Population structure of *Mycobacterium bovis* in Germany: A long-term study using Whole Genome Sequencing combined with conventional molecular typing methods

Thomas A. Kohl¹,², Katharina Kranzer³,⁴, Sönke Andres³ Thierry Wirth⁵,⁶,* Stefan Niemann¹,²,* Irmgard Moser⁷,*

¹ Molecular and Experimental Mycobacteriology, Research Center Borstel, Germany.
² German Center for Infection Research (DZIF), partner site Hamburg-Lübeck-Borstel, Germany.
³ Division of Mycobacteriology (National Tuberculosis Reference Laboratory), Research Center Borstel, Borstel, Germany.
⁴ Clinical Research Department, London School of Hygiene & Tropical Medicine
⁵ Laboratoire Biologie Intégrative des Populations, Evolution Moléculaire, EPHE, PSL University, Paris, France.
⁶ Institut de Systématicque, Evolution, Biodiversité, UMR-CNRS 7205, Muséum National d’Histoire Naturelle, Université Pierre et Marie Curie, Université des Antilles, Ecole Pratique des Hautes Etudes, Sorbonne Universités, Paris, France.
⁷ Friedrich-Loeffler-Institut, Federal Institute for Animal Health, Institute of Molecular Pathogenesis, Jena, Germany

* shared senior authorship

Key words: Tuberculosis, *Mycobacterium bovis*, human, animal, transmission, spoligotyping, MIRU-VNTR-typing, Whole Genome Sequencing

Running title: Population structure of *Mycobacterium bovis* in Germany

Corresponding author: Irmgard Moser
Abstract

*Mycobacterium bovis* (Mbov) is the primary cause of bovine tuberculosis (bTB), and also infecting a wide range of domestic animal and wildlife species and humans. In Germany, bTB still emerges sporadically in cattle herds, free-ranging wildlife, diverse captive animal species, and humans. In order to understand the underlying population structure and estimate the population size fluctuation through time, we analyzed 131 Mbov strains from animals (*n* = 38) and humans (*n* = 93) in Germany from 1999 to 2017 by whole genome sequencing (WGS), MIRU-VNTR typing, and spoligotyping. Based on WGS data analysis, 122 out of the 131 Mbov strains were classified into 13 major clades, six contained strains from both human and animal cases, and seven only from human cases. Bayesian analyses suggest that the Mbov population went through two sharp anticlimaxes, one in the middle of the 18th century and another one in the 1950's. WGS based cluster analysis grouped 46 strains into 13 clusters ranging in size from 2-11 members and involving strains from distinct host types, e.g. only cattle, and also mixed hosts. Animal strains of four clusters were obtained over a nine-year time span, pointing towards autochthonous persistent bTB infection cycles. As expected, WGS had a higher discriminatory power than spoligotyping and MIRU-VNTR typing. In conclusion, our data confirm that WGS and suitable bioinformatics is the method of choice to implement a prospective molecular epidemiological surveillance of Mbov. The population of Mbov in Germany is diverse, with subtle, but existing interactions between different host groups.

Introduction
Tuberculosis (TB) is one of the high priority infectious diseases affecting humans and animals worldwide (1, 2), and the leading cause of death by a single infectious agent in humans (2). Causative agents for TB are the members of the *Mycobacterium tuberculosis* complex (MTBC), namely *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. caprae*, *M. microti*, and *M. pinnipedii*. In addition, *M. canettii*, *M. mungi*, and *M. orygis* have been proposed as separate ecotypes. However, their taxonomic classification is still under debate (3).

*M. bovis* (Mbov) is the primary cause of bovine TB (bTB) but also affects a wide range of other domestic animal and wildlife species and even humans (4, 5, 6, 7). After time periods of high prevalence of bTB infection in cattle until the second half of the 20th century, Germany has reached the status of being officially free of bTB. Since July, 1st, 1996 (Decision 97/76/EC), 99.9 percent of the cattle herds remained officially free of bTB infection and disease for at least six consecutive years (Article 2(d) of Council Directive 64/432/EEC, 8, 9, 10). However, bTB is still emerging sporadically in cattle herds (11), free-ranging wildlife, captive animal species (12), and humans (13). Confirmed animal bTB cases are notified through an electronic national disease information system (TSN) and published annually (14). From January 1999 to December 2015, a total of 214 bTB outbreaks in cattle herds were notified in Germany, with about half of the cases caused by either *M. bovis* or *M. caprae*. In general, *M. caprae* is reported mainly in middle European countries with sporadic cases also in Asia and Peru (15,16), with cattle and wildlife cases in Germany restricted to an area at the German-Austrian border (17,18). *M. caprae* was therefore not included in this study. According to the European Food Safety Authority (EFSA) 2017, from 2013-2017, 43-56 bTB cases in humans were diagnosed annually (13). Notification rates for bTB ranged from 0.05- to 0.07 per 100,000 population. Mbov and the closely related
*M. caprae* make up about 1% of all human TB cases (5,486 cases in 2017, more than six per 100,000 population) (13, 19).

As disease transmission dynamics of Mbov within and between host groups are only partially understood (20), molecular typing methods could offer insights into transmission routes and inform pathogen surveillance (21, 22, 23). Classical genotyping methods including spoligotyping, restriction fragment length polymorphism (RFLP) and mycobacterial interspersed repetitive unit variable number of tandem repeat (MIRU-VNTR) detection allow analyzing outbreaks, assessing population structures, and performing longitudinal molecular epidemiological studies (24, 25, 26, 27, 28, 29, 30, 31).

Spoligotyping (25) is based on the analysis of CRISPR-CAS spacer sequences located in a genomic region prone to convergent evolution (21), possibly leading to uncertainty of strain relatedness. Spoligotyping patterns submitted to international databases receive unique identifiers: SITVIT (32, 33, 34) allowing for MTBC isolates from any host, and mbovis.org accepting MTBC strains from animals only (35). As of October 2018, 39,609 MTBC spoligotypes have been collected in the SITVIT database from more than 121 countries (32). At mbovis.org, 2,117 patterns are available (last update April 2020). RFLP is a method with high potential for discrimination for *M. tuberculosis* but not Mbov strains due to the small number of analyzed insertion element copies present in the respective genomes. MIRU-VNTR typing possesses a higher discriminatory power, allowing automated high throughput typing and web-based translation into a digit code identifier (29, 30, 36, 37). The method has high potential to define clusters of related strains, but cannot differentiate between closely related strains within outbreaks (38).
Next generation sequencing (NGS) allows for analysis of the nearly-complete genome of a pathogen by whole genome sequencing (WGS), providing deeper insights into the population structure, pathogen evolution, transmission chains, and biology of bacteria \((38, 39, 40, 41)\). WGS analysis facilitates the detection of recent transmission chains and monitoring re-emerging of strains after years of non-detection \((42, 43, 44, 45)\).

In this study, we used WGS, spoligotyping and MIRU-VNTR to determine the diversity of Mbov strains isolated from animals and humans in Germany and define possible transmission chains within and between different host populations over a 19-year period (1999-2017). Using Bayesian analyses, we sought insights into the dynamics of strain diversity over the last 800 years in Germany.

**Materials and Methods**

**Strain selection and DNA extraction**

In total, 131 Mbov strains were available for WGS including the reference strain Mbov BCG (DSM 43990 / ATCC 27289), with 38 strains from the Friedrich-Loeffler-Institut (FLI), Federal Institute for Animal Health, and 93 strains from the National Reference Center (NRC) for Mycobacteria in Borstel, Germany (supplementary table S1). From January 1999 to December 2015 (the study period), a total of 214 bTB outbreaks in cattle herds were notified in Germany by the electronic system implemented by the FLI to monitor bTB outbreaks, with about half of the cases in cattle caused by *M. bovis*. Mbov strains from ten cattle bTB outbreaks, from five other domestic animal species, 14 zoo animals, and wild boars were analyzed (supplementary table S2), spanning the time period from 1999–2015, and covering different regions of the country, including the known hot spot regions in the north and south. At the NRC in Borstel, all German
M. bovis strains cultured and archived from 2000 to 2017 were included. The NRC receives samples from all districts in Germany, and while it is not the only laboratory offering specialist mycobacterial diagnostics in Germany, it receives an estimated 50% of all MTBC isolates. At both institutions, strains were cultured according to standard procedures (46, 47, 48, 49), and genomic DNA was extracted using the High Pure PCR Template Preparation kit (Roche Life Science; FLI) and with the cetyltrimethylammonium bromide (CTAB) procedure (NRC), respectively (50).

Classical genotyping

Spoligotyping of animal strains was performed using a microarray format (Alere Technologies, Jena, Germany) (51). Binary codes were automatically compared with data available through SITVIT and mbovis.org to identify concordant species and lineages. For human strains, the conventional spoligotyping method was used (25). MIRU-VNTR-typing of the strains isolated from animals was performed using conventional PCR and agarose gel electrophoresis (27, 29, 52). For human strains, the automated high-throughput method was used (29). VNTR copy numbers were assessed according to allele calling tables (www.miru-vntrplus.org, EU Reference Laboratory for bovine Tuberculosis, www.visavet.es). The discriminatory power of the method was calculated according to Hunter and Gaston (53); (supplementary tables 3 and 4).

Whole genome sequencing and data analysis

Libraries for WGS were prepared from genomic DNA with a modified Illumina Nextera protocol (54) and run on the Illumina NextSeq NGS platform (Illumina, San Diego, CA, USA). We employed the MTBseq pipeline with default parameters for variant detection and a joint analysis (55), employing a threshold of 12 SNPs for cluster detection (56).
As deduced from the pairwise SNP distances distribution, we used a cutoff of 350 SNPs to detect major groups (figure 2). For all sequenced strains, mean coverage depth was at least 50-fold, and at least 95% of the reference genome fulfilled the MTBseq thresholds for variant detection. From the aligned sequences of concatenated SNP positions produced by MTBseq, we calculated a maximum likelihood tree with FastTree (57) with a general time reversible (GTR) substitution model, 1,000 resamples and Gamma20 likelihood optimization to account for rate heterogeneity among sites. The consensus tree was rooted with the “midpoint root” option in FigTree (http://tree.bio.ed.ac.uk/software/figtree), and nodes were arranged in increasing order. The resulting tree was annotated with the EvolView software (58). Additionally, we built maximum parsimony trees with the software BioNumerics version 7.5 (Applied Maths, Gent, Belgium) with default settings.

For the coalescent-based analyses, evolutionary rates and tree topologies were analyzed using the general time-reversible (GTR) and Hasegawa-Kishino-Yano (HKY) substitution models with gamma distributed among-site rate variation with four rate categories (Γ4). The substitution rate was estimated by plotting a regression line that depicts for the sole WGS clusters, in a pairwise manner, the relationship between the elapsed time and the accumulated number of SNP’s. Under this model, the slope corresponds to the mutation rate. We tested both a strict molecular clock (which assumes the same evolutionary rates for all branches in the tree) and a relaxed clock that allows different rates among branches. Constant-size, exponential and Bayesian skyline plot models, based on a general, non-parametric prior that enforces no particular demographic history were used in BEAST v1.10.4 (59). For each model, two independent chains were conducted for 200 million generations and convergence was assessed by checking ESS values for key parameters using TRACER V1.7.1 (60). We used TRACER V1.7.1 to calculate the log10 Bayes factors in order to compare the
models after a burn-in of 10% of the chain. Bayes factors represent the ratio of the
marginal likelihood of the models being compared. Approximate marginal likelihoods
for each coalescent model were calculated via importance sampling (1,000 bootstraps)
using the harmonic mean of the sampled likelihoods. A ratio between 3 and 10
indicates moderate support that one model better fits the data than another, whereas
values greater than 10 indicate strong support. For correlation with known clonal
complexes, we selected 33 strains representing the known clades contained in a recent
publication (61), and performed a joint analysis as described previously.

Data availability
All WGS data was submitted to the EMBL-EBI ENA SRA archive (supplementary table
S1).

Ethics statement
Ethical approval was not sought, as no patient data was used.

Results
In total, 131 Mbov strains, 93 of human and 38 of animal origin (supplementary table
S1) isolated in Germany from 1999–2017, including one M. bovis BCG reference
strain, were investigated by spoligotyping, MIRU-VNTR-typing, and WGS. WGS data
analysis revealed 12,726 variable SNP positions among the genomes analyzed that
were used for the calculation of a phylogenetic tree (figure 1). Interestingly, the strain
mbov-49 was clearly separated from the rest of the study collection. This strain has
been isolated at the FLI in 2000 from a Nilgau antelope (*Boselaphus tragocamelus*), which died in a German zoo, and found to be not intrinsically pyrazinamide resistant (*62*).

Overall, the median pairwise distance in distinct SNP positions of the 131 strains was 516 SNPs, and distinct peaks emerged in the frequency distribution between 0-30, 70-350, 370-620, and 780-840 distinct SNPs, agreeing with the groups of related strains found by cluster detection with a threshold of 12, 30, and 350 distinct SNPs (d12, d30, d350) between nearest group members (figure 1, figure 2). Using the d350 threshold to group strains, we found 13 cladistic groups containing 122/131 strains ranging in size from 2-35 members, with on average eight years (2-18) between the earliest and latest year of isolation.

Six of the d350 groups contained both human and animal cases, and seven only human cases. When comparing d350 groups with the known clonal complexes African 1 and 2 (Af1, Af2), European 1 and 2 (Eu1, Eu2), as well as newly determined Unknown 1-8 (*61*), we could correlate clonal complexes Af1, Eu1, Eu2, and Unknown2 with d350 groups 08, 07, 06, and 13 (supplementary figure S1, supplementary table S6). For clonal complexes Af2, Unknown1, and Unknown7, we found only one corresponding strain in our collection (mbov-118, mbov-49, mbov-119). Interestingly, three d350 groups (10, 11, 12) were attributed to clonal complex Unknown3, and four d350 groups (01, 02, 03, 04) to clonal complex Unknown4. We found no representatives of complexes Unknown5 and Unknown6 in our study, as well as correlates of d350 groups 05 and 09 among the collection of known clonal complexes.

Putative transmission clusters
We used a threshold of at most 12 distinct SNP positions to the nearest group member as indication for possible recent transmission (54), which yielded 13 d12 clusters of altogether 46 strains (figure 1, figure 3, table 1). The d12 clusters ranged in size from 2-11 members, spanned up to 15 years and involved distinct host types, with d12 clusters 5 and 12 only comprising cattle hosts, clusters 4, 7, 11, and 13 only human hosts, and the rest mixed hosts (table 1). In total, 32 of the 38 animal strains (the pair of Mbov BCG in d12 cluster 13 not counted) were grouped into WGS d12 clusters. In four of these clusters, animal strains were recovered more than nine years apart, pointing towards autochthonous persistent bTB infection cycles. In contrast, only 12 out of the 93 human strains were grouped into d12 clusters, with nine human strains forming four WGS d12 clusters of two and three members, respectively (table 1). The members of these groups were isolated within at most two years from each other. Overall, we found one cluster (cluster 8) with a putative transmission from cattle to humans with respective strains separated by two SNPs, and one cluster (cluster 6) of raccoon and human strains separated by 12 SNPs.

As the frequency distribution of pairwise SNP distances featured a peak between 0-30 SNPs (figure 2), we also clustered strains with a threshold of 30 SNPs. This yielded two new clusters of related strains with two members each, an additional member of d12 cluster 13, and d12 clusters 2 and 8 were joined together (figure 1).

Comparison with classical genotyping

The 131 strains were differentiated into 45 known spoligotypes and 11 spoligotypes not contained in the established databases (supplementary tables S1 and S5). Five or more strains each fell into four known spoligotypes: SB 120/IT0482 (35 strains), SB
Of 251 these, SB 120 and SB 121 have been reported as predominant spoligotypes circulating among animals around the world (63). Strains of these spoligotypes were present in different branches of the constructed phylogenetic tree and in different MIRU-VNTR and d12 clusters (figure 1).

Comparing the composition of the d350 groups in terms of the respective spoligotypes (figure 1), we found correlations with the well-established clonal complexes1 and 2 and Af 1 and 2, as well as with the newly determined complexes named unknown 1 – 8 (61; supplementary table S7). For example, SB0120 found in d350 groups 01, 02, 04, 05, 10, and 13 was detected in complexes Unknown 2–5. This spoligotype has been reported as predominant circulating among animals around the world (63). Seven spoligotypes present in d350 groups 01, 02, 03, and 04 were reported for complex Unknown4 (61). The 15 spoligotypes found for d350 group 06 corresponded to those for complex Eu2, and the nine spoligotypes present in d350 groups 10, 11 and 12 were found in clade Unknown3 (61). The spoligotype SB0989 found in d350 group 09 was reported for singletons not contained in a complex (61).

MIRU-VNTR analysis yielded 92 distinct patterns with 21 strain clusters ranging from two to seven members comprising altogether 62 strains. Using 121 supposedly unrelated strains, the discriminatory power index (HGD; 51) of each of the 24+1-locus MIRU-VNTR loci was determined finding allelic heterogeneity mainly restricted to 2-4 repeat copies (supplementary table S3). Allele heterogeneity of > 0.5 was found for the loci VNTR 2163a, 2163b, 2165, 2461 and 4052 (supplementary table S4). Overall, MIRU-VNTR types correlated well with both the phylogenetic tree and the d12 clusters. However, 21 strains grouped by MIRU-VNTR were not clustered by d12 analysis, and
four d12 clusters encompassed strains with different MIRU-VNTR patterns, with four distinct loci in one, and one distinct locus in three of these cases (figure 1, figure 3).

**Mutation rate estimation and demographic inference**

The geographically widespread and phylogenetically diverse nature of our strain collection did not allow implementing a Bayesian tip-dating approach. We therefore focused on the 13 d12 clusters where the measurably evolving dimension of Mbov could be captured to infer a realistic estimation of the mutation rate. A positive correlation ($r^2 = 0.682$) was found between the time elapsed between two strains and the number of accumulated SNPs (figure 4). The slope was close to 1, corresponding to the acquisition of one SNP every year between two strains and translating to a mutation rate of $1.14 \times 10^{-7}$ substitutions/nucleotide/year.

To estimate the effective population size fluctuation through time, three demographic models were compared and the best fitting evolutionary model was obtained under the Bayesian skyline model with a relaxed clock (figure 4). The relaxed clock model outperforms the constant clock model ($BF = 40$) and the Bayesian skyline was favored to its closest model, constant size ($BF = 14$). The TMRCA (TIME to Most Recent Common Ancestor) corresponding to our Mbov strain collection dated back some 950 years ago (95% HPD [highest posterior density] interval, 836-1062). According to the coalescent-based demographic reconstructions, the German Mbov population went through three successive expansions, a first twentyfold increase in the late middle age, followed by two mild expansions in the middle of 18th century and the early 20th century (figure 4).
Discussion

This investigation provides insights into population structure, persistence and population size fluctuation of Mbov strains in Germany over time and the complex interrelations in a multi-host pathogen system. In the context of a country declared officially free of bTB for more than two decades, special consideration was given to strain persistence attempting to understand recurrent outbreaks and possible links to human cases, while other publications have mainly concentrated on microevolution of strains in the context of geospatial spreading and transmission dynamics between animal reservoirs (64, 65).

The main limitation of our study is that, due to practical limitations related to access to strains, we were not able to collect a fully comprehensive set of Mbov strains from human and animal cases in Germany. Additionally, due to the restrictions set by data protection regulations, the available metadata for the strains was limited to year and host of isolation. Regrettably, this does not allow an epidemiological analysis of the WGS d12 and d30 clusters. Still, our collection covers a time span from 1999-2017 and diverse host species. While we took care to identify and remove duplicate strains from the same host, we cannot fully exclude this possibility for human strains.

We successfully performed WGS for a collection of 93 human and 38 animal Mbov strains, isolated in Germany from 1999–2017. The pairwise distance distribution and the reconstructed phylogenetic tree indicate the presence of 13 d350 groups within the study population. These encompassed the majority of strains (122/131) and represents a snapshot of Mbov sublineages historically spreading in Germany. Correlating our
phylogeny and detected groups with described clonal complexes revealed that our collection contains representatives of the well-known Mbov complexes Af1, Af2, Eu1, and Eu2, as well as of additional groups defined recently (61). Interestingly, there are at most two strains of complexes Af1, Af2, and Eu1 in our study, and we found no representatives of complexes Unknown5 and Unknown6, or correlating complexes for d350 groups 05 and 09. This might indicate a geographically uneven distribution of subgroups and that the Mbov phylogeny needs to be refined by WGS-based studies with larger, geographically diverse collections.

Using a threshold of 12 distinct SNP positions to identify strains possibly involved in recent transmission events (56), we found 32 out of the 38 animal strains and 12 out of the 93 human strains grouped into 13 d12 clusters. In four of these clusters, animal strains were recovered more than nine years apart, pointing towards autochthonous persistent bTB infection cycles. This is further supported by the combination of d12 clusters 2 and 8 into a joint group when clustering with a threshold of 30 SNPs, with the phylogenetic analysis and the number of distinct SNP positions suggesting a relatively recent common source for both clusters. Human strains within clusters were isolated within at most one-year difference and with one sole exception had at most one SNP distance, possibly indicating direct transmission.

Despite the imbalance of Mbov strains included from humans and animals, there seem to be distinct infection dynamics for animals and humans. For cattle and other animals, the majority of strains were found within d12 clusters and several strains were persistently spreading over up to 15 years, pointing towards potential reservoirs of these strains, for example in the German wildlife population. The mostly un-clustered human cases might represent progression to active disease from latently infected individuals as indicated previously (17). In general, human mobility is also higher
compared to cattle and wild animals. Here, patients having contacts to sources of infection outside Germany may contribute to the detected high diversity of strains isolated from human patients. As reported in 2003 (17), the majority of patients with Mbov disease in Germany, was over 60 years of age suggesting that they might have acquired the infection at a young age when the prevalence of bTB in cattle in Germany was much higher than today. Unfortunately, Mbov strains isolated from cattle before 1999 were not available.

Two of the d12 clusters (6 and 8) contained both animal and human strains, indicating possible recent transmission between humans and animals. The detection of only one human strain contained in a d12 cluster with cattle strains may indicate that the overall risk of human infection with Mbov is low with respect to consumption of food (milk, meat) or direct contact to indigenous cattle, while transmission can happen in outbreaks settings.

The study results show that WGS is superior in unequivocally detecting genetic relationship between strains and clarify transmission routes compared to spoligotyping and MIRU-VNTR. While spoligotyping provides some information of strain relatedness, our results demonstrate that it cannot reliably establish clusters of related strains. MIRU-VNTR typing results correlated well with WGS data. However, MIRU-VNTR cannot accurately trace gradual evolution within a transmission cluster. Twenty-one strains clustered by MIRU-VNTR were not clustered by d12 analysis, and four d12 clusters encompassed strains with distinct MIRU-VNTR patterns.

We estimated a mutation rate of $1.14 \times 10^{-7}$ substitutions/nucleotide/year for Mbov. A recent publication on the molecular clock with over 6,000 samples representing the global diversity and covering different epidemiological settings estimated a clock rate
between $1 \times 10^{-8}$ and $5 \times 10^{-7}$, while stating that sampling times below 15-20 years could be insufficient to calibrate a clock rate (67). In another study dealing explicitly with globally distributed Mbov strains, the clock rate was estimated between $6.66 \times 10^{-8}$ and $1.26 \times 10^{-7}$ (61). Our collection of 131 samples of German Mbov strains spans a time period of 19 years, maybe limiting our ability to estimate the clock rate. However, the rate we inferred is in full agreement with estimates published for *M. tuberculosis* outbreaks in Germany (37) and Eurasia (66). Estimates of the effective population size fluctuation through time according to coalescent-based demographic reconstructions suggested that, the German Mbov population went through three successive expansions, a first twentyfold increase in the late middle age, followed by two mild expansions in the mid 18th century and the early 20th century (figure 4). These expansions might be due to increasing growth and movement of human and cattle populations as well as increasing growth of human communities and of intensive animal husbandry with time. The population size sharply declined after the 1970’s, underlining the absence of ongoing epidemics in Germany and confirming the bTB free status of the country. Indirectly supporting the data, the Bayesian skyline detected an anticlimax in the 1740 to 1760 period. This observation coincides with the cattle plague outbreak (RPV virus) that severely impacted the European stocks during that period (68).

In conclusion, in this study for the first time the persistence of infectious cycles of Mbov in the officially bTB free country of Germany over more than ten years has been clearly demonstrated pointing towards the challenges controlling this pathogen. As exemplified here, WGS is definitively the method of choice for establishment of an integrated molecular surveillance of Mbov as well as for outbreak investigations.
Acknowledgements

We thank V. Mohr, F. Boysen, T. Ubben, A. Lüdemann, U. Brommer and G. Kauth for excellent technical assistance. Parts of the work have been funded by grants from German Center for Infection Research, Federal Ministry of Education and Research, Germany, from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 22167-390884018, and grants from the Leibniz Science Campus EvoLUNG.

All authors provided substantial scientific contributions, have read and approved the final manuscript and agreed to the submission. Furthermore, all authors disclose any conflicts of interest relevant to this study.

References


3. Riojas MA, McGough KJ, Rider-Riojas CJ, Rastogi N, Hazbón MH. 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


   www.rki.de › Content › Infekt › EpidBull › Archiv › 2019.


35. The *Mycobacterium bovis* Spoligotype Database. www.mbovis.org


**Figures**

**Figure 1:** Maximum likelihood tree of 131 Mbov strains built from 12,726 SNP positions, annotated with host organism, isolation year, WGS cluster, MIRU-VNTR types, and spoligotypes from the SITVIT (IT) and mbovis.org (SB) databases. Scale
bar indicates the likelihood of per-site substitution and therefore reflects a distance of 127 SNPs baring reverse mutations. Circles on nodes indicate resampling support of at least 90% (green circles) or at least 70% (black circles).

**Figure 2:** Pairwise distance distribution of SNP distances between all sequenced strains (blue) and within WGS d350 groups (red), d30 clusters (purple), and d12 clusters (yellow), with the color indicator for the respective lower thresholds superimposed. The y-axis indicates the total number of pairwise distances and x-axis the number of distinct SNPs.

**Figure 3:** A Maximum parsimony trees for the 13 WGS clusters, annotated with host of isolation. Numbers on branches indicate number of distinct SNPs, distances of 1 are not indicated. B Maximum parsimony trees for the 13 WGS clusters, annotated with MIRU-VNTR types. Numbers on branches indicate number of distinct SNPs, distances of 1 are not indicated.

**Figure 4:** Bayesian skyline plot showing the effective population size of the German Mbov sample through time, estimated from the SNP matrix. According to the coalescent-based approach, the Mbov population went through three successive expansions followed by a final decline. **Plot-in-plot** Root-to-tip genetic distances plotted against sampling dates based on 13 WGS clusters. The figure illustrates a positive correlation ($r^2 = 0.682$) of divergence with sampling date and confirms that Mbov is a measurably evolving population (MEP).
Tables

Table 1: Synopsis of the 13 d12 clusters as deduced from the maximum likelihood tree built from 131 Mbov strains. To the clusters, the number of strains, the years of isolation, spanning time, the maximum distance as indicated by the number of SNPs and the host organisms are annotated.