DNA Thermo-Protection Facilitates Whole Genome Sequencing of Mycobacteria Direct from Clinical Samples by the Nanopore Platform

Sophie George[a,b,], Yifei Xu[a,b], Gillian Rodger[a,b], Marcus Morgan[c], Nicholas D. Sanderson[a,b], Sarah J. Hoosdally[a,b], Samantha Thulborn[d], Esther Robinson[e], Priti Rathod[e], A. Sarah Walker[a,b], Timothy E. A. Peto[a,b], Derrick W. Crook[a,b], [#]Kate E. Dingle[a,b]

[a]Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford University, UK

[b]National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[c]Microbiology Department, Oxford University Hospitals NHS Trust, Oxford, UK.

[d]Respiratory Medicine Unit, Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, UK

[e]PHE National Mycobacteria Reference Service - North and Central, Birmingham Public Health Laboratory, UK

Running Head: Direct-from-sample MTB whole genome sequencing

[#] Address correspondence to kate.dingle@ndm.ox.ac.uk

Sophie George and Yifei Xu contributed equally to this work. Author order was decided on the basis of amount of time spent performing wet-lab experiments versus time spent developing bioinformatics analysis.

ABSTRACT

Mycobacterium tuberculosis (MTB) is the leading cause of death from bacterial infection. Improved rapid diagnosis and antimicrobial resistance determination, such as by whole genome sequencing, are required. Our aim was to develop a simple, low-cost method of preparing DNA for Oxford Nanopore Technologies (ONT) sequencing direct from MTB positive clinical samples (without culture). Simultaneous sputum liquefaction, bacteria heat-inactivation (99ºC/30min) and enrichment for Mycobacteria DNA was achieved using an equal volume of thermo-protection buffer (4M KCl, 0.05M HEPES buffer pH7.5, 0.1% DTT). The buffer emulated intracellular conditions found in hyperthermophiles, thus protecting DNA from rapid thermo-degradation, which renders it a poor template for sequencing. Initial validation employed Mycobacteria DNA (extracted or intracellular). Next, mock clinical samples (infection-negative human sputum spiked 0-10^5 BCG cells/ml) underwent liquefaction in thermo-protection buffer and heat-inactivation. DNA was extracted and sequenced. Human DNA degraded faster than Mycobacteria DNA, resulting in target enrichment. Four replicate experiments each demonstrated detection at 10^1 BCG cells/ml, with 31-59 MTB complex reads. Maximal genome coverage (>97% at 5x-depth) was achieved at 10^4 BCG cells/ml; >91% coverage (1x depth) at 10^3 BCG cells/ml. Final validation employed MTB positive clinical samples (n=20), revealed initial sample volumes ≥1ml typically yielded higher mean depth of MTB genome coverage, the overall range 0.55-81.02. A mean depth of 3 gave >96% one-fold TB genome coverage (in 15/20 clinical samples). A mean depth of 15 achieved >99% five-fold genome coverage (in 9/20 clinical samples). In summary, direct-from-sample sequencing of MTB genomes was facilitated by a low cost thermo-protection buffer.

INTRODUCTION

Mycobacterim tuberculosis is the leading bacterial cause of death from infection, the World Health Organization (WHO) estimating that 10 million new tuberculosis (TB) cases and 1.2 million deaths occurred worldwide in 2018 (World Health Organization Global Tuberculosis Report 2019; https://www.who.int/tb/publications/global_report/en/). In addition, 5-10% of an estimated 1.7 billion people with latent TB infections are at risk of progressing to active disease. The greatest burden occurs in under-resourced regions of South-East Asia, Africa and the Western Pacific. There are large discrepancies between the estimated annual number of new cases (10 million) and the number reported (7 million) (WHO report https://www.who.int/tb/publications/global_report/en/). Consequently, diagnostic methods for use at the point of care, to identify 'missing' cases are a global priority (1, 2). Rapid diagnosis and antimicrobial resistance determination are essential to ensure appropriate TB treatment and control, particularly in light of increasing drug resistance (3, 4). In 2018, approximately 500,000 cases of rifampicin-resistant TB were identified, 78% of which were also isoniazid resistant (multi-drug resistant) (https://www.who.int/tb/publications/global_report/en/).

The application of DNA sequencing to TB molecular diagnostics yields clinically valuable information. Its utility increases with the proportion of genome obtained; from detection, to speciation and antimicrobial resistance prediction, to phylogenetic and evolutionary insights. This allows whole genome sequencing (WGS) to out-perform other rapid molecular methods (such as GeneXpert MTB/RIF, Cepheid, Solna, Sweden) due to susceptibility predictions to multiple drugs, and its combination with classical epidemiological methods which informs transmission (5-8).

WGS of MTB from early positive cultures offers markedly faster results than traditional culture-based methods which take ≤80 days. The national implementation of Illumina sequencing of Mycobacteria from early positive cultures in England provides WGS in three to four weeks, together with antimicrobial resistance predictions (5, 9). Time to WGS could be reduced further if routine sequencing could be performed direct-from-sample. Furthermore, most DNA sequencing platforms (eg Illumina) have high capital costs, require reliable power supplies, cold chain reagent shipping and highly trained staff. In contrast, the Oxford Nanopore Technologies platform, (Oxford, UK), determines nucleotide sequences via a compact, portable device powered using a laptop USB port, which can be operated in varied and challenging locations (10-13). This offers a potential direct-from-sample sequencing platform for settings with the highest burden of TB – following the third pillar of the WHO End TB Strategy, 'intensified research and innovation' (1). However, multiple sample-preparation issues remain to be solved. Sample heat inactivation is a key health and safety requirement, but this causes DNA to degrade (14) and template of sufficient quantity and quality for nanopore sequencing is rarely recovered. Furthermore, the low proportion of Mycobacteria DNA in sputum, eg 0.01% (15) leads to poor genome coverage eg 0.002 - 0.7X by Illumina (16).

Methods published to date for direct from sputum sequencing of MTB are relatively complex and have not been widely adopted. MTB enrichment using SureSelect hybridisation and amplification (Agilent, USA) yielded 90% to complete genome assemblies, allowing antimicrobial susceptibility prediction (17, 18) and informing treatment for one patient in real-time (19). An alternative approach using kit-based depletion of non-target DNAs (16) obtained wide variation in genome coverage (<12% to >90%). Both approaches included heat inactivation, the former at 80°C for 50 min and the latter at 95°C for 30 min (15, 17). These

methods depend on commercial kits which inflate cost, shipping/storage requirements, and protocol complexity.

Our aim was to develop a simple, robust, low-cost method of preparing DNA of sufficient quality for nanopore sequencing, directly from positive sputum samples. Heat inactivation was essential, but culture and DNA amplification were excluded. The method was to be immediately transferable to two high burden settings for field testing - India and Madagascar.

METHODS

Research Ethics Statement

The protocol for this study was approved by London – Queen Square Research Ethics Committee (17/LO/1420). Human samples were collected under approval of East Midlands Research Ethics Committee (08/H0406/189) and all subjects gave written informed consent in accordance with the Declaration of Helsinki.

Mock Clinical Samples for Method Development

A model system comprising standardised mock clinical samples was established by pooling infection negative human sputum samples and spiking with enumerated *Mycobacterium bovis* (BCG) Pasteur strain at known concentrations.

(i) Culture and Enumeration of BCG Cells

Culture conditions for BCG cells were optimised to yield mostly single cells which could be stained and counted, rather than rafts comprising large numbers of cells. Freshly prepared Bactec Mycobacteria growth indicator tube (MGIT) (Becton Dickinson, Wokingham, United Kingdom), UK) were inoculated very sparsely with 10 µl BCG Pasteur frozen stock. After 30 days incubation at 37°C, the culture was vortexed vigorously. Larger clusters of BCG cells were allowed to settle for 10 min. Fresh MGIT tubes containing 0.5% Tween 80 (Acros Organics, Geel, Belgium), to encourage non-clustered cell growth (20), were inoculated using 200 µl of the 'settled' BCG culture. The tubes were incubated for 18 days incubation at 37°C, then BCG cells were harvested and counted. After vigorous vortexing, 1 ml fluid was removed and BCG cells were pelleted by centrifugation for 10 min (13,000 rpm). The pellet was resuspended in 100 µl crystal violet stain (Pro Lab Diagnostics, Birkenhead, UK). Cells were counted using a Petroff Hausser counting chamber (Hausser Scientific, Horsham, PA,

125 USA) for bacteria enumeration. The enumerated BCG stock was stored at -20°C in 1 ml
126 aliquots until required.

127 **(ii) Combining Enumerated BCG Cells and Infection Negative Sputum**

128 Negative human sputum samples were obtained anonymously from asthmatic patients (see
129 research ethics statement). Up to ten samples were pooled, then liquefied by treatment with
130 an equal volume of 2x strength thermo-protection buffer (4 M KCl, 0.05 M HEPES buffer pH
131 7.5 (Sigma Aldrich, MO, USA), 0.1% DTT (Roche, UK), nuclease free molecular biology
132 grade water) to ensure a final concentration of 2 M KCl. Fresh buffer was made weekly and
133 stored in the dark at 4°C. Sputum was incubated at 37°C with occasional vortexing, until
134 liquefaction was complete. The enumerated BCG cell stock was thawed, and a 10-fold
135 dilution series was made in PBS, from 10^5 to 10^1 cells per 200 µl. The dilution series was
136 spiked into 800 µl aliquots of the liquefied sputum in 2 ml screw cap tubes to make 1ml
137 mock clinical samples. Microscopy was performed on these mock samples using ZN staining,
138 and GeneXpert semi quantitative, cartridge based PCR (Cepheid, Solna, Sweden) for
139 MTB/RIF Ultra according to the manufacturer's instructions.
140

141 **Validation of Mycobacteria Heat Inactivation**

142 A validation experiment was performed which confirmed that viable Mycobacteria did not
143 survive heat inactivation at 99°C for 30 min in thermo-protection buffer (Table 1). This 'heat-
144 kill' validation was performed prior to using the method on *Mycobacterium tuberculosis*
145 positive clinical samples or MGIT cultures. Identical control samples prepared in parallel
146 were incubated for 30 min at room temperature (Table 1). To assess Mycobacteria viability
147 post-heating, each sample was added to a freshly prepared MGIT tube. These were checked
148 regularly for Mycobacteria growth during incubation at 37°C for 8 weeks (or until positive).
149 Löwenstein-Jensen slopes were also inoculated for the heat-treated samples.

Clinical Samples

Mycobacteria-positive clinical respiratory samples comprised sputum (n=16), bronchoalveolar lavage (BAL) (n=1) and lymph node biopsies (n=3) (the latter underwent 'beating' with large glass beads in saline solution for routine diagnostic testing prior to receipt). Samples were submitted for routine testing at Birmingham Heartlands Hospital NHS Foundation Trust, Birmingham, United Kingdom (n=6), or the Clinical Microbiology Laboratory, John Radcliffe Hospital, Oxford University NHS Foundation Trust, Oxford (n=14). Oxford samples had treatment with an equal volume of Sputasol (Oxoid Limited, Basingstoke, UK) prior to receipt, and were stored at 4°C. Samples from Birmingham comprised untreated sputum shipped overnight on ice to Oxford, after which they were stored at 4°C. Prior to receipt, Microscopy (with auramine staining), had yielded acid fast bacilli scores of +1 to +3 and, or a positive MTB/RIF Ultra GeneXpert result (Cepheid, Solna, Sweden). Samples were used only after routine diagnostic tests had been completed, therefore sample quality (available volume, storage time etc) varied. Samples were processed as soon as possible after receipt.

Clinical Sample Liquefaction and Heat Inactivation

All available clinical sample volume was used. Samples were liquefied using an equal volume of thermo-protection buffer containing DTT. 1 ml aliquots were heat-inactivated in 2 ml screw cap tubes at 99°C for 30 min. Cells were collected by centrifugation (6,000 x g, 3 min) and the supernatant discarded. Then cell pellets were combined (if >1 available per sample) in a total volume of 1 ml PBS. Cells were again collected by centrifugation (6,000 x g, 3 min) and resuspended in PBS followed by another centrifugation step. The two wash

steps aimed to reduce contamination with non-target DNA. The final cell pellet was resuspended in 100 µl PBS, then total DNA was extracted.

Total DNA Extraction

0.08-0.1 g silica beads (Lysing Matrix B, MP Biomedicals, CA, USA) were added to the heat inactivated cell suspension, which underwent two rounds of mechanical disruption using an MPBio Fast Prep-24 machine (MP Biomedicals, CA, USA) at 6.0 m/s for 40 s (5 min interval). After centrifugation at 16,000 x g for 10 min at room temperature, up to 100 µl of supernatant was transferred to a fresh tube (1.5 ml DNA LoBind, Eppendorf, Hamburg, Germany). DNA in the supernatant was purified using Agencourt AMPure XP beads (Beckman Coulter, CA, USA). An equal volume of beads was added to the supernatant and incubated on a hula mixer at room temperature for 10 min. Beads with DNA bound were magnetically separated and the supernatant was removed when clear. The beads were washed using 200 µl freshly prepared 70% ethanol which was removed after a 20 s incubation. This step was repeated once more, removing as much supernatant as possible at the end of the incubation and air drying for <1 min. DNA was eluted from the beads in 50 µl 1x TE buffer (pH 8, Sigma Aldrich, MO, USA) at 35°C for 10 min. DNA concentration was measured by Qubit Fluorometer (Invitrogen, CA, USA) and the DNA Integrity Number (DIN) and fragment size by TapeStation (Agilent, CA, USA).

ONT Library Preparation and Sequencing

Undigested DNA (up to 90 ng) was prepared for ONT sequencing using the Ligation Sequencing kit (SQK-LSK109). When samples were run multiplexed (more than one per flow cell), the Native Barcoding Expansion kit (EXP-NBD104) was used. The manufacturer's protocols 'Genomic DNA by Ligation' and 'Native barcoding genomic DNA'

were followed with minor amendments; 0.8 x volume of AMPure XP beads were used to purify the end-prep and barcode ligation reactions, incubation time with AMPure XP beads was doubled, and elution was performed at 35°C. Multiplexed sequencing libraries comprised 6 barcoded DNA samples and all libraries were sequenced using R9.4.1 SpotON flow cells on GridIONs with the MinKNOW and Guppy software versions current at the time of sequencing.

Bioinformatics

Nanopore reads were basecalled using Guppy (Oxford Nanopore Technology, Oxford, UK). When one sample was sequenced per flow cell (without multiplexing), all the reads in the sequence data were analysed. For multiplexed runs with more than one sample per flow cell, we used Porechop (v0.2.2, https://github.com/rrwick/Porechop) to perform stringent barcode demultiplexing to minimize the number of misclassified reads. Porechop searches for the presence of the barcode sequence at both the start and end of each read; reads were classified only if the same barcode was found at both ends, otherwise the read was discarded. This level of stringency was achieved by setting the "require_two_barcodes" option in Porechop and setting the threshold for the barcode score at 60. (Porechop was used because much of the sequencing was performed prior to the availability of deepbinner or guppy_barcoder).

To allow the correct identification of *Mycobacterium tuberculosis* complex (MTB complex) reads from the sequencing data, we used both taxonomical classification and mapping approaches. Firstly, reads from each sample were taxonomically classified against the refSeq database using Centrifuge v1.0.3 (21). A read was considered as candidate for MTB complex if it was uniquely assigned to a species within MTB complex or equally assigned to more than one species within MTB complex. Human reads were discarded and not retained as part of our in house CRuMPIT workflow (22). Then, reads were mapped to either BCG (GenBank

AM408590.1; the 16S rRNA region {1498360, 1499896} was masked) or TB (NC_000962.2; the 16S rRNA region {1471846, 1473382} was masked) reference sequences using Minimap2, (23) depending on the type of the sample. Reads were retained if more than 85% of the bases were mapped (ie if the length of a read is 1,000 bp, >850 bp were required to be mapped to the reference sequence). Finally, MTB complex reads were identified as those agreed by Centrifuge and mapping. Integrative Genomics Viewer was used to view the mapping profiles (24). The mapping coverage and depth across the whole genome and 22 genes associated with susceptibility/resistance to clinically important antimicrobials (25) were analysed using Samtools and Pysam (https://github.com/pysam-developers/pysam).

RESULTS

Our initial experiments focused on identifying a buffer in which Mycobacteria DNA was protected from degradation during heat-inactivation. Living organisms can survive at temperatures around the boiling point of water (26), indicating that DNA can exist intact at high temperatures. The high concentrations of KCl and $MgCl_2$ found in some hyperthermophiles are thought to help protect their DNA against thermo-degradation. This has been reproduced *in vitro* using plasmid DNA (26, 27) and formed the basis of buffer optimisation experiments.

Optimisation of DNA thermo-protection buffer composition and heating duration

Three different 118 ng DNA extracts were made: (i) BCG DNA, (ii) BCG and sputum DNA, (iii) sputum DNA. These were heated at 99°C for 30 min (Oxford Clinical Microbiology Laboratory health and safety requirement) in four different buffers; 25 mM HEPES pH 7.5 plus 0, 0.5, 1 or 2 M KCl, then DNA (ng) remaining post-heating was recorded (Fig. 1A). The mass of DNA post-heating increased with increasing KCl (Fig. 1A). Furthermore, BCG DNA was better protected than sputum DNA; at 2 M KCl minimal BCG DNA degradation occurred, while >50% of sputum DNA degraded (Fig. 1A), indicating potential for BCG enrichment.

Next, we determined the impact of heating duration (0, 15, 30, 45, 60 min at 99ºC) on the same three extracted DNAs, in 2M KCl thermo-protection buffer. The percentage decrease in amount of DNA remaining after heating was plotted relative to input DNA (Fig. 1B). DNA yield declined over time, but BCG DNA was again more heat stable than sputum DNA. The 30 min time point was identified as ideal for both BCG enrichment and met health and safety requirements.

Thermo-protection of DNA in intact BCG cells

Next, we investigated whether DNA within intact Mycobacteria cells could be protected by thermo-protection buffer. BCG cells (10^5 in total) were suspended in 1 ml thermo-protection buffer and incubated at 99°C for 0, 15, 30, 45 or 60 min. Control cells were heated for the same times in phosphate buffered saline (PBS). The experiment was performed in triplicate, then the DNA was extracted. The DNA yield from BCG cells heated in thermo-protection buffer was markedly higher than those heated in PBS (Fig. 2A) except at time 0 without heating. Here, the yield of DNA was lower than expected because the cell pellet was more diffuse when it had not been heated, and cells were more easily lost than in PBS.

In a separate experiment using intact Mycobacteria cells, 13 positive MGIT cultures (anonymised discards obtained from Oxford Clinical Microbiology Laboratory) were heated in 1ml thermo-protection buffer or in culture fluid. DNA yield was improved for the cells heated in thermo-protection buffer (Fig. 2B).

Confirmation of Mycobacteria Heat Inactivation

A validation experiment was performed to confirm that viable Mycobacteria (MTB H37Rv or BCG Pasteur) did not survive 30 min heating at 99ºC in thermo-protection buffer. After 8 weeks incubation at 37°C no growth occurred in the heated samples. In contrast, the room temperature controls remained viable (Table 1).

Direct-from-Sample Sequencing of BCG-spiked Mock Clinical Samples

Four sets of mock clinical samples were made, each set containing a ten-fold dilution series of enumerated BCG cells (10^5 – 10^1 and zero cells) in 1 ml infection negative human sputum (liquefied in thermo-protection buffer). Four different batches of pooled sputum were used, but the BCG cells were from the same enumerated batch. All four replicates (experiments A-

D) underwent heat inactivation (99°C for 30 min), DNA extraction and ONT sequencing using a single flow cell per sample. Replicates in experiments B, C and D underwent additional microscopy (ZN staining) and GeneXpert PCR.

Sequencing, Microscopy and GeneXpert PCR yielded reproducible data across the replicate experiments (Fig. 3, Table 2). The number of MTB complex sequencing reads generated per sample was linear and indicated detection down to 10^1 input BCG cells. At this concentration, 31, 49, 51, and 59 MTB complex reads were detected (Fig. 3A). At 10^3 BCG cells input, genome coverage (1x) was >90% (Table 2). The ratio of human reads to MTB complex reads was also linear and reproducible (Fig. 3B).

Bioinformatics methods were optimised to ensure reads in negative controls (such as rRNA genes from non-target bacterial species (28)) were not incorrectly assigned as BCG; prior to these improvements close to 10,000 reads were incorrectly identified as MTB complex in the negative control (Fig. S1A). After the improvements, the negative controls for experiments A, C and D contained zero MTB complex reads, however three Mycobacteria reads were present in the negative control of experiment B (Fig. 3A, Figure S1A, Table 2).

GeneXpert (Cepheid) and microscopy results also followed the concentration of BCG cells spiked into each sample (Table 2). The detection limit of GeneXpert was 10^1 BCG cells and microscopy 10^2 where cells were described as very scanty, ie 1 or 2 per 100 fields (Table 2).

Direct from Sample Sequencing using Multiplexing

Sequencing more than one sample per flow cell (multiplexing), offers both time and cost efficiencies. To assess its feasibility, a short DNA ‘barcode’ was ligated to each DNA sample, then the 24 DNAs from replicate experiments A to D were sequenced at six samples per flow cell – one per replicate experiment. After sequencing, the barcodes were identified

bioinformatically and the data were assigned to their original sample. Unfortunately, the 10^1 and 10^2 BCG spiked samples contained a similar number of MTB reads to the negative control (Fig. S1B), therefore using this approach the limit of detection declined 100 fold to 10^3 BCG cells/ml sputum. This was a result of the barcodes of the BCG positive samples being incorrectly (and unavoidably) identified as that of the negative control (Fig. S1B). The multiplexing approach was also compromised by a reduction in the total data available for analysis. Although we applied stringent barcode demultiplexing criteria, between 5.28 and 46.9% of total reads cannot be reliably assigned to an input sample.

Thermo-protection method enriches mock clinical samples for MTB DNA

The multiplexed data for the mock clinical samples in experiments A to D were used separately, to confirm whether the heat inactivation in thermo-protection buffer, enriched the samples for BCG sequences by depleting human DNA. This confirmed that heated samples were enriched for MTB DNA/depleted for human DNA (Fig. 4) when compared to equivalent controls prepared without heat inactivation and washing steps.

Direct-from-Sample Sequencing of *M. tuberculosis* Positive Clinical Samples

(i) DNA Preparation using thermo-protection method

A total of 20 MTB positive clinical samples comprised 16 sputa, three lymph node biopsies, and one bronchoalveolar lavage sample. Samples were 1–14 days old, and volumes ranged from 0.25 – 1.5 ml. Microscopy and GeneXpert results indicated variable MTB loads (Table 3). The total DNA extracted ranged from 105 – 3970 ng per sample, the DNA integrity number (DIN) from 1.8 – 6.3 and the peak fragment length 1,834 – 13,949 bp (Table 3). Each sample underwent direct-from-sample sequencing using a single R9.4.1 flow cell.

(ii) Sequence data

The total number of reads obtained per flow cell ranged from 3,197,564 to 14,576,788 (Table 3). Human reads were discarded prior to detailed analysis (ethics requirement). Among the non-human reads, mean MTB read length was up to 4.77 times longer than for non-MTB (Fig. 5A). MTB reads were detected in all twenty clinical samples (n=1,825-251,256) (Table 3), the mean depth of genome coverage ranging from 0.55 to 81.02. An initial sample volume ≥1ml, and a lower percentage of human reads was apparently associated with higher depth of coverage, although the numbers were too small for statistical analysis (Table 3).

Plotting the mean depth of coverage against the percentage of MTB genome covered once, or five times revealed that ≥96% of the genome was covered once when a mean depth of approximately three was reached, and ≥99% was covered five times after a mean depth of fifteen (Fig. 5B, C). The depth of coverage across 22 key genes used to predict susceptibility to clinically important antimicrobials (25) closely followed the mean genome coverage (Table 3) (except one rRNA gene which was bioinformatically masked), indicating the absence of bias in depth of coverage across these regions of the genome, and confirming potential for antimicrobial resistance prediction.

DISCUSSION

The potential of the Nanopore platform for sequencing human and plant pathogens in varied and challenging locations is well established (10-13). It works on the principle of nanopore sequencing of DNA strands, with read lengths of several hundred to hundreds of thousands of bases obtained in real time (29, 30). The preparation of long, high quality input DNA is therefore essential. However, since DNA degrades rapidly at high temperatures (31) heat-inactivated MTB clinical samples typically yield poor quality material for sequencing. Here, we describe a simple, low cost method which overcomes this technical challenge. Sputum liquefaction and heat-inactivation were accomplished following addition of an equal volume of thermo-protection buffer (4M KCl, 0.05M HEPES buffer pH7.5, and 0.1% DTT) which inhibited DNA degradation during incubation at 99 °C. Samples were fortuitously enriched for Mycobacteria DNA under these conditions (Fig. 1, 2, 4). Buffer addition was the only handling step involving infectious material, minimising risk to staff in settings where containment laboratories are not available.

The composition of thermo-protection buffer was designed to emulate intracellular conditions of hyperthermophiles (26). At high temperatures it is thought that intracellular salts such as KCl and $MgC1_2$ protect the DNA's N-glycosidic bonds against depurination and cleavage by hydrolysis of the adjacent phosphodiester bond (14, 41, 42). We chose K^+ over Mg^{2+} because high K^+ concentrations protect against cleavage at apurinic sites, while high Mg^{2+} concentrations stimulate this (42). Furthermore, plasmid DNA appeared better protected in KCl (Fig. 1 in 42). The choice of KCl concentration (2 M) was informed by our own data (Fig. 1) and published data (26, 27). The mechanism whereby DNA in intact Mycobacteria cells was protected during heating in thermo-protection buffer (Fig. 2) is unclear, but suggests that the cell is, or becomes permeable to K^+ during heating. Our data (Fig. 2B)

373 indicate that thermo-protection buffer can also improve the DNA yield obtainable from
374 positive MGIT cultures, as currently used clinically by Public Health England for routine
375 Illumina sequencing (43, 44). This may help to reduce numbers of samples which fail due to
376 a low DNA yield.
377
378 The Oxford clinical microbiology laboratory chooses to inactivate Mycobacteria positive
379 samples at 99ºC for 30 min, because less stringent conditions (eg 20 min at 80ºC) show
380 variable efficacy (32-38). We confirmed that Mycobacteria heated in thermo-protection
381 buffer at 99ºC for 30 min were not viable (Table 1). Direct from sample sequencing results
382 are optimal when input DNA is enriched for sequences of interest (16-18, 39). Depletion of
383 up to 99.99% human DNA from non-TB lower respiratory tract samples has been achieved
384 using saponin, osmotic shock, and 'high salt' nuclease treatments (40). However, no heat-
385 inactivation was performed, and a specialist nuclease (Salt Active Nuclease, ArcticZymes,
386 Tromsø, Norway) was required. Interestingly, we observed that during heating in thermo-
387 protection buffer for 30 min at 99ºC, non-target DNA degraded more rapidly than
388 Mycobacteria DNA, providing fortuitous enrichment (Fig. 1, 2, 4). Consistent with this, mean
389 read length obtained for MTB was longer than other non-human reads (Fig. 5A). We found
390 that sufficient non-target DNA remained in our samples to provide a useful 'carrier'. This
391 was particularly important in low titre MTB samples; our 10^1 BCG limit of detection in mock
392 clinical samples (Fig. 3A) would not have been achieved if the majority of human 'carrier'
393 DNA had been removed.
394
395 Three features of Mycobacteria DNA may have contributed to its enrichment relative to
396 human DNA (Fig. 1, 4). Firstly, Mycobacteria DNA has a higher GC content (*M.*
397 *tuberculosis* 65.6% GC, versus <50% GC for ~92% of human DNA (45, 46). Secondly, intact

MTB chromosomes are covalently closed circles - resistant to thermo-denaturation because the two single strands remain intertwined during heating (47). Finally, the MTB chromosome is negatively supercoiled (underwound – a feature potentially connected to its slow growth rate (48)), but less so than some bacterial species including *E.coli* (49). Lesser negative supercoiling reduces base exposure (50), which may reduce susceptibility to thermo-degradation, relative to human DNA. We did not assess the impact of heating on DNA from non-target microbes because the range and number of cells present was unpredictable and bioinformatics analysis cannot give a reliable taxonomic classification for approximately 20% of the total reads in each sample.

We obtained optimal results when using a single R9.4.1 flow cell per sample (Fig. 3, Table 3). This approach would be prohibitively expensive if used routinely, since a flow cell costs £380 - £720 depending on order size (1 to 300). Unfortunately, multiplexing six samples per flow cell did not provide a solution, since the inefficiency of barcode ligation and incorrect bar code identification post-sequencing reduced the limit of detection 100 fold (Fig. S1). Run-time flexibility is also incompatible with multiplexing, due to variations in sample MTB titres. A solution may be to wash, re-generate and re-use flow cells after each use (https://store.nanoporetech.com/flow-cell-wash-kit-r9.html), or adopt single-use ONT Flongles at a cost of £72.50 each (at 30/01/2020). Unfortunately, the latter currently offer only 60-70 active sequencing pores in our hands, compared to 1200 – 1500 pores per R9.4.1 flow cell.

The thermo-protection method was applied successfully to clinical samples (n=20). Higher mean depth of genome coverage appeared to reflect initial sample volume (≥1 ml being ideal), and a lower human DNA content, but did not necessarily correlate with microscopy or

GeneXpert Ct values (Table 3). This may reflect the known variation in copy numbers of GeneXpert targeted insertion sequences (IS6110 and IS1081) between BCG (used in mock samples, where limit of detection, microscopy and GeneXpert data all correlated, Table 2) and MTB (51, 52). Also, microscopy was performed using auramine staining for clinical samples, ZN staining for mock clinical samples, and the former was performed by multiple different staff members.

The accuracy of DNA consensus sequences obtained using the Nanopore platform is 99.9% when Nanopolish is used (R.R. Wick, L.M. Judd, K.E. Holt. 2018. Comparison of Oxford Nanopore basecalling tools; https://github.com/rrwick/Basecalling-comparison#references), indicating potential for antimicrobial susceptibility prediction. Further work is required to examine this aspect in detail, particularly for ribosomal RNA genes which were masked to improve the accuracy of MTB detection, but are implicated in resistance (such as the 16s rRNS rrs gene; aminoglycoside resistance). A further potential advantage of direct from sample sequencing (using SureSelect and Illumina) is the detection of more genetic diversity than sequencing from culture (53). Increased numbers of target reads from low titre samples will require innovations such as Mycobacteria cell fractionation or concentration, followed by DNA amplification. The method could be enhanced by adapting to a cartridge-based system, further simplifying its application in resource poor settings.

In summary, a simple, low-cost method was developed to prepare MTB DNA for Nanopore sequencing direct from clinical samples. Neither commercial kits, nor time consuming culture were required, but the key health and safety requirement, heat-inactivation, was retained and exploited to achieve target sequence enrichment. Available data suggest the method can yield complete MTB genome sequences direct from clinical samples, without amplification,

448 achieving up to 81 fold mean depth of coverage. The protocol is currently undergoing testing

449 by collaborators in India and Madagascar, with early data indicating reproducibility.

REFERENCES

1. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, Falzon D, Floyd K, Gargioni G, Getahun H, Gilpin C, Glaziou P, Grzemska M, Mirzayev F, Nakatani H, Raviglione M; for WHO's Global TB Programme. 2015. WHO's new End TB Strategy. Lancet 385:1799-1801.

2. Sachdeva KS, Mase SR. 2019. The end TB strategy for India. Indian J Tuberc. 66:165-166.

3. Tiberi S, Zumla A, Migliori GB. 2019. Multidrug and Extensively Drug-resistant Tuberculosis: Epidemiology, Clinical Features, Management and Treatment. Infect Dis Clin North Am. 33:1063-1085.

4. Dheda K, Gumbo T, Maartens G, Dooley KE, Murray M, Furin J, Nardell EA, Warren RM; Lancet Respiratory Medicine drug-resistant tuberculosis Commission group. 2019. The Lancet Respiratory Medicine Commission: 2019 update: epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant and incurable tuberculosis. Lancet Respir Med. 7:820-826.

5. Walker TM, Cruz ALG, Peto TE, Smith EG, Esmail H, Crook DW. 2017. Tuberculosis is changing. Lancet Infect Dis. 17:359-361.

6. CRyPTIC Consortium and the 100,000 Genomes Project, Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, Cabibbe AM, Cancino-Muñoz I, Caulfield MJ, Chaiprasert A, Cirillo DM, Clifton DA, Comas I, Crook DW, De Filippo MR, de Neeling H, Diel R, Drobniewski FA, Faksri K, Farhat MR, Fleming J, Fowler P, Fowler TA, Gao Q, Gardy J, Gascoyne-Binzi D, Gibertoni-Cruz AL, Gil-Brusola A, Golubchik T, Gonzalo X, Grandjean L, He G, Guthrie JL, Hoosdally S, Hunt M, Iqbal Z, Ismail N, Johnston J, Khanzada FM, Khor CC, Kohl TA, Kong C, Lipworth S, Liu Q, Maphalala G, Martinez E, Mathys V, Merker M, Miotto P, Mistry N, Moore DAJ, Murray M, Niemann S,

475 Omar SV, Ong RT, Peto TEA, Posey JE, Prammananan T, Pym A, Rodrigues C, Rodrigues
476 M, Rodwell T, Rossolini GM, Sánchez Padilla E, Schito M, Shen X, Shendure J, Sintchenko
477 V, Sloutsky A, Smith EG, Snyder M, Soetaert K, Starks AM, Supply P, Suriyapol P, Tahseen
478 S, Tang P, Teo YY, Thuong TNT, Thwaites G, Tortoli E, van Soolingen D, Walker AS,
479 Walker TM, Wilcox M, Wilson DJ, Wyllie D, Yang Y, Zhang H, Zhao Y, Zhu B. 2018.
480 Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J
481 Med. 379:1403-1415.
482 7. Lalor MK, Casali N, Walker TM, Anderson LF, Davidson JA, Ratna N, Mullarkey C,
483 Gent M, Foster K, Brown T, Magee J, Barrett A, Crook DW, Drobniewski F, Thomas HL,
484 Abubakar I. 2018. The use of whole-genome sequencing in cluster investigation of a
485 multidrug-resistant tuberculosis outbreak. Eur Respir J. 51pii:1702313.
486 8. Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, Parker L, Churchill S,
487 Bennett K, Golubchik T, Giess AP, Del Ojo Elias C, Jeffery KJ, Bowler ICJW, Laurenson IF,
488 Barrett A, Drobniewski F, McCarthy ND, Anderson LF, Abubakar I, Thomas HL, Monk P,
489 Smith EG, Walker AS, Crook DW, Peto TEA, Conlon CP. 2014. Assessment of
490 *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole
491 pathogen genome sequences: an observational study. Lancet Respir Med. 2:285-292.
492 9. Quan TP, Bawa Z, Foster D, Walker T, Del Ojo Elias C, Rathod P; MMM Informatics
493 Group, Iqbal Z, Bradley P, Mowbray J, Walker AS, Crook DW, Wyllie DH, Peto TEA,
494 Smith EG. 2018. Evaluation of Whole-Genome Sequencing for Mycobacterial Species
495 Identification and Drug Susceptibility Testing in a Clinical Setting: a Large-Scale Prospective
496 Assessment of Performance against Line Probe Assays and Phenotyping. J Clin Microbiol. 56
497 pii:e01480-17.
498

499 10. Quick, J, Quinlan, AR, Loman, NL. 2014. A reference bacterial genome dataset
500 generated on the MinION portable single-molecule nanopore sequencer. GigaScience 3:22.
501 11. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA,
502 Koundouno R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JH, Becker-
503 Ziaja B, Boettcher JP, Cabeza-Cabrerizo M, Camino-Sanchez A, Carter LL, Doerrbecker J,
504 Enkirch T, Dorival IGG, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M,
505 Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D,
506 Nitzsche K, Pallash E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan
507 K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Trina R, Bello A, Sall AA, Faye O, Faye O,
508 Magassouba N, Williams CV, Amburgey V, Winona L, Davis E, Gerlach J, Washington F,
509 Monteil V, Jourdain M, Bererd M, Camara A, Somlare H, Camara A, Gerard M, Bado G,
510 Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB, Gutierrez GJ, Milhano N,
511 Roger I, Williams CJ, Yattara F, Lewandowski K, Taylor J, Rachwal P, Turner D, Pollakis G,
512 Hiscox JA, Matthews DA, O'Shea MK, Johnston AM, Wilson D, Hutley E, Smit E, Di Caro
513 A, Woelfel R, Stoecker K, Fleischmann E, Gabriel M, Weller SA, Koivogui L, Diallo B,
514 Keita S, Rambaut A, Formenty P, Gunther S, Carroll MW. 2016. Real-time, portable genome
515 sequencing for Ebola surveillance. Nature 530:228-232.
516 12. de Jesus JG, Giovanetti M, Rodrigues Faria N, Alcantara LCJ. 2019. Acute Vector-
517 Borne Viral Infection: Zika and MinION Surveillance. Microbiol Spectr. 7(4).
518 13. Boykin LM, Sseruwagi P, Alicai T, Ateka E, Mohammed IU, Stanton JL, Kayuki C,
519 Mark D, Fute T, Erasto J, Bachwenkizi H, Muga B, Mumo N, Mwangi J, Abidrabo P, Okao-
520 Okuja G, Omuut G, Akol J, Apio HB, Osingada F, Kehoe MA, Eccles D, Savill A, Lamb S,
521 Kinene T, Rawle CB, Muralidhar A, Mayall K, Tairo F, Ndunguru J. 2019. Tree Lab:
522 Portable genomics for Early Detection of Plant Viruses and Pests in Sub-Saharan Africa.
523 Genes 10(9) pii: E632.

524 14. Lindahl T. 1993. Instability and decay of the primary structure of DNA. Nature.
525 362:709-715.
526 15. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K,
527 Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker
528 AS3,10, Peto TEA, Crook DW, Iqbal Z. 2017. Same-day diagnostic and surveillance data for
529 tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol.
530 55:1285-1298.
531 16. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ. 2014. Culture
532 independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum*
533 in sputum samples using shotgun metagenomics on a benchtop sequencer. PeerJ 2: e585.
534 17. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZ, Depledge
535 DP, Nikolayevskyy V, Broda A, Stone MJ, Christiansen MT, Williams R, McAndrew MB,
536 Tutill H, Brown J, Melzer M, Rosmarin C, McHugh TD, Shorten RJ, Drobniewski F, Speight
537 G, Breuer J. 2015. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates
538 directly from clinical samples. J Clin Microbiol. 53:2230-2237.
539 18. Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, Bryant JM, Chan J,
540 Creer D, Holdstock J, Kunst H, Lozewicz S, Platt G, Romero EY, Speight G, Tiberi S,
541 Abubakar I, Lipman M, McHugh TD, Breuer J. 2018. Direct Whole-Genome Sequencing of
542 Sputum Accurately Identifies Drug-Resistant *Mycobacterium tuberculosis* Faster than MGIT
543 Culture Sequencing. J Clin Microbiol. 56(8). pii: e00666-18.
544 19. Nimmo C, Doyle R, Burgess C, Williams R, Gorton R, McHugh TD, Brown M,
545 Morris-Jones S, Booth H, Breuer J. 2017. Rapid identification of a *Mycobacterium*
546 *tuberculosis* full genetic drug resistance profile through whole genome sequencing directly
547 from sputum. Int J Infect Dis. 62:44-46.

548 20. Caceres N, Vilaplana C, Prats C, Marzo E, Llopis I, Valls J, Lopez D, Cardona PJ.
549 2013. Evolution and role of corded cell aggregation in *Mycobacterium tuberculosis* cultures.
550 Tuberculosis 93:690-698.
551 21. Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: rapid and sensitive
552 classification of metagenomic sequences. Genome Res. 26:1721–1729.
553 22. Sanderson ND, Street TL, Foster D, Swann J, Atkins BL, Brent AJ, McNally MA,
554 Oakley S, Taylor A, Peto TEA, Crook DW, Eyre DW. 2018. Real-time analysis of nanopore-
555 based metagenomic sequencing from infected orthopaedic devices. BMC Genomics 19:710.
556 23. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics
557 34:3094–3100.
558 24. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G,
559 Mesirov, JP. 2011. Integrative Genomics Viewer. Nature Biotechnology 29:24–26.
560 25. Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, Iqbal Z,
561 Feuerriegel S, Niehaus KE, Wilson DJ, Clifton DA, Kapatai G, Ip CLC, Bowden R,
562 Drobniewski FA, Allix-Béguec C, Gaudin C, Parkhill J, Diel R, Supply P, Crook DW, Smith
563 EG, Walker AS, Ismail N, Niemann S, Peto TEA; Modernizing Medical Microbiology
564 (MMM) Informatics Group. 2015. Whole-genome sequencing for prediction of
565 *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study.
566 Lancet Infect Dis. 15:1193-1202.
567 26. Marguet E, Forterre P. 1994. DNA stability at temperatures typical for
568 hyperthermophiles. Nucleic Acids Res. 22:1681-1686.
569 27. Marguet E, Forterre P. 2001. Stability and manipulation of DNA at extreme
570 temperatures. Methods Enzymol. 334:205-215.
571 28. Wyllie DH, Sanderson N, Myers R, Peto T, Robinson E, Crook DW, Smith EG,
572 Walker AS. 2018. Control of Artifactual Variation in Reported Intersample Relatedness

573 during Clinical Use of a *Mycobacterium tuberculosis* Sequencing Pipeline. J Clin Microbiol.
574 56 pii: e00104-18.
575 29. Deamer D, Akeson M, Branton D. 2016. Three decades of nanopore sequencing. Nat
576 Biotechnol 34:518-524.
577 30. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION:
578 delivery of nanopore sequencing to the genomics community. Genome Biol 17: 239.
579 31. Gustafson CE, Alm RA, Trust TJ. 1993. Effect of heat denaturation of target DNA on
580 the PCR amplification. Gene 123:241-244.
581 32. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B,
582 Hermans P, Martin C, McAdam R, Shinnick TM. 1993. Strain identification of
583 *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized
584 methodology. J Clin Microbiol. 31:406-409.
585 33. Doig C, Seagar AL, Watt B, Forbes KJ. 2002. The efficacy of the heat killing of
586 *Mycobacterium tuberculosis*. J Clin Pathol. 55:778-779.
587 34. Zwadyk P Jr1, Down JA, Myers N, Dey MS. 1994. Rendering of mycobacteria safe
588 for molecular diagnostic studies and development of a lysis method for strand displacement
589 amplification and PCR. J Clin Micobiol. 32:2140-2146.
590 35. Bemer-Melchior P, Drugeon HB. 1999. Inactivation of *Mycobacterium tuberculosis*
591 for DNA typing analysis. J Clin Microbiol. 37:2350-2351.
592 36. Somerville W, Thibert L, Schwartzman, K, Behr MA. 2005. Extraction of
593 *Mycobacterium tuberculosis* DNA: a question of containment J Clin Microbiol. 43:2996-
594 2997.
595 37. Warren R, de Kock M, Engelke E, Myburgh R, Gey van Pittius N, Victor T, van
596 Helden P. 2006. Safe Mycobacterium tuberculosis DNA extraction method that does not
597 compromise integrity. J Clin Microbiol. 44:254-256.

598 38. Billard-Pomares T, Bleibtreu A, Walewski V, Dziri S, Barbat A, Zahar JR, Cruaud P,
599 Carbonnelle E. 2019. Proposition of a safe *Mycobacterium tuberculosis* complex denaturation
600 method that does not compromise the integrity of DNA for whole-genome sequencing.
601 Tuberculosis 117:62-64.
602 39. Pendleton KM, Erb-Downward JR, Bao Y, Branton WR, Falkowski NR, Newton
603 DW, Huffnagle GB, Dickson RP. 2017. Rapid Pathogen Identification in Bacterial
604 Pneumonia Using Real-Time Metagenomics. Am J Respir Crit Care Med. 196:1610-1612.
605 40. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, Rae D,
606 Grundy S, Turner DJ, Wain J, Leggett RM, Livermore DM, O'Grady J. 2019. Nanopore
607 metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat
608 Biotechnol. 37:783-792.
609 41. Lindahl T., Nyberg, B. 1972. Rate of depurination of native deoxyribonucleic acid.
610 Biochemistry. 11:3610-3618.
611 42. Marguet E and Forterre P. 1998. Protection of DNA by salts against
612 thermodegradation at temperatures typical for hyperthermophiles. Extremophiles. 2:115-122.
613 43. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker
614 TM, Quan TP, Wyllie DH, Del Ojo Elias C, Wilcox M, Walker AS, Peto TE, Crook DW.
615 2015. Mycobacterial DNA extraction for whole-genome sequencing from early positive
616 liquid (MGIT) cultures. J Clin Microbiol. 53:1137-1143.
617 44. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J,
618 Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J,
619 Rogers TR11, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie
620 D, Xu L, Crook DW; COMPASS-TB Study Group. 2016. Rapid, comprehensive, and
621 affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study.
622 Lancet Respir Med. 4:49-58.

623 45. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV,
624 Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D,
625 Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd
626 S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail
627 MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston
628 JE, Taylor K, Whitehead S, Barrell BG. 1988. Deciphering the biology of *Mycobacterium*
629 *tuberculosis* from the complete genome sequence. Nature 393:537-544.
630 46. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell
631 M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR,
632 Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J,
633 Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N,
634 Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I,
635 Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C,
636 Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R,
637 Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn
638 P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman
639 TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F,
640 Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D,
641 Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M,
642 Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L,
643 Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A,
644 Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K,
645 Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R,
646 Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes
647 C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F,

Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. 2001. The sequence of the human genome. Science 291:1304-1351.

47. Vinograd J, Lebowitz J, Watson R. 1968. Early and late helix-coil transitions in closed circular DNA. The number of superhelical turns in polyoma DNA. J Mol Biol. 33:173-197.

48. Higgins NP. 2014. RNA polymerase: chromosome domain boundary maker and regulator of supercoil density. Curr Opin Microbiol. 22:138-143.

49. Tretter EM, Berger JM. 2012. Mechanisms for defining supercoiling set point of DNA gyrase orthologs: II. The shape of the *gyrA* subunit C-terminal domain (CTD) is not a sole determinant for controlling supercoiling efficiency. J Biol Chem. 287:18645-18654.

672 50. Irobalieva RN, Fogg JM, Catanese DJ Jr, Sutthibutpong T, Chen M, Barker AK,
673 Ludtke SJ, Harris SA, Schmid MF, Chiu W, Zechiedrich L. 2015. Structural diversity of
674 supercoiled DNA. Nat Commun. 6:8440.
675 51. Gonzalo-Asensio J, Pérez I, Aguiló N, Uranga S, Picó A, Lampreave C, Cebollada A,
676 Otal I, Samper S, Martín C. 2018. New insights into the transposition mechanisms of IS6110
677 and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. PLoS
678 Genet. 14:e1007282.
679 52. van Soolingen D, Hermans PW, de Haas PE, van Embden JD. 1992. Insertion element
680 IS1081-associated restriction fragment length polymorphisms in Mycobacterium tuberculosis
681 complex species: a reliable tool for recognizing *Mycobacterium bovis* BCG. J Clin Microbiol.
682 30:1772-1777.
683 53. Nimmo C, Shaw LP, Doyle R, Williams R, Brien K, Burgess C, Breuer J, Balloux F,
684 Pym AS. 2019. Whole genome sequencing *Mycobacterium tuberculosis* directly from sputum
685 identifies more genetic diversity than sequencing from culture. BMC Genomics. 20:389.

686 **FIGURE LEGENDS**

687 **Fig. 1**

688 **Optimisation of DNA thermo-protection buffer composition and duration of heat**
689 **inactivation at 99°C**

690 (A) Extracted DNA was heated in 25 mM HEPES buffer pH 7.5 containing 0, 0.5, 1 or 2 M
691 KCl. Input DNA comprised 118 ng of (i) BCG DNA, (ii) BCG and sputum DNA, (iii)
692 sputum DNA. Each DNA type was heated at 99°C, for 30 min.
693 (B) Impact of heating duration on DNA yield. DNA remaining post heating is expressed as a
694 percentage of the input DNA for (i) 10^5 BCG cells, (ii) 1ml sputum spiked with 10^5 BCG
695 cells, or (iii) 1ml sputum. BCG DNA degraded more slowly than sputum DNA, indicating the
696 potential for enrichment relative to human DNA at earlier time points.
697

698 **Fig. 2**

699 **Thermo-protection of DNA in intact BCG cells**

700 (A) Effect on DNA yield of heating intact enumerated 'de-clumped' BCG cells in thermo-
701 protection buffer for the times shown. 10^5 BCG cells were heated at 99°C for 0, 15, 30, 45 or
702 60 min in 2 M KCl and 25 mM HEPES pH 7.5 (thermo-protection buffer) or PBS (control).
703 The experiment was performed in triplicate and DNA was extracted post heating.
704 (B) DNA yield obtained when heating intact Mycobacteria cells from positive MGIT culture
705 in thermo-protection buffer versus heating in MGIT culture fluid. Data are shown for 13
706 positive MGIT cultures. The DNA yield obtained after heating for 30 min at 99°C in thermo-
707 protection buffer, compared to heating in MGIT fluid is plotted. Each dot indicates the total
708 DNA recovered (ng) from 1 ml initial MGIT culture. Numbers above the dots indicate the
709 fold improvement in DNA yield when thermo-protection buffer was used rather than MGIT

710 culture fluid. N/A indicates a sample where no DNA was recovered after heating in MGIT
711 culture fluid, so no 'fold improvement' could be calculated.
712
713 **Fig. 3**
714 **Validation of DNA thermo-protection method using Mock Clinical Samples**
715 Mock clinical samples containing enumerated BCG cells ($0 – 10^5$) in 1ml infection negative
716 human sputum liquefied in thermo-protection buffer underwent heat-inactivation at 99°C for
717 30 min. DNA was extracted and sequenced on a ONT MinION (1 R9.4.1 flow cell per
718 sample). Reproducibility was assessed using four replicate experiments (A-D).
719 (A) Number of MTB complex reads generated per sample was linear and indicated a
720 detection limit of 10^1 BCG cells.
721 (B) Ratio of human reads to MTB complex reads.
722
723 **Fig. 4**
724 **Mock clinical samples were enriched for Mycobacteria DNA after heating in thermo-**
725 **protection buffer.** Data are shown for four replicate experiments, A to D, in which samples
726 were barcoded and run multiplexed six per flow cell. Each experiment comprised samples
727 made from a batch of infection negative sputum (liquefied using thermo-protection buffer
728 containing DTT), 1 ml aliquots of which were spiked with enumerated BCG cells at 10^5 to
729 10^1 cells, and zero BCG cells (control). Sputum batch and therefore 'background' DNA did
730 not vary within replicates A to D, only between them. The full set of replicates was set up
731 twice with heating (99°C, 30 min), and without heating. After sequencing, the numbers of
732 BCG and human derived reads was assessed and their ratio in each sample calculated. Higher
733 ratios of human : MTB reads were obtained for samples which were not heated in thermo-
734 protection buffer, indicating heated samples were enriched for MTB reads relative to human

735 reads ie human DNA was depleted. The exception to this was experiment B, which yielded
736 anomalous results because the number of reads for the unheated sample was unusually poor.
737
738 **Fig. 5**
739 **Sequence data generated direct from clinical samples - mean read lengths (MTB vs non**
740 **MTB non-human sequences) and relationship between mean depth of coverage and**
741 **complete genome coverage.**
742 (A) Comparison, for each of 20 clinical samples, of mean read length for MTB and non-MTB
743 sequences (human sequences excluded prior to analysis). Clinical samples were ranked
744 according to mean depth of coverage, indicated by numbers above the bars.
745 (B) Relationship between mean depth of coverage and percentage of the MTB genome
746 covered once. A mean depth of coverage of three is required to achieve >96% one-fold TB
747 genome coverage, achieved in 15/20 clinical samples.
748 (C) Relationship between mean depth of coverage and percentage of the MTB genome
749 covered five times. A mean depth of coverage of 15 is required to achieve >99% five-fold
750 genome coverage, achieved in 9/20 clinical samples.
751
752 **Fig. S1**
753 **Improvements to bioinformatic analysis of Nanopore sequence data generated direct-**
754 **from-sample.**
755 (A) Analysis of results generated from mock clinical samples, each sequenced on a single
756 flow cell. Samples comprised BCG spiked infection negative sputum containing 10^5 to 10^1
757 and zero BCG cells. Data are shown before (blue dots) and after (red dots) mapping
758 improvements. Prior to mapping improvements, close to 10,000 reads per sample were
759 incorrectly identified as MTB complex. After the improvements, three contaminant reads

were detected in the negative control (zero BGC cells) of experiment B only. These were 612, 788, and 1305 bases long and mapped independently and at high quality to the BCG reference genome, at positions 2798622-2799216, 282379-283169 and 2701067-2702371, the identity of these reads was further confirmed by BLASTn.

(B) Analysis of results generated from mock clinical samples, 'barcoded' and sequenced multiplexed six per flow cell. The results are shown before (blue dots) and after (red dots) mapping improvements. Even after mapping improvements, many reads persisted in the low BCG titre samples and negative controls. So, when running multiplexed samples, the limit of detection was compromised at 10^3. The reason for this was the barcodes of DNA fragments belonging to BCG positive samples were incorrectly (and unavoidably) identified as the barcode of the negative control. The multiplexing approach was also compromised by a reduction in the total data available for analysis, since a relatively high proportion of the total reads were unbarcoded.

774 **Table 1. Heat Inactivation Validation**

Sample type	Mycobacteria cells*[1]	Heat treatment	Time to positive MGIT culture
(i) 1 ml pooled negative human sputum liquefied with thermo-protection buffer containing DTT.	MTB H37Rv	Room temp 30 min (control)	10 days
	MTB H37Rv	99°C 30 min	Negative
	BCG Pasteur	Room temp 30 min (control)	18 days
	BCG Pasteur	99°C 30 min	Negative
(ii) 1 ml sputasol treated sputum*[2] to which an equal volume of thermo-protection buffer was added	MTB H37Rv	Room temp 30 min (control)	7 days
	MTB H37Rv	99°C 30 min	Negative
	BCG Pasteur	Room temp 30 min (control)	8 days
	BCG Pasteur	99°C 30 min	Negative
(iii) 0.5 ml positive MGIT culture plus equal volume thermo-protection buffer	MTB H37Rv	Room temp 30 min (control)	2 days
	MTB H37Rv	99°C 30 min*[3]	Negative
	BCG Pasteur	Room temp 30 min (control)	3 days
	BCG Pasteur	99°C 30 min*[3]	Negative
(iv) 1 ml positive MGIT culture spun down. Pellet resuspended in 1 ml thermo-protection buffer	MTB H37Rv	Room temp 30 min (control)	4 days
	MTB H37Rv	99°C 30 min	Negative
	BCG Pasteur	Room temp 30 min (control)	3 days
	BCG Pasteur	99°C 30 min	Negative

775
776 The final concentration of KCl used in each heat inactivation experiment (i) to (iv) was 2M.
777
778 *[1]Spiking inoculum for sputum comprised live cultured BCG or *M. tuberculosis* H37Rv cells
779 prepared by pelleting cells from 1 ml MGIT culture by centrifugation at 13,000 rpm for 10
780 minutes, then resuspending in PBS (1 ml). One drop was used as the inoculum.
781
782 *****[2]Sputum samples received by the Clinical Microbiology Laboratory, John Radcliffe
783 Hospital, Oxford, without a request for TB testing were decontaminated by treatment with
784 4% NaOH (E and O Laboratories Ltd, Bonnybridge, Scotland), neutralised, spun down, and
785 resuspended in 1ml sputasol. They were then spiked with 1 drop inoculum*[1].
786
787 *[3]A precipitate formed on heating with thermo-protection buffer, possibly comprising
788 salt/antibiotics/media components.
789

Table 2. Reproducibility and detection limits of Microscopy, GeneXpert, and direct from sample ONT Sequencing.

Input comprised mock clinical samples. These were pooled infection negative human sputum samples, liquefied using thermo-protection buffer and spiked using enumerated BCG cells.

BCG cells	Microscopy	GeneXpert		Nanopore			
Experiment B	**ZN Stain**	**Ct value**	**Detection**	**MTB complex Reads (n)**	**Mean Depth**	**Genome covered 1x (%)**	**Genome covered 5x (%)**
10^5	+++	16.4	High	701,436	261.99	98.73	98.12
10^4	+++	16.5	High	142,918	55.0	98.30	97.75
10^3	+	17.1	Medium	17,798	7.17	97.86	83.56
10^2	(+)	19.3	Low	1,280	0.56	44.26	0.04
10^1	-	23.0	Very Low	51	0.03	2.92	0
0	-	-	Negative	3	0	0.06	0
Experiment C							
10^5	+++	16.2	High	738,605	225.73	98.48	97.91
10^4	+++	16.3	High	74,731	20.19	97.76	97.12
10^3	+	17.1	Medium	9,086	2.75	91.84	17.98
10^2	+	17.6	Medium	672	0.21	19.11	0
10^1	-	22.7	Low	31	0.01	1.38	0
0	-	-	Negative	0	0	0	0
Experiment D							
10^5	+++	16.2	High	333,937	99.92	98.15	97.62
10^4	++	16.4	High	57,824	16.28	97.66	97.13
10^3	+	17.0	Medium	9,533	3.37	93.97	29.65
10^2	+	18.9	Low	494	0.19	17.51	0
10^1	-	23.0	Very Low	59	0.02	2.29	0
0	-	-	Negative	0	0	0	0

Table 3. Direct ONT Sequencing of MTB Positive Clinical Samples

Sample Data					Parallel Laboratory Tests		Extracted DNA		
ID		Type	Volume	Age	Microscopy	GeneXpert	Total DNA	Peak Fragment Length	DNA Integrity Number
			(ml)	(days)		(Ct)	(ng)	(bp)	DIN
T15211	O	LN	solid	3	+++	16.2	3590	1834	3.3
19.0609294	B	Sputum	1	2	+	28.4	422	10760	5.9
19.0609025	B	Sputum	1	8	NT	16.0	138	12266	6.0
L34626	O	Sputum	0.5	1	+++	15.9	912	8757	6.2
L32975	O	BAL	1	14	+	16.2	241	9189	6.1
19.0608818	B	Sputum	1	12	NT	16.2	126	13192	6.1
L87135_1	O	Sputum	1	14	+++	16.3	3187	7993	4.0
19.0608426	B	Sputum	1	7	+++	NT	3970	2403	4.4
L87135	O	Sputum	0.5	11	+++	16.3	926	6859	3.6
L87133	O	Sputum	1	14	+++	NT	3686	7000	4.4
19.0609396	B	Sputum	1.5	2	+	29.3	3206	6900	5.6
L11276	O	Sputum	0.25	1	+++	16.1	1735	8443	5.7
L99052	O	LN	solid	3	++	16.4	105	Could not be determined	1.8
L99521	O	Sputum	1	1	++	16.0	3600	13949	6.3
L37997	O	Sputum	0.75	3	+	16.1	228	11879	4.0
19.0608494	B	Sputum	0.5	6	NT	16.0	2610	1957	3.5
L91635	O	Sputum	0.5	2	+	17.2	2909	9044	6.2
L96231	O	Sputum	1	8	+	NT	1133	8566	4.6
L11990	O	Sputum	0.25	13	+	18.5	270	6243	5.7
W63114	O	LN	solid	2	++	NT	265	1990	4.7

O: Microbiology Department, Oxford University Hospitals NHS Trust, Oxford, UK.
B: PHE National Mycobacteria Reference Service - North and Central, Birmingham Public Health Laboratory, UK
LN: Lymph node, solid piece of tissue, disrupted by bead beating in saline before receipt.
Microscopy: +++ large numbers of cells, strong positive, ++ moderate numbers, + scanty weakly positive.
Volume: equivalent volume of initial clinical sample received.
NT: Not tested.

Table 3. Direct ONT Sequencing of MTB Positive Clinical Samples – Continued.

Sample Data			Sequence Data										
ID		Type	Total Reads	Human Reads		MTB Complex Reads		Mean Depth of Coverage		Genome covered 1x (%)		Genome covered 5x (%)	
			(n)	(n)	(%)	(n)	(%)	Genome	R genes	Genome	R genes	Genome	R genes
T15211	O	LN	3755361	2806761	74.74	251256	6.69	81.02	85.59	99.60	100	99.53	100
19.0609294	B	Sputum	14576788	8472212	58.12	80357	0.55	45.01	45.30	99.61	100	99.49	100
19.0609025	B	Sputum	9108521	6675047	73.28	104748	1.15	36.40	38.08	99.59	100	99.51	100
L34626	O	Sputum	3197564	2315352	72.41	56421	1.76	24.19	24.37	99.58	100	99.38	100
L32975	O	BAL	4119806	2497293	60.62	52536	1.28	23.02	25.24	99.84	100	99.75	100
19.0608818	B	Sputum	5979711	3730501	62.39	43148	0.72	22.30	22.97	99.61	100	99.28	100
L87135_1	O	Sputum	7617356	5034871	66.10	38989	0.51	19.77	14.94	99.60	100	99.41	100
19.0608426	B	Sputum	6635435	6404796	96.52	74672	1.13	15.74	16.62	99.40	100	98.74	100
L87135	O	Sputum	5637884	3569630	63.32	26216	0.46	14.58	14.94	99.60	100	99.10	100
L87133	O	Sputum	7824904	6971470	89.09	12740	0.16	6.50	6.78	99.38	99.63	77.08	75.35
19.0609396	B	Sputum	6719776	6022913	89.63	14491	0.22	5.89	6.23	99.05	100	71.39	77.42
L11276	O	Sputum	5514521	5397278	97.87	9505	0.17	4.75	3.85	98.39	99.48	53.37	32.96
L99052	O	LN	13507355	13268163	98.23	7124	0.05	3.36	2.90	96.03	92.19	27.27	22.47
L99521	O	Sputum	7521636	7438072	98.89	4701	0.06	3.27	3.69	95.81	100	25.06	23.82
L37997	O	Sputum	8014834	7714434	96.25	4650	0.06	2.89	2.83	93.14	95.02	19.39	18.37
19.0608494	B	Sputum	11811075	10915937	92.42	8091	0.07	2.06	2.03	86.48	83.73	7.24	6.25
L91635	O	Sputum	4271540	4243731	99.35	1975	0.05	1.57	1.76	78.83	76.27	2.91	2.35
L96231	O	Sputum	8829259	8572310	97.09	1545	0.02	1.14	1.32	67.35	61.46	1.24	3.01
L11990	O	Sputum	4200973	145749	3.47	1898	0.05	0.86	0.73	57.53	48.04	0.34	0
W63114	O	LN	7408231	7351842	99.24	1825	0.02	0.55	0.71	42.97	48.74	0.05	0

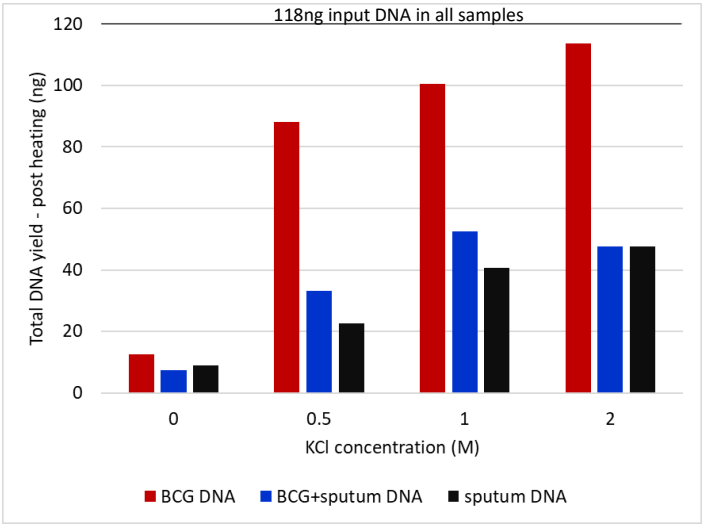
R genes: resistance genes – coverage across 22 genes associated with susceptibility/resistance to clinically important antimicrobials (25).

807 **FUNDING**

808 The study was funded by the NIHR Oxford Biomedical Research Centre. Computation used
809 the Oxford Biomedical Research Computing (BMRC) facility, a joint development between
810 the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health
811 Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed
812 in this publication are those of the authors and not necessarily those of the NHS, the National
813 Institute for Health Research, the Department of Health or Public Health England.

Figure 1

A


118ng input DNA in all samples
Total DNA yield - post heating (ng)
120
100
80
60
40
20
0
0
0.5
1
2
KCl concentration (M)
BCG DNA
BCG+sputum DNA
sputum DNA

B

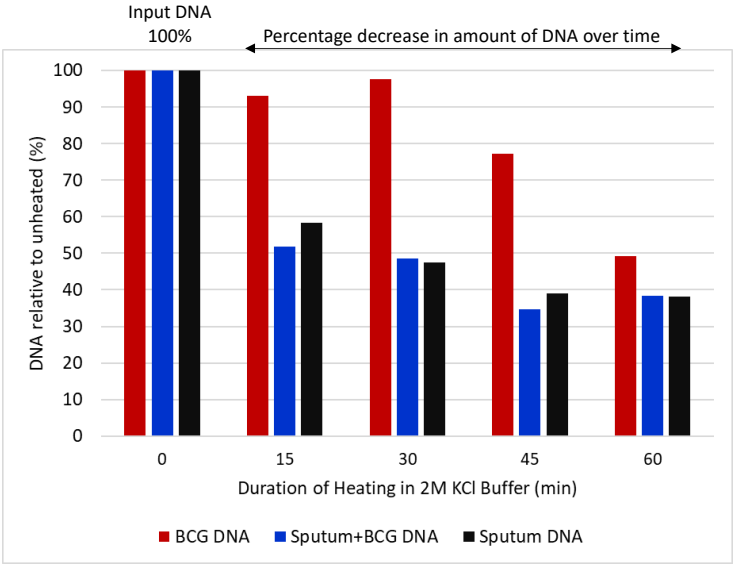
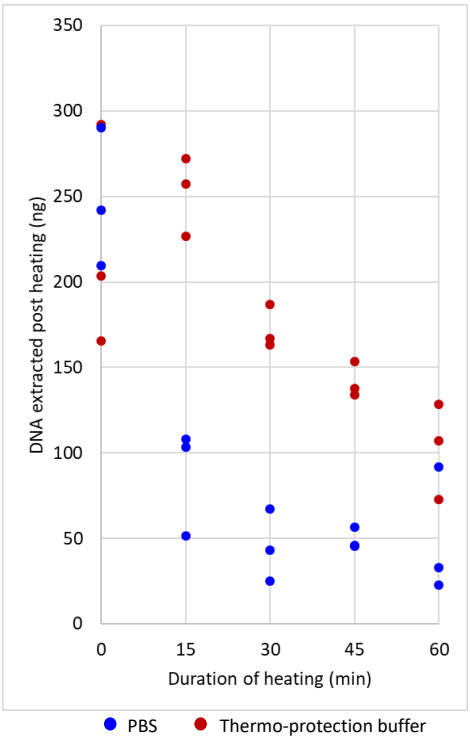

Input DNA
100%
Percentage decrease in amount of DNA over time
DNA relative to unheated (%)
100
90
80
70
60
50
40
30
20
10
0
0
15
30
45
60
Duration of Heating in 2M KCl Buffer (min)
BCG DNA
Sputum+BCG DNA
Sputum DNA

Figure 2

A


350
300
250
200
150
100
50
0
DNA extracted post heating (ng)
0
15
30
45
60
Duration of heating (min)
PBS
Thermo-protection buffer

B

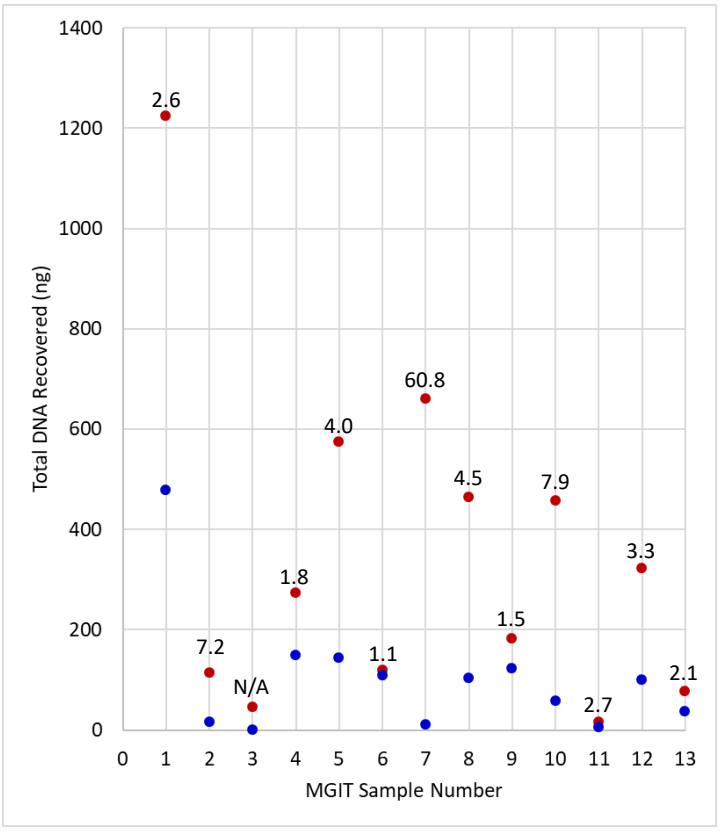

1400
1200
1000
800
600
400
200
0
Total DNA Recovered (ng)
2.6
7.2
N/A
1.8
4.0
1.1
60.8
4.5
1.5
7.9
2.7
3.3
2.1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
MGIT Sample Number
DNA yield after heating in:
MGIT culture fluid
Thermo-protection buffer

Figure 3

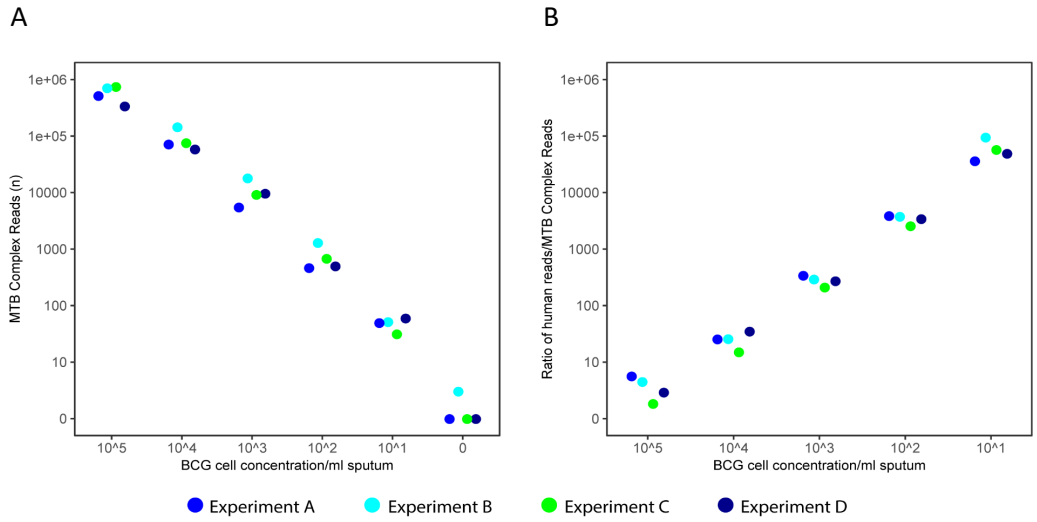

A
B
MTB Complex Reads (n)
Ratio of human reads/MTB Complex Reads
1e+06
1e+05
10000
1000
100
10
0
10^5
10^4
10^3
10^2
10^1
0
BCG cell concentration/ml sputum
Experiment A
Experiment B
Experiment C
Experiment D

Figure 4

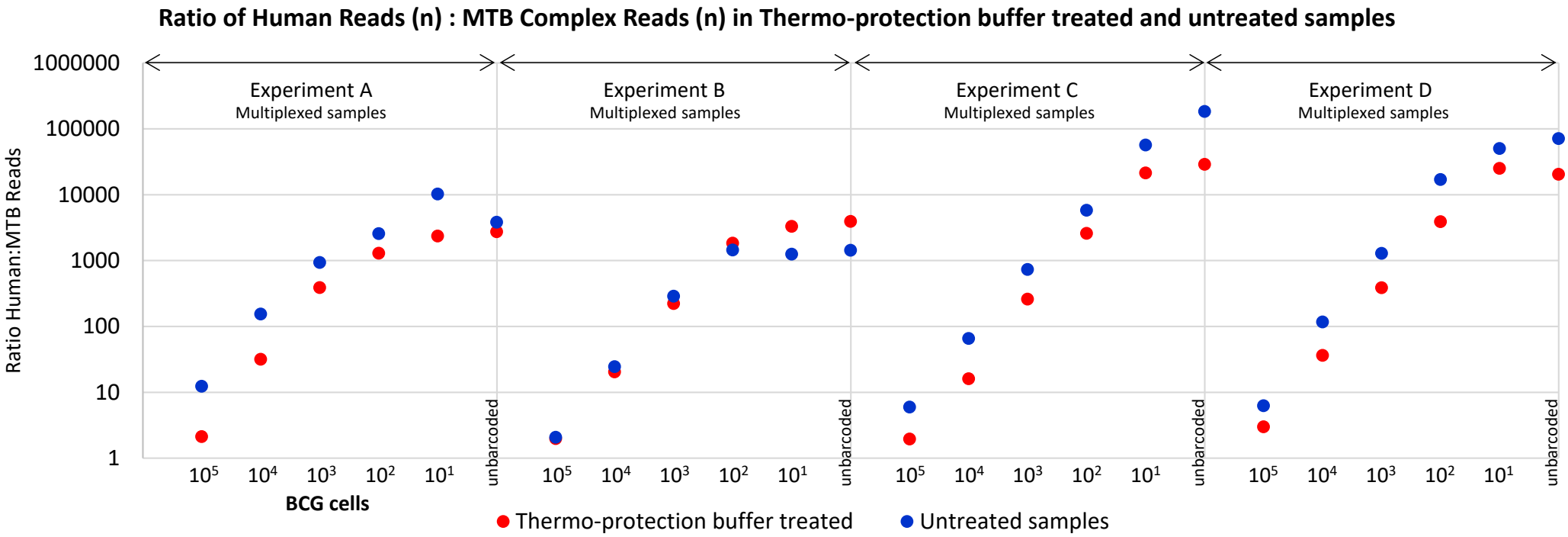

Ratio of Human Reads (n) : MTB Complex Reads (n) in Thermo-protection buffer treated and untreated samples
Ratio Human:MTB Reads
1000000
100000
10000
1000
100
10
1
Experiment A
Multiplexed samples
Experiment B
Multiplexed samples
Experiment C
Multiplexed samples
Experiment D
Multiplexed samples
unbarcoded
10^5 10^4 10^3 10^2 10^1
BCG cells
Thermo-protection buffer treated
Untreated samples

Figure 5

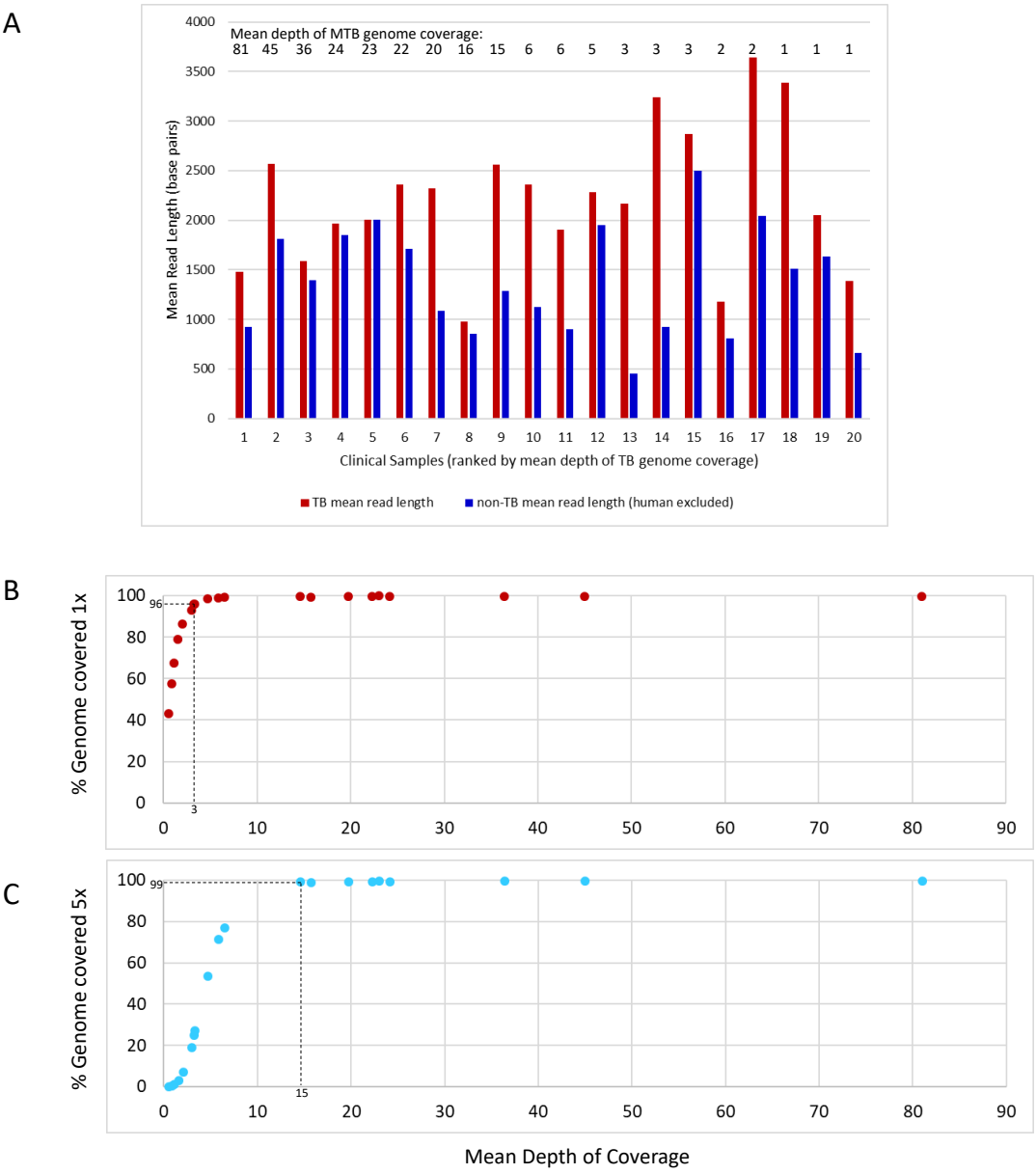
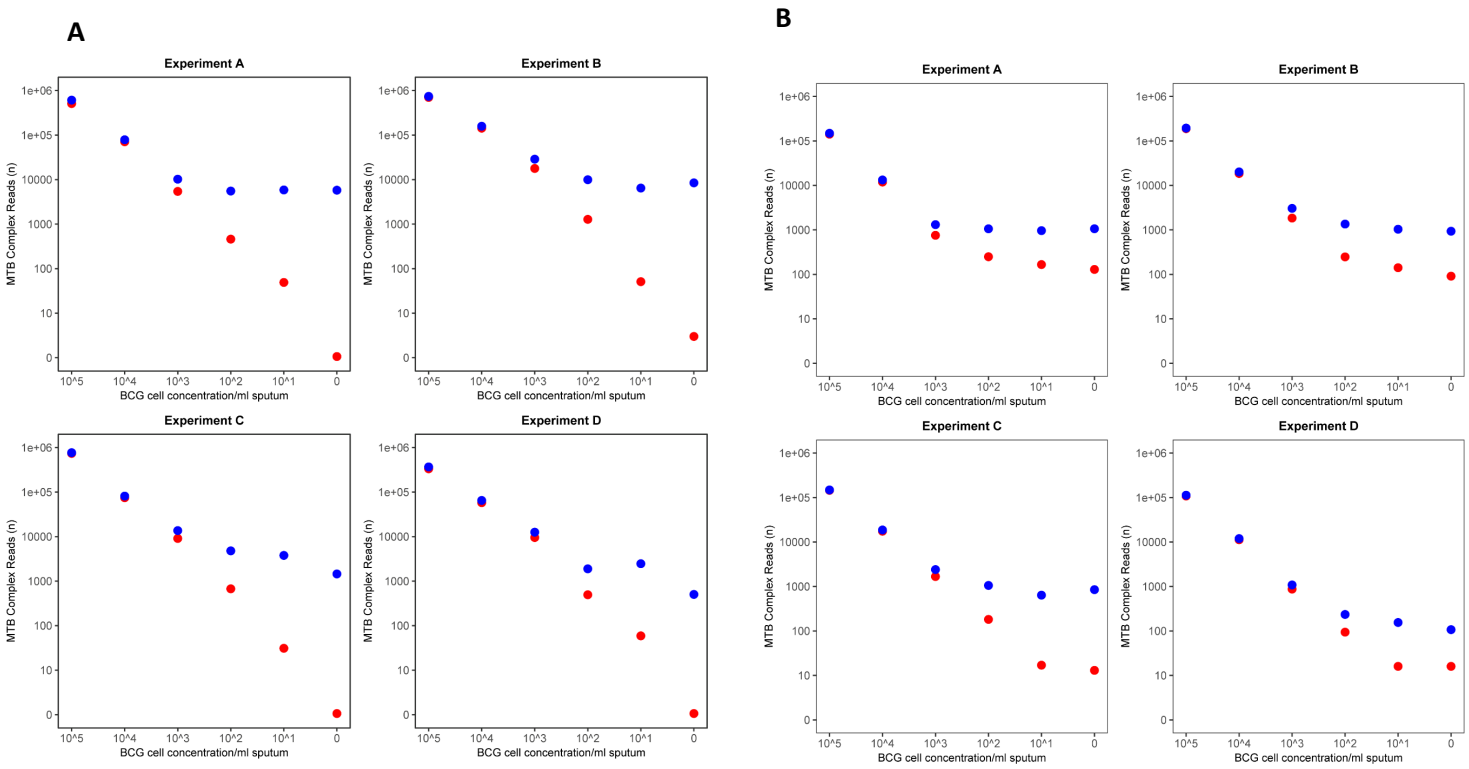

A
Mean depth of MTB genome coverage:
81 45 36 24 23 22 20 16 15 6 6 5 3 3 3 2 2 1 1 1
Mean Read Length (base pairs)
Clinical Samples (ranked by mean depth of TB genome coverage)
TB mean read length
non-TB mean read length (human excluded)
B
% Genome covered 1x
96
3
C
% Genome covered 5x
99
15
Mean Depth of Coverage

Fig. S1



- MTB Complex reads detected before mapping improvements
- MTB Complex reads detected after mapping improvements