# Progenitor strain introduction of *Mycobacterium bovis* at the wildlife-livestock interface can lead to clonal expansion of the disease in a single ecosystem

Anzaan Dippenaar[a@], Sven David Charles Parsons[a], Michele Ann Miller[a], Tiny Hlokwe[b], Nicolaas Claudius Gey van Pittius[a], Sabir Abdu Adroub[d], Abdallah Musa Abdallah[d], Arnab Pain[d], Robin Mark Warren[a], Anita Luise Michel[c], Paul David van Helden[a]

[a]DST/NRF Centre of Excellence in Biomedical Tuberculosis Research / SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Francie van Zijl Drive, Cape Town 8000, South Africa;

[b]Tuberculosis Laboratory, ARC-Onderstepoort Veterinary Institute, South Africa;

[c]Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, South Africa;

[d]Pathogen Genomics Laboratory, BESE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

**Institute at which laboratory work was done:** Stellenbosch University, Tygerberg, South Africa, ARC-Onderstepoort Veterinary Institute, Pretoria, South Africa and King Abdullah University of Technology, Thuwal, Saudi Arabia.

**Email addresses:**

A Dippenaar: adippenaar@sun.ac.za

SDC Parsons: sparsons@sun.ac.za

MA Miller: miller@sun.ac.za

T Hlokwe: HlokweT@arc.agric.za

NC Gey van Pittius: ngvp@sun.ac.za

SA Adroub: sabir.adroub@kaust.edu.sa

AM Abdallah: abdallah.abdallah@kaust.edu.sa

A Pain: arnab.pain@kaust.edu.sa

RM Warren: rw1@sun.ac.za

AL Michel: anita.michel@up.ac.za

PD van Helden: pvh@sun.ac.za

**[@]Corresponding Author:**

Anzaan Dippenaar

DST/NRF Centre of Excellence in Biomedical Tuberculosis Research

SAMRC Centre for Tuberculosis Research

Division of Molecular Biology and Human Genetics

Faculty of Medicine and Health Sciences

Stellenbosch University

PO Box 241

Francie van Zijl Drive

Cape Town

8000

South Africa

E-mail address: adippenaar@sun.ac.za

Telephone: +27 21 9389073

## HIGHLIGHTS

• 17 M. BOVIS isolates from various locations in South Africa were whole genome sequenced.

• Introduction of a single M. BOVIS strain type in the Kruger National Park lead to disease in multiple host species.

• M. BOVIS strains analyzed here seem to be geographic rather than host species-specific.

**Abstract:**

*Mycobacterium bovis* infects multiple wildlife species and domesticated cattle across South Africa, and negatively impacts on livestock trade and movement of wildlife for conservation purposes. *M. bovis* infection was first reported in the Kruger National Park (KNP) in South Africa during the 1990s, and has since spread to infect numerous animal host species throughout the park and across South Africa. Whole genome sequencing data of 17 *M. bovis* isolates were analyzed to investigate the genomic diversity among *M. bovis* isolates causing disease in different animal host species from

various locations in South Africa. *M. bovis* strains analyzed in this study are geographic rather than host species-specific. The clonal expansion of *M. bovis* in the KNP highlights the effect of an introduction of a transmissible infectious disease leading to a rising epidemic in wildlife, and emphasizes the importance of disease control and movement restriction of species that serve as disease reservoirs. In conclusion, the point source introduction of a single *M. bovis* strain type in the KNP ecosystem lead to an *M. bovis* outbreak in this area that affects various host species and poses an infection risk in neighboring rural communities where HIV prevalence is high.

**Keywords:**

Kruger National Park, *Mycobacterium bovis*, whole genome sequence, bovine tuberculosis

## Introduction

*Mycobacterium bovis* infects a wide variety of wildlife species and domesticated cattle across South Africa, which negatively impacts on international and inter-regional livestock trade and movement of wildlife for conservation purposes (Michel et al., 2006). The presence of *M. bovis* wildlife reservoir hosts add considerably to the difficulties regarding the control and eradication of this disease (Briones et al., 2000; Corner, 2006). Additionally, *M. bovis* is a zoonotic disease threatening rural communities where there is a close animal-human interface and high prevalence of human immunodeficiency virus (HIV) infection (Hlavsa et al., 2008; Müller et al., 2013; Rodwell et al., 2008).

Retrospective analysis of the state veterinary records suggests that *M. bovis* was first introduced into the Kruger National Park (KNP) in the late 1950s or early 1960s, presumably as a result of primarily, cattle to buffalo transmission (Michel et al., 2009). In less than 3 decades, *M. bovis* infection has moved through the ecosystem to infect multiple host species, including buffaloes, lions, kudus, leopards and baboons, and has spread northwards, covering the long axis (360 km) of the park in less than 3 decades (Bengis et al., 1996; Michel et al., 2009). Previously, no host species-specific variants in *M. bovis* isolates were found using the *Mycobacterium tuberculosis* complex-specific insertion sequence, IS*6110*, and polymorphic GC-rich sequences (PGRS) restriction length polymorphism

(RFLP) typing (Michel et al., 2009). However, a putative evolutionary reconstruction of *M. bovis* in the KNP region and Hluhluwe-iMfolozi Park, roughly 400 km south of KNP, suggests that the bovine tuberculosis epidemics in these two conservation areas have been caused by different progenitor strains (Hlokwe et al., 2011). In one report, a single genotype of *M. bovis* was found in different animal species, challenging the hypothesis that evolution may be accelerated during cross-over into new species, or that species-specific variants exist (Michel et al., 2009).

In this study, whole genome sequencing data of 17 *M. bovis* isolates were analyzed to investigate the genomic diversity among *M. bovis* isolates causing disease in different animal host species from various locations in South Africa. We used this data to derive an evolutionary scenario for the population structure of *M. bovis* from various locations in South Africa. We suggest that *M. bovis* infection in the KNP resulted from point source introduction followed by clonal expansion in multiple host species.

## 2. Materials and Methods

Additional information on the workflow is provided in the supplementary material (Supplemental Methods).

### *2.1. Isolate selection and whole genome sequencing*

Seventeen *M. bovis* isolates from various host species from different locations in South Africa (KNP, KZN, Western Cape, North-West) were selected from an in-house sample bank (Michel et al., 2009; Mostowy et al., 2005). *M. bovis* isolates were cultured and DNA was extracted as previously described (Warren et al., 2006). Whole genome sequencing was done on either the Illumina HiSeq2000 or MiSeq platform (Illumina, California, USA) using a paired-end approach (2 x 100 bp) with ~500 base fragment sizes which resulted in insert sizes between 250 and 550 bases. One microgram of DNA was used to prepare libraries for sequencing per the manufacturer's instructions using either the Illumina TruSeq DNA LT or the Nextera DNA sample preparation kit (Illumina, Inc, San Diego, CA).

### *2.2. Data analyses*

*Pre-processing, alignment and variant calling*

The Illumina paired-end reads published in this study and 89 genomes published previously or available in public databases (Supplemental Data 1, Table S1) were analysed with open source software as described previously (Black et al., 2015; Dippenaar et al., 2015). Briefly, reads were trimmed from the 3' end based on quality with Trimmomatic (Bolger et al., 2014). The sequencing reads were aligned to the reference genome (*Mycobacterium tuberculosis* H37Rv, GenBank NC000962.2 and *M. bovis*, GenBank NC002945.3) to detect genomic variants. Three alignment algorithms were used, namely, Novoalign (Novocraft), Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), and SMALT (Ponstingl and Ning, 2010). An average of 100x depth of coverage was obtained for all *M. bovis* isolates sequenced for this study. The Genome Analysis Tool Kit (GATK) (McKenna et al., 2010) was used to identify single nucleotide variants in all the alignment files from the different mapping algorithms used. Only variants identified by the GATK in all three alignments that overlap in both position and base identity were used and further filtered to exclude variants in *pe/ppe* family genes, repeat regions, insertion sequences and phages, as described previously (Coscolla et al., 2013). Deleted regions in the *M. bovis* genomes with respect to the *M. tuberculosis* H37Rv genome were detected with DELLY and alignments of these regions were visually inspected (Rausch et al., 2012). SpolPred was used to predict the spoligotype patterns of all *M. bovis* clinical isolates (Coll et al., 2012).

*Phylogenetic analysis*

Concatenated sequences containing 36 788 high-confidence variable sites (coding and non-coding single nucleotide variants) were used to construct a maximum likelihood phylogeny of the isolates included in this analysis with RaxML with 1000 bootstrap pseudo-replicates (Stamatakis, 2006, 2014).

**Table 1. *M. bovis* isolate information**

| Sample ID | **Location** | **Animal** | **IS6110 classification** (Michel et al., 2009) | **Year of isolation** | ***In silico* spoligotype** | **SB number** |
|---|---|---|---|---|---|---|
| 1474 | KNP | Lion | C8 | 1998 |  | SB0121 |
| 1771 | KNP | Leopard | C8v | 2000 |  | SB0121 |
| 1067(2) | St. Lucia | Buffalo | C2 | 1997 |  | SB0130 |
| 1081(1) | KNP | Kudu | C8v | 1997 |  | SB0121 |
| 1457(4) | KNP | Buffalo | C8v | 1999 |  | SB0121 |
| 150(3) | KNP | Buffalo | C8v | 1998 |  | SB0121 |
| 1531(10) | Spioenkop (KZN) | Buffalo | Not done | 1999 |  | SB0140 |
| 1865(A) | KNP | Buffalo | C8v | 2000 |  | SB0121 |
| 2837(33) | KNP | Buffalo | C8v | 2001 |  | SB0121 |
| 3912 (B) | Mpumalanga | Cattle | C1 | 2003 |  | SB0130 |
| 440(5) | KNP | Buffalo | C8v | 1998 |  | SB0121 |
| 659(A) | KNP | Buffalo | C8v | 1996 |  | SB0121 |
| 734(16) | KNP | Baboon | C8v | 1996 |  | SB0121 |
| 747(11) | KNP | Buffalo | C8 | NA |  | SB0121 |
| 9969-1 | Cape Town | Cattle | C8v | NA |  | SB0267 |
| 1595-F(24) | KNP | Lion | C8v | 1999 |  | SB0121 |
| 1199 B(19) | MGR | Lion | C2 | 1998 |  | SB0130 |

KNP = Kruger National Park, KZN = KwaZulu-Natal, MGR = Madikwe Game Reserve.

## 3. Results

### *3.1. In silico spoligotyping*

*In silico* predicted spoligotype patterns agreed with spoligotyping done on isolates included in a study by Mostowy *et al*., 2005 (659, 1595F, 734, 9969-1). Strains analyzed in our study did not have the RDAf1 deletion or deletion of spacer 30 in the standard spoligotyping scheme, in accordance with a study by Müller *et al.* (2009). Isolate information and *in silico* spoligotyping results are summarized in Table 1.

### *3.2. Phylogenetic analysis*

A maximum likelihood phylogenetic reconstruction was done and included isolates sequenced for this study as well as previously published *M. tuberculosis* isolates, representative of lineage 1 – 4, and -7, *M. africanum* isolates (lineage 5,6), and various animal-adapted species of the *Mycobacterium* genus (Figure 1). The phylogeny confirmed the clustering of *M. bovis* isolates cultured from different animal hosts from the KNP. Variant (single nucleotide polymorphisms (SNPs) and small insertions and deletions) distances between isolates in this clade (indicated in green in Figure 1) were between 5 and 25 variants. When considering all isolates from KNP, this group had 246 unique SNVs with respect to the *M. tuberculosis* H37Rv reference genome, compared to all other *M. tuberculosis* complex isolates analyzed here, with the majority of changes occurring in genes encoding for proteins involved in the functional categories of cell wall and cell processes, intermediary metabolism and respiration.

### *3.3. Regions of difference*

*M. bovis* isolates from various animal hosts in the KNP have a distinct deletion profile (which is consistent among all isolates) as opposed to *M. bovis* isolates from other regions in South Africa which show more diverse deletion profiles, suggestive of regional genomic differences (Supplemental Data 2, Table S2). All isolates from KNP have RDbovis(c)_Kruger (affecting Rv1355c-Rv1374c) deleted and RD17 present in comparison to other *M. bovis* isolates included here which have the RDbovis(c)_Kruger region present and RD17 (affecting Rv1563c) deleted. These distinct genomic

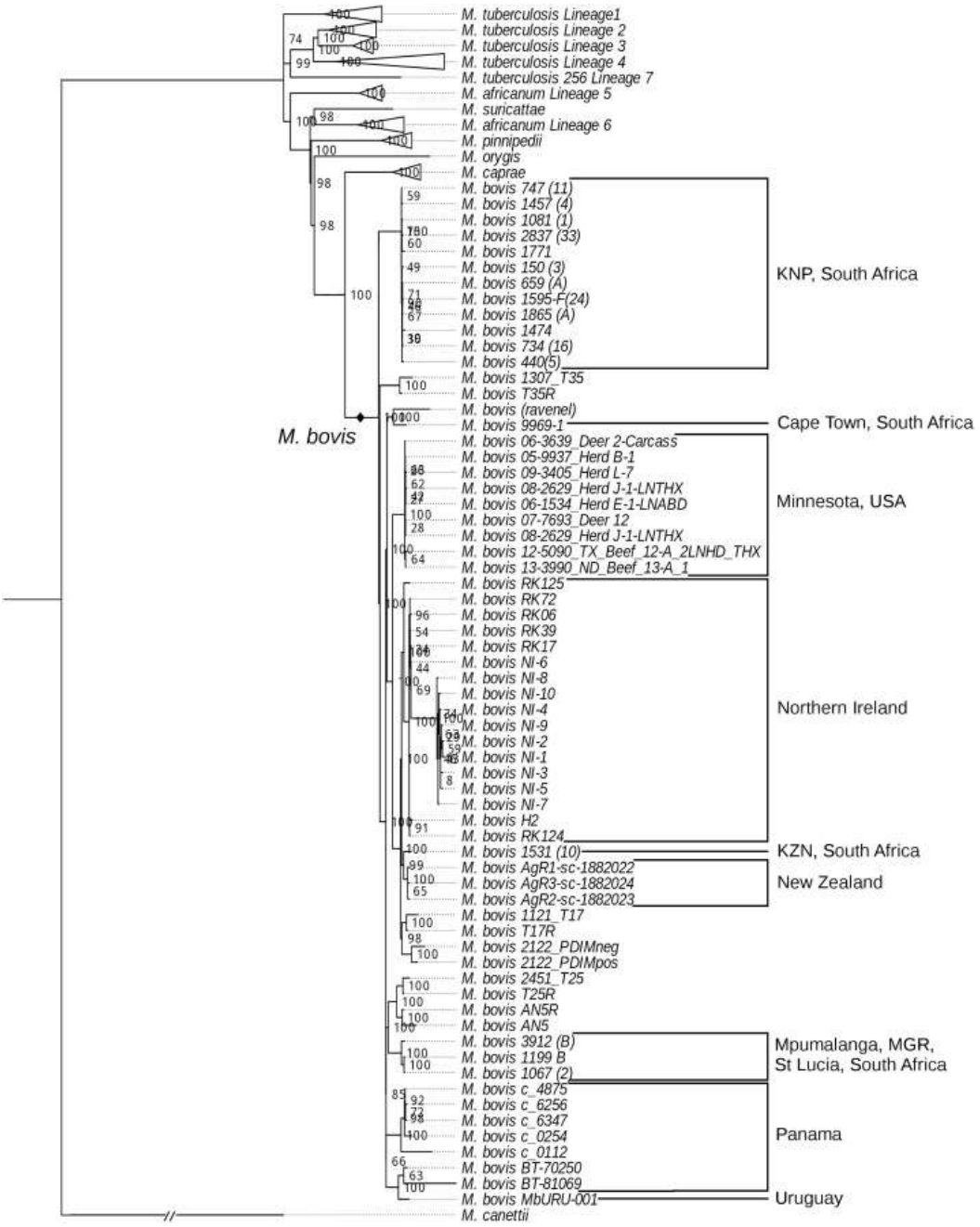


**Figure 1. Molecular Phylogenetic analysis by Maximum Likelihood method.** The bootstrap consensus tree based on 36 788 variable positions inferred from 1000 replicates (Felsenstein, 1989) is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1989). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the nodes. The tree shown was produced by RaxML and is based on variable positions identified with respect to *M. tuberculosis* H37Rv (Stamatakis, 2006, 2014). The *M. bovis* clade is indicated with a diamond (♦) on phylogenetic tree.

profiles suggest that regional genomic characteristics can be used to track the movement of the disease within the country.

## 4. Concluding remarks

The clear clustering of *M. bovis* isolates from the KNP and the evolutionary distance between these strains and *M. bovis* isolates from elsewhere in South Africa suggests that evolution from the common ancestors occurred in South Africa after multiple strains of *M. bovis* were introduced to the country on different occasions and these variants then became endemic to different areas. Furthermore, we show that the *M. bovis* strains analyzed in this study seem to be geographic rather than host species-specific, considering the genomic characteristics analyzed here. The clonal expansion shows how, following the introduction of *M. bovis* to a naïve ecosystem, this pathogen can spread relatively rapidly to multiple host species and across a large geographic range. This emphasizes the importance of disease control and movement restriction of species that may serve as disease reservoirs. The point source introduction of a single *M. bovis* strain type in the KNP ecosystem thus lead to an *M. bovis* outbreak in this area which affects multiple host species and poses an infection risk in neighboring rural communities where HIV infection is high.

## Conflict of interest

The authors declare that they have no competing interests of either a financial or nonfinancial nature regarding the work described in the present manuscript and its publication.

## Sequence data

The whole genome sequencing data were deposited in the European Nucleotide Archive with the following accession number: PRJEB18668 and are available at: http://www.ebi.ac.uk/ena/data/view/ PRJEB18668.

**Acknowledgements**

This study was enabled by funding from the South African Medical Research Council, the National Research Foundation, and faculty baseline funding from King Abdullah University of Science and Technology (KAUST, Thuwal, Saudi Arabia) awarded to A Pain. MA Miller is funded by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (NRF) of South Africa, award number UID 86949. Any opinion, finding and conclusion or recommendation expressed in this material is that of the author(s) and the NRF does not accept any liability in this regard. The authors declare that no conflict of interest exists.

**References**

Bengis, R.G., Kriek, N.P., Keet, D.F., Raath, J.P., de Vos, V., Huchzermeyer, H.F., 1996. An outbreak of bovine tuberculosis in a free-living African buffalo (Syncerus caffer--sparrman) population in the Kruger National Park: a preliminary report. Onderstepoort J. Vet. Res. 63, 15–18.

Black, P.A., Vos, M. de, Louw, G.E., Merwe, R. van der, Dippenaar, A., Streicher, E.M., Abdallah, A.M., Sampson, S.L., Victor, T.C., Dolby, T., Simpson, J.A., Helden, P. van, Warren, R.M., Pain, A., 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates. BMC Genomics 16, 857. doi:10.1186/s12864-015-2067-2

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics btu170. doi:10.1093/bioinformatics/btu170

Briones, V., de Juan, L., Sánchez, C., Vela, A.I., Galka, M., Montero, Goyache, J., Aranaz, A., Domìnguez, L., 2000. Bovine tuberculosis and the endangered Iberian lynx. Emerg. Infect. Dis. 6, 189–191.

Coll, F., Mallard, K., Preston, M.D., Bentley, S., Parkhill, J., McNerney, R., Martin, N., Clark, T.G., 2012. SpolPred: rapid and accurate prediction of Mycobacterium tuberculosis spoligotypes from short genomic sequences. Bioinforma. Oxf. Engl. 28, 2991–2993. doi:10.1093/bioinformatics/bts544

Corner, L.A.L., 2006. The role of wild animal populations in the epidemiology of tuberculosis in domestic animals: How to assess the risk. Vet. Microbiol., 4th International Conference on Mycobacterium bovis 4th International Conference on Mycobacterium bovis 112, 303–312. doi:10.1016/j.vetmic.2005.11.015

Coscolla, M., Lewin, A., Metzger, S., Maetz-Rennsing, K., Calvignac-Spencer, S., Nitsche, A., Dabrowski, P.W., Radonic, A., Niemann, S., Parkhill, J., Couacy-Hymann, E., Feldman, J., Comas, I., Boesch, C., Gagneux, S., Leendertz, F.H., 2013. Novel *Mycobacterium tuberculosis* Complex Isolate from a Wild Chimpanzee. Emerg. Infect. Dis. 19, 969–976. doi:10.3201/eid1906.121012

Dippenaar, A., Charles Parsons, S.D., Sampson, S.L., Gerhard van der Merwe, R., Drewe, J.A., Abdallah, A.M., Siame, K.K., van Pittius, N.C.G., van Helden, P.D., Pain, A., Warren, R.M., 2015. Whole genome sequence analysis of Mycobacterium suricattae. Tuberculosis. doi:10.1016/j.tube.2015.10.001

Felsenstein, J., 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164–166.

Hlavsa, M.C., Moonan, P.K., Cowan, L.S., Navin, T.R., Kammerer, J.S., Morlock, G.P., Crawford, J.T., Lobue, P.A., 2008. Human tuberculosis due to Mycobacterium bovis in the United States,

1995-2005. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 47, 168–175. doi:10.1086/589240

Hlokwe, T.M., Jenkins, A.O., Streicher, E.M., Venter, E.H., Cooper, D., Godfroid, J., Michel, A.L., 2011. Molecular characterisation of Mycobacterium bovis isolated from African buffaloes (Syncerus caffer) in Hluhluwe-iMfolozi Park in KwaZulu-Natal, South Africa. Onderstepoort J. Vet. Res. 78, 39–44.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25, 1754–1760. doi:10.1093/bioinformatics/btp324

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. doi:10.1101/gr.107524.110

Michel, A.L., Bengis, R.G., Keet, D.F., Hofmeyr, M., Klerk, L.M. de, Cross, P.C., Jolles, A.E., Cooper, D., Whyte, I.J., Buss, P., Godfroid, J., 2006. Wildlife tuberculosis in South African conservation areas: implications and challenges. Vet. Microbiol. 112, 91–100. doi:10.1016/j.vetmic.2005.11.035

Michel, A.L., Coetzee, M.L., Keet, D.F., Maré, L., Warren, R., Cooper, D., Bengis, R.G., Kremer, K., van Helden, P., 2009. Molecular epidemiology of Mycobacterium bovis isolates from free-ranging wildlife in South African game reserves. Vet. Microbiol. 133, 335–343. doi:10.1016/j.vetmic.2008.07.023

Mostowy, S., Inwald, J., Gordon, S., Martin, C., Warren, R., Kremer, K., Cousins, D., Behr, M.A., 2005. Revisiting the evolution of Mycobacterium bovis. J. Bacteriol. 187, 6386–6395. doi:10.1128/JB.187.18.6386-6395.2005

Müller, B., Dürr, S., Alonso, S., Hattendorf, J., Laisse, C.J.M., Parsons, S.D.C., van Helden, P.D., Zinsstag, J., 2013. Zoonotic *Mycobacterium bovis* –induced Tuberculosis in Humans. Emerg. Infect. Dis. 19, 899–908. doi:10.3201/eid1906.120543

Ponstingl, H., Ning, Z., 2010. SMALT - A new mapper for DNA sequencing reads. F1000Posters 1.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., Korbel, J.O., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinforma. Oxf. Engl. 28, i333–i339. doi:10.1093/bioinformatics/bts378

Rodwell, T.C., Moore, M., Moser, K.S., Brodine, S.K., Strathdee, S.A., 2008. Tuberculosis from Mycobacterium bovis in binational communities, United States. Emerg. Infect. Dis. 14, 909–916. doi:10.3201/eid1406.071485

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics btu033. doi:10.1093/bioinformatics/btu033

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. doi:10.1093/bioinformatics/btl446

Warren, R., de Kock, M., Engelke, E., Myburgh, R., Gey van Pittius, N., Victor, T., van Helden, P., 2006. Safe Mycobacterium tuberculosis DNA extraction method that does not compromise integrity. J. Clin. Microbiol. 44, 254–256. doi:10.1128/JCM.44.1.254-256.2006