



HAL
open science

Newly Identified *Mycobacterium africanum* Lineage 10, Central Africa

Christophe Guyeux, Gaetan Senelle, Adrien Le Meur, Philip Supply, Cyril Gaudin, Jody E Phelan, Taane G Clark, Leen Rigouts, Bouke de Jong, Christophe Sola, et al.

► **To cite this version:**

Christophe Guyeux, Gaetan Senelle, Adrien Le Meur, Philip Supply, Cyril Gaudin, et al.. Newly Identified *Mycobacterium africanum* Lineage 10, Central Africa. *Emerging Infectious Diseases*, 2024, 30 (3), 10.3201/eid3003.231466 . hal-04489082

HAL Id: hal-04489082

<https://hal.science/hal-04489082>

Submitted on 4 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Submitted 10/28/2023

Accepted 12/27/2023

23-1466 Dispatch

DOI: <https://doi.org/10.3201/eid3003.231466>

Table: 1

Figure: 1

Appendices: 2

TOC title: Newly Identified *Mycobacterium africanum* Lineage 10, Central Africa

Running head: *M. africanum* Lineage 10, Central Africa

Keywords: *Mycobacterium tuberculosis*, tuberculosis and other mycobacteria, TB, respiratory infections, bacteria, pathogen diversification, tuberculosis emergence, lineage 10, L10, central Africa, Democratic Republic of the Congo

Suggested citation for this article: Guyeux C, Senelle G, Le Meur A, Supply P, Gaudin C, Phelan JE, et al. Newly identified *Mycobacterium africanum* lineage 10, central Africa. *Emerg Infect Dis.* 2024 Mar [*date cited*]. <https://doi.org/10.3201/eid3003.231466>

Newly Identified *Mycobacterium africanum* Lineage 10, Central Africa

Christophe Guyeux, Gaetan Senelle, Adrien Le Meur, Philip Supply, Cyril Gaudin, Jody E. Phelan, Taane G Clark, Leen Rigouts, Bouke de Jong, Christophe Sola, Guislaine Refrégier

Author affiliations: University Bourgogne Franche-Comté (UBFC), Besançon, France (C. Guyeux, G. Senelle); Université Paris-Saclay – AgroParisTech, Gif-sur-Yvette, France (A. Le Meur, G. Refrégier); Institut Pasteur de Lille Center for Infection and Immunity of Lille, Lille, France (P. Supply, C. Gaudin); London School of Hygiene and Tropical Medicine, London, UK (J.E. Phelan, T.G. Clark, L. Rigouts, B. de Jong); Université Paris-Saclay, Saint-Aubin, France (C. Sola); Université Paris Cité, Paris (C. Sola)

Abstract

Analysis of genome sequencing data from >100,000 genomes of *Mycobacterium tuberculosis* complex using TB-Annotator software revealed a previously unknown lineage, proposed name L10, in central Africa. Phylogenetic reconstruction suggests L10 could represent a missing link in the evolutionary and geographic migration histories of *M. africanum*.

The traditional view of restricted diversity among bacterial agents causing human and animal tuberculosis is being revised thanks to wide use of whole genome sequencing (WGS). Besides *Mycobacterium canettii*, representing exceptional, nonclonal, early-evolution branching lineages of tubercle bacilli in eastern Africa, several previously unknown lineages of the *M. tuberculosis* complex have been identified in Africa over the past decade. *M. tuberculosis* complex lineage 7 (L7) was discovered in the Horn of Africa and L8 in the African Great Lakes region (1,2). *M. africanum* L9 was found only in Djibouti and Somalia. In contrast, 2 other major *M. africanum*-affiliated lineages contributing substantially to the tuberculosis burden, L5 and L6, are found mostly in western Africa (3). The pathway between eastern and western Africa in the evolutionary history of the bacillus remains unclear. We describe a newly identified sister lineage of L6 and L9, associated with central Africa, and discuss implications for determining the evolutionary history of related *M. africanum* lineages L5, L6, and L9. We based research on publicly available data and thus required no ethical approval.

The Study

We used the TB-Annotator platform (G. Senelle, unpub. data, <https://www.biorxiv.org/content/10.1101/2023.06.12.526393v1>) to integrate WGS data from 102,001 *M. tuberculosis* complex isolates in the National Center for Biotechnology Information (NCBI) public domain. This platform identifies genetic variations, including single nucleotide polymorphisms (SNPs), regions of difference (RDs), and *IS6110* insertions, differentiating selected genomes from *M. tuberculosis* H37Rv; the TB-Annotator database also contains information on genotypic drug resistance and geographic location of variant isolation.

SNPs from an exploratory set comprising 15,699 isolates largely from African origin were used to build a phylogenetic tree. Our analysis identified a sister to *M. africanum* L6 and L9, branching between these lineages and the animal lineage A1 (La_A1) (3). The newly identified lineage is represented by only 2 genomes: ERR2707158, obtained from a strain

isolated in 2008 from a patient residing in Kinshasa, Democratic Republic of the Congo (DRC), now incorporated under reference ITM-501386 (CT2008–03226) in the Coordinated Collections of microorganisms of the Institute of Tropical Medicine in Antwerp, Belgium; and ERR2516384, obtained from a strain isolated in Belgium in 2013 (V. Mathys, pers. comm., July 5th 2023, email). The genomes of the new lineage carried none of the SNP markers described in the latest *M. tuberculosis* complex lineage classification scheme (4) and no SNP that confers drug resistance.

To confirm the phylogenetic position of those 2 genomes, we identified SNPs from 132 isolates covering the genetic and geographic diversity of L5 and L6 and including representatives of all other lineages using the Genotube pipeline (Le Meur, pers. comm., September 15th, 2023, email) and TB-Profler (5). Resulting phylogenetic reconstruction confirmed the clustering of ERR2707158 and ERR2516384 in a branch between L6 and L9 and animal lineage La_A1 (Figure). The newly designated L10 samples shared 375 specific SNPs with isolates from our selected set of 132 samples; 243/375 specific SNPs were not detected in any of the 102,001 genomes included in TB-Annotator. Among these specific SNPs, 91 were synonymous (Appendix 1; <https://wwwnc.cdc.gov/EID/article/30/1/23-1466-App1.xlsx>). The pairwise distance between the 2 samples of interest was 382 SNPs (SNPs outside of repetitive regions, manually checked when discordant between 2 pipelines), much shorter than the distance to the other samples of our selection (minimum = 1,137 SNPs; average = 1,591 ±222 SNPs) (Appendix 2, Figure 1; <https://wwwnc.cdc.gov/EID/article/30/1/23-1466-App1.pdf>).

We next explored other features of the genomes to corroborate SNP-based phylogenetic inferences. In addition to the deletion of RD9 shared with the L5/L6 branch and animal-associated lineages, the 2 L10 genomes lacked RD7, RD8, and RD10 (3). However, they did not show the RD702 (L6/L9) or RD713 (L5) deletions. In contrast, the 2 unclassified genomes harbored the same specific large 9,134 nt deletion (*Rv0613c–Rv0622*) in *M. tuberculosis* H37Rv [NC_000962.3:706602–715736] not observed in any other lineage. This segment included the toxin/antitoxin gene pair *vapB29/vapC29*. Two other shared deletions encompassed *eis* and *dnaE2* (Appendix 1), potentially limiting the ability to acquire aminoglycoside resistance (6) and possibly affecting some mutational properties (7) of these *M. africanum* strains. The 2 genomes also shared 4 IS6110 copies at a position found in no other lineage (Appendix 1). In the CRISPR

locus of the 2 L10 genomes, reconstructed using CRISPRbuilder-TB (8), we found the same absence of spacers 7 and 9 (43-spacer spoligotype format) seen in L6, L9, and La_A1 (Table) and all last spacers starting from spacer 22 (ERR2516384) or 26 (ERR2707158; Table).

The genetic features of those strains, combining outlying phylogenetic position, genetic distance from the L6/L9 branch and other known *M. tuberculosis* lineages, distinctive regions of deletions and IS6110 insertions, and specific spoligotype signature, led us to propose their classification in a newly designated L10 lineage. We propose 3 synonymous SNPs (*gyrA* G7901T, *recN* C1920096T, and *dnaG* C2621730T) compared with the H37Rv 000962.3 reference sequence in housekeeping genes to identify the new lineage.

To evaluate potential regional and global circulation of L10 strains, we searched for similar spoligotype patterns using SITVIT2, which accumulates spoligotypes from over 110,000 isolates from 131 countries (9). We identified a single instance, BEL04200301729, showing the same spoligotype pattern as ERR2516384, which might represent a third occurrence of L10. Of note, that strain was isolated in the Republic of the Congo, a country neighboring DRC where ERR2707158 was collected (Appendix 2, Appendix Figure 2). We also browsed spoligotyping results from next generation sequencing data, collected from ≈1,500 isolates from a 2016–2017 national survey in DRC, targeted using Deeplex Myc-TB (Deeplex, <https://www.deeplex.com>) (10) but detected no similar pattern. Thus, both global (TB-Annotator and SITVIT2) and local (10) datasets suggested that L10 strains are rare at the worldwide level, and aside from migratory dissemination, likely restricted to central Africa. Of note, mapping *M. africanum* diversity in Africa showed that in addition to L10, central Africa also hosts a relatively large diversity of L5 strains (Appendix 2, Appendix Figure 2).

Despite the rarity of L10, its specific phylogenetic positioning and presence in central Africa provide new elements to the complex evolutionary history of *M. africanum*. Currently, the most likely scenario favors western Africa as the place of origin of all *M. africanum* variants (3). This scenario implies that, although L5 and L6 ancestors emigrated from eastern Africa and diversified in western Africa, L9 migrated back to eastern Africa. Finding L10 in central Africa with intermediate branching between L5 and L6/L9, in addition to animal-associated lineages with less clear geographic ranges, along with L10 migration from western Africa to central Africa, make this scenario plausible. Alternatively, *M. africanum* could have emerged close to

central Africa and subsequently migrated westwards and eastwards. This alternative scenario however would require greater sampling in central regions of Africa to gain real support.

Conclusions

Through the extensive mining of WGS and genotyping databases, we newly identified a thus far rare *M. tuberculosis* complex lineage, L10 (proposed), present in central Africa. The lineage is characterized by a new region of deletion, IS6110 insertions, and 243 SNPs, including *gyrA* G7901T, *recN* C1920096T, and *dnaG* C2621730T. L10 represents a sister clade to L6, mainly in western Africa, and L9, specifically in eastern Africa, and reveals a putative previously missing piece in the evolutionary history and migrations of *M. africanum*. Our findings extend the known diversity of *M. africanum* in Africa.

Acknowledgments

We thank the Institute of Tropical Medicine of Antwerp (Belgium) and Genoscreen (France) for sharing their SRA data on public databases. We thank Vanessa Mathys (Sciensano, Brussel, Belgium) for providing location of isolation for ERR2516384 sample.

References

1. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. Nat Rev Microbiol. 2018;16:202–13. [PubMed https://doi.org/10.1038/nrmicro.2018.8](https://doi.org/10.1038/nrmicro.2018.8)
2. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfaia O, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. Nat Commun. 2020;11:2917. [PubMed https://doi.org/10.1038/s41467-020-16626-6](https://doi.org/10.1038/s41467-020-16626-6)
3. Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. Microb Genom. 2021;7:00047. [PubMed https://doi.org/10.1099/mgen.0.000477](https://doi.org/10.1099/mgen.0.000477)
4. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, et al. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. Genome Med. 2020;12:114. [PubMed https://doi.org/10.1186/s13073-020-00817-3](https://doi.org/10.1186/s13073-020-00817-3)
5. Phelan JE, O’Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-

6. Chen W, Biswas T, Porter VR, Tsodikov OV, Garneau-Tsodikova S. Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB. Proc Natl Acad Sci U S A. 2011;108:9804–8. [PubMed https://doi.org/10.1073/pnas.1105379108](https://doi.org/10.1073/pnas.1105379108)
7. Dupuy P, Ghosh S, Adefisayo O, Buglino J, Shuman S, Glickman MS. Distinctive roles of translesion polymerases DinB1 and DnaE2 in diversification of the mycobacterial genome through substitution and frameshift mutagenesis. Nat Commun. 2022;13:4493. [PubMed https://doi.org/10.1038/s41467-022-32022-8](https://doi.org/10.1038/s41467-022-32022-8)
8. Guyeux C, Sola C, Noûs C, Refrégier G. CRISPRbuilder-TB: “CRISPR-builder for tuberculosis.” Exhaustive reconstruction of the CRISPR locus in *Mycobacterium tuberculosis* complex using SRA. PLOS Comput Biol. 2021;17:e1008500. [PubMed https://doi.org/10.1371/journal.pcbi.1008500](https://doi.org/10.1371/journal.pcbi.1008500)
9. Couvin D, David A, Zozio T, Rastogi N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. Infect Genet Evol. 2019;72:31–43. [PubMed https://doi.org/10.1016/j.meegid.2018.12.030](https://doi.org/10.1016/j.meegid.2018.12.030)
10. Kayomo MK, Mbula VN, Aloni M, André E, Rigouts L, Boutachkourt F, et al. Targeted next-generation sequencing of sputum for diagnosis of drug-resistant TB: results of a national survey in Democratic Republic of the Congo. Sci Rep. 2020;10:10786. [PubMed https://doi.org/10.1038/s41598-020-67479-4](https://doi.org/10.1038/s41598-020-67479-4)
11. 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. [PubMed https://doi.org/10.1038/ng.590](https://doi.org/10.1038/ng.590)

Address for correspondence: Guislaine Refrégier, Ecologie Systématique et Evolution, 12 rue 128, 91190, Gif-sur-Yvette, France; email: guislaine.refregier@universite-paris-saclay.fr

Table. Spoligotype patterns of main strains of the study: L10 strains, representative strains of L6, L9 and A1 lineage. Note that L6, L9 and both L10 samples harbor the same losses of spacer 7 and of spacer 9.*†

IsoNumber or other ID	Source	No.	Spoligotype binary	SIT#	Country of isolation
L10-BEL04200301729	SITVITWEB	1	███████████	Orph	Congo Republic
L10-ERR2516384	TB- Annotator	1	███████████	Orph	Belgium (origin unknown)
L10-ERR2707158	TB- Annotator	1	███████████	Orph	Democratic Republic of Congo
L6_SIT181	SITVITWEB	208	███████████	181	Gambia,...
L9	Coscolla 2021	3	███████████	Orph	Somalia
L9	Coscolla 2021	1	███████████	U	Djibouti
L9_FXX01199901706	SITVITWEB	1	███████████	Orph	France
NLD009501731	SITVITWEB	2	███████████	Orph	The Netherlands
A1_Dassie Bacillus	TB- Annotator	1	███████████	U	South Africa
A1_ <i>M. mungi</i>	mbovis.org, TB- Annotator	1	███████████	SB 1960	Botswana
A1_chimpanzee bacillus	Coscolla 2013, TB- Annotator	1	███████████	U	Ivory Coast

*L6, L9, and both L10 samples harbor the same losses of spacer 7 and of spacer 9. DRC, Democratic Republic of the Congo; orph, orphan (single reported occurrence); SIT, spoligotype international type; U, undesignated.

†Coscolla 2021, (3); Coscolla 2013, (12); SITVITWEB, (9); TB-Annotator, G. Senelle, unpub. data,

<https://www.biorxiv.org/content/10.1101/2023.06.12.526393v1>.

‡Origin unknown.

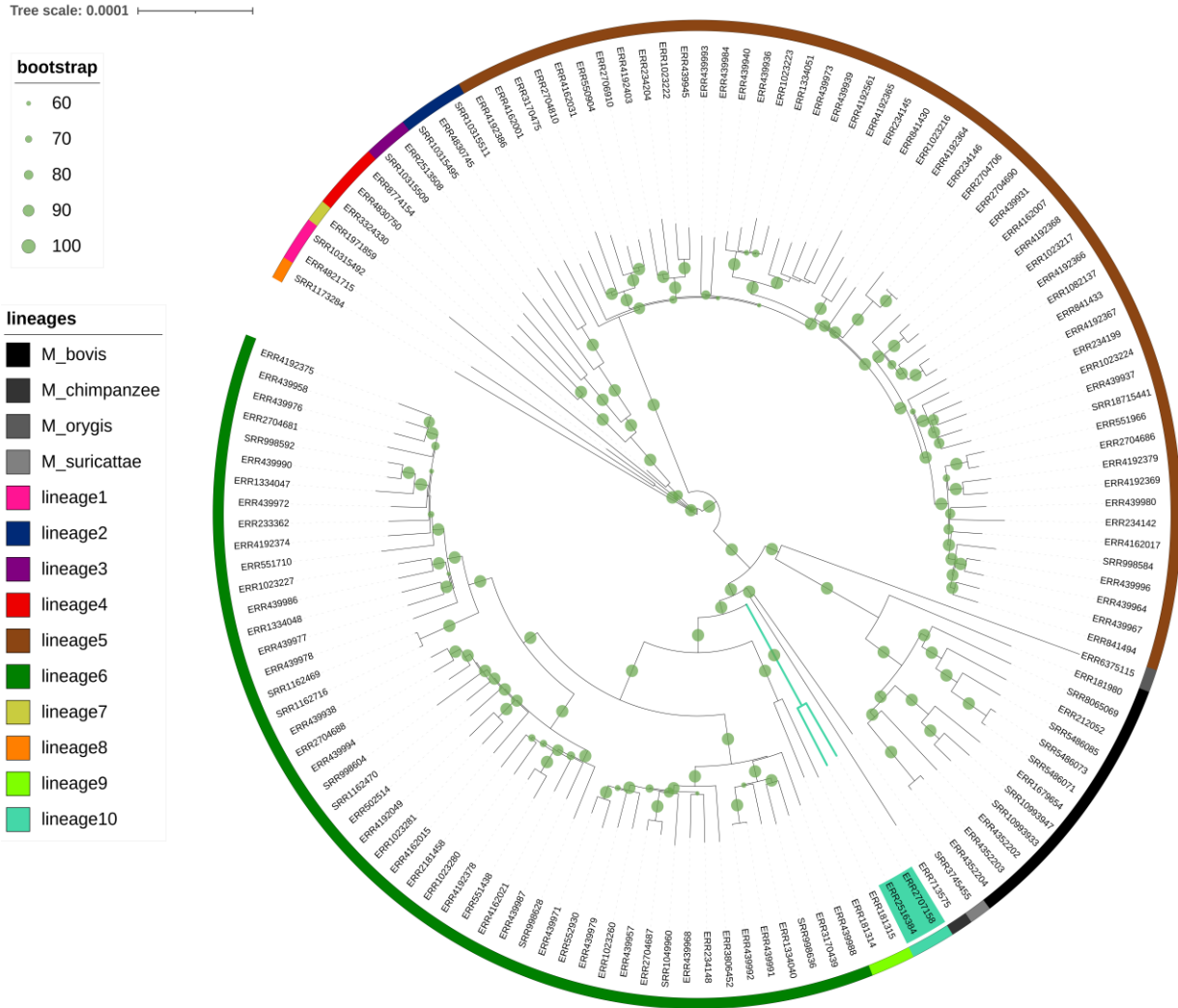
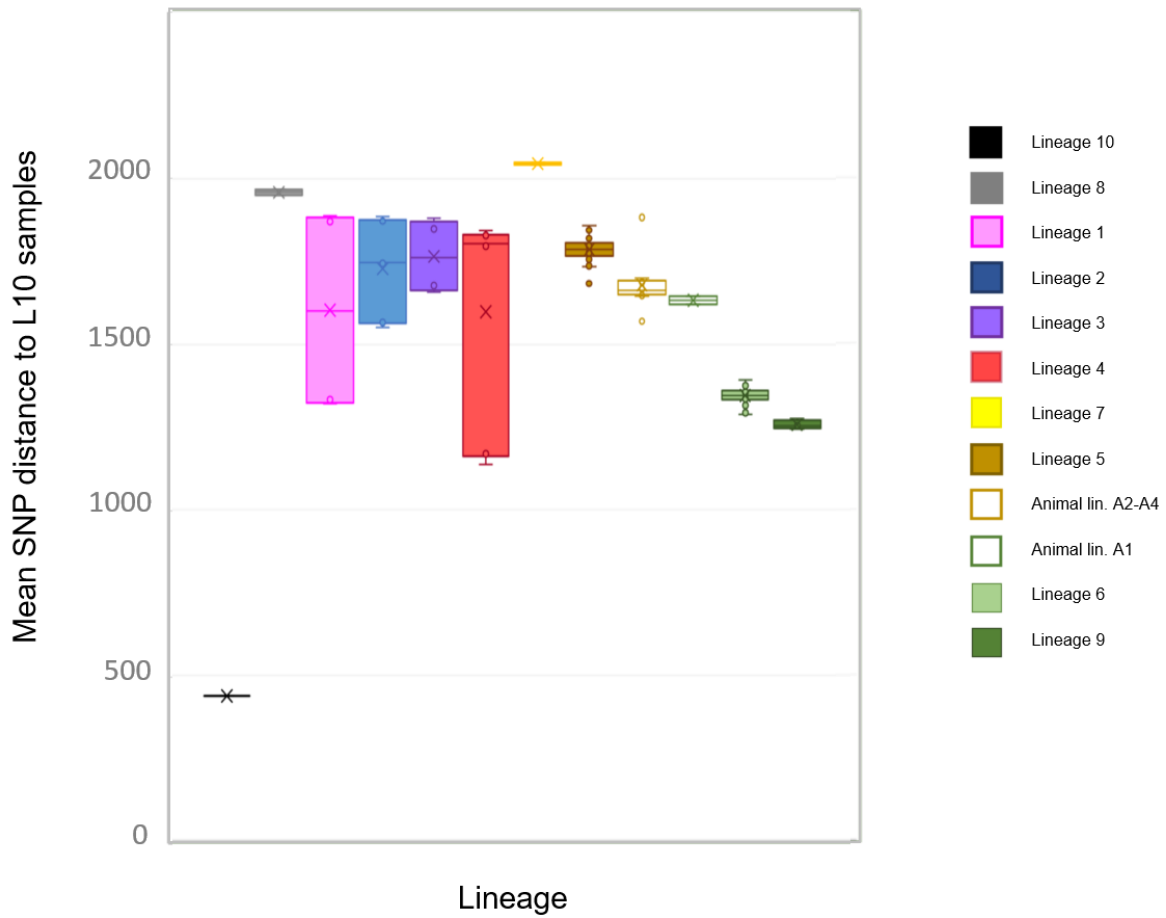


Figure. Global *M. tuberculosis* phylogeny including the new L10 strains. We chose *M. africanum* samples for harboring RD9 deletion, having documented country of origin (for the purpose of additional analyses, Appendix 2, Figure 2), and made a selection to retain a sole representative of each sublineage for each country representing the genetic and geographic diversity of *M. africanum* in Africa. Representatives of other lineages were included for comparison. Specifically for this phylogenetic reconstruction, SNPs were identified in comparison with an *M. tuberculosis* ancestor (11) and reincorporated into the whole genome to avoid biases in the molecular model or need for Lewis correction. Phylogeny was rooted with *M. canettii*, subsequently removed for better visualization. Bootstrap support was computed using 100 replicates and shown when ≥ 0.6 . Circles confirm the large support of almost all branches, especially of L10 and its sister branches. L10 branching point lies between L9 and the La_A1 lineage grouping chimpanzee and Dassie bacillus. The scale bar indicates the number of nucleotide substitutions per site.

Appendix 1. Additional information about genomic details for newly identified *M. africanum* lineage L10.

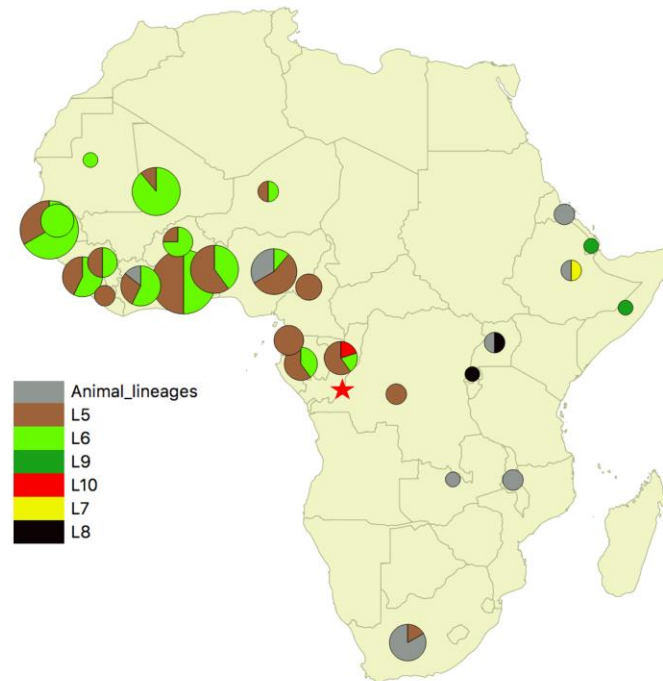
S1 Fig



Suppl. Fig. 1 – Intra- versus inter-lineage genetic distance of L10 strains. The distances between L10 strains and all other samples shown in the tree of Fig. 1 were computed using snp-dist tool excluding PE-PPE regions. The legend is provided in the same order as the one in which they appear on the X-axis. Note that L10 strains are much more similar one to the other as they are to strains from other lineages.

Appendix 2. Additional information about phylogenetic relatedness and geographic circulation of newly identified *M. africanum* lineage L10.

S2 Fig



Suppl. Fig. 2 - Africa map featuring the distribution of *M. africanum* sublineages. This map locates the African samples from our 132 samples diversity set (Fig. 1), as annotated by Genotube. Animal lineages include La_A1 and other animal lineages. The sampling of these lineages should not be considered as representative because of lack of sublineage definition among animal lineages in Genotube. In contrast, as one sample from each known *M. africanum* sublineages was retained for each country, the sampling can be considered as representative of L5 and L6 diversity. The map was built with QGIS® (v3.4.6-Madeira on Mac-OS10.13.6, using an open source shapefile with African boundaries obtained from <https://hub.arcgis.com/>). The single localized L10 strain reported in NCBI SRA database is shown in red (in Democratic Republic of Congo). The red asterisk shows The Republic of Congo where another L10 strain was isolated according to SITVITWEB database. Of note, Belgium where the third L10 strain was collected has a history of colonization with DRC; the strain was isolated from a patient carrying an African name which is not indicative but compatible with the actual origin of this strain in DRC, *i. e.* also in Central Africa. This map also highlights the presence of L5 in Central Africa.