*. After many decades of assigning species based upon serotyping, *Salmonella* nomenclature had become increasingly disorganized and disjointed. This led various groups to propose that a reorganization of the genus was necessary. After the five species listed in the Approved Lists [36] were found to form a single coherent genomospecies in DNA hybridization experiments, it was proposed that the various species be consolidated into a single species with a previously unused specific epithet, *enterica* [37]. In 1999, it was further proposed that the previously recognized type species *S. choleraesuis* had become ambiguous and was a source of confusion and should be rejected; instead it was proposed that *S. enterica* become the neotype species [38]. By 2005, the commonly used nomenclature had become independent of that recognized by official nomenclatural guidelines; however, both were still in use, causing even greater confusion. This led the Judicial Commission of the International Committee for Systematics of Prokaryotes to issue Opinion No. 80 which established *Salmonella enterica* as the type species of the genus *Salmonella* and conserved the specific epithet *enterica* over all earlier epithets applied to the species [39]. One of the effects of these changes was to combine both pathogenic and nonpathogenic species of the genus *Salmonella* into a single subspecies, *S. enterica* subsp. *enterica*. The previously existing specific epithets however were retained as the designations for the various serovars, e.g. *Salmonella typhi*, the etiologic agent of typhoid fever, became *S. enterica* subsp. *enterica* serovar Typhi.

The taxonomic and nomenclatural changes made to the genus *Salmonella* are in some ways analogous to the changes to the *Mycobacterium tuberculosis* Complex proposed in the current work. Whole-genome sequencing has shown that the currently recognized species of the MTBC actually constitute a single genomospecies. Although this has been suspected taxonomically for several decades and proven for a variety of non-type strains, the current work has shown that the type strains of each of the species fall well within the circumscription of *M. tuberculosis*. So, the next prudent step is to align the relevant nomenclature with the taxonomy, in much the same way as with *Salmonella*.

As a final note, it is prudent to discuss the potential implications of changing the associated nomenclature to reflect better the genomic/taxonomic reality. According to Rule 56a of the ICNP, care should be exercised when considering 'a proposed change in the specific epithet of a nomenspecies that is widely recognized to be contagious, virulent, or highly toxigenic' [34]. The species currently named *M. africanum*, *M. bovis*, *M. caprae*, *M. microti*, and *M. pinnipedii* are pathogenic, and changing these specific epithets could hypothetically result in a perilous name (*nomen periculosum*). However, the specific epithet *tuberculosis* is even more well-known and associated with human disease. If all the species within the MTBC were reclassified as *M. tuberculosis*, it seems unlikely that the application of

this epithet would be 'likely to lead to accidents endangering health or life or both, or serious economic consequences' [34]. Changing the name of these species to *M. tuberculosis* – an extremely recognizable name that immediately conveys a significant biohazard – would in fact make this much less likely.

Another important benefit to the proposed nomenclatural scheme is that biosafety regarding the handling of these organisms is likely to be enhanced. For example, current International Air Transport Association (IATA) regulations for shipping infectious substances classify only certain bacterial species as Category A 'Infectious substances affecting humans' (UN 2814). This list of items includes only *M. tuberculosis* and excludes the other currently recognized species of the MTBC, despite their ability to cause essentially the same disease in humans and/or animals. The explicit interpretation of these regulations suggests that other members of the MTBC can be shipped under the less stringent 'Biological substance Category B' (UN 3373) regulatuions, which is inappropriate. Although the regulations specify that the list is not exhaustive and professional judgement should be used to assign infectious substances to Category A [40], this introduces a potential ambiguity that could result in avoidable infections. Reclassification of the existing species of the MTBC as *M. tuberculosis* would result in all such bacteria explicitly being assigned to Category A (whether or not the variant was specified), resulting in safer shipment and handling of MTBC-associated materials.

Finally, as a preemption of potential criticism, it is recognized that there are numerous important facets to the classification of organisms, e.g. clinical, epidemiological, phylogeographic, legal/biosafety risk groups and prokaryotic systematics. The present study was in no way intended to downplay the coherence of the currently accepted lineages, differential phenotypic or genotypic characteristics of various strains, or the importance of maintaining the concept of such groups. Rather, the discussion within the current work is primarily restricted to prokaryotic systematics and the application of the most modern technological and bioinformatical methodologies to the taxonomic and nomenclatural classification of bacteria of the MTBC. Uniting the MTBC as *M. tuberculosis* with the currently recognized lineages as variants allows the reality of their overall genomic similarity to guide their systematics (as it should), whilst simultaneously retaining classification according to specific characteristics that are of clinical, epidemiological, phylogeographic, or host preference relevance.

Emended description of *Mycobacterium tuberculosis* (Zopf 1883) Lehmann and Neumann 1896 (Approved Lists 1980)

M. africanum, *M. bovis*, *M. caprae*, *M. microti* and *M. pinnipedii* are reclassified as later heterotypic synonyms of *M. tuberculosis*. The strains described with the effectively but not validly published names '*M. canettii*' [6], '*M. mungi*' [7], and '*M. orygis*' [8] are likewise reclassified as belonging to the species *M. tuberculosis*. The phylogenetic groups that correspond to the previously named species should be considered infrasubspecific subdivisions, i.e. variants. Thus, the characteristics described for the above previously named species are now included in the description of *M. tuberculosis*. Although infrasubspecific subdivisions are not governed by the rules of prokaryotic nomenclature, it is recommended that these infrasubspecific subdivisions retain the previous specific epithet as the infrasubspecific designations, e.g. *M. tuberculosis* var. bovis (see Table 4).

Morphological, biochemical, genetic (e.g. spoligotypes, RD patterns), and host-preference characteristics of the respective variants are as previously described for the previously named species. Thus, the current differential characteristics remain unchanged, but their specificity is now applied at a lower taxonomic level (variety) than previously accepted (species).

M. tuberculosis has a broad host range. The known variants likely represent host-adapted ecotypes and generally correlate with host range, though most are generally also capable of causing human disease, particularly among immunocompromised individuals. *M. tuberculosis* var. tuberculosis, *M. tuberculosis* var. africanum, and *M. tuberculosis* var. canettii are typically isolated from humans. *M. tuberculosis* var. bovis is typically isolated from cattle, other bovids, or humans. *M. tuberculosis* var. caprae is typically associated with goats, *M. tuberculosis* var. microti is typically isolated from voles and other rodents, *M. tuberculosis* var. pinnipedii is typically isolated from marine mammals, *M. tuberculosis* var. mungi is typically isolated from mongooses, and *M. tuberculosis* var. orygis is typically isolated from antelope species (oryxes). Although each of these variants causes the disease tuberculosis in the affected host species, variants may cause little or no disease outside of their adapted host.

As the 16S rRNA sequences of the various *M. tuberculosis* variants are essentially identical, molecular methods with greater resolution must be used to differentiate the variants from each other. Such methods include MLST, spoligotyping, MIRU-VNTR, and whole-genome sequencing-based comparisons such as dDDH or ANI.

The type strain of *Mycobacterium tuberculosis* is H37Rv^T (=ATCC 27294^T=NCTC 13114^T). For the variants of *M. tuberculosis*, the genomes range in size from approximately 4.2–4.5 Mbp with a DNA G+C content of 65.0–65.6 mol%.

Funding information
The authors received no specific grant from any funding agency.

Acknowledgements
The authors wish to thank Kim Clem and Debra Hendrickson for DNA extraction, Stephen King for genome sequencing, Hunter Brigman for submission of the draft genome sequences into GenBank, and Yann Reynaud for constructive critique of the manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
No research was conducted on humans or animals.

References

1. Magee JG, Ward AC. Genus I. *Mycobacterium* Lehmann and Neumann 1896, 363^{AL}. In: Goodfellow M, Kampfer P, Busse HJ, Trujillo ME, Suzuki K *et al.* (editors). *Bergey's Manual of Systematic Bateriology*, 2nd ed, vol. 5: *The Actinobacteria, Part A*. New York: Springer; 2012. pp. 312–375.
2. Parte AC. LPSN–list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res* 2014;42:D613–D616.
3. Aranaz A, Liébana E, Gómez-Mampaso E, Galán JC, Cousins D *et al.* *Mycobacterium tuberculosis* subsp. *caprae* subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. *Int J Syst Bacteriol* 1999;49:1263–1273.
4. Niemann S, Richter E, Rüsch-Gerdes S. Biochemical and genetic evidence for the transfer of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz *et al.* 1999 to the species *Mycobacterium bovis* Karlson and Lessel 1970 (Approved Lists 1980) as *Mycobacterium bovis* subsp. *caprae* comb. nov. *Int J Syst Evol Microbiol* 2002;52:433–436.
5. Aranaz A, Cousins D, Mateos A, Domínguez L. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz *et al.* 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *Int J Syst Evol Microbiol* 2003;53:1785–1789.
6. Pfyffer GE, Auckenthaler R, van Embden JD, van Soolingen D. *Mycobacterium canettii*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. *Emerg Infect Dis* 1998;4:631–634.
7. Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD *et al.* Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis* 2010;16:1296–1299.
8. van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R *et al.* Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerg Infect Dis* 2012;18:653–655.
9. Koch R. Die äetiologie der Tuberkulose. *Berl klin Wchschr* 1882;19:221–230.
10. Reed G. Genus *Mycobacterium* (species affecting warm-blooded animals except those causing leprosy). In: Breed RS, Murray EGD and Smith NR (editors). *Bergey's Manual of Determinative Bacteriology*, 7th ed. Baltimore: Williams & Wilkins; 1957. pp. 703–704.
11. Castets M, Rist N, Boisvert H. La variété africaine du bacille tuberculeux humain. *Médecine d'Afrique Noire* 1969;16:321–322.
12. Karlson AG, Lessel EF. *Mycobacterium bovis* nom. nov. *Int J Syst Bacteriol* 1970;20:273–282.
13. Rodriguez-Campos S, Smith NH, Boniotti MB, Aranaz A. Overview and phylogeny of *Mycobacterium tuberculosis* complex organisms: implications for diagnostics and legislation of bovine tuberculosis. *Res Vet Sci* 2014;97:S5–S19.
14. Ei PW, Aung WW, Lee JS, Choi GE, Chang CL. Molecular strain typing of *Mycobacterium tuberculosis*: a review of frequently used methods. *J Korean Med Sci* 2016;31:1673–1683.
15. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA* 2002;99:3684–3689.
16. Shabbeer A, Cowan LS, Ozcaglar C, Rastogi N, Vandenberg SL *et al.* TB-lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. *Infect Genet Evol* 2012;12:789–797.
17. Coscolla M, Gagneux S. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin Immunol* 2014;26:431–444.
18. Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* 2007;17:R373–R386.
19. Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR *et al.* Ecotypes of the *Mycobacterium tuberculosis* complex. *J Theor Biol* 2006;239:220–225.
20. Wayne LG. Microbiology of tubercle bacilli. *Am Rev Respir Dis* 1982;125:31–41.
21. Garcia-Betancur JC, Menendez MC, Del Portillo P, Garcia MJ. Alignment of multiple complete genomes suggests that gene rearrangements may contribute towards the speciation of Mycobacteria. *Infect Genet Evol* 2012;12:819–826.
22. Wayne LG, Moore WEC, Stackebrandt E, Kandler O, Colwell RR *et al.* Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Evol Microbiol* 1987;37:463–464.
23. Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010;2:117–134.
24. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 2013;14:60.
25. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V *et al.* Complete genome sequence of DSM 30083^{T}, the type strain (U5/41^{T}) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 2014;9:2.
26. Liu Y, Lai Q, Göker M, Meier-Kolthoff JP, Wang M *et al.* Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci Rep* 2015;5:14082.
27. Colston SM, Fullmer MS, Beka L, Lamy B, Gogarten JP *et al.* Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. *MBio* 2014;5:e02136.
28. Chosewood LC, Wilson DE Centers for Disease Control and Prevention (U.S.), National Institutes of Health (U.S.). *Biosafety in Microbiological and Biomedical Laboratories (BMBL)*, 5th ed. Washington, DC: U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institutes of Health; 2009. pp. xxii, 415.
29. Lee I, Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.
30. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 2009;106:19126–19131.
31. Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* 2015;32:2798–2800.
32. Xia E, Teo YY, Ong RT. *SpoTyping*: fast and accurate *in silico* Mycobacterium spoligotyping from sequence reads. *Genome Med* 2016;8:19.
33. Demay C, Liens B, Burguière T, Hill V, Couvin D *et al.* SITVITWEB–a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 2012;12:755–766.
34. Parker CT, Tindall BJ, Garrity GM. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 2015.
35. Rosselló-Móra R, Amann R. Past and future species definitions for *Bacteria* and *Archaea*. *Syst Appl Microbiol* 2015;38:209–216.
36. Skerman VBD, Sneath PHA, Mcgowan V. Approved lists of bacterial names. *Int J Syst Evol Microbiol* 1980;30:225–420.
37. Le Minor L, Popoff MY. Designation of *Salmonella enterica* sp. nov., nom. rev., as the type and only species of the genus *Salmonella*: request for an opinion. *Int J Syst Bacteriol* 1987;37:465–468.
38. Euzéby JP. Revised *Salmonella* nomenclature: designation of *Salmonella enterica* (*ex* Kauffmann and Edwards 1952) Le Minor and Popoff 1987 sp. nov., nom. rev. as the neotype species of the genus *Salmonella* Lignieres 1900 (Approved Lists 1980), rejection of the name *Salmonella choleraesuis* (Smith 1894) Weldin 1927 (Approved

Lists 1980), and conservation of the name *Salmonella typhi* (Schroeter 1886) Warren and Scott 1930 (Approved Lists 1980). Request for an opinion. *Int J Syst Bacteriol* 1999;49:927–930.

39. **Judicial Commission of the International Committee on Systematics of Prokaryotes.** The type species of the genus *Salmonella* Lignieres 1900 is *Salmonella enterica* (*ex* Kauffmann and Edwards 1952) Le Minor and Popoff 1987, with the type strain LT2^T, and conservation of the epithet *enterica* in *Salmonella enterica* over all earlier epithets that may be applied to this species. Opinion 80. *Int J Syst Evol Microbiol* 2005;55:519–520.
40. IATA Dangerous Goods Regulations, 3.6.2.2.2.1 Category A. Section 3.6.2.2 Classification of infectious substances; 2015. https://www.iata.org/whatwedo/cargo/dgr/Documents/infectious-substance-classification-DGR56-en.pdf.