

1 **Full title:** A pragmatic pipeline for drug resistance identification in *Mycobacterium tuberculosis* using  
2 whole genome sequencing

3

4 **Short title:** A sequencing pipeline for diagnosing *Mycobacterium tuberculosis*

5

6 **Authors:**

7 Linzy Elton<sup>1</sup>, Alp Aydin<sup>2</sup>, Neil Stoker<sup>1</sup>, Sylvia Rofael<sup>1,3</sup>, Leticia Muraro Wildner<sup>1</sup>, Jabar Babatunde

8 Pacome Agbo Achimi Abdul<sup>4</sup>, John Tembo<sup>5</sup>, Muzamil Abdel Hamid<sup>6</sup>, Mfoutou Mapanguy Claujens

9 Chastel<sup>7,8</sup>, Julio Ortiz Canseco<sup>9</sup>, Ronan Doyle<sup>10</sup>, Giovanni Satta<sup>1,11</sup>, Justin O'Grady<sup>2</sup>, Adam Witney<sup>12</sup>,

10 Francine Ntoumi<sup>7,8</sup>, Alimuddin Zumla<sup>1,13</sup>, Timothy D McHugh<sup>1</sup>

11

12 **Institutional affiliations:**

13 1. Centre for Clinical Microbiology, University College London, UK

14 2. Oxford Nanopore Technologies plc. Quadram Institute, Rosalind Franklin Road, Norwich

15 Research Park, Norwich, UK

16 3. Faculty of Pharmacy, Alexandria University, Egypt

17 4. Centre de Recherches Médicales de Lambaréné, Gabon

18 5. HerpeZ, University Teaching Hospital, Lusaka, Zambia

19 6. Institute for Endemic Diseases, University of Khartoum, Khartoum, Sudan

20 7. Fondation Congolaise pour la Recherche Médicale, Republic of Congo

21 8. Université Marien Ngouabi, Brazzaville Republic of Congo

22 9. Francis Crick Institute, London, UK

23 10. London School of Hygiene and Tropical Medicine, University of London, London, UK

24 11. University College London Hospitals NHS Foundation Trust, London, UK

25 12. St George's, University of London, London, UK

26 13. National Institute for Health Research Biomedical Research Centre, University College

27 London Hospitals NHS Foundation Trust, London, UK

28

29 Corresponding author: Dr Linzy Elton, Centre for Clinical Microbiology, University College London,

30 UK. Email: [linzy.elton@ucl.ac.uk](mailto:linzy.elton@ucl.ac.uk) Tel: +44 (0)20 8016 8189

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52 **Word count:**

53 Abstract: 215

54 Manuscript: 3501

55

56 **Tables:**

57 Table 1. Comparison of basecalling algorithms

58 Table 2. Concordance of different basecalling models, assuming greatest accuracy using SUP

59 basecalling

60 Table 3. Summary comparison of the TB sequencing data analysis software Mykrobe, as compared

61 against TB-Profiler

62 Table 4. Table of pipeline steps and overall recommended choice for each section of the pipeline

63

64 **Figures:**

65 Figure 1. Genomic DNA yield and integrity obtained from MDR-TB and DS-TB isolate cultures

66 Figure 2. Precipitation and spin-column CTAB DNA extraction method comparison

67 Figure 3. DNA shearing to 10 kb and 20 kb compared to 50 kb

68 Figure 4. Mean read depth and genome coverage when comparing each of the three kits

69 Figure 5. Comparison of BAM file read depth of *rpoB*

70 Figure 6. Drug resistance and lineage profile concordance resulting from phenotypic DSTs, ONT and

71 Illumina sequencing

72 Figure 7. The recommended ONT WGS TB pipeline, including culturing conditions, DNA extraction

73 methodology, optimal DNA input library preparation kit, basecalling and data analysis

74

75 **Abstract**

76 **Background:** Delays in accurate diagnosis of drug resistant tuberculosis (DR-TB) can hinder

77 treatment. Whole genome sequencing (WGS) provides more information than standard molecular

78 and phenotypic testing, but commonly used platforms are expensive to implement, and data  
79 interpretation requires significant expertise.

80

81 **Aims:** We aimed to optimise a TB WGS diagnostic pipeline balancing user-friendliness, cost-  
82 effectiveness and time to results, whilst ensuring accuracy.

83

84 **Materials and methods:** Growth conditions, DNA extraction protocols and Oxford Nanopore  
85 Technologies (ONT) library preparation kits were compared. Software for basecalling and analysis  
86 were evaluated to find the most accurate resistance SNP and lineage predictor.

87

88 **Results:** Optimally, a spin-column CTAB DNA extraction method was combined with the RBK110.96  
89 library preparation kit, high accuracy basecalling and data analysis using TB-Profiler. Compared with  
90 Illumina, the pipeline was concordant for 16/17 (94%) isolates (lineage) and for 17/17 (100%)  
91 isolates (resistance SNPs). Our pipeline was 71% (12/17) concordant with phenotypic drug  
92 susceptibility test (DST) results. Time-to-diagnosis was around four weeks.

93

94 **Conclusions:** This optimised TB sequencing pipeline requires less time expertise to run and analyse  
95 than Illumina, takes less time than phenotypic DSTs and the results are comparable with Illumina.  
96 The cost per sample is comparable with other methods. These features make it an important tool for  
97 incorporating into routine DR-TB diagnostic pipelines and larger scale drug resistance surveillance in  
98 all settings.

99

100 **Keywords:** tuberculosis, drug resistance, whole genome sequencing, Oxford Nanopore, MinION,  
101 diagnosis

102

103 **Abbreviations**

CTAB	Cethyl trimethyl ammonium bromide
CLI	Command line interface
CPU	Central processing unit
DR-TB	Drug resistant tuberculosis
DS-TB	Drug sensitive tuberculosis
DST	Drug sensitivity test
GPU	Graphics processing unit
GUI	Graphical user interface
HAC	High accuracy (basecalling)
LMIC	Low- and middle-income country
MDR	Multi-drug resistant
ONT	Oxford Nanopore Technologies
QC	Quality control
RoC	Republic of Congo
RR	Rifampicin-resistant
SD	Standard deviation
SUP	Super-high accuracy (basecalling)
TB	Tuberculosis
WGS	Whole genome sequencing
XDR	Extensively-drug resistant

104

## 105 Introduction

106

107 Treatment for drug-resistant tuberculosis (DR-TB) is intensive and can take up to 20 months to  
108 complete, compliance can be low due to drug toxicity, and there is increased risk of mortality (1).

109 Misdiagnosis and delays to correct diagnosis, both between drug sensitive (DS-TB), DR-TB and multi-  
110 drug resistant (MDR-TB) strains, and the differentiation between *M. tuberculosis* and non-  
111 tuberculosis mycobacteria (NTMs), can complicate and delay appropriate treatment. Delayed  
112 treatment is associated with high early mortality and the increase in drug resistance prevalence (2).

113 Whilst the prevalence of DR-TB has remained stable over the last decade, the top 30 countries with  
114 the highest TB prevalence identified by WHO are low- and middle-income countries (LMICs) (3). DR-  
115 TB has been most commonly reported from Southern America and Central Asia, although data is not  
116 available from many African countries (4,5). Phenotypic drug sensitivity testing (DST) tends to be  
117 restricted to reference facilities or specialist centres, but the roll out of molecular assays such as the  
118 GeneXpert MTB/RIF Ultra and MTB/XDR (Cepheid) or the line probe assay GenoType® MTBDR

119 (Bruker Corporation) has allowed more localised testing. However, these are incomprehensive  
120 targeted methods which are limited to detecting specific resistance-encoding mutations to certain  
121 drugs (6,7).

122

123 Whole genome sequencing (WGS) has emerged as a diagnostic tool for DR-TB cases (8–10). WGS is a  
124 comprehensive molecular diagnostic test that can identify mutations for all drugs in one assay  
125 including novel drug resistance mutations and novel drug targets. It can predict resistance through  
126 gene inactivation caused by multiple mutations, frameshift mutations or large structural variations in  
127 the genome [9]. WGS can also identify TB lineages and mixed strain infections (11). In outbreak  
128 situations, WGS is a powerful tool to identify transmission chains and clusters (12). In cases of  
129 negative treatment outcomes, WGS can differentiate between cases of relapse, due to treatment  
130 failure, and reinfection, which might be due to failure to develop protective immunity, through  
131 calculating the genetic distance between pre- and post-treatment *M. tuberculosis* genomes (13).  
132 Despite the increasing availability of sequencing platforms in LMICs, the lack of bioinformaticians,  
133 data analysis training opportunities, and high-performance computing facilities have been a  
134 bottleneck to the roll out of WGS globally (14).

135

136 Despite the potential of WGS, *M. tuberculosis* is a challenging organism to work with. DNA extraction  
137 is complicated by the complex cell wall structure (15). Many laboratories utilise the cethyl trimethyl  
138 ammonium bromide (CTAB) method (16), which generally produces a higher yield of genomic DNA  
139 of good integrity and purity for sequencing applications compared to commercial kits, and the  
140 reagents for which are mostly available and commonly used in TB laboratories in LMICs. DR-TB  
141 isolates tend to be slower to grow than DS-TB ones and generally produce less biomass, which can  
142 make extracting sufficient amounts of genomic DNA more difficult (17).

143

144 The mycobacterial genome has a high GC content (>60%) and multiple repeat regions. These  
145 structural features create challenges when considering sequencing library preparation and data  
146 analysis (18,19). On the other hand, an advantage in the analysis of the DR genotype for *M.*  
147 *tuberculosis* is that mycobacterial resistance is mainly dependent on the evolution of mutations such  
148 as single nucleotide polymorphisms (SNPs) and insertions or deletions (INDELs) within chromosomal  
149 genes rather than horizontal gene transmission (20).

150

151 The employment of short read, high throughput sequencing platforms, such as the Illumina MiSeq, is  
152 established for DR-TB genotyping (21–23), but these platforms have high resource requirements and  
153 highly skilled operators, and in consequence such large scale sequencing is not always available,  
154 especially in resource-constrained settings. In contrast, long read sequencing offered by ONT using  
155 the MinION; a small, portable device, is accessible to resource-constrained environments because of  
156 lower set-up costs and less complex training needs. Long read sequencing can additionally be  
157 advantageous for sequencing complex genomes which are GC rich or have many repeated regions.

158

159 There is little research from resource-constrained settings on the use of ONT platforms for TB  
160 diagnosis. Those publishing have often done so as individual research studies (24–29). Previous  
161 studies often made use of the ligation-based ONT library preparation (25,28). These require costly  
162 third-party reagents and high DNA input (1 µg), which can be hard to routinely obtain. If optimised  
163 appropriately, use of the rapid barcoding kits, employing multiplexing, should improve cost  
164 effectiveness (24).

165

166 Whilst phenotypic DSTs are the benchmark and used as a standard as there are international  
167 standards, diagnosis utilising WGS may help to provide more accurate results and therefore a faster  
168 correct drug regimen. Genomics is especially important in countries with no DST capabilities, or for  
169 new drugs like bedaquiline, where there is no other diagnostic. Currently, there does not appear to

170 be a TB sequencing protocol that could be integrated into diagnostic pipelines in resource  
171 constrained settings to aid DR-TB diagnosis. We aimed to develop a pipeline, optimising  
172 methodology to create a tool that had an optimal combination of accuracy, speed, cost-effectiveness  
173 and user-friendliness.

174

## 175 [Materials and methods](#)

176

177 The full manual for this pipeline, including protocols and tutorial videos can be found on the  
178 PANDORA-ID-NET Global Health Network hub (30).

179

### 180 [M. tuberculosis samples and ethics statement](#)

181 17 *M. tuberculosis* isolates from an archive maintained by the Centre for Clinical Microbiology (CCM)  
182 at University College London (UCL), which have metadata and phenotypic drug sensitivity data  
183 (streptomycin, isoniazid, rifampicin, ethambutol plus pyrazinamide) were utilised. Eight DS-TB  
184 isolates, five isoniazid mono-resistant isolates and four MDR-TB isolates were selected. The study did  
185 not include live participants. The study used archived *M. tuberculosis* isolates and no patient data  
186 was included in the study. There is no means to associate the isolates with the original source and  
187 ethical approval was not required.

188

### 189 [Growth of M. tuberculosis](#)

190 *M. tuberculosis* isolates were incubated in MGIT tubes (Becton Dickson), using the BD BACTEC™  
191 MGIT™ automated mycobacterial detection system or on Middlebrook 7H11 slopes (EO laboratories)  
192 and incubated at 37°C in a static incubator. To identify the optimal growth time to achieve the  
193 greatest DNA yield, two representative MDR-TB and two DS-TB samples were selected from the  
194 collection. 500 µL of MGIT subcultures were seeded onto 7H11 slopes and into MGIT tubes and left



195 to grow. Once the MGIT flagged positive, each isolate was further cultured for 2, 3 or 4 weeks, after  
196 which DNA was extracted.

197

## 198 DNA extraction

199 Two different CTAB methods were compared. For each extraction method, 400  $\mu$ L of liquid culture  
200 taken from a MGIT tube was used.

201

202 Method 1: The precipitation CTAB DNA extraction method, as described in Kent *et al.* (16) was  
203 optimised, with the addition of a 15-minute sonication step after heat killing and 3  $\mu$ L yeast tRNA  
204 (Fisher Scientific) at the isopropanol precipitation step. DNA was eluted into 25  $\mu$ L molecular grade  
205 water.

206

207 Method 2: A spin-column CTAB method, developed by ONT, was compared. Briefly, samples were  
208 heat killed, then glass beads and lysozyme (10mg/mL) added. The samples were incubated in a  
209 shaking incubator at 37°C for 15 min at 800 RPM. Proteinase K and CTAB buffer were added and  
210 then incubated at 56°C for 30 min at 1000 RPM. 5M NaCl was added and the debris pelleted by  
211 centrifugation, before a modified version of the Monarch Genomic Purification Kit (New England  
212 Biosciences) was followed. The full protocol for spin-column CTAB is in supplementary materials S1.

213

## 214 DNA quantification

215 DNA quantified using spectrophotometry and the Qubit™ dsDNA BR Assay Kit (Thermo Fisher) and  
216 DNA size and integrity confirmed using electrophoresis with the Genomic DNA ScreenTape and  
217 reagents on the TapeStation 4150 (Agilent Technologies Inc.).

218

## 219 DNA shearing

220 DNA was sheared mechanically using g-TUBES (Covaris), following manufacturer's instructions (31).  
221 Samples with an average size of 10 kb, 20 kb and ~50 kb were sequenced for 2 hours each and the  
222 output number of reads, N50 (the sequence length of the shortest contig at 50% coverage of the  
223 total genome assembly) and bases compared.

224

## 225 Library preparation

226 DNA libraries were prepared using either the Rapid Barcoding Kit (SQK-RBK004) (ONT) with a DNA  
227 input of ~400 ng, the newer Rapid Barcoding Kit 96 (SQK-RBK110.96) (ONT) with a DNA input of  
228 between 50-200 ng, or the PCR Barcoding kit (SQK-RPB004) (ONT) with DNA inputs of 1-5 ng,  
229 following the respective ONT kit protocols (32–34). DNA libraries were prepared from the same spin-  
230 column DNA extractions for Illumina sequencing using the DNA Prep, (M) Tagmentation (24 Samples,  
231 IPB) library preparation kit (Illumina) and the MiSeq Reagent Kit v3 (600-cycle) (Illumina) (35,36).

232

## 233 Sequencing and basecalling

234 Twelve or 24 barcoded samples were run together on a flow cell version R9.4.1 (ONT) using a  
235 MinION device for 72 hours, using the default parameters on the MinKNOW software (v23.04.6).  
236 Basecalling was performed either by the MinKNOW software alongside sequencing or using the  
237 Guppy basecalling software (v6.5.7) (ONT) (37), using either the flip-flop Fast, high accuracy (HAC),  
238 or super-high accuracy (SUP) algorithm. Basecalling was undertaken on a 32 GB RAM CPU computer  
239 (for specifications, see Supplementary materials S2).

240

## 241 Data analysis

242 Fastq files were quality checked (QC) using FastQC (v0.21.1) and MultiQC (v1.15) (38,39). Barcodes  
243 were demultiplexed and trimmed from the reads using Guppy. To view depth and coverage of the

244 sequences, fastq files were aligned to the H37Rv reference genome (NCBI Reference Sequence  
245 NC\_000962.3) using minimap2 (v2.26) (40), then sorted and indexed using SAMtools (v1.17) (41).  
246 Alignments were visualised using Artemis (v18.1.0) (42). Depth and coverage were calculated using  
247 SAMtools. Fastq files were input to the resistance gene SNP and lineage calling software configured  
248 for ONT data, Mykrobe (v0.10.0) (43) and TB-Profiler (v4.4.0) (44) graphical user interfaces (GUIs) to  
249 compare accuracy. ONT results were compared to Illumina data obtained from the same extraction.  
250 Drug resistance SNPs with a confidence grading of  $\geq 3$ , as defined by the WHO were included in the  
251 comparison (45). Data were analysed using Prism (v10.0.0.153) (GraphPad). Sequence data were  
252 deposited under Project accession number PRJEB68143 on the European Nucleotide Archive and  
253 outlined in supplementary materials S3.

254

## 255 Role of the funding source

256 The study sponsors had no role in the design, data collection, analysis or write up of this study. No  
257 funding (real or in kind) was received from ONT.

258

## 259 Results

260

261 DNA yield and fragment size for cultures sampled over time are displayed in Figure 1. The mean DNA  
262 yield for the MDR-TB isolates was 688 ng (SD=413) (from MGIT liquid culture) and 966 ng (SD=841)  
263 (from 7H11 slopes), whilst for the DS-TB isolates it was 1414 ng (SD=1265) (MGIT) and 1712 ng  
264 (SD=1327) (7H11). The quantity of DNA obtained from 7H11 was significantly higher than from  
265 MGITs especially for the MDR-TB isolates ( $p < 0.0001$ , unpaired t-test). For MDR-TB isolates, both  
266 the MGIT and 7H11 yields did not change significantly over time. For DS-TB isolates, for both the  
267 MGIT and 7H11 cultures, the yields decreased over time.

268

269 Figure 1. Genomic DNA yield and integrity obtained from MDR-TB and DS-TB isolate cultures.  
270 Overall, 7H11 slopes produced the greatest yield of DNA, whilst MGIT tubes produced the greatest  
271 DNA size. A) DNA yields extracted from MGIT tubes over time; there was no change in MDR-TB DNA  
272 yield, but the DS-TB isolate yield decreased. B) DNA amounts extracted from 7H11 slopes over time;  
273 there was no change in MDR-TB DNA yields, but DS-TB isolate yields decreased. C) Size of DNA  
274 extracted from MGIT tubes over time; both MDR-TB and DS-TB isolates increased over time. D) Size  
275 of DNA extracted from 7H11 slopes over time; the MDR-TB isolates peaked at week 3 and then  
276 decreased ( $p=0.0025$ ), the DS-TB isolates peaked at weeks 3 and 4 ( $p=0.01$ ). Significance was  
277 calculated using One-Way ANOVA and Tukey's post ad hoc analysis.

278  
279 The largest DNA fragments for both MDR-TB and DS-TB isolates was obtained from MGIT cultures.  
280 The mean size for the MDR-TB isolates was 56,150 bp (Standard Deviation (SD)=6171) (MGIT) and  
281 30,439 bp (SD=13,890) (7H11), whilst for the DS-TB isolates it was 54,776 bp (SD=5028) (MGIT) and  
282 18,618 bp (SD=7134) (7H11). There were no significant differences between the means of DNA sizes  
283 for isolates grown in MGITs or on 7H11 for either MDR-TB or DS-TB isolates. For the MDR-TB  
284 isolates, when cultured in MGIT, the DNA size increased over time, whereas on 7H11, the DNA size  
285 peaked at week 3 ( $p=0.01$ , one-way ANOVA). For the DS-TB isolates, the increased over time in both  
286 the MGIT and 7H11 cultures.

287  
288 DNA yield and DNA size results from the precipitation and spin-column CTAB extraction method  
289 were compared for MGIT and 7H11 cultures and stratified into resistance categories (Figure 2).

290  
291 Figure 2. Precipitation and spin-column CTAB DNA extraction method comparison. A) DNA yield  
292 measured using Qubit dsDNA broad spectrum kit, B) Size (bp) measured using the TapeStation  
293 genomic DNA kit. Apart from spin-column CTAB from MGIT tubes, all other methods provided the  
294 minimum required yield, although precipitation based CTAB protocols provided higher yields. Spin-

295 column CTAB from MGIT media also provided the smallest size. Significance was calculated using  
296 One-Way ANOVA and Tukey's post ad hoc analysis.

297

298 We carried out MinION-based sequencing of selected samples. Firstly we looked at fragment size, as  
299 in our experience, sequencing coverage can be reduced both by fragments that are too short and  
300 too long. Long fragment lengths can especially inhibit read output for the mycobacteria. To do this  
301 we generated shorter DNA fragments (20 and 10 kb average size) which we compared with the  
302 original (50 kb) samples. A size of 20 kb resulted in the greatest number of reads, bases and N50  
303 (Figure 3). The obtained N50 in sequencing was significantly higher with the 20 kb input DNA  
304 fragments compared to the other two sizes ( $p=0.01$ , one-way ANOVA). Although there was variation  
305 in the number of reads and bases obtained between the DNA sizes tested, the differences were not  
306 significant ( $p>0.05$ , one-way ANOVA).

307

308 Figure 3. DNA shearing to 10 kb and 20 kb compared to 50 kb. A) number of reads (no significant  
309 difference) B) N50 (overall  $p=0.01$ ) and C) total number of bases (no significant difference). One-Way  
310 ANOVA and Tukey's multiple comparison test was applied.

311

312 We compared different library preparation kits, all three of which are rapid, transposase-based kits.  
313 SQK-RPB004 used a PCR amplification step resulting in fragments of around 4000 bp, the other two  
314 were different versions of kits that utilise a transposase that cuts native DNA to add the sequencing  
315 adapter, and has a major impact on fragment length. When 12 randomly selected isolates were  
316 barcoded and multiplexed and sequenced on one flow cell, the PCR-based SQK-RPB004 kit gave a  
317 greater mean read depth (mean=394x, SD=762) than the native kits, but the genome coverage was  
318 low (mean=67%, SD=19). Whilst the coverage was improved using the older native DNA-based kit  
319 (SQK-RBK004) (mean=80%, SD=38), the read depth was lower (mean=4x, SD=3). The newer native  
320 DNA-based kit (SQK-RBK110.96) provided an improved depth for both precipitation DNA extraction

321 method (mean=13x, SD=9) and spin-column CTAB method (mean=49x, SD=36), and improved  
322 genome coverage (mean=84.6%, SD=31 and 99.8%, SD=2 respectively). Figure 4 highlights the  
323 quality scores for each of the kits. When 24 samples were run together, the mean depth was  
324 reduced to 18x (SD=25).

325

326 Figure 4. A) Mean read depth and B) genome coverage when comparing each of the three kits.  
327 Statistical significance calculated using One-Way ANOVA and Tukey's multiple comparisons test.  
328 pCTAB = precipitation CTAB, cCTAB = spin column CTAB. Numbers in brackets denote the number of  
329 samples multiplexed per run.

330

331 When the assembled BAM files were viewed for *rpoB*, where most rifampicin-resistance SNPS map  
332 (as an example), the PCR-based SQK-RPB004 kit showed isolated, highly amplified regions and other  
333 regions with no coverage (see Figure 5). Native library preparation kits demonstrated a better  
334 coverage for the whole *rpoB* gene.

335

336 Figure 5. Comparison of BAM file read depth of *rpoB* for A) the PCR-based SQK-RPB004 kit, showing  
337 low coverage, B) SQK-RBK004 kit showing greater coverage and C) SQK-RBK110.96 resulted in the  
338 best coverage and depth, using Artemis.

339

340 Here we compared three different bioinformatic tools Fast basecalling was compared with high  
341 accuracy (HAC) and super high accuracy (SUP) basecalling. The mean number of .fast5 input files was  
342 6.4 (SD=1.5) (Table 1). Whilst SUP basecalling was the slowest, it did provide significantly better  
343 quality (Q) scores.

344

	Fast	HAC	SUP	One-way ANOVA significance
Mean time (minutes) (SD)	15 (5.9)	205 (65.4)	>36 hours	$p < 0.0001$

Mean Q score (SD)	17.1 (0.2)	19.7 (0.3)	20.8 (0.3)	$p < 0.0001$
Mean accuracy (SD)	98.02% (0.06)	98.94% (0.05)	99.13% (0.05)	$p < 0.0001$

345 Table 1. Comparison of basecalling algorithms. One-way ANOVA was used to test statistical  
 346 significance. When Tukey’s multiple comparisons test was applied, there was a significant difference  
 347 of  $p < 0.0001$  between each of the basecalling algorithms for time, mean Q score and accuracy.

348

349 The DNA samples came from strains with known DR phenotypes and genotypes. Fast, HAC and SUP-  
 350 basecalled versions of each of the files were analysed using TB-Profiler to compare resistance SNP  
 351 and lineage identification (Table 2). All isolate’s median read depths were  $>40x$ . Compared with SUP  
 352 basecalling, Fast basecalling agreed for 15/17 (88%) samples for lineage and all resistance SNPs. The  
 353 sample for which Fast basecalling missed a SNP (streptomycin *gidC* 102de1G) had a depth of 47x.  
 354 The sample for which Fast basecalling called lineage 3, and HAC and SUP called lineage 4.3, had a  
 355 depth of 85x. HAC and SUP basecalling agreed for 17/17 (100%) samples. For the complete  
 356 basecalling comparison dataset, see Supplementary materials S4.

357

	Average read depth	Fast vs SUP			HAC vs SUP		
		Lineage	Drug resistance	Total concordance	Lineage	Drug resistance	Total concordance
DS-TB	90	7/8 (88%)	8/8 (100%)	7/8 (88%)	8/8 (100%)	8/8 (100%)	8/8 (100%)
Isoniazid mono	111	5/5 (100%)	4/5 (80%)	4/5 (80%)	5/5 (100%)	5/5 (100%)	5/5 (100%)
MDR-TB	72	4/4 (100%)	4/4 (100%)	4/4 (100%)	4/4 (100%)	4/4 (100%)	4/4 (100%)
Total isolates	85	16/17 (94%)	16/17 (94%)	15/17 (88%)	17/17 (100%)	17/17 (100%)	17/17 (100%)

358 Table 2. Concordance of different basecalling models, assuming greatest accuracy using SUP  
 359 basecalling. HAC and SUP were 100% concordant, Fast basecalling called one lineage differently and  
 360 missed one drug resistance SNP, compared to HAC and SUP basecalling. Isoniazid mono-resistance is  
 361 defined by phenotypic DSTs.

362

363 Using phylogenetic DSTs as a guide, HAC-basecalled data were put through TB-Profler and Mykrobe.  
364 Mykrobe called a mixed population of lineage 3 and 4.6.2.2 for the sample where Fast, HAC and SUP  
365 disagreed in TB-Profler (Table 3). The complete dataset can be found in Supplementary materials S5.  
366

<b>Mykrobe vs TB-Profler</b>	<b>Lineage</b>	<b>Drug resistance</b>	<b>Total concordance</b>
DS-TB	7/8 (88%)	7/8 (88%)	7/8 (88%)
Isoniazid mono-resistance	5/5 (100%)	5/5 (100%)	5/5 (100%)
MDR-TB	4/4 (100%)	1/4 (25%)	1/4 (25%)
Total isolates	16/17 (94%)	13/17 (76%)	13/17 (76%)

367 Table 3. Summary comparison of the TB sequencing data analysis software Mykrobe, as compared  
368 against TB-Profler.

369

370 HAC-basecalled consensus sequence fastq files were compared with the equivalent Illumina fastq  
371 files for each isolate, to identify lineage and resistance SNP concordance using TB-Profler. ONT data  
372 was concordant with phenotypic DSTs in 12/17 (71%) samples, taking into account phenotypic DST  
373 testing only included rifampicin, isoniazid, ethambutol and streptomycin, and pyrazinamide for MDR  
374 samples. ONT picked up SNPs in one DS-TB sample, one isoniazid mono-resistant samples two MDR-  
375 TB samples which phenotypic DSTs concluded were sensitive. Both ONT and Illumina showed two  
376 phenotypically pyrazinamide-resistant isolates as sensitive. The full list of phenotypic and genotypic  
377 drug resistances found, including full SNP calls are in Supplementary materials S6.

378

379 The ONT pipeline was overall 94% (16/17) concordant (16/17 for lineage and 17/17 for all resistance  
380 SNPs) with Illumina data. For lineage, one DS-TB sample appeared to be a mixed population of  
381 lineage 3 (the only one picked up by Illumina) and 4.6 (picked up by HAC and SUP-basecalling in the  
382 ONT pipeline, but not Fast). An overview of these results is in Figure 6.

383

384 Figure 6. Drug resistance and lineage profile concordance resulting from phenotypic DSTs, ONT and  
385 Illumina sequencing. P = phenotypic DST results, O = Oxford Nanopore results, I = Illumina results, M  
386 = mixed (lineage 3 and 4), S = sensitive, R = resistant, NT = not tested.



387

388 Table 4 outlines each component analysed, which tested component performed the best for each

389 category, and the overall recommended pipeline options.

390

Step	Culturing (media)	Culturing (length of time)	DNA extraction	Library preparation	Basecalling algorithm	Analysis software
<b>Reasoning for testing</b>	Cost, DNA quality and quantity	Optimal DNA yield	Optimisation of DNA quality and quantity (reduce inhibitors)	Input DNA vs output data	Accuracy vs speed (dependent on computing power)	Accuracy and detail vs ease of use
<b>Comparators</b>	MGIT, 7H11	2 weeks, 3 weeks and 4 weeks post-MGIT flagging	Precipitation CTAB, Spin-column CTAB	RPB004 (PCR), RBK004 (native), RBK110.96 (native)	Fast, HAC and SUP	TB-Profiler, Mykrobe
<b>Cheapest</b>	7H11	n/a	Precipitation CTAB	RBK110.96	n/a	n/a
<b>Fastest</b>	n/a	2 weeks post-MGIT flagging	Spin-column CTAB	RBK004	Fast	Computing power/server speed dependent
<b>Most accurate</b>	n/a	2 or 3 weeks post-MGIT flagging	n/a	RBK110.96	HAC or SUP	TB-Profiler
<b>Most user-friendly</b>	7H11 (if bought ready made)	n/a	Both protocols are similar	RBK004	n/a	GUIs are user friendly
<b>Recommended option</b>	7H11	2 weeks post-MGIT flagging	Spin-column CTAB	RBK110.96	HAC	TB-Profiler GUI

391 Table 4. Table of pipeline steps and overall recommended choice for each section of the pipeline.

392

## 393 Discussion

394

395 We conclude that a WGS pipeline using ONT technology is feasible for the identification of SNPs that

396 determine *M. tuberculosis* drug resistance and lineage and is suitable for our aim to create a

397 pragmatic pipeline for use in LMICs. Taking into consideration the data presented here, we propose

398 the following pipeline (Figure 7) as the optimized route for TB WGS.

399

400 Figure 7. The recommended ONT WGS TB pipeline, including culturing conditions, DNA extraction

401 methodology, optimal DNA input library preparation kit, basecalling and data analysis. Images taken

402 from Windows 10 Education and Shutterstock.

403

404 In our study, the ONT pipeline showed the same accuracy as Illumina but showed less concordance  
405 with phenotypic DSTs. This is likely due to the complex relationship between genotypic and  
406 phenotypic resistance. All phenotypic and genotypic discordances in this study were seen in three  
407 drugs; streptomycin, ethambutol and pyrazinamide, each of which are known to have multifaceted  
408 resistance mechanisms, which are not fully understood, thus further work is needed (46).

409

410 We identified isolates with phenotypic sensitivity and genotypic resistance, which may be a result of  
411 subpopulations and heteroresistance, which is a growing concern. When phenotypic resistance and  
412 genotypic sensitivity was seen, this may have been a result of the specific resistance SNPs causing  
413 genotypic resistance not being included in the analysis software databases. Ultimately, the standard  
414 will be which drugs are clinically relevant, and whilst phenotypic DSTs are the current benchmark,  
415 with the accumulation of genotypic resistance knowledge, this may shift to a molecular benchmark  
416 in the future.

417

418 Whilst a reduced fragment size did improve sequencing read yield, the Covaris g-TUBEs added a  
419 significant cost to the pipeline. We found that extracting from 7H11 plates in conjunction with spin-  
420 column CTAB DNA extraction, produced a fragment length of <30 kb. Thus, we recommend that g-  
421 TUBEs could be used as an optional extra within the pipeline, if different media and extraction  
422 methodologies are utilised.

423

424 Whilst the precipitation CTAB method does provide adequate DNA size and yield for use with the  
425 new native ONT library preparation kit (SQK-RBK110.96), lower numbers of reads and bases were  
426 generally achieved compared with spin-column CTAB extraction, although the latter produces lower  
427 DNA yields. We hypothesise that the co-precipitation of inhibitors, possibly from the mycobacterial

428 cell wall, may be interfering with the binding of the transposase complex used in native library  
429 preparation kits, so that less DNA is adapted and therefore is unable to be sequenced.

430

431 The cost of WGS has decreased greatly in the last few years, making it more accessible for resource  
432 constrained settings. Whilst the cost per sample may be lower using Illumina due to the higher  
433 throughput, if the samples have to be sent away due to a lack of local facilities, extra costs are likely  
434 to be incurred. If batching is used to reduce costs, then this can increase the time to results in  
435 smaller laboratories with fewer samples. Costs in this pipeline could be reduced by substituting  
436 products or equipment, although automated systems and high-speed computers may be beneficial  
437 in the long term.

438

439 The bottleneck of this pipeline is likely to be basecalling, unless high-computing power set ups are  
440 available. The time from culture to drug resistance and lineage results could be within four weeks if  
441 the appropriate human resources and computers are available, with 2-3 weeks of culturing and a  
442 week for extraction, quantification, sequencing and analysis. Whilst this is not as fast as other  
443 methods such as GeneXpert or line probe assays which are commonly available in local regional  
444 laboratories, it is comparable with phenotypic DSTs. In terms of the ease-of-use of this pipeline, if  
445 users opt to use Guppy within the MinKNOW GUI, minimal CLI knowledge is required, meaning less  
446 training would be required for routine laboratory staff, and should therefore be faster to implement  
447 and more widely usable.

448

449 Some complications were encountered in this study, and in sequencing of the mycobacteria in  
450 general. This pipeline requires cultured isolates, and thus the use of a biosafety level 3 (BSL3)  
451 laboratory, which are not widely available. The possibility of utilising another site's BSL3 facilities to  
452 grow cultures, heat killing and then transferring samples to the sequencing laboratory is an  
453 alternative route but may increase the time to results.

454

455 With a greater understanding of genomic drug resistances, plus ever-evolving bioinformatics  
456 software, the accuracy TB diagnostics will improve. To increase the speed of time to results,  
457 sequencing direct from sputum would cut out culturing time, but is yet to be standard practice in  
458 routine diagnostics. Targeted kits, such as the Illumina and GenoScreen Deeplex® Myc-TB Combo Kit  
459 are an alternative to WGS but provide less information (47). An ONT targeted kit for DR-TB has  
460 recently been approved by the WHO (48).

461

462 Whilst the necessity of culture in this pipeline does mean time to results is in weeks not minutes or  
463 days, the resolution of resistance and lineage detail obtained from ONT WGS sequencing is greater  
464 than that obtained from standard diagnostic tests such as GeneXpert, line probe assays or  
465 phenotypic DSTs, and is comparable to Illumina sequencing. There is a need for genotypic drug  
466 resistance data, not only for routine diagnostics, but also for geographical drug resistance profiling  
467 and surveillance to aid correct prescribing, and it is important that data is collected as broadly as  
468 possible (9). The ability of smaller, localised laboratories to participate in this is vital and the  
469 optimised ONT pipeline presented here will help to facilitate WGS roll out within these sites.

470

### 471 [Conflict of interest](#)

472 This project obtained technical support from Oxford Nanopore Technologies Ltd (ONT), but ONT did  
473 not contribute financially to the study.

474

### 475 [Data sharing statement](#)

476 All genetic fasta data obtained in this study is available under the Project accession number  
477 PRJEB68143 on the European Nucleotide Archive upon publication. The full manual, including  
478 further information and tutorial videos, relating to this protocol is available on the PANDORA-ID-NET

479 Global Health Network hub pages: [https://pandora.tghn.org/sequencing/tuberculosis-sequencing-](https://pandora.tghn.org/sequencing/tuberculosis-sequencing-pipeline/)  
480 [pipeline/](https://pandora.tghn.org/sequencing/tuberculosis-sequencing-pipeline/). All data and additional material is Open Access.

481

## 482 References

- 483 1. Thwaites G, Nguyen N V. Linezolid for Drug-Resistant Tuberculosis. *New England Journal of*  
484 *Medicine*. 2022 Sep 1;387(9):842–3.
- 485 2. Adamu AL, Gadanya MA, Abubakar IS, Jibo AM, Bello MM, Gajida AU, et al. High mortality  
486 among tuberculosis patients on treatment in Nigeria: A retrospective cohort study. *BMC*  
487 *Infect Dis*. 2017 Feb 23;17(1).
- 488 3. World Health Organization. *Global Tuberculosis Report*. 2019.
- 489 4. World Health Organization. *Global Tuberculosis Report*. 2016.
- 490 5. World Health Organization. *Global Tuberculosis report*. 2021.
- 491 6. Hain Lifescience GmbH. GenoType MTBDRsl VER 1.0 and VER 2.0 - Your Important Assistance  
492 for Detetion of XDR-TB [Internet]. [cited 2022 Sep 26]. Available from: [https://www.hain-](https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdsl.html)  
493 [lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdsl.html](https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdsl.html)
- 494 7. Hain Lifescience GmbH. GenoType MTBDRplus VER 2.0 - Your Test System for a Fast and  
495 Reliable Way to detect MDR-TB [Internet]. [cited 2022 Sep 26]. Available from:  
496 [https://www.hain-](https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdplus.html)  
497 [lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-](https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdplus.html)  
498 [mtbdplus.html](https://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdplus.html)
- 499 8. Satta G, Atzeni A, Mchugh TD. Mycobacterium tuberculosis and whole genome sequencing: a  
500 practical guide and online tools available for the clinical microbiologist. *Clinical Microbiology*  
501 *and Infection* [Internet]. 2017;23(2):69–72. Available from:  
502 <http://dx.doi.org/10.1016/j.cmi.2016.09.005>
- 503 9. Mahomed S, Naidoo K, Dookie N, Padayatchi N. Whole genome sequencing for the  
504 management of drug-resistant TB in low income high TB burden settings: Challenges and

- 505           implications. *Tuberculosis* [Internet]. 2017;107:137–43. Available from:  
506           <https://doi.org/10.1016/j.tube.2017.09.005>
- 507 10. Takiff HE, Feo O. Clinical value of whole-genome sequencing of *Mycobacterium tuberculosis*.  
508           *Lancet Infect Dis* [Internet]. 2015;15(9):1077–90. Available from:  
509           [http://dx.doi.org/10.1016/S1473-3099\(15\)00071-7](http://dx.doi.org/10.1016/S1473-3099(15)00071-7)
- 510 11. Sobkowiak B, Glynn JR, Houben RMGJ, Mallard K, Phelan JE, Guerra-Assunção JA, et al.  
511           Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data.  
512           *BMC Genomics*. 2018;19(1):1–15.
- 513 12. Nikolayevskyy V, Kranzer K, Niemann S, Drobniowski F. Whole genome sequencing of  
514           *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: A  
515           systematic review. *Tuberculosis*. 2016;98:77–85.
- 516 13. Witney AA, Bateson ALE, Jindani A, Phillips PPJ, Coleman D, Stoker NG, et al. Use of whole-  
517           genome sequencing to distinguish relapse from reinfection in a completed tuberculosis  
518           clinical trial. *BMC Med*. 2017 Mar 29;15(1).
- 519 14. Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. *Mycobacterium tuberculosis*  
520           and whole-genome sequencing: how close are we to unleashing its full potential? *Clinical*  
521           *Microbiology and Infection* [Internet]. 2018;24(6):604–9. Available from:  
522           <https://doi.org/10.1016/j.cmi.2017.10.030>
- 523 15. Brennan PJ. Structure, function, and biogenesis of the cell wall of *Mycobacterium*  
524           *tuberculosis*. *Tuberculosis*. 2003;83:91–7.
- 525 16. Kent L, Hugh TDMC, Billington O, Dale JW. Demonstration of Homology between IS6110 of  
526           *Mycobacterium tuberculosis* and DNAs of Other *Mycobacterium* spp. *J Clin Microbiol*.  
527           1995;33(9):2290–3.
- 528 17. Levin BR, Lipsitch M, Perrot V, Schrag S, Antia R, Simonsen L, et al. The Population Genetics of  
529           Antibiotic Resistance A Model of the Population Genetics of Antibiotic Resistance in the  
530           Commensal Flora. *Clinical Infectious Diseases*. 1997;24(1):9–16.

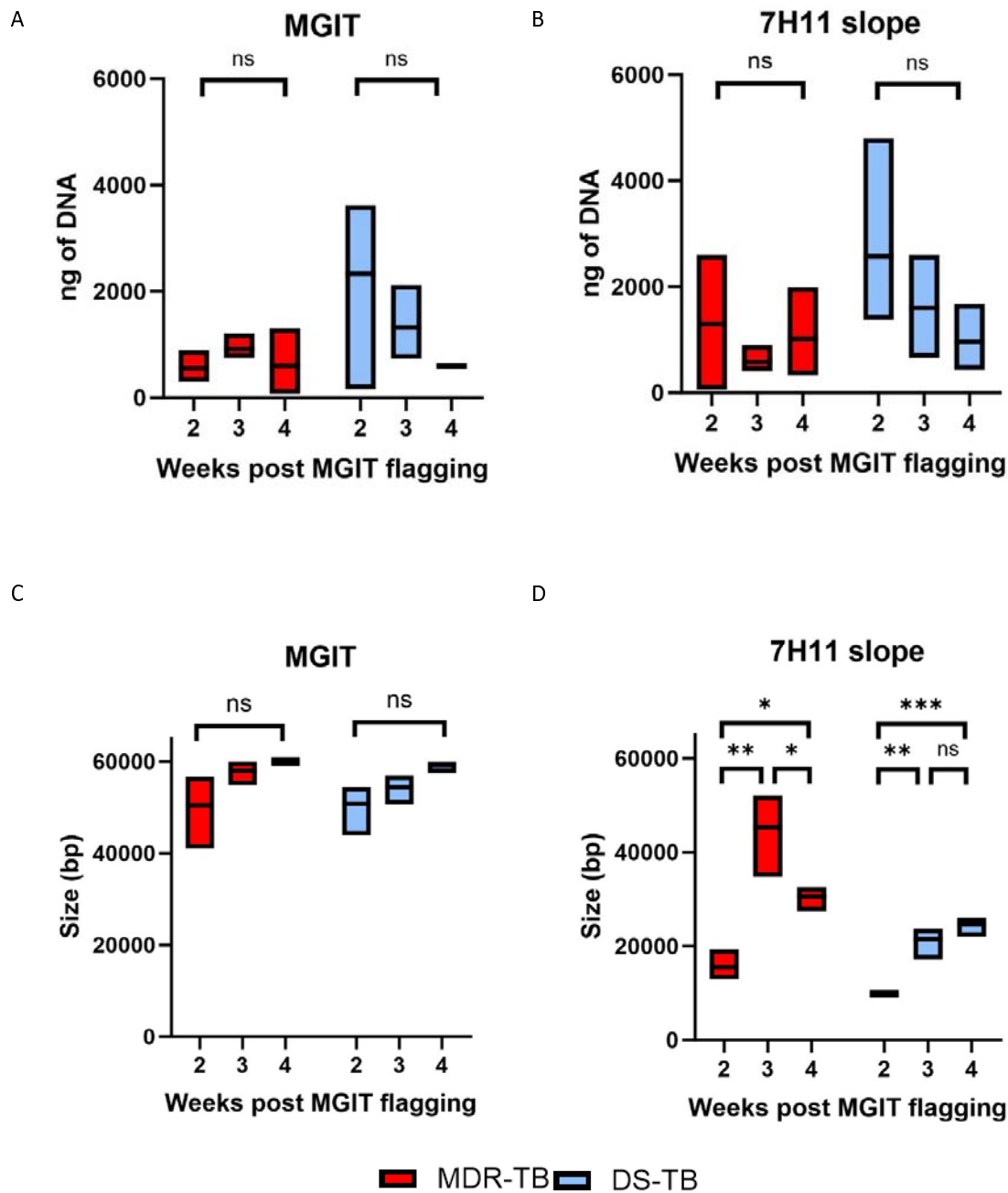
- 531 18. Modlin SJ, Robinhold C, Morrissey C, Mitchell SN, Sarah M. Exact mapping of Illumina blind  
532 spots in the Mycobacterium tuberculosis genome reveals platform-wide and workflow-  
533 specific biases. 2020;
- 534 19. Assal N, Lin M. PCR procedures to amplify GC-rich DNA sequences of Mycobacterium bovis. J  
535 Microbiol Methods. 2020;181:1–8.
- 536 20. Papaventsis D, Casali N, Kontsevaya I, Drobniowski F, Cirillo DM, Nikolayevskyy V. Whole  
537 genome sequencing of Mycobacterium tuberculosis for detection of drug resistance: a  
538 systematic review. Vol. 23, Clinical Microbiology and Infection. Elsevier B.V.; 2017. p. 61–8.
- 539 21. Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et al. Rapid,  
540 comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A  
541 prospective study. Lancet Respir Med. 2016;4(1):49–58.
- 542 22. Colman RE, Mace A, Seifert M, Hetzel J, Mshael H, Suresh A, et al. Whole-genome and  
543 targeted sequencing of drug-resistant Mycobacterium tuberculosis on the iSeq100 and  
544 MiSeq: A performance, ease-of-use, and cost evaluation. PLoS Med. 2019;16(4):1–13.
- 545 23. The CRyPTIC Consortium. Genome-wide association studies of global Mycobacterium  
546 tuberculosis resistance to 13 antimicrobials in 10,228 genomes identify new resistance  
547 mechanisms. PLoS Biol. 2022;1–27.
- 548 24. Cervantes J, Yokobori N, Hong BY. Genetic identification and drug-resistance characterization  
549 of mycobacterium tuberculosis using a portable sequencing device. A pilot study. Antibiotics.  
550 2020;9(9):1–15.
- 551 25. Bainomugisa A, Duarte T, Lavu E, Pandey S, Coulter C, Marais BJ, et al. A complete high-  
552 quality MinION nanopore assembly of an extensively drug-resistant Mycobacterium  
553 tuberculosis Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions.  
554 Microb Genom. 2018;4(7).

- 555 26. Bates M, Polepole P, Kapata N, Loose M, O’Grady J. Application of highly portable MinION  
556 nanopore sequencing technology for the monitoring of nosocomial tuberculosis infection. *Int*  
557 *J Mycobacteriol.* 2016;5:S24.
- 558 27. Yu G, Shen Y, Zhong F, Zhou L, Chen G, Fang L, et al. Diagnostic accuracy of nanopore  
559 sequencing using respiratory specimens in the diagnosis of pulmonary tuberculosis.  
560 *International Journal of Infectious Diseases.* 2022;0–33.
- 561 28. Smith C, Halse TA, Shea J, Modestil H, Fowler RC, Musser KA, et al. Assessing Nanopore  
562 Sequencing for Clinical Diagnostics: a comparison of Next-Generation Sequencing (NGS) for  
563 *Mycobacterium tuberculosis.* *J Clin Microbiol.* 2021;59(1):1–14.
- 564 29. Foster D, Coles S, Wyllie D, Robinson E, Ahyow L, Cox S, et al. *Mycobacterium tuberculosis*  
565 whole-genome sequencing and cluster investigation handbook [Internet]. 2022 [cited 2022  
566 Oct 4]. Available from: [https://www.gov.uk/government/publications/tb-strain-typing-and-](https://www.gov.uk/government/publications/tb-strain-typing-and-cluster-investigation-handbook/mycobacterium-tuberculosis-whole-genome-sequencing-and-cluster-investigation-handbook)  
567 [cluster-investigation-handbook/mycobacterium-tuberculosis-whole-genome-sequencing-](https://www.gov.uk/government/publications/tb-strain-typing-and-cluster-investigation-handbook/mycobacterium-tuberculosis-whole-genome-sequencing-and-cluster-investigation-handbook)  
568 [and-cluster-investigation-handbook](https://www.gov.uk/government/publications/tb-strain-typing-and-cluster-investigation-handbook/mycobacterium-tuberculosis-whole-genome-sequencing-and-cluster-investigation-handbook)
- 569 30. Elton L, Stoker N, Rofael S. Tuberculosis whole genome sequencing diagnostics pipeline (using  
570 Oxford Nanopore Technologies) v2.0. 2023.
- 571 31. Covaris. USER MANUAL: g-TUBE covaris. 2016;(10):1–9.
- 572 32. Oxford Nanopore Technologies. Rapid Barcoding kit protocol (SQK-RBK004). 2019. p. 1–4.
- 573 33. Oxford Nanopore Technologies. Rapid sequencing gDNA - barcoding (SQK-RBK110.96)  
574 [Internet]. 2022 [cited 2022 Dec 12]. Available from:  
575 [https://community.nanoporetech.com/docs/prepare/library\\_prep\\_protocols/rapid-](https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-barcoding-kit-96-sqk-rbk110-96/v/rbk_9126_v110_rev_m_24mar2021)  
576 [barcoding-kit-96-sqk-rbk110-96/v/rbk\\_9126\\_v110\\_rev\\_m\\_24mar2021](https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-barcoding-kit-96-sqk-rbk110-96/v/rbk_9126_v110_rev_m_24mar2021)
- 577 34. Oxford Nanopore Technologies. Rapid PCR Barcoding Kit (SQK-RPB004) protocol [Internet].  
578 2020 [cited 2021 Jun 28]. p. 1–4. Available from:  
579 [https://community.nanoporetech.com/protocols/rapid-pcr-barcoding/checklist\\_example.pdf](https://community.nanoporetech.com/protocols/rapid-pcr-barcoding/checklist_example.pdf)

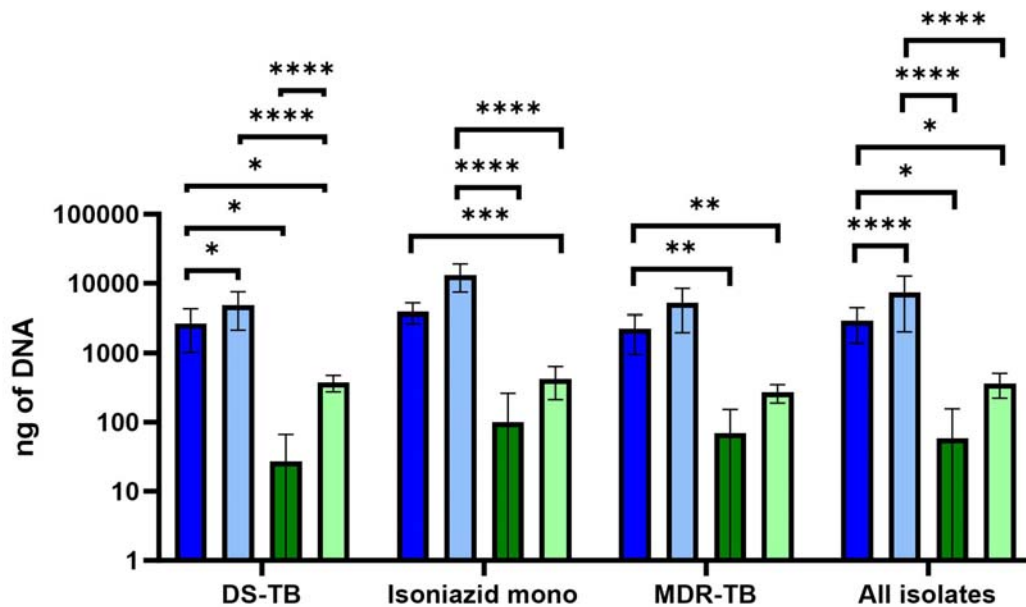


- 580 35. Illumina. Illumina DNA Prep Reference Guide [Internet]. 2020. Available from:  
581 [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).
- 582 36. Illumina. MiSeq<sup>®</sup> Reagent Kit v3 Reagent Preparation Guide FOR RESEARCH USE ONLY  
583 Revision History 3 MiSeq Reagent Kits 4 Kit Contents and Storage Requirements 5 Flow Cell  
584 Overview 6 Reagent Cartridge Overview 7 Prepare Reagents 9 Next Steps 12 Technical  
585 Assistance. 2013.
- 586 37. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford  
587 Nanopore sequencing. *Genome Biol.* 2019;20(1):1–10.
- 588 38. Babraham Bioinformatics. FastQC [Internet]. 2019 [cited 2021 Jun 28]. Available from:  
589 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- 590 39. Ewels P, Magnusson M, Lundin S, Källner M. MultiQC: Summarize analysis results for multiple  
591 tools and samples in a single report. *Bioinformatics.* 2016;32(19):3047–8.
- 592 40. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics.*  
593 2018;34(18):3094–100.
- 594 41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence  
595 Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
- 596 42. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: An integrated platform for  
597 visualization and analysis of high-throughput sequence-based experimental data.  
598 *Bioinformatics.* 2012;28(4):464–9.
- 599 43. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, et al. Antibiotic resistance  
600 prediction for *Mycobacterium tuberculosis* from genome sequence data with mykrobe.  
601 *Wellcome Open Res.* 2019;4.
- 602 44. Phelan JE, O’Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, et al. Integrating  
603 informatics tools and portable sequencing technology for rapid detection of resistance to  
604 anti-tuberculous drugs. *Genome Med.* 2019;11(1):1–7.

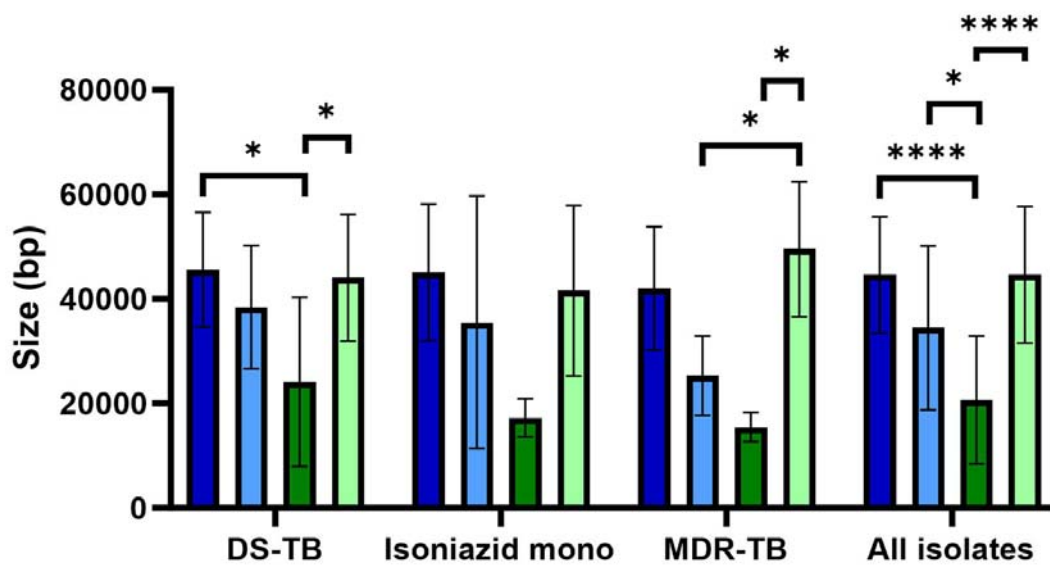
- 605 45. World Health Organization. Catalogue of mutations in Mycobacterium tuberculosis complex  
606 and their association with drug resistance [Internet]. 2021. Available from:  
607 <https://www.who.int/publications/i/item/9789240028173>
- 608 46. Khawbung JL, Nath D, Chakraborty S. Drug resistant Tuberculosis: A review. Vol. 74,  
609 Comparative Immunology, Microbiology and Infectious Diseases. Elsevier Ltd; 2021.
- 610 47. Illumina. Illumina and GenoScreen Deeplex® Myc-TB Combo Kit [Internet]. [cited 2023 May  
611 11]. Available from: [https://emea.illumina.com/products/by-type/sequencing-kits/library-  
612 prep-kits/deeplex-myctb-combo-kit.html](https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits/deeplex-myctb-combo-kit.html)
- 613 48. World Health Organization. Use of targeted next-generation sequencing to detect drug-  
614 resistant tuberculosis [Internet]. 2023 [cited 2023 Sep 1]. 1–8 p. Available from:  
615 <https://apps.who.int/iris/rest/bitstreams/1521320/retrieve>  
616  
617



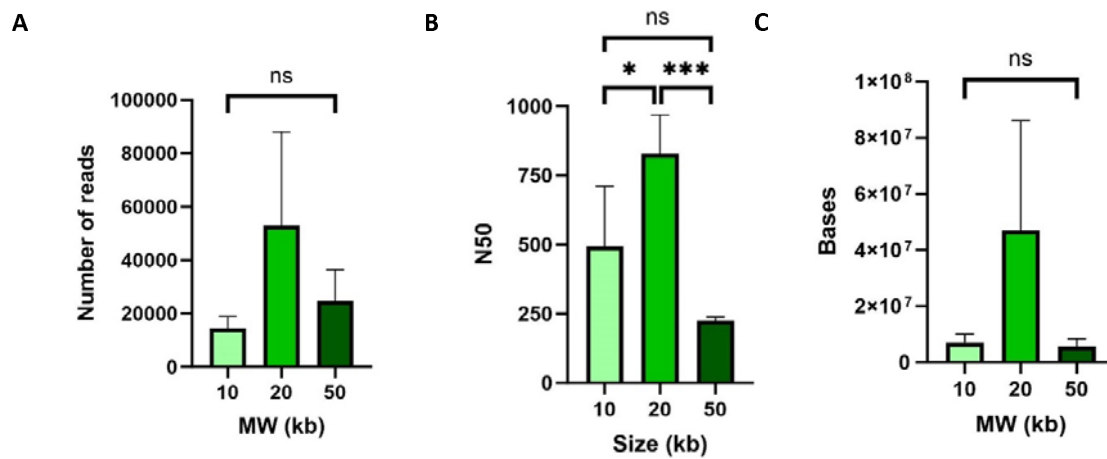
A

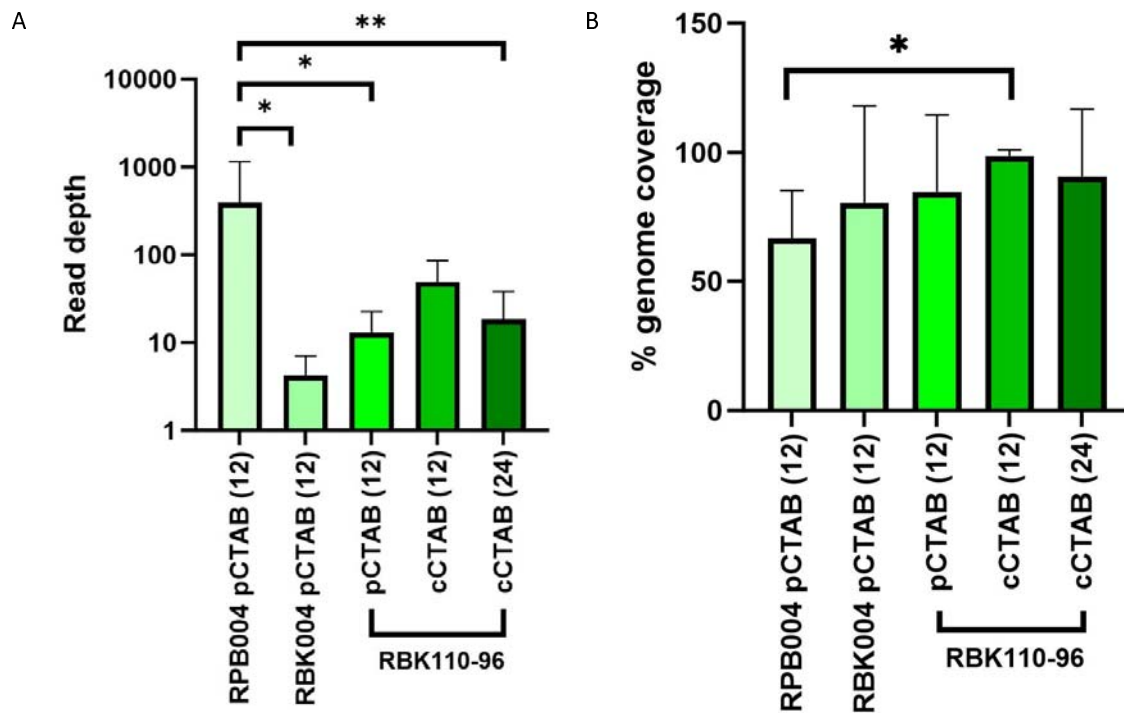


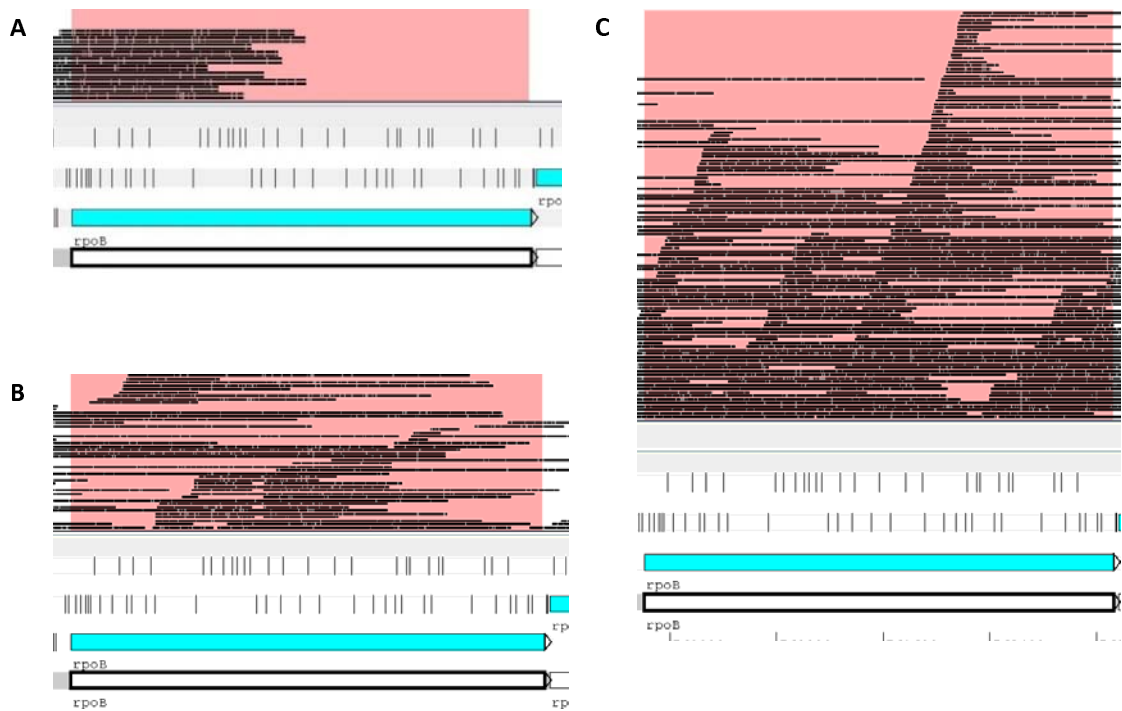
B



■ Precipitation CTAB, MGIT  
■ Precipitation CTAB, 7H11 slope  
■ Spin-column CTAB, MGIT  
■ Spin-column CTAB, 7H11 slope







Lineage	DS-TB																							
	S1			S2			S3			S4			S5			S6			S7			S8		
	-	2	2	-	4	4	-	4	4	-	M	M	-	4	4	-	2	2	-	4	4	-	1	1
Rifampicin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
Isoniazid	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
Ethambutol	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
Streptomycin	S	S	S	S	R	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		
Moxifloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Ofloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Amikacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Capreomycin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Kanamycin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Pyrazinamide	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Ethionamide	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Ciprofloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Fluoroquinolones	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
PAS	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Cycloserine	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Delaminid	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
	P	O	I	P	O	I	P	O	I	P	O	I	P	O	I	P	O	I	P	O	I	P	O	I

Lineage	Isoniazid mono-resistance														
	I1			I2			I3			I4			I5		
	-	4	4	-	4	4	-	4	4	-	4	4	-	4	4
Rifampicin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Isoniazid	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Ethambutol	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Streptomycin	S	S	S	S	R	R	S	S	S	S	S	S	S	S	S
Moxifloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Ofloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Amikacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Capreomycin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Kanamycin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Pyrazinamide	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Ethionamide	NT	R	R	NT	R	R	NT	R	R	NT	R	R	NT	R	R
Ciprofloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Fluoroquinolones	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
PAS	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Cycloserine	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Delaminid	NT	S	S	NT	S	S	NT	S	S	NT	S	S	NT	S	S
	P	O	I	P	O	I	P	O	I	P	O	I	P	O	I

Lineage	MDR-TB											
	M1			M2			M3			M4		
	-	4	4	-	2	2	-	2	2	-	4	4
Rifampicin	R	R	R	R	R	R	R	R	R	R	R	R
Isoniazid	R	R	R	R	R	R	R	R	R	R	R	R
Ethambutol	S	R	R	S	R	R	S	R	R	S	S	S
Streptomycin	R	R	R	R	S	S	R	S	S	S	S	S
Moxifloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Ofloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Amikacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Capreomycin	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Kanamycin	NT	S	S	NT	R	R	NT	R	R	NT	S	S
Pyrazinamide	R	R	R	R	S	S	R	S	S	S	S	S
Ethionamide	NT	S	S	NT	R	R	NT	R	R	NT	R	R
Ciprofloxacin	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Fluoroquinolones	NT	S	S	NT	S	S	NT	S	S	NT	S	S
PAS	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Cycloserine	NT	S	S	NT	S	S	NT	S	S	NT	S	S
Delaminid	NT	S	S	NT	S	S	NT	S	S	NT	S	S
	P	O	I	P	O	I	P	O	I	P	O	I



