



OPEN ACCESS

EDITED BY

Axel Cloeckaert,
Institut National de Recherche pour
l'Agriculture, l'Alimentation et l'Environnement
(INRAE), France

REVIEWED BY

Lorraine Michelet,
Agence Nationale de Sécurité Sanitaire de
l'Alimentation, de l'Environnement et du Travail
(ANSES), France
Amandine Hauer,
University of Gdańsk, Poland
Linzy Elton,
University College London, United Kingdom
Ramon A. Juste,
NEIKER, Basque Institute for Agricultural
Research and Development, Spain

*CORRESPONDENCE

Wynand J. Goosen
✉ wjgoosen@sun.ac.za

†These authors have contributed equally to this
work and share first authorship

†These authors share senior authorship

RECEIVED 05 October 2023

ACCEPTED 23 October 2023

PUBLISHED 21 November 2023

CITATION

Ghielmetti G, Loubser J, Kerr TJ, Stuber T,
Thacker T, Martin LC, O'Hare MA, Mhlophe SK,
Okunola A, Loxton AG, Warren RM,
Moseley MH, Miller MA and Goosen WJ (2023)
Advancing animal tuberculosis surveillance
using culture-independent long-read whole-
genome sequencing.
Front. Microbiol. 14:1307440.
doi: 10.3389/fmicb.2023.1307440

COPYRIGHT

© 2023 Ghielmetti, Loubser, Kerr, Stuber,
Thacker, Martin, O'Hare, Mhlophe, Okunola,
Loxton, Warren, Moseley, Miller and Goosen.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Advancing animal tuberculosis surveillance using culture-independent long-read whole-genome sequencing

Giovanni Ghielmetti^{1,2†}, Johannes Loubser^{1†}, Tanya J. Kerr¹,
Tod Stuber³, Tyler Thacker³, Lauren C. Martin¹,
Michaela A. O'Hare¹, Sinegugu K. Mhlophe¹, Abisola Okunola¹,
Andre G. Loxton¹, Robin M. Warren¹, Mark H. Moseley⁴,
Michele A. Miller^{1‡} and Wynand J. Goosen^{1*†}

¹Division of Molecular Biology and Human Genetics, South African Medical Research Council Centre for Tuberculosis Research, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, ²Section of Veterinary Bacteriology, Institute for Food Safety and Hygiene, Vetsuisse Faculty, University of Zurich, Zurich, Switzerland, ³National Veterinary Services Laboratories, Veterinary Services, Animal and Plant Health Inspection Service, United States Department of Agriculture, Ames, IA, United States, ⁴School of Biological Sciences, University of Aberdeen, Aberdeen, United Kingdom

Animal tuberculosis is a significant infectious disease affecting both livestock and wildlife populations worldwide. Effective disease surveillance and characterization of *Mycobacterium bovis* (*M. bovis*) strains are essential for understanding transmission dynamics and implementing control measures. Currently, sequencing of genomic information has relied on culture-based methods, which are time-consuming, resource-demanding, and concerning in terms of biosafety. This study explores the use of culture-independent long-read whole-genome sequencing (WGS) for a better understanding of *M. bovis* epidemiology in African buffaloes (*Syncerus caffer*). By comparing two sequencing approaches, we evaluated the efficacy of Illumina WGS performed on culture extracts and culture-independent Oxford Nanopore adaptive sampling (NAS). Our objective was to assess the potential of NAS to detect genomic variants without sample culture. In addition, culture-independent amplicon sequencing, targeting mycobacterial-specific housekeeping and full-length 16S rRNA genes, was applied to investigate the presence of microorganisms, including nontuberculous mycobacteria. The sequencing quality obtained from DNA extracted directly from tissues using NAS is comparable to the sequencing quality of reads generated from culture-derived DNA using both NAS and Illumina technologies. We present a new approach that provides complete and accurate genome sequence reconstruction, culture independently, and using an economically affordable technique.

KEYWORDS

adaptive sampling, African buffaloes, culture-independent, *Mycobacterium bovis*, next-generation sequencing, whole-genome sequencing

Introduction

Animal tuberculosis (TB) caused by *Mycobacterium bovis* (*M. bovis*) is a significant infectious disease affecting a wide range of domesticated and wild animals, representing a worldwide public health concern (Muller et al., 2013; Meiring et al., 2018). In South Africa, animal TB poses a serious threat to wildlife, including the iconic African buffalo (*Syncerus caffer*) (Bernitz et al., 2019). These animals serve as a reservoir for *M. bovis* and can transmit the pathogen to domestic cattle and other susceptible species (Sichewo et al., 2019). The implications of *M. bovis* infection in wildlife extend beyond the health of individual animals, as it can have far-reaching consequences for conservation efforts, public health, and the livestock industry (Muller et al., 2013).

Efficient and accurate surveillance and management strategies are crucial for mitigating the impact of animal TB in wildlife populations (Ortiz et al., 2021; Perea et al., 2021). Currently, detection and characterization of *M. bovis* has relied on culture-based methods, where bacterial isolates are obtained from infected animals and subjected to phenotypic and genotypic analyses (Goosen et al., 2022). However, this approach has inherent limitations, particularly when dealing with wildlife populations in general and foot-and-mouth virus endemic regions where movements of animals or samples are not permitted (NAHE, 2022). In addition, the isolation and culture of *M. bovis* from wildlife samples are challenging due to the complex microbiota and the presence of non-culturable or fastidious strains (Bull et al., 2016; Goosen et al., 2022). Moreover, the culture step is time-consuming and labor-intensive, impeding real-time surveillance and response (Perea et al., 2021).

In recent years, whole-genome sequencing (WGS) has revolutionized the field of infectious disease genomics, offering unprecedented resolution and insights into pathogen biology, transmission dynamics, and antimicrobial resistance (Meehan et al., 2019). WGS enables the comprehensive characterization of bacterial genomes, including the identification of strain types, the detection of genetic variations, and the prediction of drug resistance profiles (Rossi et al., 2023). In the context of *M. bovis*, WGS has proven invaluable in understanding the transmission patterns and evolution of the pathogen, guiding control measures, and investigating outbreaks (Rossi et al., 2023).

In South Africa, the implementation of *M. bovis* WGS in wildlife populations holds great promise for improved understanding of animal TB epidemiology and management. By obtaining high-quality genomic data directly from infected animals, it becomes possible to overcome the limitations of culture-based methods and obtain a more accurate representation of *M. bovis* diversity and population structure (Martin et al., 2022; Su et al., 2023). Furthermore, WGS allows for the identification of specific genetic markers associated with virulence, host adaptation, and antimicrobial resistance, providing valuable insights for targeted interventions (Pan et al., 2011).

This study explored a novel sequencing approach that has the potential to address and bridge the existing knowledge gap by comparing two next-generation sequencing (NGS) platforms for *M. bovis* detection and characterization in South African buffalo populations. The generation of whole-genome sequences for transmission and surveillance studies of animal TB based on culture extracts can be impeded by coinfection or contamination of samples with other microorganisms that outcompete the target agents causing animal TB (Maiga et al., 2012; Riello et al., 2016; Gopalaswamy et al.,

2020). Co-infections can also modulate disease outcomes, particularly in infections with closely related microorganisms that share antigenic properties and mimic immunological escape mechanisms (Gorsich et al., 2018). Recent studies have demonstrated the advantages of using full-length 16S rRNA amplicon sequencing for taxonomic classification, as opposed to short-read amplicon sequencing (Earl et al., 2018; Callahan et al., 2019; Johnson et al., 2019). Nonetheless, previous methods for high-throughput full-length 16S rRNA gene amplicon sequencing using Oxford Nanopore Technologies (ONT) often relied solely on reference database alignment (Curry et al., 2022; Zorz et al., 2023) rather than employing *de novo* generation of sequence features, such as ASVs (amplicon sequence variants) or OTUs (operational taxonomic units). A major drawback for ONT has been its base-calling error rate. However, the recently released sequencing chemistry and base-calling software have been significantly improved, placing ONT as an attractive and reliable alternative to the Illumina platform (Oude Munnink et al., 2020; Smith et al., 2021). By using reference-free bioinformatic analyses of multiple housekeeping genes amplified using target specific primers on the ONT platform, we aimed to detect the presence of microbial contamination, including non-tuberculous mycobacteria (NTM), in each sample (Macovei et al., 2015; Vierstraete and Braeckman, 2022). We compared the performance of Illumina and Nanopore adaptive sampling (NAS) WGS on cultured extracts and NAS directly on DNA extracted from tissue homogenates. This study aimed to assess the accuracy and resolution of each method and explore the potential of NAS combined with an appropriate data analysis pipeline as a culture-independent alternative for comprehensive genomic surveillance studies. Based on these data, the ability of the NAS approach to distinguish target sequences from host genomic material was evaluated.

This research presents novel culture-independent, long-read based, and more accessible NGS approaches for a deeper understanding of the epidemiology and transmission dynamics of animal TB. The findings will inform evidence-based strategies for disease management, risk assessment, and control measures in wildlife populations, with implications for both animal health and human health. Furthermore, this study showcases the value of WGS as a powerful tool in wildlife disease surveillance, highlighting its potential for improving conservation efforts, preventing zoonotic transmission, and safeguarding the livestock industry.

Methods

Sample collection and culture

In 2018, a *M. bovis* test-and-slaughter program was conducted in Hluhluwe-iMfolozi Park (HiP), South Africa (SA), on a herd of African buffaloes (Bernitz et al., 2020; Smith et al., 2021). All animals were captured, immobilized and whole blood collected as previously described (Parsons et al., 2011). All animals of the herd ($n = 50$) were euthanized due to their high cell-mediated immunological (CMI) responses towards *M. bovis* specific antigens (>50% of the animals). Various tissue samples (head, thorax, lung lymph nodes and lung tissue) were collected during in-depth necropsies and lesion scores captured as previously described (Table 1; Bernitz et al., 2019). Mycobacterial cultures were performed on all samples using the BACTEC™ MGIT™ 960 Mycobacterial Detection System (Becton Dickinson). For each sample, parallel uncultured tissue homogenate

aliquots were obtained. In brief, approximately 10 g of tissue were homogenized in 50-mL skirted tubes (Becton Dickinson) containing eight 4.8-mm metal beads and 4 mL of sterile PBS using a blender (Bullet Blender 50; Next Advance) for 15 min at maximum speed as previously described (Goosen et al., 2022). The homogenates were subsequently preserved (-80°C) for subsequent DNA extractions, repeat culturing and/or future sequencing. All culture-positive culture crude extracts were subjected to speciation using a PCR targeting genetic regions of difference to confirm *M. bovis* infections, as previously described (Warren et al., 2006). Further genetic speciation using spoligotyping was performed (Belakehal et al., 2022). Thereafter, 1 mL of each *M. bovis*-positive MGIT was further inoculated into 5 mL Middlebrook 7H9 media containing sodium pyruvate and incubated for an additional 21 days at 37°C for DNA extraction (Figure 1).

Sample selection

For this pilot study, frozen native tissue homogenates from seven African buffaloes were retrospectively selected for downstream analysis. Selection criteria were based on the (1) quality of whole-genome sequences generated using the Illumina MiSeq instrument from culture, (2) tissue types, (3) lesion scores, and (4) varying amounts target genomic DNA. In detail, seven culture extracts obtained from tissue homogenates and sequenced with Illumina were selected based on the sequencing quality (coverage $>99\%$, mean

depth $>60\times$, and $>99\%$ of mapped reads to the reference genome). In total, 3 lung samples, 1 retropharyngeal, 1 tonsil, 1 parotid, and 1 pool of the head lymph nodes presenting different lesion scores ranging from 0–3 were included (Bernitz et al., 2019; Table 1). The amount of *M. bovis* genomic DNA present in each native tissue homogenate was assed based on insertion elements *IS6110* and *IS1081* (combined detection) detected by Cepheid's GeneXpert MTB/RIF Ultra (Ct values: 15.1–22.5), as previously described (Hlokwe and Mogano, 2020). Additionally, for each sample selected, a corresponding aliquot of the crude 7H9 mycobacterial culture was included for DNA extraction and NAS whole-genome sequencing alongside the corresponding native tissue homogenate.

Whole-genome sequencing processing

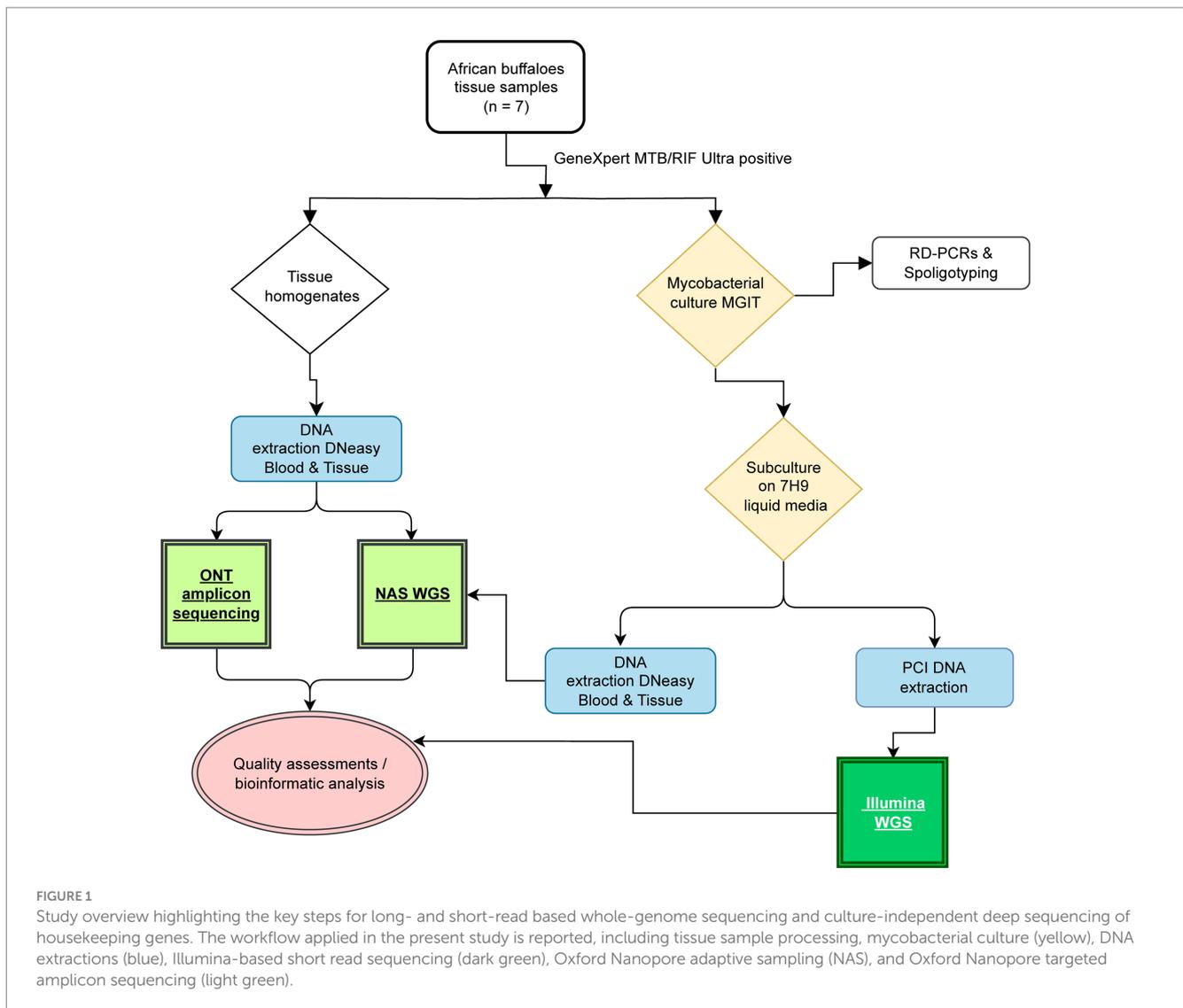
Illumina platform

Genomic DNA was extracted from heat-inactivated (98°C for 45 min) culture pellets and WGS was performed as previously described (van Soolingen et al., 1991). Briefly, libraries were prepared with the Nextera DNA Flex Library Prep Kit (Illumina). The quality of each library was assessed with a FragmentAnalyzer Automated CE System (Agilent) using a Next Generation Sequencing Fragment Kit (1–6,000 bp; Agilent). Libraries were paired-end sequenced in 250 bp reads on an Illumina MiSeq System using the MiSeq Reagent Kit (version 3, 600-cycle) (Ortiz et al., 2021).

TABLE 1 Metadata of the seven African buffaloes included in the present study.

Culture ID	Tissue type	Lesions score	Age	Gender	Mycobacterial culture		Culture-independent PCR	
					Final culture RD-PCR result	Spoligotyping	Tissue - GeneXpert MTB/RIF ultra	<i>IS6110</i> / <i>IS1081</i> Ct-values
NB18193	Lung	2	Adult	Female	<i>M. bovis</i>	SB1474	MTB DETECTED HIGH	15.8
NB18254	Retropharyngeal	2	Juvenile	Female	<i>M. bovis</i>	SB1474	MTB DETECTED HIGH	15.1
NB18212	Lung	2	Juvenile	Female	<i>M. bovis</i>	SB1474	MTB DETECTED LOW	22.5
NB18191	Parotid	1	Adult	Female	<i>M. bovis</i>	SB1474	MTB DETECTED MEDIUM	16
NB18211	Tonsil	1	Adult	Female	<i>M. bovis</i>	SB1474	MTB DETECTED MEDIUM	16.8
NB18246	Pooled head	0	Juvenile	Female	<i>M. bovis</i>	SB1474	MTB DETECTED HIGH	15.6
NB18249	Lung	3	Adult	Female	<i>M. bovis</i>	SB1474	MTB DETECTED MEDIUM	16.3

Sample source, speciation using conventional molecular methods, and results of GeneXpert MTB/RIF Ultra applied on tissue homogenates is reported.



Oxford Nanopore

Total DNA was extracted using a modification of the DNeasy Blood and Tissue kit (Qiagen) directly from (a) 1 mL uncultured tissue homogenate aliquots ($n = 7$) and (b) paired 1 mL 7H9 crude mycobacterial culture pellets ($n = 7$) as previously described (Stanton et al., 2010). In brief, subsequently to heat-inactivation (98°C for 45 min) and centrifugation (1,500 \times g for 10 min), 300 μ L of Buffer ATL was added to the cell pellet. A total of 25 μ L Proteinase K was added and left for digestion at 56°C at 600 rpm in a thermo mixer overnight. The solution was centrifuged at 5,500 \times g for 5 min, 500 μ L of the supernatant recovered, and transferred to a 1.5 mL tube. Thereafter, 400 μ L of buffer AL and 400 μ L of ethanol were added to the solution and transferred to a Mini Spin Column. Finally, the exposed DNA was purified using washbuffers AW1 and AW2 and eluted in 60 μ L of buffer AE prewarmed at 54°C. DNA concentrations were determined using the Qubit 1x dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific), according to the manufacturer's instructions (Supplementary Table S1).

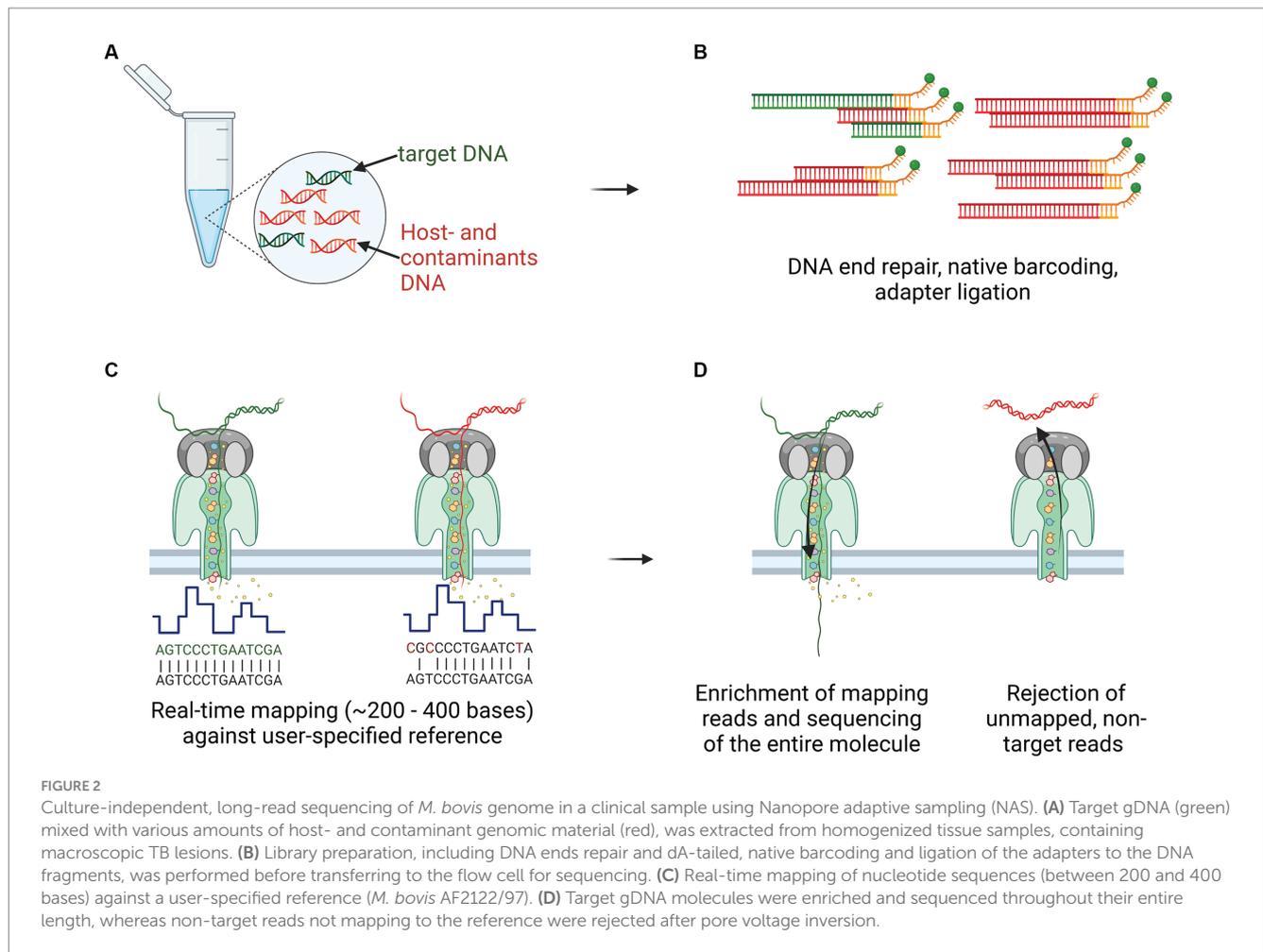
Nanopore adaptive sampling

The DNA from culture-independent tissue homogenates and crude 7H9 *M. bovis* positive culture extracts from the same animals

were all first diluted with EDTA (pH = 8) to a final concentration of 34 ng/ μ L. Thereafter, 12 μ L (~400 ng) were FFPE repaired and end-prepped using NEBNext FFPE DNA Repair Mix and NEBNext Ultra II End Repair/da-tailing module (New England Biolabs, Inc.). Each sample was individually barcoded using NEB Blunt/TA ligation (New England Biolabs, Inc.) and the Native Barcoding Kit 96 V14 (SQK-NBD114.96), as described by the manufacturer (ONT). Adapter ligations were performed on each barcoded DNA library using Quick T4 DNA Ligase in the NEBNext Quick Ligation Module and NEBNext Quick Ligation Reaction Buffer (New England Biolabs, Inc.). Thereafter, each DNA library was loaded onto separate R10.4.1 flow cells (>1,250 pores available, FLO-MIN114, LOT 11002667, SN B024009663), adaptive whole genome sequenced using reference genome *M. bovis* AF2122/97 (Accession NC_002945.4) for enrichment on the MinION Mk1C device for 22–27 h (Figure 2).

Detection of contaminants using ONT targeted amplicon sequencing

The amplification and deep-sequencing of three housekeeping genes, namely *hsp65* (441 bp) (Telenti et al., 1993), *rpoB* (680 bp) and the full-length 16S rRNA gene (~1,500 bp) (Johnson et al., 2019), was performed culture-independently directly on DNA extracted from



buffalo tissue homogenates as previously described (Clarke et al., 2022). For the *rpoB* gene, newly designed primers were applied (F1-CGTGTGTATGTGGCTCAGAA and R1-GTGTATCGGACTTGATGGT). A 25 μ L reaction contained 14 μ L Q5 High-Fidelity 2X Master Mix (New England Biolabs, Inc.), 0.5 μ L of each 50 μ M primer stock solution, 6 μ L sterile, nuclease free water and 2 μ L undiluted extracted DNA. PCR cycling conditions were as follows: 1 cycle initial denaturation at 98°C for 15 min, followed by 40 cycles of denaturation (98°C for 30s), annealing (62.5°C for 30s) and elongation (72°C for 2 min). Final elongation took place at 72°C for 5 min. Presence of the amplified products was confirmed by 1% agarose gel electrophoresis. Thereafter, amplicons were end-repaired, individually barcoded, pooled into a single library, native adapter ligated, loaded onto a single R10.4.1 flow cell (>1,250 pores), and sequenced using the MinION mk1C device (Figure 3).

Bioinformatics

Illumina and ONT whole-genome sequencing

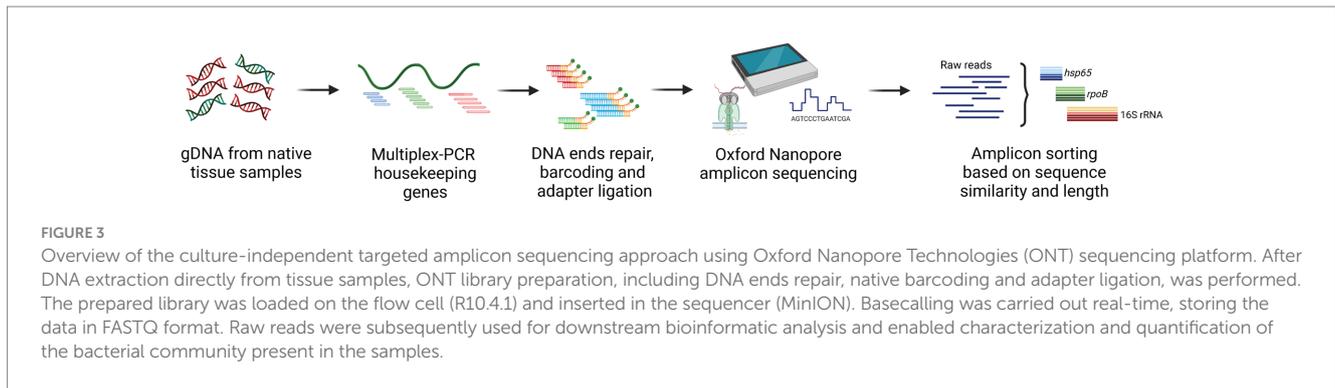
Data analysis was applied to 7 datasets generated with the Illumina MiSeq and 14 datasets generated with MinION mk1C. For the NAS sequencing datasets, once the data acquisition was stopped after 22–27 h (Table 2), Guppy [v6.4.6] was used for basecalling (260 bps – High-Accuracy), demultiplexing and trimming of the barcodes

TABLE 2 Overview of Nanopore adaptive sampling (NAS) sequencing runs and basecall summary.

	NAS direct	NAS culture
Estimated bases	2.75 Gb	2.42 Gb
Data produced	42.41 GB	44.77 GB
Reads generated	579.85k	1.64 M
Estimated N50	6.07 kb	4.97 kb
Elapsed time	22 h 20 min	27 h 18 min

(Wick et al., 2019; Steinig and Coin, 2022). Reads with a Q score of <8 were discarded. Quality control, filtering, and summary reports for nanopore reads were generated using nanoq v 0.10.0 (Wick et al., 2019; Steinig and Coin, 2022). The quality of sequencing reads was evaluated using FastQC (Version 0.11.9)¹ and pycoQC (v2.5.0.23)² (Leger and Leonardi, 2019). Thereafter, sample contamination was evaluated via taxonomic classification using Kraken2 (v2.1.1)³ (Wood et al., 2019).

- <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- <https://github.com/tleonardi/pycoQC>
- <https://github.com/DerrickWood/kraken2>



Genome sequences generated with both platforms were assessed with the validated vSNP3 pipeline⁴ tool of the US Department of Agriculture-Veterinary Services (Supplementary material S1). The evaluation of each isolate's sequencing run performance was carried out using a synopsis of quality metrics generated with default settings. Finally, sequencing coverage and mapping quality across the reference genome was performed in QualiMap [v.2.2.2-dev] and visualized in GraphPad Prism (v10.0.0) (Parsons et al., 2011; Smith et al., 2021). After performing various filtering steps to remove error prone region variant positions (Supplementary material S1), informative SNPs were used to create SNPs tables and phylogenetic trees. Single indels' accuracy was manually confirmed on an isolate-by-isolate basis using bam files generated in Step 1 and visualized in Artemis [v18.1.0]. Only validated SNPs were retained for downstream analysis. The validated SNPs were used to generate an SNP alignment file, which was utilized to build a maximum likelihood phylogenetic tree using RAXML software [v8.2; GTR CAT model] (Stamatakis et al., 2005). To confirm the assigned *in silico* spoligotypes and determine the animal lineage of the sequences, the TBprofiler pipeline using H37Rv as a reference was used to analyze short- and long-reads using the flag `-platform` Illumina and ONT, respectively (Phelan et al., 2019).

To evaluate the quality of the genome mapping, two parameters were monitored: (i) genome completeness, and (ii) sufficient sequencing depth. To assess the first parameter, the percentage of the total number of covered positions of the reference sequence (*M. bovis* AF2122/97) was determined after mapping with BWA and Minimap2. For the second parameter, the percentage of genome positions with a sequencing depth below 20X, defined as N%, was retrieved from the bam files using BEDTools (Quinlan and Hall, 2010) in accordance with previous minimal sequencing depth observations. Comparison of mean numbers of no coverage bases between Illumina, ONT direct and ONT from culture was assessed using multiple unpaired t-test (GraphPad Prism 10.0.0).

Ultimately, to conduct a comparison between the acquired whole genome sequences produced using the ONT platform and sequences from various *M. bovis* strains generated through the same method, the US National Center for Biotechnology Information (NCBI) was searched for publicly available genomes. The following search terms were applied: “*Mycobacterium tuberculosis* variant *bovis* [organism]” AND “Nanopore.” A total of 98 genome sequences concordant with

our search terms were available (1st of August 2023). From these genomes, we excluded those that met the following criteria prior to downstream analysis: genomes registered as bacillus Calmette-Guérin (BCG) and average whole-genome coverage <15X. One genome sequence from *M. bovis* BCG and *M. tuberculosis* H37Rv each were included as outliers. The retrieved sequences were subsequently analyzed using the vSNP3 pipeline as described above and the acquired phylogenetic tree was visualized using iTOL (v5) (Letunic and Bork, 2021). Pairwise SNP distance matrixes from SNP alignments were computed using snp-dists (version 0.8.2)⁵ and used to draw heatmaps combined with the corresponding dendrograms using the pheatmap package.⁶

In conjunction with the vSNP3 pipeline, MINTyper (Hallgren et al., 2021), designed to compare sequences from both the Illumina and ONT platform, were implemented to determine genetic distance between the sequences and estimate a distance matrix.

Distribution analysis of insertion sequences IS6110 and IS1081

The ISMapper (v2.0) pipeline (Hawkey et al., 2015) was applied on genome short-read sequences for identification of IS6110 and IS1081 using the AF2122/97 reference genome. The identified genomic positions were subsequently manually checked for their completeness using the bam files and Artemis (v18.1.0). Once the presence and location of the ISs were confirmed, an *in silico* PCR with previously published primers was performed (CLC Genomics Workbench v22; Qiagen) to estimate the quantity of genomic DNA load present in the native material used for adaptive sampling and as culture inoculum.

Detection of contaminants using ONT targeted amplicon sequencing

For targeted amplicon sequencing, base-calling was performed in real-time using Guppy [v6.4.6] (Wick et al., 2019) and the 400 bps – Low-Accuracy option. Data acquisition and base-calling were stopped after 23 h. Quality control, filtering, and summary report were performed as above. Reads with a Q score of <12 were discarded. Thereafter, reference-free reads sorting, based on similarity and length, was performed using the amplicon sorter tool [v2023-06-19]

4 <https://github.com/USDA-VS/vSNP3/tree/main>

5 <https://github.com/tseemann/snp-dists>

6 <https://github.com/raivokolde/pheatmap>

(Vierstraete and Braeckman, 2022). A total of 100,000 randomly chosen reads with minimum and maximum lengths of 300 bp and 2000 bp, respectively, were selected for each barcode. Consensus sequences were grouped for each species and amplified target gene, and relative abundances were retrieved based on the representative pool of reads analyzed (Figure 3). Finally, a reference database for Mycobacteria was built using the nsdpy python script.⁷ Briefly, for each target gene, the nsdpy (NCBI sequence Downloader) was used to retrieve publicly available sequences using the following commands `nsdpy -r "Mycobacteriaceae[Organism] AND gene_of_interest[Title]";` with `gene_of_interest` defined as *hsp65*, *rpoB*, and 16S rRNA, respectively. The downloaded sequences were used to build a database for each gene using ABRicate⁸ and the same tool was applied for the screening of consensus sequences and summarizing the report files. The distribution of consensus sequences generated per target gene was visualized using the R package ggplot2.

For all bioinformatics tools, we used the default settings unless stated otherwise.

Results

All seven buffaloes sampled in the present study were confirmed to be *M. bovis* infected and the presence of the *M. bovis* region of difference 4 (RD4) signature was confirmed from culture positive crude extracts (Warren et al., 2006). Spacer oligonucleotide typing hybridization assay (spoligotyping) revealed profile SB1474 for all the tested samples, based on mycobacterial culture extracts. Results from the GeneXpert MTB/RIF Ultra, using tissue homogenates, revealed the presence of MTBC genomic DNA in every sample tested, at low ($n=1$), medium ($n=3$) and high ($n=3$) relative concentrations (Table 1).

Comparison of Illumina, direct and culture derived NAS sequences using vSNP3 and MINTyper pipelines

The results of the sequencing comparison among the three groups of DNA sequencing sources (ONT direct and ONT culture – sequences generated using NAS from DNA extracted from tissues and crude liquid culture, respectively, and Illumina – sequences generated from DNA extracted from the same liquid culture as ONT culture) are presented in Supplementary Table S1. The Illumina platform generated the highest mean read count with approximately 1,682,963 reads, followed by the ONT direct group with a mean read count of 47,105, and the ONT culture group with the lowest mean read count at 42,229 using comparable amounts of DNA and similar multiplexing. The average read length of the NAS generated reads was 3,470 bp and 4,648 bp for the ONT culture and ONT direct groups, respectively, whereas the mean size of the longest reads per each group was 58,177 bp for the ONT culture derived reads and 50,928 bp for the

ONT direct group. Both the ONT direct and ONT culture groups showed similar mean values for passing Q20 ($\geq 77\%$) and Q30 ($\geq 51\%$). Sequences generated using the Illumina platform had higher mean values for passing Q20 (94.01%) and Q30 (90.27%) and demonstrated the highest mean read quality with a value of 31.4, while the ONT groups had a mean read quality of 16.9 from culture and 16.4 directly from tissue as shown in Figure 4A for sample NB18191 and in Supplementary Table S1.

The Illumina reads exhibited a significantly ($p < 0.0001$) lower mean reference genome coverage of 99.76%, compared to mean coverage of $>99.99\%$ for ONT culture and ONT direct groups (Figure 4B and Supplementary materials S2, S3). In other words, the Illumina group had the highest number of bases with no coverage (depth=0) with 0.24% ($M = 10,279$ positions, $SD = 1,369.24$), followed by the two ONT datasets with similar results and only 0.006% of the reference genome not being covered (ONT culture $M = 275$ positions, $SD = 126.2$ and ONT direct $M = 271$ positions, $SD = 148.73$) (Figure 4B). Mean depths of coverage for Illumina were 62X - 132X, followed by the ONT direct group with means of 27X - 74X, and the ONT culture group with mean depths of coverage of 16X - 49X. For ONT generated sequences, high-quality genomes (defined as $>99\%$ of genomic positions with a sequencing depth above 20X) were retrieved from data showing at least 36X mean depth of coverage (Figure 4C). The number of positions showing the minimal sequencing depth (defined as 20X) dropped significantly with mean depth of coverage below 30X, showing only 91 and 60% of positions fulfilling the criteria in genomes with 27X and 21X mean depth of coverage, respectively. For the Illumina data, the number of genome positions with minimum sequencing depth was very stable, independently from the mean depth of coverage, and ranged between 97.3–98.4% (Figure 4C).

The obtained sequences were classified as spoligotype SB1474, belonging to the EU1 clonal complex, group Mbovis-09 (Supplementary Table S1). The *in silico* spoligotyping result was confirmed using the TBProfiler pipeline, which additionally defined all the sequences as animal lineage La1.8.1 (Zwyer et al., 2021).

A total of 11 variants discriminating between the 21 analyzed genomes were identified using vSNP3 pipeline (Supplementary Table S2). These were manually inspected and excluded from the comparison due to the following reasons. The first SNP reported by vSNP3 was located within a 30-base deletion between positions 1,122,168 and 1,122,198 of the reference strain (Supplementary material S4). All reads from both platforms were inaccurately mapped in this region. Another large deletion of 147 bases, located between positions 3,082,830 and 3,082,977, led to two incorrect variant calls, defined as “A,” “T,” “C,” “G,” and “N.” The pipeline incorrectly assigned variant nucleotides in 3 out of 7 genomes for Illumina mapped reads. For long reads mapped to the reference genome, only 9 out of 28 calls were correctly assigned as deletions. At position 4,194,117, listed as a putative transposase gene, 6 erroneously assigned variants were observed for the mapped short reads, while 14 correct nucleotide assignments were made for the ONT generated reads. Additionally, the remaining 7 variants were in PPE and PKS encoding genes and were excluded from the comparison due to known difficulties with short reads in resolving these genomic regions. Upon manual investigation of the reported variants compared to the reference genome, all 98 reported SNPs from the 14 ONT sequences were found to be correct. Overall, the

⁷ <https://nsdpy.readthedocs.io/en/latest/main.html#instructions-for-use-from-the-terminal>

⁸ <https://github.com/tseemann/abricate>

7 genomes were defined to be clonal and by including comparable positions using short- and long-reads data mapped over the entire genome, no SNPs differences were retrieved. Outputs from the MINTyper analysis were in concordance with the vSNP3 analysis and a SNP difference of 1 was detected between the Illumina and ONT sequences.

Finally, using publicly available sequences generated with the ONT platform, the ability of the vSNP3 pipeline to compare long-read data for transmission investigation purposes was evaluated. A summary of the quality metrics for the retrieved genomes is provided in [Supplementary Table S3](#). Mean depth of coverage ranged between 23X and 1,949X. The three *M. bovis* genomes obtained from clinical isolates showed distinct profiles when subjected to *in silico* spoligotyping (SB0971, SB0673, and SB0130). The pipeline was able to accurately retrieve a total of 3,508 discriminatory SNPs compared to the reference sequence ([Supplementary Table S4](#)). In the phylogenetic tree, the clonal position of four selected *M. bovis* genomes from African buffaloes was confirmed, and a clear distinction between these sequences and the publicly available sequences was evident ([Figures 5A,B](#)).

Distribution analysis of insertion sequences IS6110 and IS1081

By analyzing the occurrence of insertion sequences (IS) IS6110 and IS1081, we were able to determine that all genomes contained a single copy of IS6110 and six copies of IS1081, respectively. Of the latter six copies, one copy was truncated ([Charles et al., 2023](#)) and not amplifiable using the GeneXpert MTB/RIF Ultra primers. The genomic positions of the IS were conserved across the sequences analyzed. Similarly to *M. bovis* BCG for which an equal number of IS6110 and IS1081 copies was described ([Chakravorty et al., 2017](#)), we estimated a detection limit for GeneXpert MTB/RIF Ultra of 143.4 CFU/mL (95% CI, 106.2 to 243.7 CFU/mL).

Detection of contaminants using ONT targeted amplicon sequencing

The *hsp65*, *rpoB*, and 16S rRNA gene regions were successfully amplified for each sample ([Table 3](#)). The number of reads with Q8

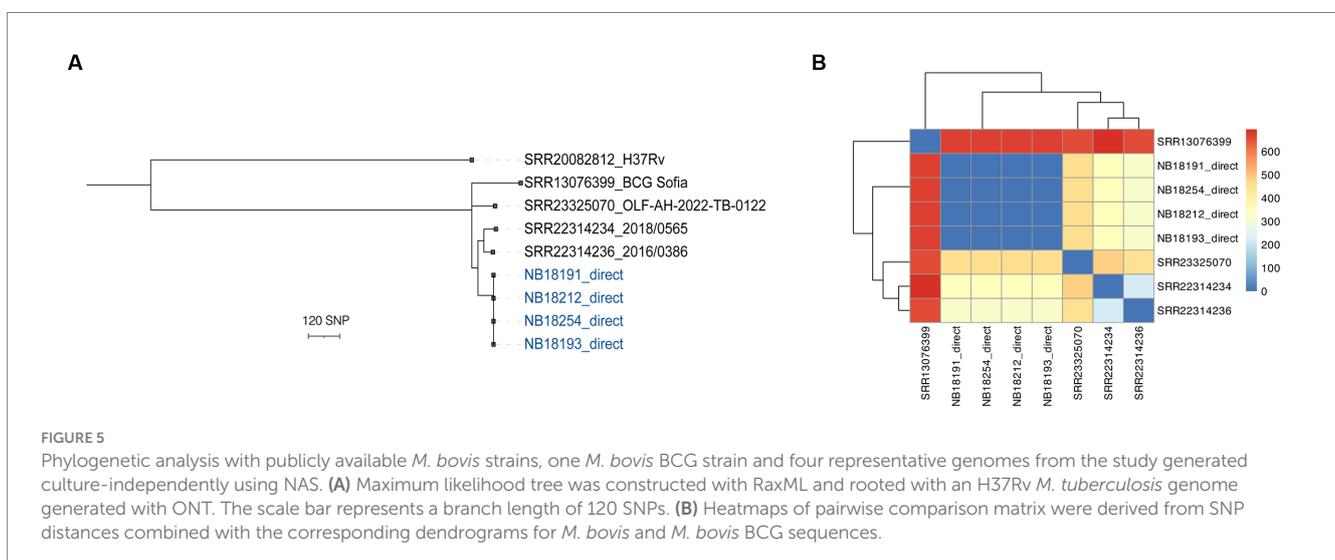
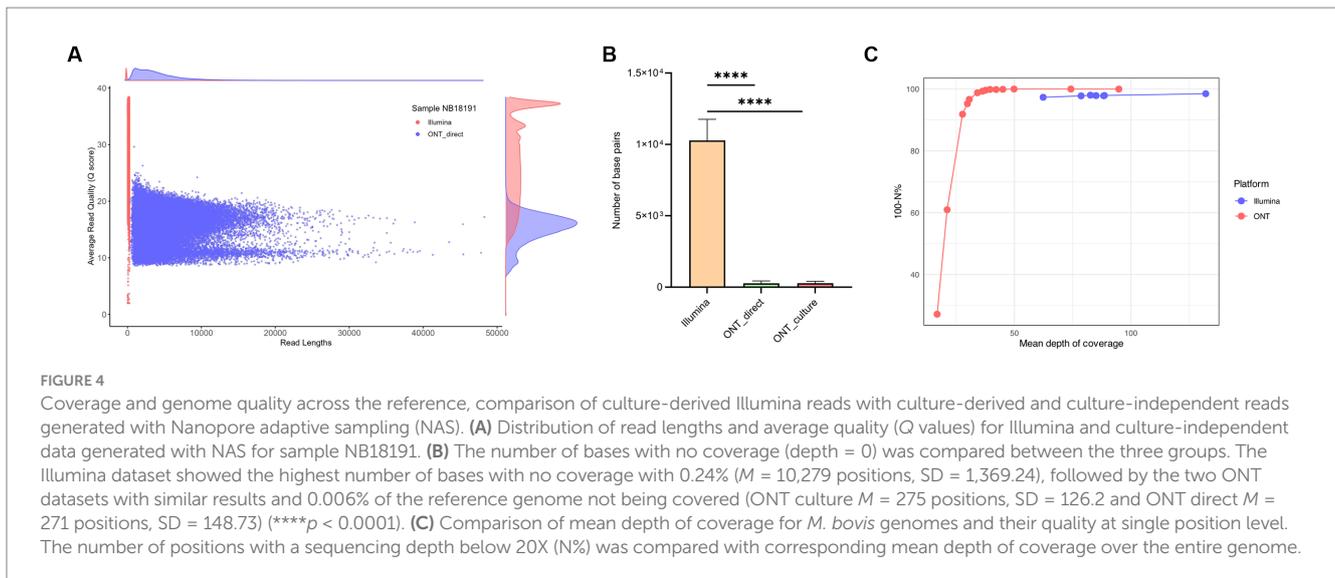


TABLE 3 Summary of data output for targeted sequencing from PCR-derived amplicons.

Culture ID	<i>hsp65</i>	<i>rpoB</i>	16S rRNA	Number of reads	Number of bases	N50 read length	Mean read length	Mean read quality
NB18193	+	+	+	638,707	408,202,430	568	639	13.7
NB18254	+	+	+	570,988	359,167,245	568	629	13.73
NB18212	+	+	+	563,289	322,558,743	563	572	13.65
NB18191	+	+	+	756,804	464,220,866	565	613	13.6
NB18211	+	+	+	676,936	419,171,718	566	619	13.68
NB18246	+	+	+	363,303	229,970,748	567	632	13.64
NB18249	+	+	+	283,064	172,922,714	567	610	13.32

“+” successful PCR amplification.

ranged from 283,064 to 756,804 ($M = 550,441$ reads, $SD = 157,423$), and the total number of bases sequenced ranged from 172,922,714 to 464,220,866 ($M = 339,459,209$ bases, $SD = 97,863,859$). The N50 read length, which represents the median read length of the longest contigs, varied from 563 to 568 bases. The mean read length for the samples ranged from 572 to 639 bases. The mean read quality, as indicated by the Phred score, was consistently above 13.0 for all samples. After quality-filtering the raw reads at Q12, the number of reads was reduced by 80.5%. Of these, 100,000 reads showing a mean quality of 14.3 ($SD = 0.07$) were randomly selected for downstream analysis (Figure 6A). This mean quality threshold was chosen due to the usually sterile nature of these body sites and the significant computational resources required to analyze a larger number of reads without additional informative value.

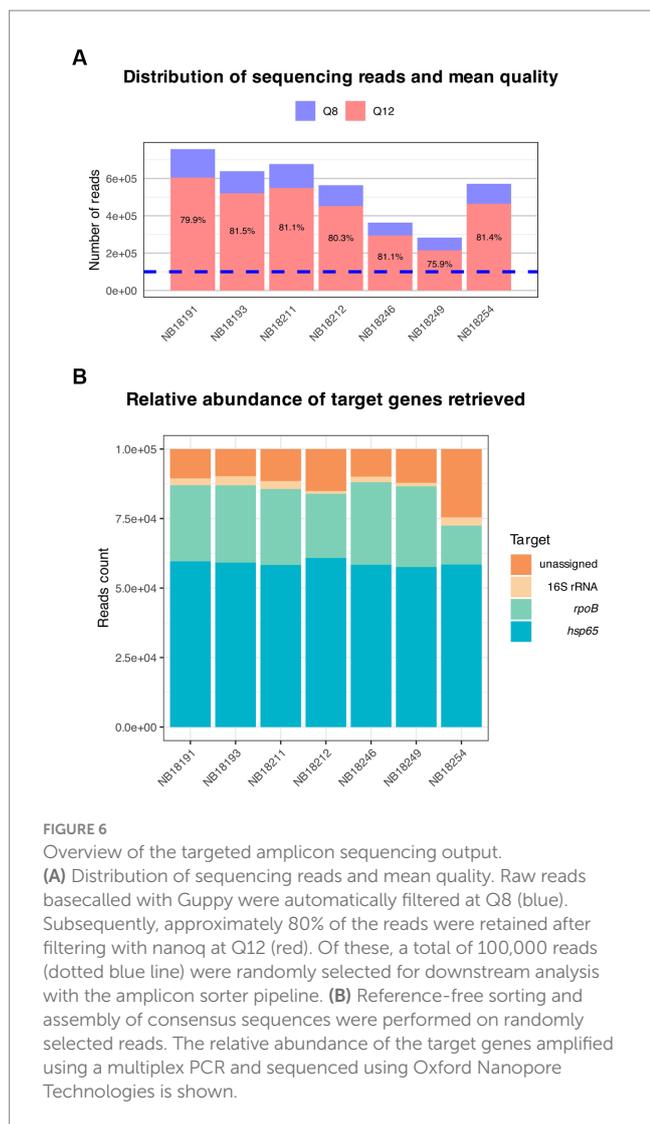
One consensus sequence per target in each sample was generated. Consensus sequences for *hsp65* were obtained after combining >31,000 reads per sample ($M = 55,024$ reads, $SD = 9,706$), >23,000 ($M = 27,337$ reads, $SD = 1963$) for *rpoB*, and between 980 and 3,309 reads ($M = 2,279$ reads, $SD = 815$) for the 16S rRNA gene (Figure 6B). Finally, consensus sequences were queried against curated databases generated for each gene and the output was evaluated for sequence coverage and percentage identities. All consensus sequences matched to reference sequences from the MTBC for the respective genes with coverage and identity of above 99.9 and 99.7%, respectively (Supplementary Table S5), excluding the presence of additional microorganisms, including other *Mycobacterium* spp.

Discussion

Whole-genome sequencing (WGS) has emerged as the gold standard for bacterial outbreak investigation and pathogen typing, offering a level of resolution that conventional molecular methods cannot achieve (Bogaerts et al., 2021). The continuous evolution of sequencing methods provided by platforms based on different techniques brings forth new possibilities and increased accessibility to data, particularly in under-resourced areas or limited research settings. By applying a culture-independent approach and comparing the results with conventional culture-based Illumina WGS, we were able to retrieve crucial data for epidemiological purposes. In South Africa, there is a lack of information regarding the transmission and spread of animal TB, especially in complex ecosystems such as Natural Parks where multiple susceptible animal species share the

same habitat. By providing evidence-based data regarding the burden of animal TB to the competent authorities, the use of *in vivo* tests with increased sensitivity or vaccination of susceptible animal populations for TB control will be facilitated. Therefore, there is a need for tools that enable unequivocal characterization of *M. bovis* in food-and-mouth disease (FMD) affected areas where culture of suspicious tissues samples is not possible (Arnot and Michel, 2020; Davey, 2023). Not only conventional markers, including spoligotyping and MIRU-VNTR markers, were retrieved from long reads generated by the ONT MinION Mk1C device, but high-quality SNPs used for transmission investigation at the deepest level were obtained and compared with short-read data derived from the Illumina MiSeq instrument. We report the presence of a clonal strain of *M. bovis* appertaining to the Eu1 clonal complex and showing spoligotype SB1474. The latter is a commonly reported spoligotype in South African buffaloes and has been circulating among this wildlife species for at least 15 years, being reported for the first time in 2007, and added to the Mbovis.org spoligotype database in 2008 (Hlokwe et al., 2011). According to the new nomenclature for livestock associated MTBC ecotypes, the sequences were classified as La1.8.1, which is one of the eight described for *M. bovis* (Zwyer et al., 2021). These sublineages exhibit significant differences in their geographic occurrence, with some being confined to specific regions while others being prevalent worldwide, suggesting different adaptation abilities and possibly host predilections. La1.8.1 is one of the most well-studied clusters of *M. bovis* strains, which is known to be widespread in the UK and regions that have historical trading ties with the UK (Zwyer et al., 2021). Hence, it is likely that the strain was introduced to South Africa through infected livestock originating from the UK and subsequently spread to wildlife, particularly African buffaloes in this specific case. More data, however, are necessary to support this hypothesis.

The NAS, also known as selective sequencing, is a newly developed method that relies on software-controlled enrichment, providing a potential solution without physical manipulation. A programming interface called ReadUntil allows for precise control over individual nanopores, enabling the software to request the removal of the currently sequenced molecule from a specific pore (Martin et al., 2022; NASCarD, 2023). By examining the initial portion of a molecule, typically a few hundred bases, a determination can be made if the molecule is the desired “target” or not. If a molecule is deemed “off target,” the current across the pore is reversed, rejecting the molecule, and allowing the pore to capture a new one. It is crucial for this rejection process to occur quickly to ensure the molecule is ejected



before a significant portion of it is sequenced. The speed at which decisions are made and rejections occur directly impacts the potential level of enrichment achievable (Martin et al., 2022).

Short-read sequencing exhibits a pronounced bias towards GC content, but it excels in accurately detecting single nucleotide variants and small indels (Chen et al., 2013). In contrast, long-read sequencing offers the advantage of resolving structural variations and variants found within repetitive regions (Chen et al., 2023). For instance, the PE/PPE gene families in *M. tuberculosis*, which constitute around 10% of its coding regions, have been implicated in virulence and their association with drug resistance warrants further exploration (McEvoy et al., 2012; Dippenaar et al., 2022). Unfortunately, short-read sequencing struggles to accurately resolve these regions, resulting in their exclusion from bioinformatics studies on *M. tuberculosis*. However, long-read sequencing holds the potential to provide a more comprehensive understanding of these regions, including their impact on resistance phenotypes and strain pathogenesis. It is important to note that until recently, long-read nanopore sequencing faced accuracy limitations due to error-prone homopolymer regions (Dippenaar et al., 2022). The introduction of R.10 flow cells,

specifically designed to optimize translocation speed for homopolymer sequences within pores, along with improvements in base-callers such as Guppy, has significantly enhanced sequence accuracy (Dippenaar et al., 2022). Consequently, variant calling and subsequent drug-resistance prediction have also improved.

Precise characterization of bacterial communities present in infected patients is crucial to understanding their diversity and potential role in the disease and can help identify microbial biomarkers, expected to be consistently enriched or depleted in patient cohorts (Duvall et al., 2017). In this study, we investigated the capabilities of Nanopore sequencing to deliver high-accuracy housekeeping gene profiles for the genus *Mycobacterium*, along with *de novo* generated consensus sequences using the full-length 16S rRNA gene amplicon sequencing. Upon applying a deep-sequencing culture-independent approach on native DNA extracted directly from TB confirmed tissues, we excluded the presence of other microorganisms potentially interfering with NAS. The abundance distributions of amplicons obtained for the three target genes should be regarded with caution since the affinity of the primer pairs used may differ significantly. It is the authors' opinion that, by implementing the outlined method, future studies on samples presenting multiple mycobacterial species and possibly from body locations presenting microbial colonization, such as the upper respiratory tract, will demonstrate the specificity of the NAS and facilitate a deeper understanding of the interactions and within-host coexistence of microbiome and TB causing agents.

Animal TB is difficult to control and eradicate in part due to the costly nature of surveillance and poor sensitivity of *ante-* and *post-mortem* diagnostic tests (Meiring et al., 2018). The disease has a significant impact on agriculture, biodiversity, and the financial situation of farmers, particularly in low-income countries. Detection of TB in livestock results in a loss of profit for the farm, mainly due to the animal's slaughter and replacement (Perez-Morote et al., 2020). Recent studies have shown that animal TB causes increased mortality, reduced milk and meat productivity, reduced fertility, and organ or carcass condemnation (Tschopp et al., 2021). National and supranational surveillance and monitoring programs have reduced TB prevalence in cattle, but spillover at the livestock-wildlife-human interface represents a challenge for eradication and leads to pathogen maintenance, both in animals and in the environment (Fitzgerald and Kaneene, 2013).

According to the WHO tuberculosis laboratory biosafety manual, any manipulation of clinical specimens suspected of containing *M. tuberculosis* should be conducted in high-risk or TB-containment laboratories (WHO, 2012; Ssengooba et al., 2015). Depending on local and national regulations, similar biosafety restrictions are applied to mammal samples suspected to be infected with animal tuberculosis. Although the establishment and maintenance of TB-containment laboratories are financially challenging in low-income countries, the escalating prevalence of disease is expected to drive a higher need for these facilities (Ssengooba et al., 2015). Retrieving high-quality whole-genome sequences from native material (independent of culture), will not only shorten the necessary time for data generation and subsequent implementation of restriction measurements but will also provide accessibility to crucial data in resource-limited settings, without the need for establishing the costly biosafety laboratory settings necessary for culture and manipulation of MTBC members.

It is important to note, however, that manipulation of samples suspected to be infected with *M. bovis*, and especially procedures that lead to aerosol formation should be undertaken using the necessary biosafety precautions, including inactivation steps at the initial stage of DNA extraction. Finally, the risk of infection for laboratory personnel associated with MTBC culturing will be significantly lower.

Currently, there are no established recommendations regarding the minimum coverage threshold for whole-genome sequences obtained from long-reads for investigating transmission and outbreaks. According to the literature, it is suggested that a minimum threshold of 10X-30X coverage for short-read generated sequences is required to identify SNPs associated with antimicrobial resistance (AMR) in *M. tuberculosis* (Ellington et al., 2017; Bogaerts et al., 2021). We investigated the effect of a 20X threshold on short- and long-read generated sequences at a single position level, in order to determine the number of “high-quality” positions over the entire genome. For this study, and because of the small number of available genomes, all obtained sequences were analyzed using the vSNP3 pipeline and included in downstream analysis. In the future, to more accurately decide whether to exclude or include a genome for downstream analysis and transmission investigation, additional data are required to further evaluate the accuracy of the ONT platform. When comparing the number of genomic positions covered with a minimum sequencing depth of 20X between genomes generated using the Illumina and ONT platforms, it became evident that, regardless of the mean coverage depth achieved, short-read data from the Illumina platform cannot match the reference genome coverage achieved by long-read data, likely due to the inability of short reads to resolve repetitive regions of the genome.

Finally, for the purpose of this pilot study, only samples originating from presumably sterile tissue locations were included. Based on previous semiquantitative measurements using GeneXpert MTB/RIF Ultra (Chakravorty et al., 2017), different amounts of genomic DNA were detected in the tested samples, varying from low to high bacterial loads and corresponding to an estimated 10^3 – 10^7 CFU/mL for the seven samples from tissue homogenates included in the present study. Interestingly, the tissue homogenate with low detection in GeneXpert MTB/RIF Ultra presented with the highest genome coverage of the tested samples. The presence of contaminating agents or co-infection with NTM was excluded after targeted amplicon sequencing. The performance of this culture-independent approach on more challenging samples needs to be explored and reported by including samples from microbially-complex upper respiratory tract or sputum specimens.

Conclusion

By applying a culture-independent approach using NAS on DNA extracted from infected African buffalo tissue samples, we were able to generate high quality *M. bovis* whole-genome sequences. Significant differences were observed in the reference genome position coverage between long- and short-reads, which warrant further investigation. By including comparable genomic positions between sequencing platforms and based on SNP analysis, we confirmed a clonal spread of the pathogen in the investigated buffalo herd. The described approach will be applied to obtain crucial information regarding the epidemiology of animal TB in complex ecosystems at the livestock–wildlife–human interface.

Data availability statement

The datasets presented in this study can be found in online repositories available at: ncbi.nlm.nih.gov/bioproject/PRJEB63553 under project reference number PRJEB63553.

Ethics statement

The animal study was approved by Stellenbosch University Animal Care and Use Committee (SU-ACUD15-00072), Stellenbosch University Biological and Environmental Safety Research Ethics Committee (SU-BEE-2021-22561) and Department of Agriculture, Land Reform and Rural Development (ref: 12/11/1/7/6 and 12/11/1/7/2). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

GG: Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. JL: Data curation, Formal analysis, Writing – review & editing. TK: Data curation, Writing – review & editing. TS: Validation, Writing – review & editing. TT: Data curation, Investigation, Writing – review & editing. LM: Visualization, Writing – review & editing. MO'H: Visualization, Writing – review & editing. SM: Writing – review & editing. AO: Data curation, Writing – review & editing. AL: Funding acquisition, Resources, Writing – review & editing. RW: Conceptualization, Funding acquisition, Resources, Writing – review & editing. MHM: Conceptualization, Investigation, Methodology, Resources, Writing – review & editing. MAM: Conceptualization, Methodology, Resources, Supervision, Writing – review & editing. WG: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Wellcome Foundation (grant #222941/Z/21/Z), the South African Medical Research Council, American Association of Zoo Veterinarians Wild Animal Health Fund [S005651 and S007355], the National Research Foundation South African Research Chair Initiative [grant #86949], and MHM was supported by Wellcome Trust (grant #216634/Z/19/Z). AGL is supported by the EDCTP TESA III network (CSA2020NoE-3104).

Acknowledgments

Some of the figures (Figures 4–6 and Supplementary Material S1) were generated using BioRender and draw.io, respectively.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1307440/full#supplementary-material>

References

- Arnot, L. F., and Michel, A. (2020). Challenges for controlling bovine tuberculosis in South Africa. *Onderstepoort J. Vet. Res.* 87, e1–e8. doi: 10.4102/ojvr.v87i1.1690
- Belakehal, F., Barth, S. A., Menge, C., Mossadak, H. T., Malek, N., and Moser, I. (2022). Evaluation of the discriminatory power of spoligotyping and 19-locus mycobacterial interspersed repetitive unit-variable number of tandem repeat analysis (MIRU-VNTR) of *Mycobacterium bovis* strains isolated from cattle in Algeria. *PLoS One* 17:e0262390. doi: 10.1371/journal.pone.0262390
- Bernitz, N., Kerr, T. J., de Waal, C., Cooper, D. V., Warren, R. M., Van Helden, P. D., et al. (2020). Test characteristics of assays to detect *Mycobacterium bovis* infection in high-prevalence African buffalo (*Syncerus caffer*) herds. *J. Wildl. Dis.* 56, 462–465.
- Bernitz, N., Kerr, T. J., Goosen, W. J., Higgitt, R. L., de Waal, C., Clarke, C., et al. (2019). Impact of *Mycobacterium bovis*-induced pathology on interpretation of QuantiFERON(R)-TB gold assay results in African buffaloes (*Syncerus caffer*). *Vet. Immunol. Immunopathol.* 217:109923. doi: 10.1016/j.vetimm.2019.109923
- Bogaerts, B., Winand, R., Van Braekel, J., Hoffman, S., Roosens, N. H. C., De Keersmaecker, S. C. J., et al. (2021). Evaluation of WGS performance for bacterial pathogen characterization with the Illumina technology optimized for time-critical situations. *Microb. Genom.* 7:000699. doi: 10.1099/mgen.0.000699
- Bull, T. J., Munshi, T., Mikkelsen, H., Hartmann, S. B., Sorensen, M. R., Garcia, J. S., et al. (2016). Improved culture medium (TiKa) for *Mycobacterium avium* subspecies paratuberculosis (MAP) matches qPCR sensitivity and reveals significant proportions of non-viable MAP in lymphoid tissue of vaccinated MAP challenged animals. *Front. Microbiol.* 7:2112. doi: 10.3389/fmicb.2016.02112
- Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., et al. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* 47:e103. doi: 10.1093/nar/gkz569
- Chakravorty, S., Simmons, A. M., Rownecki, M., Parmar, H., Cao, Y., Ryan, J., et al. (2017). The new Xpert MTB/RIF ultra: improving detection of *Mycobacterium tuberculosis* and resistance to rifampin in an assay suitable for point-of-care testing. *MBio* 8:e00812-17. doi: 10.1128/mBio.00812-17
- Charles, C., Conde, C., Vorimore, F., Cochard, T., Michelet, L., Boschiroli, M. L., et al. (2023). Features of *Mycobacterium bovis* complete genomes belonging to 5 different lineages. *Microorganisms*. 11:177. doi: 10.3390/microorganisms11010177
- Chen, Y. C., Liu, T., Yu, C. H., Chiang, T. Y., and Hwang, C. C. (2013). Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS One* 8:e62856. doi: 10.1371/journal.pone.0062856
- Chen, Y., Wang, A. Y., Barkley, C. A., Zhang, Y., Zhao, X., Gao, M., et al. (2023). Deciphering the exact breakpoints of structural variations using long sequencing reads with DeBreak. *Nat. Commun.* 14:283. doi: 10.1038/s41467-023-35996-1
- Clarke, C., Cooper, D. V., Miller, M. A., and Goosen, W. J. (2022). Detection of *Mycobacterium tuberculosis* complex DNA in oronasal swabs from infected African buffaloes (*Syncerus caffer*). *Sci. Rep.* 12:1834. doi: 10.1038/s41598-022-05982-6
- Curry, K. D., Wang, Q., Nute, M. G., Tyshaieva, A., Reeves, E., Soriano, S., et al. (2022). Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* 19, 845–853. doi: 10.1038/s41592-022-01520-4
- Davey, S. (2023). Challenges to the control of *Mycobacterium bovis* in livestock and wildlife populations in the south African context. *Ir. Vet. J.* 76:14. doi: 10.1186/s13620-023-00246-9
- Dippenaar, A., Goossens, S. N., Grobbelaar, M., Oostvogels, S., Cuypers, B., Laukens, K., et al. (2022). Nanopore sequencing for *Mycobacterium tuberculosis*: a critical review of the literature, new developments, and future opportunities. *J. Clin. Microbiol.* 60:e0064621. doi: 10.1128/JCM.00646-21
- Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., and Alm, E. J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* 8:1784. doi: 10.1038/s41467-017-01973-8
- Earl, J. P., Adappa, N. D., Krol, J., Bhat, A. S., Balashov, S., Ehrlich, R. L., et al. (2018). Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific biosciences sequencing of full-length 16S rRNA genes. *Microbiome*. 6:190. doi: 10.1186/s40168-018-0569-2
- Ellington, M. J., Ekelund, O., Aarestrup, F. M., Canton, R., Doumith, M., Giske, C., et al. (2017). The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST subcommittee. *Clin. Microbiol. Infect.* 23, 2–22. doi: 10.1016/j.cmi.2016.11.012
- Fitzgerald, S. D., and Kaneene, J. B. (2013). Wildlife reservoirs of bovine tuberculosis worldwide: hosts, pathology, surveillance, and control. *Vet. Pathol.* 50, 488–499. doi: 10.1177/0300985812467472
- Goosen, W. J., Kleynhans, L., Kerr, T. J., Buss, P., and Miller, M. A. (2022). Culture-independent PCR detection and differentiation of mycobacteria spp. in antemortem respiratory samples from African elephants (*Loxodonta africana*) and rhinoceros (*Ceratotherium simum*, *Diceros bicornis*) in South Africa. *Pathogens* 11:709. doi: 10.3390/pathogens11060709
- Goosen, W. J., Kleynhans, L., Kerr, T. J., van Helden, P. D., Buss, P., Warren, R. M., et al. (2022). Improved detection of *Mycobacterium tuberculosis* and *M. bovis* in African wildlife samples using cationic peptide decontamination and mycobacterial culture supplementation. *J. Vet. Diagn. Investig.* 34, 61–67. doi: 10.1177/10406387211044192
- Gopalaswamy, R., Shanmugam, S., Mondal, R., and Subbian, S. (2020). Of tuberculosis and non-tuberculous mycobacterial infections - a comparative analysis of epidemiology, diagnosis and treatment. *J. Biomed. Sci.* 27:74. doi: 10.1186/s12929-020-00667-6
- Gorsich, E. E., Etienne, R. S., Medlock, J., Beechler, B. R., Spaan, J. M., Spaan, R. S., et al. (2018). Opposite outcomes of coinfection at individual and population scales. *P Natl Acad Sci USA*. 115, 7545–7550. doi: 10.1073/pnas.1801095115
- Hallgren, M. B., Overballe-Petersen, S., Lund, O., Hasman, H., and Clausen, P. (2021). MINType: an outbreak-detection method for accurate and rapid SNP typing of clonal clusters with noisy long reads. *Biol. Methods Protoc.* 6:bpab008. doi: 10.1093/biomethods/bpab008
- Hawkey, J., Hamidian, M., Wick, R. R., Edwards, D. J., Billman-Jacobe, H., Hall, R. M., et al. (2015). ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC Genomics* 16:667. doi: 10.1186/s12864-015-1860-2
- Hlokwe, T. M., Jenkins, A. O., Streicher, E. M., Venter, E. H., Cooper, D., Godfroid, J., et al. (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.* 78:232. doi: 10.4102/ojvr.v78i1.232
- Hlokwe, T. M., and Mogano, R. M. (2020). Utility of xpert(R) MTB/RIF ultra assay in the rapid diagnosis of bovine tuberculosis in wildlife and livestock animals from South Africa. *Prev. Vet. Med.* 177:104980. doi: 10.1016/j.prevetmed.2020.104980
- Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10:5029. doi: 10.1038/s41467-019-13036-1
- Leger, A., and Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore sequencing. *JOSS* 4, 1236.
- Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Macovei, L., McCafferty, J., Chen, T., Teles, F., Hasturk, H., Paster, B. J., et al. (2015). The hidden 'mycobacteriome' of the human healthy oral cavity and upper respiratory tract. *J. Oral Microbiol.* 7:26094. doi: 10.3402/jom.v7.26094
- Maiga, M., Siddiqui, S., Diallo, S., Diarra, B., Traore, B., Shea, Y. R., et al. (2012). Failure to recognize nontuberculous mycobacteria leads to misdiagnosis of chronic pulmonary tuberculosis. *PLoS One* 7:e36902. doi: 10.1371/journal.pone.0036902

- Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M. D., and Leggett, R. M. (2022). Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* 23:11. doi: 10.1186/s13059-021-02582-x
- McEvoy, C. R., Cloete, R., Muller, B., Schurch, A. C., van Helden, P. D., Gagneux, S., et al. (2012). Comparative analysis of *Mycobacterium tuberculosis* pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. *PLoS One* 7:e30593. doi: 10.1371/journal.pone.0030593
- Meehan, C. J., Goig, G. A., Kohl, T. A., Verboven, L., Dippenaar, A., Ezewudo, M., et al. (2019). Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat. Rev. Microbiol.* 17, 533–545. doi: 10.1038/s41579-019-0214-5
- Meiring, C., van Helden, P. D., and Goosen, W. J. (2018). TB control in humans and animals in South Africa: a perspective on problems and successes. *Front. Vet. Sci.* 5:298. doi: 10.3389/fvets.2018.00298
- Muller, B., Durr, S., Alonso, S., Hattendorf, J., Laisse, C. J., Parsons, S. D., et al. (2013). Zoonotic *Mycobacterium bovis*-induced tuberculosis in humans. *Emerg. Infect. Dis.* 19, 899–908. doi: 10.3201/eid1906.120543
- NAHF. (2022) Guiding document | cattle movement ban | foot and mouth disease outbreak. Available at: <https://nahf.co.za/guiding-document-cattle-movement-ban-foot-and-mouth-disease-outbreak-16-august-20222022>
- NASCarD (Nanopore adaptive sampling with carrier DNA: A rapid, PCR-free method for whole genome sequencing of pathogens in clinical samples. (2023). Available at: <https://www.medrxiv.org/content/10.1101/2023.03.10.23287094v1> (Accessed September 13, 2023).
- Ortiz, A. P., Perea, C., Davalos, E., Velazquez, E. F., Gonzalez, K. S., Camacho, E. R., et al. (2021). Whole genome sequencing links *Mycobacterium bovis* from cattle, cheese and humans in Baja California. *Mexico. Front Vet Sci.* 8:674307. doi: 10.3389/fvets.2021.674307
- Oude Munnink, B. B., Nieuwenhuijse, D. F., Stein, M., O'Toole, A., Haverkate, M., Mollers, M., et al. (2020). Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* 26, 1405–1410. doi: 10.1038/s41591-020-0997-y
- Pan, Y., Yang, X., Duan, J., Lu, N., Leung, A. S., Tran, V., et al. (2011). Whole-genome sequences of four *Mycobacterium bovis* BCG vaccine strains. *J. Bacteriol.* 193, 3152–3153. doi: 10.1128/JB.00405-11
- Parsons, S. D., Cooper, D., McCall, A. J., McCall, W. A., Streicher, E. M., le Maitre, N. C., et al. (2011). Modification of the QuantiFERON-TB gold (in-tube) assay for the diagnosis of *Mycobacterium bovis* infection in African buffaloes (*Syncerus caffer*). *Vet. Immunol. Immunopathol.* 142, 113–118. doi: 10.1016/j.vetimm.2011.04.006
- Perea, C., Ciaravino, G., Stuber, T., Thacker, T. C., Robbe-Austerman, S., Allepuz, A., et al. (2021). Whole-genome SNP analysis identifies putative *Mycobacterium bovis* transmission clusters in livestock and wildlife in Catalonia, Spain. *Microorganisms* 9:1629. doi: 10.3390/microorganisms9081629
- Perez-Morote, R., Pontones-Rosa, C., Gortazar-Schmidt, C., and Munoz-Cardona, A. I. (2020). Quantifying the economic impact of bovine tuberculosis on livestock farms in South-Western Spain. *Animals (Basel)* 10:243. doi: 10.3390/ani10122433
- Phelan, J. E., O'Sullivan, D. M., Machado, D., Ramos, J., Oppong, Y. E. A., Campino, S., et al. (2019). Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med.* 11:41. doi: 10.1186/s13073-019-0650-x
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Riello, F. N., Brigido, R. T., Araujo, S., Moreira, T. A., Goulart, L. R., and Goulart, I. M. (2016). Diagnosis of mycobacterial infections based on acid-fast bacilli test and bacterial growth time and implications on treatment and disease outcome. *BMC Infect. Dis.* 16:142. doi: 10.1186/s12879-016-1474-6
- Rossi, G., Shih, B. B., Egbe, N. F., Motta, P., Duchatel, F., Kelly, R. F., et al. (2023). Unraveling the epidemiology of *Mycobacterium bovis* using whole-genome sequencing combined with environmental and demographic data. *Front Vet Sci.* 10:1086001. doi: 10.3389/fvets.2023.1086001
- Sichewo, P. R., Etter, E. M. C., and Michel, A. L. (2019). Prevalence of *Mycobacterium bovis* infection in traditionally managed cattle at the wildlife-livestock interface in South Africa in the absence of control measures. *Vet. Res. Commun.* 43, 155–164. doi: 10.1007/s11259-019-09756-w
- Smith, K., Bernitz, N., Cooper, D., Kerr, T. J., de Waal, C. R., Clarke, C., et al. (2021). Optimisation of the tuberculin skin test for detection of *Mycobacterium bovis* in African buffaloes (*Syncerus caffer*). *Prev. Vet. Med.* 188:105254. doi: 10.1016/j.prevetmed.2020.105254
- Smith, C., Halse, T. A., Shea, J., Modestil, H., Fowler, R. C., Musser, K. A., et al. (2021). Nanopore sequencing for clinical diagnostics: a comparison of next-generation sequencing (NGS) methods for *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* 59:e00583-20. doi: 10.1128/JCM.00583-20
- Ssengooba, W., Gelderbloem, S. J., Mboowa, G., Wajja, A., Namaganda, C., Musoke, P., et al. (2015). Feasibility of establishing a biosafety level 3 tuberculosis culture laboratory of acceptable quality standards in a resource-limited setting: an experience from Uganda. *Health Res Policy Syst.* 13:4. doi: 10.1186/1478-4505-13-4
- Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21, 456–463. doi: 10.1093/bioinformatics/bti191
- Stanton, J. J., Zong, J. C., Latimer, E., Tan, J., Herron, A., Hayward, G. S., et al. (2010). Detection of pathogenic elephant endotheliotropic herpesvirus in routine trunk washes from healthy adult Asian elephants (*Elephas maximus*) by use of a real-time quantitative polymerase chain reaction assay. *Am. J. Vet. Res.* 71, 925–933. doi: 10.2460/ajvr.71.8.925
- Steinig, E., and Coin, L. (2022). Nanoq: ultra-fast quality control for nanopore reads. *JOSS* 7:2991.
- Su, J., Lui, W. W., Lee, Y., Zheng, Z., Siu, G. K., Ng, T. T., et al. (2023). Evaluation of *Mycobacterium tuberculosis* enrichment in metagenomic samples using ONT adaptive sequencing and amplicon sequencing for identification and variant calling. *Sci. Rep.* 13:5237. doi: 10.1038/s41598-023-32378-x
- Telenti, A., Marchesi, F., Balz, M., Bally, F., Bottger, E. C., and Bodmer, T. (1993). Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis. *J. Clin. Microbiol.* 31, 175–178. doi: 10.1128/jcm.31.2.175-178.1993
- Tschopp, R., Conlan, A. J. K., Gemechu, G., Almw, G., Hattendorf, J., Zinsstag, J., et al. (2021). Effect of bovine tuberculosis on selected productivity parameters and trading in dairy cattle kept under intensive husbandry in Central Ethiopia. *Front Vet Sci.* 8:698768. doi: 10.3389/fvets.2021.698768
- van Soolingen, D., Hermans, P. W., de Haas, P. E., Soll, D. R., Goulart, L. R., and van Embden, J. D. (1991). Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J. Clin. Microbiol.* 29, 2578–2586. doi: 10.1128/jcm.29.11.2578-2586.1991
- Vierstraete, A. R., and Braeckman, B. P. (2022). Amplicon_sorter: a tool for reference-free amplicon sorting based on sequence similarity and for building consensus sequences. *Ecol. Evol.* 12:e8603. doi: 10.1002/ece3.8603
- Warren, R. M., Gey van Pittius, N. C., Barnard, M., Hesselting, A., Engelke, E., de Kock, M., et al. (2006). Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference. *Int. J. Tuberc. Lung Dis.* 10, 818–822.
- WHO. Tuberculosis laboratory biosafety manual. Guidelines approved by the guidelines review committee. (2012).
- Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20:129. doi: 10.1186/s13059-019-1727-y
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0
- Zorz, J., Li, C., Chakraborty, A., Gittins, D. A., Surcon, T., Morrison, N., et al. (2023). SituSeq: an offline protocol for rapid and remote Nanopore 16S rRNA amplicon sequence analysis. *ISME Commun.* 3:33. doi: 10.1038/s43705-023-00239-3
- Zwyer, M., Cavusoglu, C., Ghielmetti, G., Pacciarini, M. L., Scaltriti, E., Van Soolingen, D., et al. (2021). A new nomenclature for the livestock-associated *Mycobacterium tuberculosis* complex based on phylogenomics. *Open Res Eur.* 1:100. doi: 10.12688/openreseurope.14029.2