

---

# Novel *Mycobacterium tuberculosis* Complex Isolate from a Wild Chimpanzee

Mireia Coscolla, Astrid Lewin, Sonja Metzger, Kerstin Maetz-Rennsing, Sébastien Calvignac-Spencer, Andreas Nitsche, Piotr Wojtek Dabrowski, Aleksandar Radonic, Stefan Niemann, Julian Parkhill, Emmanuel Couacy-Hymann, Julia Feldman, Iñaki Comas, Christophe Boesch, Sebastien Gagneux,<sup>1</sup> and Fabian H. Leendertz<sup>1</sup>

Tuberculosis (TB) is caused by gram-positive bacteria known as the *Mycobacterium tuberculosis* complex (MTBC). MTBC include several human-associated lineages and several variants adapted to domestic and, more rarely, wild animal species. We report an *M. tuberculosis* strain isolated from a wild chimpanzee in Côte d'Ivoire that was shown by comparative genomic and phylogenomic analyses to belong to a new lineage of MTBC, closer to the human-associated lineage 6 (also known as *M. africanum* West Africa 2) than to the other classical animal-associated MTBC strains. These results show that the general view of the genetic diversity of MTBC is limited and support the possibility that other MTBC variants exist, particularly in wild mammals in Africa. Exploring this diversity is crucial to the understanding of the biology and evolutionary history of this widespread infectious disease.

**T**uberculosis (TB) is caused by closely related acid-fast bacteria known as the *Mycobacterium tuberculosis* complex (MTBC) (1). MTBC includes the typical

---

Author affiliations: Swiss Tropical and Public Health Institute, Basel, Switzerland (M. Coscolla, J. Feldman, S. Gagneux); University of Basel, Basel (M. Coscolla, J. Feldman, S. Gagneux); Centro Superior de Investigación en Salud Pública, Valencia, Spain (M. Coscolla, I. Comas); Robert Koch-Institut, Berlin, Germany (A. Lewin, S. Metzger, S. Calvignac-Spencer, A. Nitsche, P. Wojtek Dabrowski, A. Radonic, F.H. Leendertz); Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany (S. Metzger, C. Boesch); German Primate Center, Goettingen, Germany (K. Maetz-Rennsing); Research Centre Borstel, Borstel, Germany (S. Niemann); Wellcome Trust Sanger Institute, Cambridge, UK (J. Parkhill); Laboratoire Nationale de la Pathologie Animale, Bingerville, Côte d'Ivoire (E. Couacy-Hymann); and CIBER in Epidemiology and Public Health, Barcelona, Spain (I. Comas)

human-associated pathogens *M. tuberculosis* and *M. africanum* (2); *M. canettii* and other so-called “smooth TB bacilli” (3), the actual host range of which remains unknown; and several lineages adapted to different mammal species that include *M. bovis*, *M. microti*, *M. caprae*, *M. orygis*, and *M. pinnipedii* (4–6). Because of the wider host range of animal-associated MTBC, the common view until a decade ago was that human TB strains had evolved from *M. bovis*, the typical agent of bovine TB. Recent comparative genomic analyses have challenged this view by showing that animal MTBC strains nest within the genetically more diverse human MTBC strains (4,7–9). These results not only contradict the hypothesis of an animal origin for human MTBC but also promote an alternative scenario for a human origin of animal MTBC (10). However, little is known about MTBC diversity in domestic animals, and even less about MTBC diversity in wildlife, including our phylogenetically closest relatives, the great apes. Of note, novel members of MTBC affecting wild mammals in Africa have recently been discovered (11,12), a finding that suggests animal MTBC is more diverse than previously thought.

We report microbiologically confirmed MTBC infection in a wild chimpanzee. We show that this infection was caused by a divergent MTBC strain that does belong to the clade that includes *M. bovis* and all other animal-associated members of MTBC but is more closely related to human-associated lineage 6 (also known as *M. africanum* West Africa type 2 [WA2]). This finding highlights critical gaps in knowledge of MTBC diversity and indicates that African wildlife, and more particularly nonhuman primates, are potential hosts of novel MTBC variants.

---

<sup>1</sup>These authors contributed equally to this article.

## Materials and Methods

### Investigation of Wild Chimpanzee Death

In the course of a long-term study comprising behavioral observations and disease investigations of wild chimpanzees habituated to humans in Taï National Park, Côte d'Ivoire, necropsies are performed routinely on any chimpanzee or other mammal found dead. Detailed analyses are performed to identify the causes of death of every animal (13).

On August 5, 2009, an adult female chimpanzee of one of the study communities was found dead; lesions on the throat and alarm calls by other members of the community under observation at the time indicated that the animal had been killed by a leopard. The chimpanzee was one of the oldest females of the group (estimated age 52 years), and her body condition had deteriorated over the years. Necropsy was performed and tissue samples were frozen and fixed in formalin. Frozen organ material was submerged in 70% ethanol, rinsed twice in 0.85% NaCl, shredded with a scalpel, and streaked onto Löwenstein-Jensen PACT agar (Oxoid, Cambridge, UK). Bacteria were then cultivated on Middlebrook 7H11 agar supplemented with OADC or in Middlebrook 7H9 broth supplemented with OADC (Becton Dickinson, Franklin Lakes, NJ, USA) without shaking at 37°C.

### Investigation of MTBC in Other Chimpanzees

To investigate the possible presence of MTBC strains in other chimpanzees, samples were collected from 28 chimpanzees, many from the same community, that died in the same area within the previous 10 years. DNA was extracted from 115 tissue samples (lung, spleen, liver, lymph nodes, and small intestines) from these 28 chimpanzees by using the QIAGEN DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany). All tissues were tested in duplicate by using the primers MTC\_IAC Fw and MTC\_IAC Rv and MTC Probe as described (14). We performed the PCR-RFLP of *gyrB* using the primers MTUB-f (5'-TCGGACGCGTATGCGATATC-3') and MTUB-r (5'-ACATACAGTTCGGACTTGCG-3') and an annealing temperature of 65°C for the PCR. We used the DreamTaq DNA Polymerase Kit and Fermentas restriction enzymes (Thermo Scientific, Waltham, MA, USA).

### Genome Sequencing of MTBC Isolates

Mycobacterial DNA was isolated by using the CTAB method as described (15). The DNA was used to generate libraries for 454 and Illumina sequencing (Illumina, Inc., San Diego, CA, USA). For both libraries, the DNA was sheared to a size of 400–500 bp by using a Covaris S2 (Covaris, Inc., Woburn, MA, USA). The 454 library was generated by using the Rapid Library Kit and sequenced with

Titanium chemistry on a 454 FLX instrument (Roche, Penzberg, Germany). The paired-end library for Illumina sequencing was generated by using the TruSeq DNA Sample Preparation Kit (Illumina). Cluster generation was done by using TruSeq PE Cluster Kit version 2.5 (Illumina) on a c-bot. Sequencing was performed on a HiScanSQ instrument and TruSeq SBS Kit–HS chemistry (Illumina) to generate  $2 \times 100$  bases long paired-end reads.

Mycobacterial strains (as defined in [7,16,17]) were cultured from single colonies. Genomic DNA was extracted by using a standard kit (QIAGEN) and sequenced with an Illumina Genome Analyzer. Sequencing libraries were constructed by using standard kits from Illumina, according to the manufacturer's instructions. Libraries for each strain were loaded into a single lane of a flow cell. SYBR green assays were used to test flow cells for optimal cluster density.

### Single-nucleotide Polymorphism Calling and Genome Assembly

#### Illumina Sequencing Reads

We used BWA (18) to map Illumina reads from the 10 genome sequences published in this study ([www.ebi.ac.uk/ena/data/view/ERP001571](http://www.ebi.ac.uk/ena/data/view/ERP001571)) and 24 genomes published previously (18) or available in public databases (online Technical Appendix Table 1, [wwwnc.cdc.gov/EID/article/19/6/12-1012-Techapp1.pdf](http://wwwnc.cdc.gov/EID/article/19/6/12-1012-Techapp1.pdf)) against the MTBC reference genome. The reference genome used was an inferred common ancestor of all MTBC lineages (19). BWA outputs were analyzed with SAMtools (20). We applied heuristic filters to remove problematic positions and set Phred-scaled probability at 20. SNP lists for individual strains were combined in a single, nonredundant dataset, and the corresponding base call was recovered for each strain. After excluding single-nucleotide polymorphisms (SNPs) in genes annotated as PE/PPE, integrase, transposase, or phage and SNPs that showed an ambiguous base call, we kept 12,920 high-confidence variable positions for downstream analysis. Lineage 6 strains showed an average sequencing depth of between 80 and 204-fold whereas the chimpanzee strain was sequenced at 4428-fold coverage.

#### 454 Sequencing Reads

A combined mapping and de novo assembly was performed on the 454 reads obtained during the initial sequencing. Mapping of reads to the genome of *M. tuberculosis* strain CCDC5180 resulted in reference coverage of 98.39% and a total of 90 contigs. Mapping to the genome of strain H37Rv resulted in reference coverage of 98.3% and 87 contigs. Newbler 2.5 (Roche) and MIRA 3.0.0 (21) were used for de novo assembly. The parameters (minimum overlap identity and length, seed length, step and count, and alignment difference and identity scores for newbler; minimum

overlap identity and length; and clip length and stringency for MIRA) were automatically optimized for contig length by using a genetic algorithm with the default parameters used as starting points, a population size of 10, and 10 generations. The best set of parameters resulted in 2,538 contigs with a maximum length of 14,183 bp and a mean length of 1,738 bp. Reassembly of the contigs obtained from both mappings and from the de novo assemblies was performed by using Geneious 5.0 (Biomatters Ltd., Auckland, New Zealand) and yielded 33 contigs with a maximum length of 435,720 bp, a mean length of 130,500 bp, and a total length of 4,306,842 bp, compared with the ≈4.4 Mbp of the reference strains. The raw reads were mapped against these contigs to eliminate assembly errors.

A total of 33 contigs resulting from assembling 454 reads were aligned respective to the MTBC reconstructed ancestor genome using MAUVE (22). SNP lists obtained from Illumina sequencing were verified with the 454 contig sequences.

**Phylogenetic Analysis**

Phylogenetic analysis was performed on the basis of 13,480 high-confidence variable positions, specifying *M. canettii* as the outgroup (Figure 1). Both coding and noncoding SNPs were included. The SNPs were used to infer the phylogenetic relationships between strains by using neighbor-joining (Figure 1), maximum-likelihood (ML; online Technical Appendix Figure 2), and Bayesian (online Technical Appendix Figure 2) methods. Because of the low number of homoplasies expected (18), a neighbor-joining tree was obtained by using MEGA5 (23), with observed number of substitutions as a measure of genetic distance. We used the Akaike information criterion as implemented in jModelTest version 0.1 (24) to select the best-fit model of nucleotide substitution for the ML and Bayesian analyses. The ML tree was obtained by using PhyML version 3 (25), assessing branch robustness through bootstrapping (1,000 pseudo-replicates). The Bayesian summary tree was obtained by summarizing posterior tree samples generated along two 1 million generation-long Metropolis-coupled Markov chain Monte Carlo runs of 4 chains, which were performed in MrBayes version 3.1 (26). Convergence of the chains was assessed visually in Tracer version 1.5 (<http://tree.bio.ed.ac.uk/software/tracer/>), and all parameters were checked to have an effective sample size of ≥100 in the combined run. Branch robustness was assessed through their posterior probabilities (i.e., the proportion of trees in the posterior sample in which the considered branches appeared).

**Spoligotyping, Deletion, and Principal Component Analyses**

Spoligotyping was performed as described and compared with data published in SITVITWEB (27). Deleted

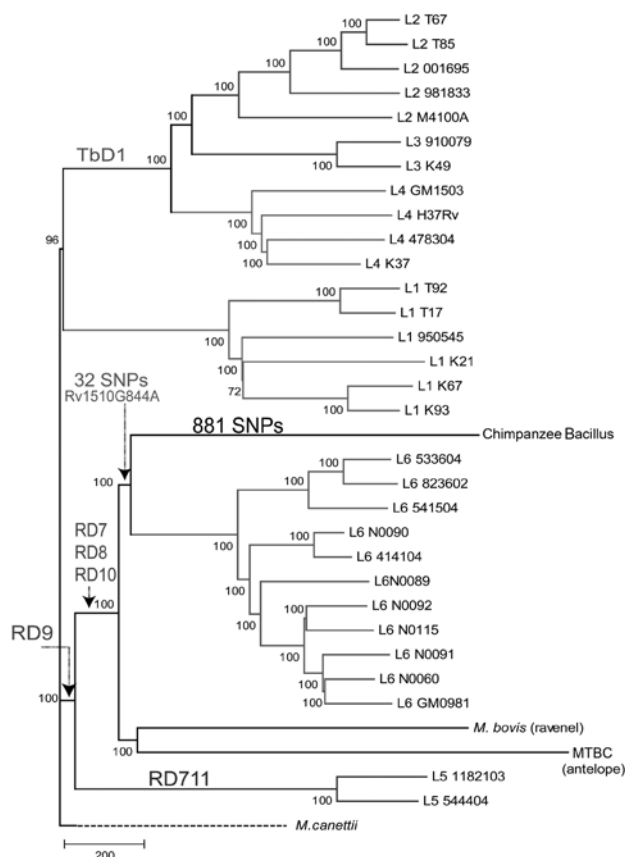


Figure 1. Neighbor-joining phylogenetic tree constructed on the basis of 13,480 variable common nucleotide positions across 36 human and animal *Mycobacterium tuberculosis* complex (MTBC) genome sequences, including 21 previously published genomes (18) and the MTBC strain isolated from an adult female chimpanzee that was found dead in Taï National Park, Côte d'Ivoire, on August 5, 2009 (Chimpanzee Bacillus). The tree is rooted with *M. canettii*, the closest known outgroup. Node support after 1,000 bootstrap replications is indicated. Genomic deletions identified in (7) are indicated. The number of single-nucleotide polymorphisms (SNPs) exclusive of the chimpanzee strain is indicated in the respective branch, and the number of SNPs shared with the most closely related group of strains is indicated in the common branch. Scale bar indicates number of SNPs. This tree is congruent with the maximum-likelihood phylogeny shown in Technical Appendix Figure 2 ([wwwnc.cdc.gov/EID/article/19/6/12-1012-Techapp1.pdf](http://wwwnc.cdc.gov/EID/article/19/6/12-1012-Techapp1.pdf)).

regions in the chimpanzee genome with respect to H37Rv genome were inferred as regions showing a mean coverage of <50 (1% mean coverage of the genome) by using awk scripts. Principal component analysis was conducted by using BioNumerics 6.6 ([www.applied-maths.com/bionumerics](http://www.applied-maths.com/bionumerics)) with the 13,480 high-confidence variable positions used for the phylogeny. The first 3 principal components accounted for 28%, 15%, and 8% of the variability and were used to generate a 3-dimensional scatter plot (online Technical Appendix Figure 4).

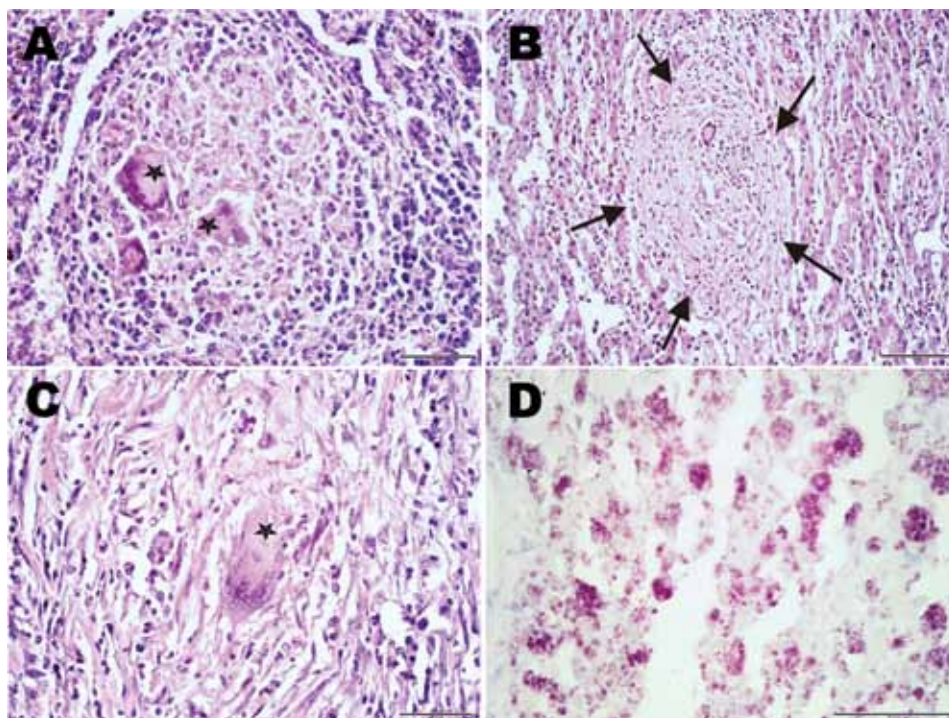


Figure 2. Histopathologic examination of tissue samples from adult female chimpanzee that was found dead in Taï National Park, Côte d'Ivoire, on August 5, 2009. A) Hematoxylin and eosin (H&E) stain of the spleen shows focal granulomatous inflammation with central accumulation of multinucleated Langhans giant cells (stars). B, C) H&E stain of the liver shows focal granulomatous inflammation within liver parenchyma (B, arrows) and large granulomatous alteration demarcated by fibrous connective tissue infiltrated by Langhans giant cells (C). D) Ziehl-Neelsen stain of the liver shows aggregates of acid-fast bacilli within a large granuloma. Results were consistent with *Mycobacterium tuberculosis* complex infection.

## Results

Necropsy of a wild chimpanzee found dead in Taï National Park, Côte d'Ivoire revealed a large, yellow-white granuloma of  $5 \times 6 \times 3$  cm in the liver and several smaller ones in the spleen parenchyma and the mesenteric lymph nodes. All other organs appeared unaffected; the lungs could not be evaluated in full because the leopard had consumed most of the tissue. 16S rDNA testing of the frozen tissue samples from the dead wild chimpanzee indicated the presence of a *Mycobacterium* sp. in various tissues. PCR–restriction fragment length polymorphism analysis confirmed MTBC in DNA preparations from spleen and mesenteric lymph nodes (28). MTBC was also confirmed by real-time PCR (29) in lung, spleen, liver, and colon abscesses.

Histopathologic examination of the liver, spleen, and lymph nodes revealed a chronic granulomatous inflammation within the altered tissues (Figure 2, panels A–C). Multiple unencapsulated granulomas of varying sizes were observed in the spleen (Figure 2, panel A) and liver (Figure 2, panel B). These lesions were composed of epithelioid macrophages, few granulocytes, and multinucleated Langhans giant cells. Larger tuberculoid lesions in the liver and the lymph nodes contained a prominent central necrotic core surrounded by epithelioid cells and a few scattered Langhans giant cells. The periphery of the granulomas was demarcated by variable amounts of fibrous connective tissue and infiltrates of lymphocytes interspersed with few Langhans giant cells (Figure 2, panel C). Both

intra- and extracellular acid-fast bacilli were present in the lesions (Figure 2, panel D). Taken together, these lesions were characteristic of TB and indicative of hematogenous spread and generalization of the disease.

After 23 days' incubation, the lymph node preparations yielded typical mycobacterial colonies. The isolated MTBC strain exhibited a slow growth on Middlebrook 7H11 agar and yielded colonies after 43 days, compared with 27 days for *M. tuberculosis*. The rough surface of the colonies and the irregular spreading margins were typical features of MTBC (online Technical Appendix Figure 1).

Whole-genome sequencing was conducted by using the 454 and Illumina platforms (online Technical Appendix Table 1). Phylogenetic reconstruction using previously published MTBC genomes representative of the MTBC's global diversity (18) confirmed that the chimpanzee strain belonged to MTBC but not to any of the known phylogenetic lineages (Figure 1). Specifically, the chimpanzee strain grouped with strains from the human-associated lineage 6, sharing 32 SNPs with this lineage, but was separate from the lineage leading to most of the animal-adapted MTBC. To further test whether the chimpanzee strain represented a new lineage rather than a variant within lineage 6, we sequenced the genome of 9 lineage 6 clinical strains from TB patients originating from different West-African countries (online Technical Appendix Table 1). Our phylogenomic analysis revealed that the chimpanzee strain harbored 881 exclusive SNPs, not found anywhere else in the global MTBC phylogeny, even when including these additional

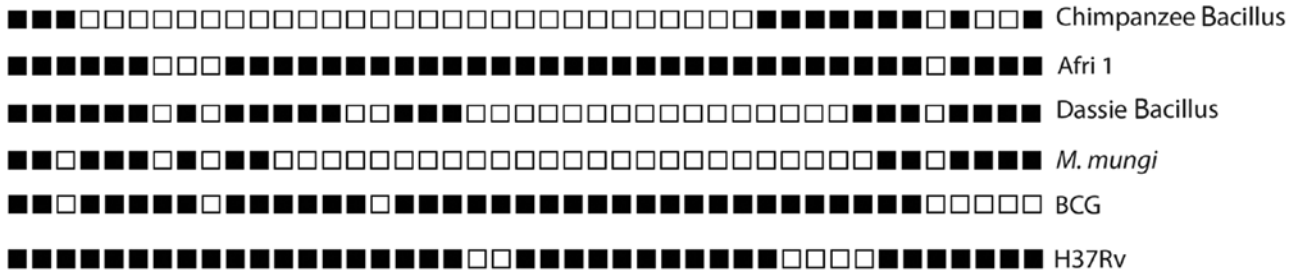


Figure 3. Comparison of the spoligotype of the *Mycobacterium tuberculosis* complex chimpanzee strain isolated from an adult female chimpanzee that was found dead in Tai National Park, Côte d'Ivoire, on August 5, 2009 (Chimpanzee Bacillus), with the Afri 1 spoligotype found in the most closely related human strain and the Dassie Bacillus and *M. mungi* spoligotypes described in (12). Spoligotypes are also shown for *M. bovis* strain BCG and human lineage 4 strain H37Rv.

lineage 6 strains. Moreover, the maximum number of SNPs between the 2 most divergent lineage 6 strains was only about half (783 SNPs) of the minimum number of differences between the chimpanzee strain and the most closely related lineage 6 strain (1,405 SNPs). When all MTBC lineages were considered, pairwise SNP distances among any 2 strains belonging to a particular lineage were always markedly lower than the minimum number of differences between the chimpanzee strain and the most closely related lineage 6 strain (online Technical Appendix Figure 3).

To investigate the distinctiveness of the chimpanzee strain we isolated, we used a principal component analysis as an additional clustering method. The first 3 principal components were used to generate a 3-dimensional scatter plot (online Technical Appendix Figure 4). This analysis confirmed that the chimpanzee strain did not group with lineage 6. Taken together, these results strongly suggest that the chimpanzee strain belongs to a distinct MTBC population, separate from the human-associated lineage 6.

To further confirm the uniqueness of the chimpanzee strain, we compared the chimpanzee spoligotype (Figure 3) with 2 large international databases that encompass >8,702 spoligotyping profiles corresponding to >58,187 MTBC isolates from global sources [33]; www.mbovis.org). We found that none of the spoligotyping patterns included in these databases (including 64 spoligotyping profiles from Côte d'Ivoire) matched the pattern of the chimpanzee strain (30) (Figure 3).

Large-sequence polymorphisms have been used as phylogenetic markers for MTBC (4,7,8). We found that the region of difference (RD) 9 was absent in the chimpanzee strain (4) (Figure 1). In addition, this strain harbored deletions in RD7, RD8, and RD10, which supports its phylogenetic relationship with lineage 6 (online Technical Appendix Table 2 and Figure 2). However, the chimpanzee strain harbored only 1 of 14 lineage 6-specific deletions and did not contain the lineage 6-specific region RD900 (online Technical Appendix Table 3) (29,31). Hence, this deletion-based analysis also supports a related, yet

separate, phylogenetic position of the chimpanzee strain relative to lineage 6.

To investigate the possible presence of MTBC in other chimpanzees, 115 tissue samples from 28 chimpanzees that died in the same area (many from the same community) within the previous 10 years were tested by real-time PCR (14). However, no test results were positive. Consistent with the molecular analyses, necropsies and pathological evaluation of these animals revealed no signs suggestive of TB.

**Discussion**

Chimpanzees are known to be susceptible to TB, and MTBC strains have been reported previously in captive chimpanzees (32,33). These 2 studies concluded that the infecting strains belonged to *M. africanum* and *M. tuberculosis*, respectively. Direct comparison of those isolates was not possible because of the limitations of typing techniques at that time. However, close contact with humans (i.e., the persons caring for these animals) suggests those captive chimpanzees were infected with human strains, as reported for other captive nonhuman primates.

By contrast, several lines of evidence support the view that the chimpanzee strain we report was not acquired from humans. First, its position on the MTBC phylogeny strongly suggests it belongs to a novel lineage. This notion is sustained by the fact that the minimum genetic distance between the chimpanzee strain and any of the nearest human strains (i.e., lineage 6) was larger than the corresponding distance between any 2 strains from the same lineage. Second, our PCA analysis showed that the chimpanzee strains did not group with lineage 6. Third, spoligotyping revealed a novel pattern among 58,187 clinical isolates from 102 countries, including Côte d'Ivoire. Fourth, genome deletion analyses corroborated the distinct phylogenetic position of the chimpanzee strain compared with known MTBC lineages. Fifth, researchers and their assistants who are in proximity of the chimpanzees at Tai National Forest are regularly screened for TB, but none has ever had a positive test result.

Two other animal-associated members of MTBC are known to cluster with lineage 6 rather than with the classical animal-adapted lineages: *M. mungii* and the Dassie Bacillus, which infect African mongooses and hyraxes, respectively (11,12). Whole genome analyses are not available for these organisms, but genomic deletion data have been reported (34). The chimpanzee strain we isolated did not harbor any of the specific deletions found in *M. mungii* or Dassie Bacillus (online Technical Appendix Table 2) (34). Moreover, spoligotyping confirmed that the chimpanzee strain was distinct from *M. mungii* and Dassie Bacillus and from any other MTBC strain genotyped to date (30). However, 1 of the 32 SNPs shared between the chimpanzee genome and the lineage 6 strains also occurred in *M. mungii* or Dassie Bacillus. On the basis of this 1 SNP in Rv1510, which has been reported before (5), one could hypothesize that the chimpanzee strain and *M. mungii* and Dassie Bacillus might be related. However, genome-wide data will be necessary to define the exact phylogenetic position of *M. mungii* and Dassie Bacillus, and their relationship with the chimpanzee strain, in the global MTBC tree.

Even though chimpanzees maintain close social contacts with other members of their group, extensive necropsies and molecular screening of 28 chimpanzees from the same region yielded no additional case of TB infection, which suggests TB is rare in this chimpanzee population. This low prevalence could have several explanations. While we can likely disregard a human origin of the chimpanzee strain described here, we cannot exclude the possibility that this strain was acquired from another unidentified animal host, including other primates; chimpanzees are known to hunt other animals, including monkeys and small antelopes. The chimpanzee strain we isolated shared  $\geq 1$  SNP with *M. mungii* and the Dassie Bacillus, which are pathogens of 2 other small African mammals. On a more speculative note, and if it is assumed that the MTBC strain described is indeed chimpanzee-specific, this MTBC variant might be relatively attenuated and only marginally affect a chimpanzee's health and longevity. This would enable sustained transmission and persistence of the pathogen in small host populations (35).

Although more work is needed to establish the prevalence, diversity, and clinical outcome of MTBC infection in wild chimpanzees and other great apes, from a conservation point of view, MTBC may join Ebola virus, *Bacillus cereus* biovar *anthracis*, and simian immunodeficiency viruses as a microorganism capable of threatening great apes in the wild (36). Our results suggest the effect of this MTBC strain on chimpanzee populations might be limited, but small outbreaks or single deaths can have a strong influence on the viability of isolated populations, particularly in great apes, which exhibit a slow reproductive rate and a high juvenile mortality rate (37).

Our study also sheds new light on the overall diversity of MTBC with implications for understanding the evolution of this pathogen. Together with other recent reports (11,12), our work suggests wider MTBC diversity, particularly among African mammals. Moreover, these data indicate that the theory that MTBC originally evolved as a human pathogen and jumped into animals is overly simplistic and may apply mainly to domestic animals. These data indicate 1 possible model could be that the common ancestor of MTBC was a generalist capable of infecting many mammals, including humans. From here, only few descendants spread around the world through human and animal migrations, creating the human-dominated phylogenetic picture we see today. We may well expect to find a much higher diversity of this MTBC, extending well outside the human-associated MTBC strains that infect various species, represented mainly through wildlife, including our closest relatives, the great apes. *M. canettii* and the other smooth TB bacilli show a high genetic diversity and are largely limited to the Horn of Africa, although whether these bacilli should be formally considered part of MTBC is controversial (3). Hence, they have been proposed to be part of the mycobacterial population that gave rise to the classical members of MTBC. Together with the apparent lack of human-to-human transmission of *M. canettii* (38), this suggestion would be consistent with a wider host range and/or environmental reservoir for the original ancestor of MTBC.

In conclusion, we report a microbiologically confirmed case of TB in a wild chimpanzee. Our molecular data show that the chimpanzee strain described here belongs to a novel lineage, more closely related to human-associated lineage 6 than to the other classical animal MTBC and possibly related to *M. mungii* and the Dassie Bacillus. This strain could represent a chimpanzee-specific pathogen or an MTBC variant acquired from another source. Because of our limited understanding of the ecology of this microbe, we propose at this stage to name it "Chimpanzee Bacillus" rather than to develop a dedicated species or subspecies name. Further studies are warranted, not only to better understand the natural history of TB in great apes and the biology of the Chimpanzee Bacillus, but also to estimate a possible risk for transmission of new types of MTBC to humans (e.g., through hunting and consumption of bushmeat). Moreover, further characterization of MTBC diversity will be crucial for understanding the origins of TB and the potential for the emergence of new strains through proximity between humans and wildlife.

#### Acknowledgments

We thank the Ivorian authorities for long-term support, especially the Ministry of the Environment and Forests and the Ministry of Research, the directorship of the Taï National Park,



and the Swiss Research Center in Abidjan. We also thank S. Becker for support in isolation of mycobacteria from chimpanzee tissue and U. Thiesen, K. Merkel, E. Kamal, and E. Radam for excellent technical assistance.

M.C. was supported by a fellowship from Programa VALi+d per a investigadors en fase postdoctoral de la Comunitat Valenciana. I.C. is supported by European Union funding from the Marie Curie FP7 actions (project number 272086). Work in S.G.'s laboratory is supported by the Swiss National Science Foundation (grant number PP0033-119205) and the National Institutes of Health (AI090928 and HHSN266200700022C). DNA sequencing was partially supported by core funding of the Wellcome Trust Sanger Institute.

Dr Coscollá is a postdoctoral researcher at the Department of Medical Parasitology and Infection Biology of the Swiss Tropical and Public Health Institute, Basel, Switzerland. Her research focuses on pathogen evolution and molecular epidemiology of infectious diseases using population genomics and molecular systematics.

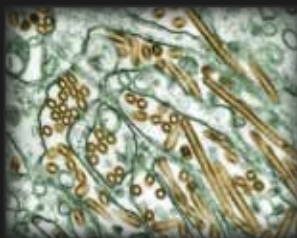
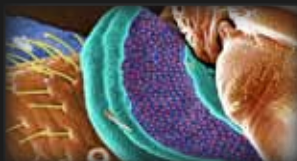
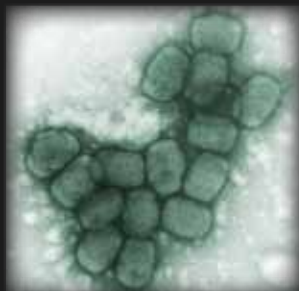
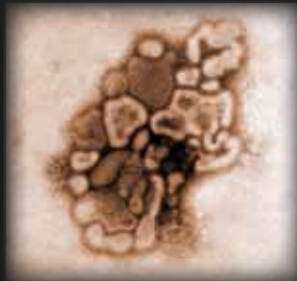
## References

- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393:537–44. <http://dx.doi.org/10.1038/31159>
- de Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*—review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis*. 2010;4:e744. <http://dx.doi.org/10.1371/journal.pntd.0000744>
- Gutierrez MC, Brisse S, Brosch R, Fabre M, Omas B, Marmiesse M, et al. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog*. 2005;1:e5. <http://dx.doi.org/10.1371/journal.ppat.0010005>
- Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*. 2002;99:3684–9. <http://dx.doi.org/10.1073/pnas.052548299>
- Huard RC, Fabre M, de Haas P, Claudio Oliveira Lazzarini L, van Soolingen D, Cousins D, et al. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis*. *J Bacteriol*. 2006;188:4271–87. <http://dx.doi.org/10.1128/JB.01783-05>
- van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, et al. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerg Infect Dis*. 2012;18:653–5. <http://dx.doi.org/10.3201/eid1804.110888>
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*. 2006;103:2869–73. <http://dx.doi.org/10.1073/pnas.0511240103>
- Mostowy S, Cousins D, Brinkman J, Aranaz A, Behr M. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *J Infect Dis*. 2002;186:74–80. <http://dx.doi.org/10.1086/341068>
- Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A*. 2003;100:7877–82. <http://dx.doi.org/10.1073/pnas.1130426100>
- Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, et al. Ecotypes of the *Mycobacterium tuberculosis* complex. *J Theor Biol*. 2006;239:220–5. <http://dx.doi.org/10.1016/j.jtbi.2005.08.036>
- Cousins DV, Peet RL, Gaynor WT, Williams SN, Gow BL. Tuberculosis in imported hyrax (*Procavia capensis*) caused by an unusual variant belonging to the *Mycobacterium tuberculosis* complex. *Vet Microbiol*. 1994;42:135–45. [http://dx.doi.org/10.1016/0378-1135\(94\)90013-2](http://dx.doi.org/10.1016/0378-1135(94)90013-2)
- Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, et al. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis*. 2010;16:1296–9. <http://dx.doi.org/10.3201/eid1608.100314>
- Leendertz FH, Pauli G, Maetz-Rensing K, Boardman W, Nunn C, Ellerbrok H, et al. Pathogens as drivers of population declines: the importance of systematic monitoring in great apes and other threatened mammals. *Biol Conserv*. 2006;131:325–37. <http://dx.doi.org/10.1016/j.biocon.2006.05.002>
- Reddington K, O'Grady J, Dorai-Raj S, Niemann S, van Soolingen D, Barry T. A novel multiplex real-time PCR for the identification of *Mycobacteria* associated with zoonotic tuberculosis. *PLoS ONE*. 2011;6:e23481. <http://dx.doi.org/10.1371/journal.pone.0023481>
- Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al. *Current protocols in molecular biology*. New York: Greene Publishing Associated and Wiley-Interscience; 1987.
- Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis*. 2007;7:328–37. [http://dx.doi.org/10.1016/S1473-3099\(07\)70108-1](http://dx.doi.org/10.1016/S1473-3099(07)70108-1)
- Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol*. 2008;6:e311. <http://dx.doi.org/10.1371/journal.pbio.0060311>
- Li H, Durbin R. Fast and accurate short read alignment with Burrows Wheeler transform. *Bioinformatics*. 2009;25:1754–60. <http://dx.doi.org/10.1093/bioinformatics/btp324>
- Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet*. 2010;42:498–503. <http://dx.doi.org/10.1038/ng.590>
- Li H, Handsaker B, Wysoker B, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9. <http://dx.doi.org/10.1093/bioinformatics/btp352>
- Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: *Computer science and biology: proceedings of the German Conference on Bioinformatics*. Braunschweig (Germany): GBF Braunschweig Department of Bioinformatics; 1999. p. 45–56.
- Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics*. 2009;25:2071–3. <http://dx.doi.org/10.1093/bioinformatics/btp356>
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24:1596–9. <http://dx.doi.org/10.1093/molbev/msm092>
- Posada D. jModelTest: Phylogenetic model averaging. *Mol Biol Evol*. 2008;25:1253–6. <http://dx.doi.org/10.1093/molbev/msn083>
- Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res*. 2005;33:W557–9. <http://dx.doi.org/10.1093/nar/gki352>

26. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4. <http://dx.doi.org/10.1093/bioinformatics/btg180>
27. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. 1997;35:907–14.
28. Goh KS, Fabre M, Huard RC, Schmid S, Sola C, Rastogi N. Study of the *gyrB* gene polymorphism as a tool to differentiate among *Mycobacterium tuberculosis* complex subspecies further underlines the older evolutionary age of *Mycobacterium canettii*. *Mol Cell Probes*. 2006;20:182–90. <http://dx.doi.org/10.1016/j.mcp.2005.11.008>
29. Bentley SD, Comas IÁ, Bryant JM, Walker D, Smith NH, Harris SR, et al. The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. *PLoS Negl Trop Dis*. 2012;6:e1552. <http://dx.doi.org/10.1371/journal.pntd.0001552>
30. Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, et al. SITVITWEB—a publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol*. 2012;12:755–66. <http://dx.doi.org/10.1016/j.meegid.2012.02.004>
31. Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic analysis distinguishes *Mycobacterium africanum*. *J Clin Microbiol*. 2004;42:3594–9. <http://dx.doi.org/10.1128/JCM.42.8.3594-3599.2004>
32. Thorel MF. Isolation of *Mycobacterium africanum* from monkeys. *Tubercle*. 1980;61:101–4. [http://dx.doi.org/10.1016/0041-3879\(80\)90018-5](http://dx.doi.org/10.1016/0041-3879(80)90018-5)
33. Michel AL, Venter L, Espie IW, Coetzee ML. *Mycobacterium tuberculosis* infections in eight species at the national zoological gardens of South Africa, 1991-2001. *J Zoo Wildl Med*. 2003;34:364–70. <http://dx.doi.org/10.1638/02-063>
34. Mostowy S, Cousins D, Behr MA. Genomic interrogation of the Dassie Bacillus reveals it as a unique RD1 mutant within the *Mycobacterium tuberculosis* complex. *J Bacteriol*. 2004;186:104–9. <http://dx.doi.org/10.1128/JB.186.1.104-109.2003>
35. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature*. 2007;447:279–83. <http://dx.doi.org/10.1038/nature05775>
36. Calvignac-Spencer S, Leendertz SAJ, Gillespie TR, Leendertz FH. Wild great apes as sentinels and sources of infectious disease. *Clin Microbiol Infect*. 2012;18:521–7. <http://dx.doi.org/10.1111/j.1469-0691.2012.03816.x>
37. Boesch C, Boesch-Achermann H. The chimpanzees of the Taý Forest: behavioural ecology and evolution. Oxford/New York: Oxford University Press; 2000.
38. Koeck JL, Fabre M, Simon F, Daffe M, Garnotel E, Matan AB, et al. Clinical characteristics of the smooth tubercle bacilli *Mycobacterium canettii* infection suggest the existence of an environmental reservoir. *Clin Microbiol Infect*. 2011;17:1013–9. <http://dx.doi.org/10.1111/j.1469-0691.2010.03347.x>

Address for correspondence: Fabian H. Leendertz, Emerging Zoonoses, Robert Koch-Institut, Nordufer 20, 13353 Berlin, Germany; email: [leendertzf@rki.de](mailto:leendertzf@rki.de)

## The Public Health Image Library (PHIL)



The Public Health Image Library (PHIL), Centers for Disease Control and Prevention, contains thousands of public health-related images, including high-resolution (print quality) photographs, illustrations, and videos.

PHIL collections illustrate current events and articles, supply visual content for health promotion brochures, document the effects of disease, and enhance instructional media.

PHIL Images, accessible to PC and Macintosh users, are in the public domain and available without charge.

Visit PHIL at <http://phil.cdc.gov/phil>.



# Novel *Mycobacterium tuberculosis* Complex Isolate from a Wild Chimpanzee

## Technical Appendix

Technical Appendix Table 1. Sequences of *Mycobacterium tuberculosis* complex isolates from previous studies compared with isolates from this study

Isolate identification	Origin study	Lineage	GenBank accession no.
T92	(1)	L1	SRX003589
T17	(1)	L1	SRX005394
95_0545	(1)	L1	SRX007721
K21	(1)	L1	SRX002001
K67	(1)	L1	SRX002004
K93	(1)	L1	SRX002005
T67	(1)	L2	SRX007715
T85	(1)	L2	SRX003590
00_1695	(1)	L2	SRX007716
98_1833	(1)	L2	SRX007718
M4100A	(1)	L2	SRX007719
91_0079	(1)	L3	SRX007720
K49	(1)	L3	SRX002002
GM_1503	(1)	L4	SRX012272
4783_04	(1)	L4	SRX007723
K37	(1)	L4	SRX002003
544404	(1)	L5	SRX007725
11821_03	(1)	L5	SRX007724
4141_04	(1)	L6	SRX007726
GM_0981	(1)	L6	SRX007722
<i>M. canettii</i>	(1)	Outgroup	SRX002429
<i>M. tuberculosis</i> complex/ <i>M. orygis</i> *	SRA EBI	Animal	ERR015582
<i>M. bovis</i> (ravenel)	SRA EBI	Animal	SRR022532
541504	This study	L6	ERR031459
N0092	This study	L6	ERR031480
N0115	This study	L6	ERR031484
N0091	This study	L6	ERR031479
533604	This study	L6	ERS153831
823602	This study	L6	ERS153832
N0060	This study	L6	ERS125978
N0089	This study	L6	ERS125979
N0090	This study	L6	ERS125980
H37Rv	This study	L4	ERS153830
Chimpanzee bacillus	This study		Pending
Chimpanzee bacillus	This study		Pending

\*Isolate from an antelope, recently *M. orygis* proposed for bacilli isolated from members of the Bovidae family, i.e., oryxes, gazelles, deer, antelope, and waterbucks (2). SRA EBI, European Bioinformatics Institute ([www.ebi.ac.uk](http://www.ebi.ac.uk)).

Technical Appendix Table 2. Genomic deletions respect to H37Rv genome inferred by more than 100-fold decreased coverage in the chimpanzee bacillus genome

Genome positions	Mean coverage in genome	Homologous genomic feature in H37Rv genome <sup>a</sup>	Genomic feature	Overlapping deletion already described
265477..265506	0	*IG223_Rv0221-Rv0222		RD10†
265507..266295	0	Rv0222	enoyl-CoA hydratase	
266296..266300	0	*IG224_Rv0222-Rv0223c		
889072..889398	1.6875	Rv0795	transposase IS6110 phage	
889395..890333	0	Rv0796	transposase IS6110 phage	
1541994..1542878	0	Rv1369c	transposase IS6110 phage	
1542879..1542928	0	*IG1392_Rv1369c-Rv1370c		
1542929..1543255	1.96429	Rv1370c	transposase IS6110 phage	
1606147..1606385	0	*IG1452_Rv1429-Rv1430		
1718726..1719970	1.4898	Rv1524	glycosyltransferase	
1719971..1720016	0	*IG1552_Rv1524-Rv1525		
1720017..1720802	0	Rv1525	<i>wbbL2</i>	
1720780..1722060	0	Rv1526c	glycosyltransferase	
1722061..1722082	0	*IG1554_Rv1526c-Rv1527c		
1779299..1779313	0	*IG1600_Rv1572c-Rv1573		RD3†
1779314..1779724	1	Rv1573	phiRV1 phage	
1779725..1779929	0	*IG1601_Rv1573-Rv1574		
1779930..1780241	0	Rv1574	phiRV1 phage	
1780199..1780699	0	Rv1575	phiRV1 phage	
1780643..1782064	0	Rv1576c	phiRV1 phage	
1782065..1782071	0	*IG1604_Rv1576c-Rv1577c		
1782072..1782584	1	Rv1577c	phiRV1 phage	
1782585..1782757	1	*IG1605_Rv1577c-Rv1578c		
1782758..1783228	1	Rv1578c	phiRV1 phage	
1783229..1783308	0	*IG1606_Rv1578c-Rv1579c		
1783309..1783623	0	Rv1579c	phiRV1 phage	
1783620..1783892	0	Rv1580c	phiRV1 phage	
1783893..1783905	0	*IG1608_Rv1580c-Rv1581c		
1783906..1784301	0	Rv1581c	phiRV1 phage	
1784302..1784496	0	*IG1609_Rv1581c-Rv1582c		
1784497..1785912	0	Rv1582c	phiRV1 phage	
1785912..1786310	0	Rv1583c	phiRV1 phage	
1786307..1786528	0	Rv1584c	phiRV1 phage	
1786529..1786583	0	*IG1612_Rv1584c-Rv1585c		
1786584..1787099	0	Rv1585c	phiRV1 phage	
1787099..1788505	0	Rv1586c	phiRV1 phage	
1987745..1988629	0	Rv1756c	putative transposase	
1988630..1988679	0	*IG1786_Rv1756c-Rv1757c		
1988680..1989006	1.8169	Rv1757c	putative transposase	
1989007..1989041	1	*IG1787_Rv1757c-Rv1758		
1996479..1996528	0	*IG1793_Rv1763-Rv1764		
1996529..1997413	0	Rv1764	putative transposase	
1997414..1997417	0	*IG1794_Rv1764-Rv1765c		
2208498..2208506	0	*IG1995_Rv1964-Rv1965		RD7†
2208507..2209322	0	Rv1965	<i>yrbE3b</i>	
2209323..2209326	0	*IG1996_Rv1965-Rv1966		
2209327..2210604	0	Rv1966	<i>mce3A</i>	
2210601..2211629	0	Rv1967	<i>mce3b</i>	
2211626..2212858	0	Rv1968	<i>mce3C</i>	
2212855..2214126	0	Rv1969	<i>mce3D</i>	
2214123..2215256	0	Rv1970	<i>lprM</i>	
2215257..2216570	0	Rv1971	<i>mce3F</i>	
2216571..2216591	0	*IG2002_Rv1971-Rv1972		
2216592..2217167	0	Rv1972	<i>mce</i> associated membrane protein	
2217164..2217646	0	Rv1973	<i>mce</i> associated membrane protein	
2217647..2217658	0	*IG2004_Rv1973-Rv1974		
2217659..2218036	0	Rv1974	hypothetical protein	
2218037..2218051	0	*IG2005_Rv1974-Rv1975		
2218052..2218717	0	Rv1975	hypothetical protein	
2218718..2218843	0	*IG2006_Rv1975-Rv1976c		
2218844..2219251	0	Rv1976c	hypothetical protein	
2219252..2219753	0	*IG2007_Rv1976c-Rv1977		

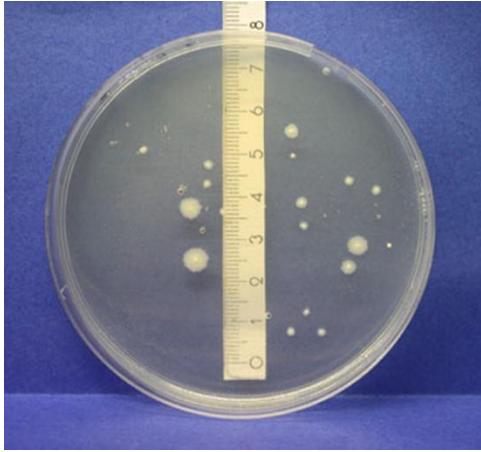
Genome positions	Mean coverage in genome	Homologous genomic feature in H37Rv genome <sup>a</sup>	Genomic feature	Overlapping deletion already described
2330147..2330213	1	*IG2104_Rv2072c-Rv2073c		RD9 <sup>†</sup>
2330214..2330963	0	Rv2073c	shortchain dehydrogenase	
2330964..2330992	0	*IG2105_Rv2073c-Rv2074		
2330993..2331406	0	Rv2074	hypothetical protein	
2331407..2331415	0	*IG2106_Rv2074-Rv2075c		
2365465..2365791	0	Rv2105	transposase phage	
2365788..2366726	1	Rv2106	transposase phage	
2430159..2431199	1	Rv2167c	transposase phage	
2431094..2431420	0	Rv2168c	transposase phage	
2550030..2550064	0	*IG2310_Rv2277c-Rv2278		
2550065..2550391	0	Rv2278	transposase phage	
2550388..2551326	0	Rv2279	transposase phage	
2635593..2635627	0	*IG2392_Rv2353c-Rv2354		
2635628..2635954	1	Rv2354	transposase phage	
2635951..2636889	1	Rv2355	transposase phage	
2784657..2785697	0	Rv2479c	transposase phage	
2785592..2785918	1.23077	Rv2480c	transposase phage	
2972160..2972486	1	Rv2648	transposase IS6110 phage	
2972435..2973421	0	Rv2649	transposase IS6110 phage	
3120566..3121504	0	Rv2814c	transposase phage	
3121501..3121827	0	Rv2815c	transposase phage	
3551281..3551607	0	Rv3184	transposase phage	
3551604..3552542	0	Rv3185	transposase phage	
3552543..3552763	70.2409	*IG3235_Rv3185-Rv3186		
3552764..3553090	0	Rv3186	transposase phage	
3553087..3554025	0	Rv3187	transposase phage	
3710756..3711694	0	Rv3326	transposase phage	
3795100..3795984	1	Rv3380c	transposase phage	
3795985..3796034	0	*IG3438_Rv3380c-Rv3381c		
3796035..3796361	23.0909	Rv3381c	transposase phage	
3890830..3891156	1	Rv3474	transposase IS6110 phage	
3891051..3892091	0	Rv3475	transposase IS6110 phage	
4057733..4058701	0	Rv3617	EphA	RD8 <sup>†</sup>
4058698..4059885	0	Rv3618	monooxygenase	
4059886..4059983	3	*IG3678_Rv3618-Rv3619c		
4060592..4060647	0	*IG3680_Rv3620c-Rv3621c		
4060648..4061889	0	Rv3621c	PPE65	
4061890..4061898	0	*IG3681_Rv3621c-Rv3622c		
4061899..4062198	0	Rv3622c	PE32	
4062199..4062526	0	*IG3682_Rv3622c-Rv3623		
4343314..4345035	1.31683	Rv3868	hypothetical protein	
4345036..4345038	0	*IG3933_Rv3868-Rv3869		

<sup>a</sup>IG account for intergenic regions.

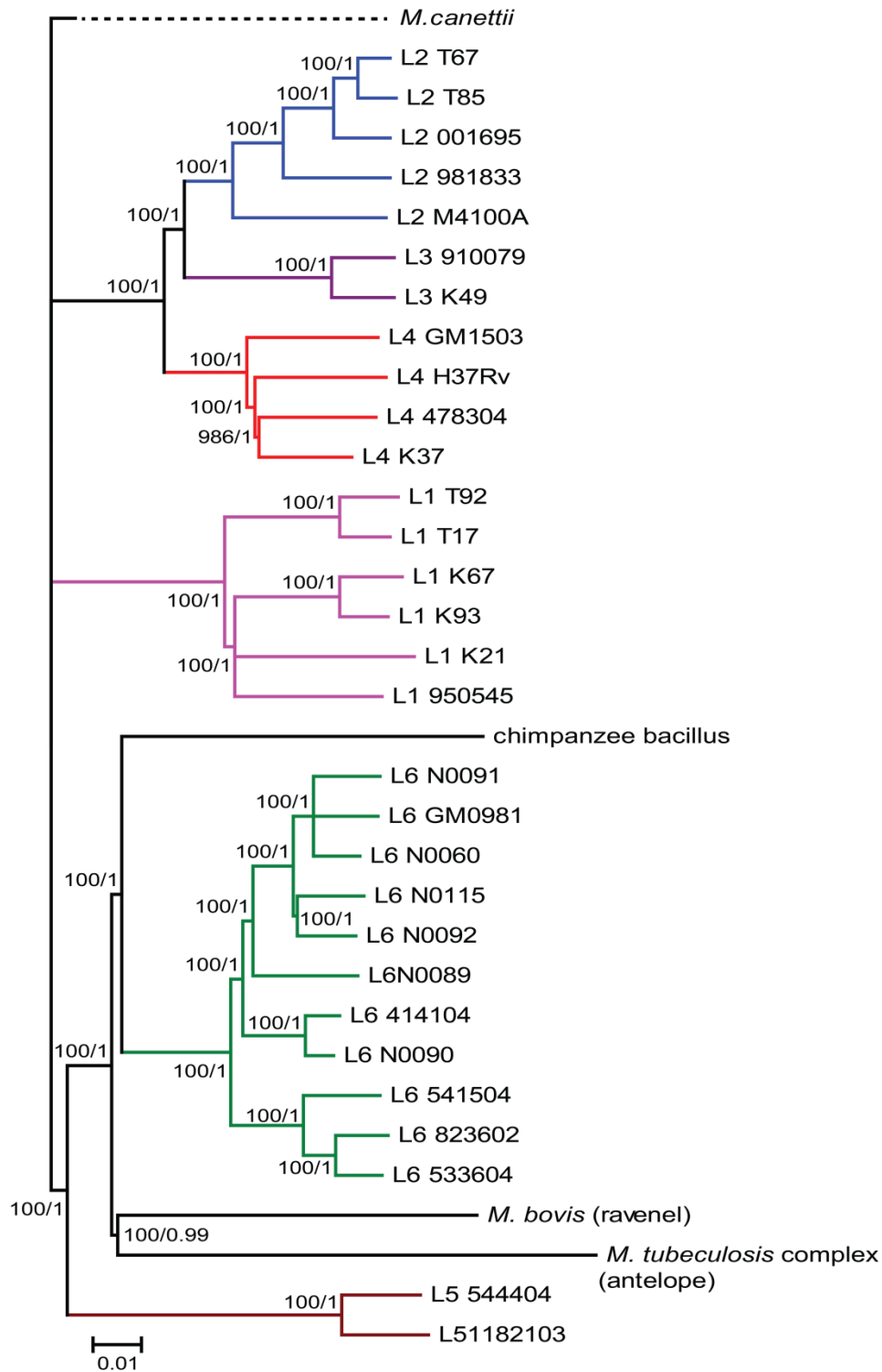
<sup>†</sup>Genomic deletion identified in Gagneux et al. (3).

Technical Appendix Table 3. Coverage of the chimpanzee bacillus sequence reads in *M. africanum* specific deletions described in Mostowy et al. (4)

Deletion	Mean coverage
RD720	4546.4117
RD721	4728.18
RD701	1440.11
RD702	3991
RD722	4310.79
RD711	4695.3543
RD742	3709.58
RD713	3252.15
RD743	4693.5488
RD724	4145.295
RD728	4749.474
RD715	0.6
RD727	4559.3533
RD735	5045.54

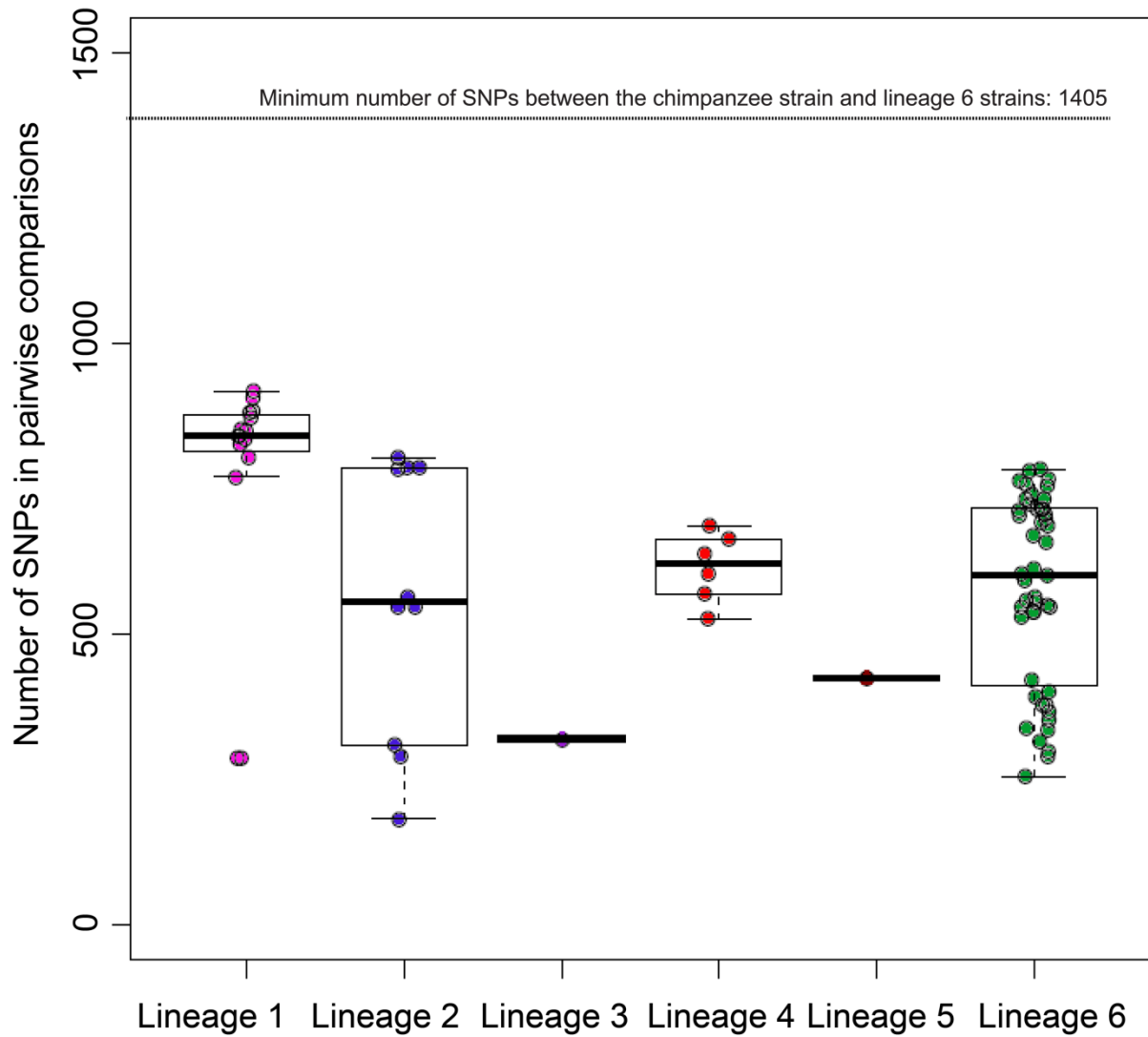


Technical Appendix Figure 1. Broth culture of *Mycobacterium tuberculosis* complex isolate, diluted, plated onto Middlebrook 7H11 agar, and photographed after 76 days.

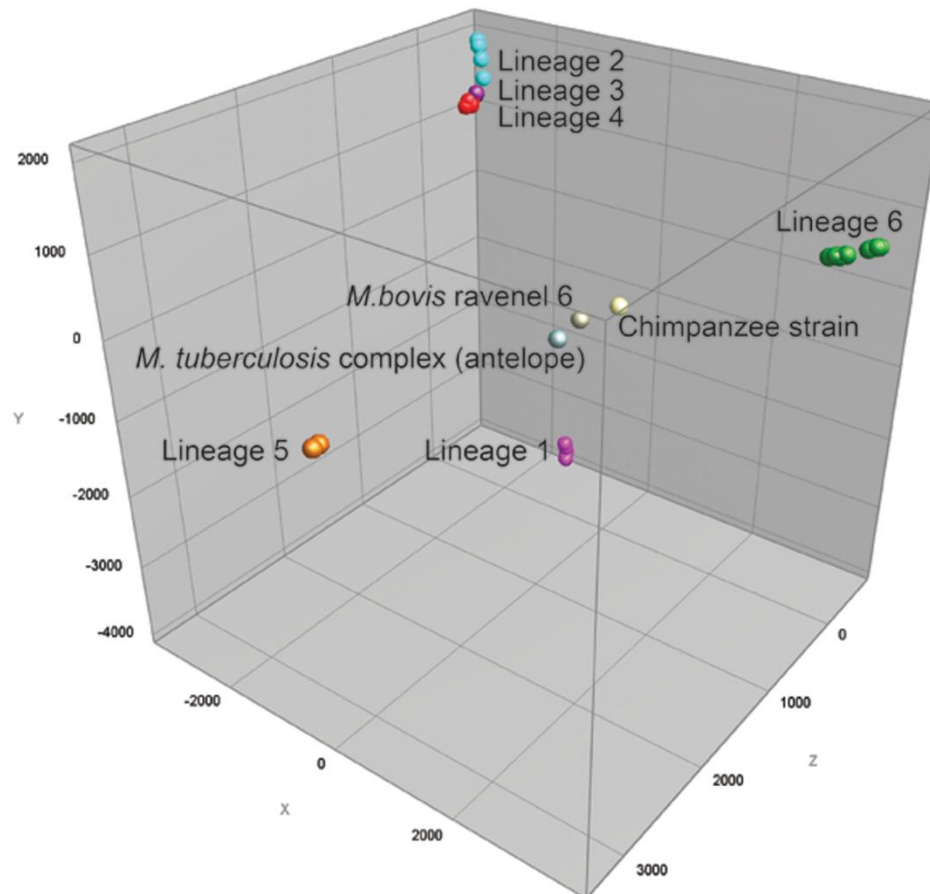


Technical Appendix Figure 2. Maximum-likelihood phylogeny of *Mycobacterium tuberculosis* complex isolate, obtained by using PhymL (5) and 1,000 bootstrap pseudoreplicates for clade support. Posterior probabilities for each node by Bayesian-inference tree (6) are place next to the bootstrap values.





Technical Appendix Figure 3. Observed pairwise distance between strains within lineages of *Mycobacterium tuberculosis*. For box plots: median (horizontal line), interquartile range (box), minimum and maximum values (whiskers), and outliers (circles). The minimum number of single-nucleotide polymorphisms (SNPs) between the chimpanzee strain and the closest group (Lineage 6/*M. africanum* WA2) is shown as a dotted line; note that this distance is larger than the maximum pairwise distance within any of the lineages.



Technical Appendix Figure 4. PCA scatter plot of *Mycobacterium tuberculosis* complex and related strains, based on 13,480 variable nucleotide positions. Only the first three principal components from the PCA were used. Strains are grouping according to lineages, and the chimpanzee strain is not grouping with Lineage 6.

## References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows Wheeler transform. *Bioinformatics*. 2009;25:1754–60. [PubMed http://dx.doi.org/10.1093/bioinformatics/btp324](http://dx.doi.org/10.1093/bioinformatics/btp324)
2. van Ingen J, Rahim Z, Mulder A, Boeree MJ, Simeone R, Brosch R, et al. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerg Infect Dis*. 2012;18:653–5. [PubMed http://dx.doi.org/10.3201/eid1804.110888](http://dx.doi.org/10.3201/eid1804.110888)
3. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*. 2006;103:2869–73. [PubMed http://dx.doi.org/10.1073/pnas.0511240103](http://dx.doi.org/10.1073/pnas.0511240103)

4. Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, et al. Genomic analysis distinguishes *Mycobacterium africanum*. J Clin Microbiol. 2004;42:3594–9. [PubMed](#)  
<http://dx.doi.org/10.1128/JCM.42.8.3594-3599.2004>
5. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 2005;33:W557–9. [PubMed](#)  
<http://dx.doi.org/10.1093/nar/gki352>
6. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19:1572–4. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btg180>