Typhi Mykrobe: fast and accurate lineage identification and antimicrobial resistance genotyping directly from sequence reads for the typhoid fever agent *Salmonella* Typhi

Authors:

Danielle J. Ingle[1], Jane Hawkey[2], Martin Hunt[3,4,5,6], Zamin Iqbal[3,7], Jacqueline A. Keane[8] Ayorinde O. Afolayan[9,10], Niyaz Ahmed[11], Saadia Andleeb[12], Philip M. Ashton[13,14], Isaac I. Bogoch[15], Megan E. Carey[16,17], Marie Anne Chattaway[18], John A. Crump[19], Paula Diaz Guevara[20], Benjamin P. Howden[1,21,22], Hidemasa Izumiya[23], Jobin John Jacob[24], Louise M. Judd[1,22,], Arti Kapil[25], Karen H. Keddy[26], Justin Y. Kim[27], Myron M.Levine[28], Masatomo Morita[23], Satheesh Nair[18], Sophie Octavia[29], Iruka N. Okeke[9], Precious E. Osadebamwen[9], Sadia Isfat Ara Rahman[30], Assaf Rokney[31], David A. Rasko[28,32], Varun Shamanna[33], Michael J. Sikorski[28,32,], Anthony M. Smith[34,35], Gabriel T. Sunmonu[9], Kaitlin A. Tagg[27], Ryan R. Wick[1], Zoe A. Dyson[16,36]*, Kathryn E. Holt[2,16]* and Global Typhoid Genomics Consortium

* Indicates equal contribution

Affiliations:

1 The Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia

2 Department of Infectious Diseases, School of Translational Medicine, Monash University, Melbourne, Victoria 3004, Australia

3 European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom

4 Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

5 National Institute of Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Headley Way, Oxford, United Kingdom

6 Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, United Kingdom.

7 Milner Centre for Evolution, University of Bath, United Kingdom

8 Department of Medicine, University of Cambridge, Cambridge Institute of Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, United Kingdom

9 Faculty of Pharmacy and College of Medicine, University of Ibadan, Ibadan, Nigeria

34 10 Institute for Infection Prevention and Control, Freiburg University Medical Centre, Freiburg,
35 Germany
36 11 Pathogen Biology Laboratory, Department of Biotechnology and Bioinformatics, University of
37 Hyderabad, Hyderabad, India
38 12 Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and
39 Technology, Pakistan
40 13 Malawi Liverpool Wellcome Programme, Blantyre, Malawi
41 14 Institute of Infection, Ecology and Veterinary Sciences, University of Liverpool, Liverpool, UK
42 15 Department of Medicine, University of Toronto, Canada
43 16 Department Infection Biology, London School of Hygiene and Tropical Medicine, London, UK
44 17 IAVI, Chelsea & Westminster Hospital, London, United Kingdom
45 18 Gastrointestinal Bacteria Reference Unit, United Kingdom Health Security Agency, London,
46 United Kingdom
47 19 Centre for International Health, University of Otago, Dunedin, New Zealand
48 20 Grupo de Microbiologia, Instituto Nacional de Salud, Bogota, Colombia
49 21 Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology &
50 Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and
51 Immunity, Melbourne, Australia
52 22 Innovation Hub, Centre for Pathogen Genomics, University of Melbourne (Doherty Institute)
53 23 Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan
54 24 Department of Clinical Microbiology, Christian Medical College, Vellore, India
55 25 All India Institute of Medical Sciences, New Delhi, India
56 26 Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of
57 Pretoria, South Africa
58 27 Centers for Disease Control and Prevention, Atlanta, United States of America
59 28 Center for Vaccine Development and Global Health, University of Maryland School of
60 Medicine, Baltimore, Maryland, United States of America
61 29 School of Biotechnology and Biomolecular Sciences, University of New South Wales,
62 Sydney, Australia
63 30 International Centre for Diarrhoeal Disease Research Bangladesh (icddr,b), Bangladesh
64 31 Ministry of Health, Jerusalem, Israel
65 32 Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore,
66 Maryland, United States of America
67 33 Central Research Laboratory, KIMS, Bengaluru, India

34 Centre for Enteric Diseases, National Institute for Communicable Diseases, Johannesburg, South Africa
35 Department of Medical Microbiology, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa
36 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

Correspondence:
Danielle J. Ingle: danielle.ingle@unimelb.edu.au
Zoe A. Dyson: zoe.dyson@lshtm.ac.uk
Kathryn E. Holt: kat.holt@lshtm.ac.uk

Key words: Typhoid fever, *Salmonella* Typhi, Antimicrobial Resistance (AMR), Whole Genome Sequencing (WGS), pathogen sequencing, genotyping, lineage

Abstract

Background

Typhoid fever results from systemic infection with *Salmonella enterica* serovar Typhi (Typhi) and causes 10 million illnesses annually. Disease control relies on prevention (water, sanitation, and hygiene interventions or vaccination) and effective antimicrobial treatment. Antimicrobial resistant (AMR) Typhi lineages have emerged and become established in many parts of the world. Knowledge of local pathogen populations informed by genomic surveillance, including of lineages (defined by the GenoTyphi scheme) and AMR determinants, is increasingly used to inform local treatment guidelines and to inform vaccination strategy. Current tools for genotyping Typhi require multiple read alignment or assembly steps and have not been validated for analysis of data generated with Oxford Nanopore Technologies (ONT) long-read sequencing devices. Here, we introduce Typhi Mykrobe, a command line software tool for rapid genotyping of Typhi lineages, AMR determinants, and plasmid replicons direct from sequencing reads.

Results

We validated Typhi Mykrobe lineage genotyping by comparison with the current standard read mapping-based approach and demonstrated 99.8% concordance across nearly 13,000 genomes sequenced with Illumina platforms. For the few isolates with discordant calls, we show that Typhi Mykrobe results are better supported by the evidence from raw sequence read data than the results generated using the mapping-based approach. We also demonstrate 99.9% concordance for detection of AMR determinants compared with the current standard assembly-based approach, with similar results for plasmid marker detection. Typhi Mykrobe predicts clinical resistance categorisation (S/I/R) for eight drug classes, and we show strong agreement with phenotypic categorisations generated from reference laboratory minimum inhibitory concentration (MIC) data for n=1,572 Illumina-sequenced isolates (>99% agreement within one doubling dilution). Using n=100 isolates sequenced with both Illumina and ONT, we show strong concordance (>96% for genotype and >98% for AMR and plasmid) between calls made from ONT reads and those made from Illumina reads. Typhi Mykrobe takes less than a minute per sample and is available at https://github.com/typhoidgenomics/genotyphi.

Conclusions

Typhi Mykrobe provides rapid and sensitive genotyping of Typhi genomes direct from Illumina and ONT reads. It demonstrated accurate assignment of GenoTyphi lineage, detection of AMR determinants and prediction of corresponding AMR phenotypes, and identification of plasmid replicons.

Background

Typhoid fever is caused by systemic infection with the human-restricted bacterium *Salmonella enterica* subspecies *enterica* serovar Typhi (Typhi) [1]. More than 10 million typhoid fever illnesses occur annually (mainly in south Asia), associated with 100,000 deaths, and the risk of complications and death increases without effective antimicrobial therapy [2]. The World Health Organization AWaRe (Access, Watch, Reserve) Antibiotic guidance recommends typhoid fever is treated with ciprofloxacin, unless resistance is prevalent locally, in which case azithromycin is recommended to treat uncomplicated disease and intravenous ceftriaxone is recommended for treatment of severe disease [3,4]. A high proportion of typhoid fever is caused by antimicrobial resistant (AMR) Typhi, and nearly all Typhi isolated in south Asia have been ciprofloxacin non-susceptible for over two decades [5,6]. Extensively drug resistant (XDR, resistant to amoxicillin/ampicillin, chloramphenicol, trimethoprim-sulfamethoxazole, ciprofloxacin, and ceftriaxone) Typhi, for which the only treatment options are oral azithromycin or intravenous carbapenems, have been circulating in Pakistan for nearly a decade with travel-associated infections reported globally [7–9]. Concerningly, XDR Typhi with no travel links has been reported in USA and China [7,10], and XDR isolates with additional resistance to azithromycin and carbapenems were recently reported in Pakistan [11].

Tracking the emergence and spread of AMR Typhi through whole genome sequencing (WGS) is important to inform public health control measures such as deployment of typhoid conjugate vaccines (TCVs) [12], investigating and controlling outbreaks [10], and informing the development of antimicrobial treatment guidance [3], including management of travel-associated infections in non-endemic countries [4]. Genetic determinants of AMR are well-understood in Typhi [13,14], and include both acquired genes that are plasmid-borne or chromosomally integrated [15], and point mutations in core genes. These include *acrB* for azithromycin resistance [16], and quinolone resistance-determining regions (QRDR) of *gyrA*, *gyrB,* and *parC* for ciprofloxacin non-susceptibility and resistance [13,17]. Recently, carbapenem-resistant Typhi was described, due to acquisition of a plasmid-borne carbapenemase gene [11]. Typhi is highly clonal, which has hampered subtyping in the past; however, its population structure has been well defined using WGS [15], and the GenoTyphi framework for genotyping and variant nomenclature first proposed in 2016 [18] has been widely adopted by the global community to facilitate identification of emerging variants and communication across laboratories and settings [6,14,19–21].

149 The GenoTyphi scheme was originally defined based on nearly 2,000 WGS isolates from over
150 60 countries [18]. It uses a hierarchical framework based on the global phylogeny, stratifying the
151 Typhi population structure into four primary lineages, 16 clades, and 49 subclades (henceforth
152 referred to as genotypes), based on lineage-specific marker single nucleotide variants (SNVs).
153 Under the GenoTyphi scheme, the lineage known as 'H58', which has been associated with the
154 spread of multidrug resistance (MDR, resistant to amoxicillin/ampicillin, chloramphenicol,
155 trimethoprim-sulfamethoxazole) Typhi throughout Asia and into Eastern Africa [15], was
156 designated genotype 4.3.1. The scheme was later extended to provide further discriminatory
157 power for identifying additional subpopulations of epidemiological importance [21–27], such as
158 the XDR Typhi lineage associated with outbreaks in Pakistan (genotype 4.3.1.1.P1) [26,27]. The
159 scheme currently includes 87 genotypes [28], and its future maintenance and development will
160 be managed by a working group within the Global Typhoid Genomics Consortium
161 (https://www.typhoidgenomics.org/).
162
163 The GenoTyphi scheme specification consists of a set of marker SNVs defining the hierarchical
164 genotypes, which in principle can be used by any bioinformatics tool to assign genotypes to new
165 genomes. The original GenoTyphi paper [18] was accompanied by a Python-based pipeline
166 (available at: https://github.com/typhoidgenomics/genotyphi) designed for typing Illumina WGS
167 data against the scheme. This implementation requires users to first map short reads or
168 assemblies to the reference sequence of Typhi strain CT18 (accession AL513382.1) [29], then
169 provide as input the resulting binary alignment map (BAM) or variant call file (VCF). Marker
170 SNVs were identified from these input files and used to assign genomes to genotypes.
171 Functionality was extended to include detection of mutations associated with reduced
172 susceptibility to ciprofloxacin or azithromycin [22]. This mapping-based approach had various
173 limitations, largely due to the need for users to pre-map their data to a specific reference, which
174 was cumbersome (e.g. some users had issues with mapping tools creating incompatible output
175 files, the BAM files are large and inefficient to process), and AMR analysis was restricted to
176 mutation detection. The GenoTyphi scheme has also been implemented in other genotyping
177 tools, including Typhi PathogenWatch, [13] which works by analysing assembled genomes, and
178 the split-*k*-mer-based BioHansel tool, which does not include AMR detection [30].
179
180 Here, we present a novel approach to rapid genotyping of both AMR and GenoTyphi lineages
181 from reads, implemented in the Mykrobe open-source software framework [31], dubbed Typhi
182 Mykrobe. Input into this new tool accepts raw WGS reads (FASTQ files), and has been tested

183 with reads generated by ONT and Illumina platforms. The tool performs rapid (i) lineage
184 genotyping using the GenoTyphi framework; (ii) detection of a broad range of clinically relevant
185 molecular determinants of AMR, including both acquired genes and mutations; and (iii)
186 detection of plasmid replicon markers common to Typhi. We validated GenoTyphi lineage
187 assignment from reads by comparison with the original mapping-based implementation, using
188 12,839 Typhi genomes [14]. We assessed AMR genotype calls by comparison with assembly-
189 based genotype calls from PathogenWatch, and with antimicrobial susceptibility phenotypes for
190 4,018 isolates. Finally, we assessed performance on long read data generated using Oxford
191 Nanopore Technologies (ONT) MinION devices, using 93 isolates with matched ONT and
192 Illumina data.

193 Implementation

194 Genotyping targets

195 The current scope of the GenoTyphi scheme as discussed here (n=87 genotypes) is illustrated
196 in **Supplementary Figure 1**. This includes five new genotypes (3.5.4.1, 3.5.4.2, 3.5.4.3,
197 4.3.1.2.1, 4.3.1.2.1.1) recently described in a technical report [28]. The full scheme specification
198 is available in the GenoTyphi repository (https://github.com/typhoidgenomics/genotyphi) in the
199 file Genotype_specification.csv.
200
201 The AMR and plasmid targets included in Typhi Mykrobe are summarised in **Figure 1**. These
202 include sequence probes targeted to detect presence/absence of epidemiologically relevant
203 mobile AMR determinants [13,14], conferring resistance to the currently recommended drugs
204 ciprofloxacin, ceftriaxone, and azithromycin, as well as older drugs ampicillin, chloramphenicol,
205 trimethoprim-sulfamethoxazole, and tetracycline, as well as carbapenems. Fourteen SNVs
206 associated with resistance to ciprofloxacin and azithromycin are also included. In addition, Typhi
207 Mykrobe includes probes to detect the presence or absence of 13 plasmid replicons previously
208 reported in Typhi [14], including 11 associated with AMR plasmids, the cryptic plasmid pHCM2,
209 and linear plasmid pBSSB1, which carries the alternative flagellin z66 [32]. A marker SNV to
210 identify the plasmid sequence type 6 (PST6) lineage of the dominant IncHI1 plasmid is also
211 included [33]. The full list of AMR targets is available in the GenoTyphi repository
212 (https://github.com/typhoidgenomics/genotyphi) in the file AMR_genes_mutations_plasmids.csv.

213 **Implementation in Mykrobe**

214 The Mykrobe software, first developed in 2015 for *Mycobacterium tuberculosis* and
215 *Staphylococcus aureus* [34], uses a *k*-mer-based approach to identify marker SNVs and the
216 presence or absence of target genes using 'probe' sequences to screen against [31]. These
217 data are then used to assign hierarchical genotypes and report AMR determinants. Probe sets
218 have been previously developed and implemented for other bacterial pathogens including
219 *Mycobacterium tuberculosis* [34–36], *Staphylococcus aureus* [31] and *Shigella sonnei* [37].
220 Mykrobe is available at https://github.com/Mykrobe-tools/mykrobe.
221
222 Mykrobe probe sets were created for genotype and AMR marker SNVs using the *mykrobe*
223 *variant make-probes* command using a *k*-mer size of 21 and the Typhi CT18 reference genome
224 (accession: AL513382.1). The nested hierarchical relationships of the marker SNVs were
225 specified and Mykrobe used this information to identify the best supported genotype call (e.g., a
226 call of '4.3.1.1' is typically supported by detection of three nested markers that demarcate
227 different levels of the hierarchy: 4, 4.3.1, 4.3.1.1). Briefly, probe sets for gene presence or
228 absence were created using a Python script that extracted the relevant AMR sequences. Details
229 of the specific commands and Python scripts are provided in
230 https://github.com/typhoidgenomics/genotyphi. Probe sets were also created to confirm input
231 read sets as *S. enterica* using the presence of marker gene *invA*, involved in invasion of
232 intestinal epithelium cells by *S. enterica*, and serovar Typhi based on the 26 known multi-locus
233 sequence types (STs) (see **Figure 1**). Further details of the preliminary testing and
234 implementation of the Typhi probe sets are available in the technical report
235 https://zenodo.org/doi/10.5281/zenodo.7407984. The Typhi Mykrobe probe sets and typing
236 panel are available at https://doi.org/10.6084/m9.figshare.27102334.

237 **Summarising Mykrobe outputs**

238 To support interpretation of the Mykrobe output, we developed a Python-based parser script
239 *parse_typhi_mykrobe.py,* which takes as input a set of JavaScript Object Notation (JSON) files
240 output by Mykrobe, and summarises them in a tab-delimited table with one row per genome,
241 and columns to indicate the genotype calls (see **Figure 1**). Also included are several quality-
242 control fields reported by Mykrobe, including read-level support for each marker SNV
243 contributing to the final genotype call ('node support'), and for any additional markers detected.
244 This information is helpful for identifying contaminated samples (see examples in Results). The
245 parser script also assigns a summary 'confidence' level in the final GenoTyphi lineage call,

246 based on the Mykrobe quality scores for each nested component marker that makes up the final
247 genotype call. In brief, a 'weak' confidence call is assigned when one or more marker SNVs are
248 of low quality, defined as Mykrobe quality score of 0 or Mykrobe quality score of 0.5 with
249 minority support (i.e., <50% of reads) for the derived allele. A 'moderate' confidence call is
250 assigned if one (and only one) marker SNV has a Mykrobe quality score of 0.5 but with majority
251 support (i.e. ≥50% of reads) for the derived allele. Otherwise, the genotype call is assigned
252 'strong' confidence.
253
254 The parser script summarises the AMR markers detected by Mykrobe in the form of a pseudo-
255 antibiogram with one column per drug. Each drug column contains either 'S' (indicating no
256 markers detected) or 'R' (detected markers). For ciprofloxacin, it is useful to distinguish between
257 low-level resistance (MIC 0.06-0.5 mg/L), which is associated with a single mutation or acquired
258 gene and high-level resistance (MIC >0.5 mg/L), which is associated with combinations of two
259 or more determinants [14,17], because while both groups are associated with increased fever
260 clearance times and probability of treatment failure compared with susceptible strains, high-level
261 resistance is associated with worse outcomes than low-level resistance [38]. Current guidelines
262 and terminology for these groups vary, but for simplicity, the Mykrobe parser reports these
263 groups as 'I' (for intermediate) and 'R', respectively. This matches the current CLSI guidelines
264 (https://clsi.org/standards/products/microbiology/documents/m100/). However, it should be
265 noted that (i) EUCAST defines the clinical breakpoint for 'R' in *Salmonella* spp. as MIC >0.06
266 mg/L (https://www.eucast.org/clinical_breakpoints), which includes both low-level and high-level
267 resistance; and (ii) previous literature has referred to the low-level range as 'decreased
268 ciprofloxacin susceptibility' rather than 'I' [39]. In addition to this antibiogram-style summary,
269 individual AMR and plasmid markers are reported in their own columns, coded as 1=present
270 and 0=absent (as seen in **Supplementary Table 1,** available at
271 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe).
272
273 All code and instructions for genotyping isolates with GenoTyphi using Typhi Mykrobe are freely
274 available in the GitHub repository at https://github.com/typhoidgenomics/genotyphi (DOI:
275 10.5281/zenodo.13859721).

276 **Validation data**

277 Validation on Illumina short-read data was conducted using the Global Typhoid Genomics
278 Consortium dataset described in Carey et al [14]. Reads were mapped to the CT18 reference

279 genome (accession: AL513382) [29] using bwa-mem v0.7.17 via the Centre for Genomic
280 Pathogen Surveillance (CGPS) mapping pipeline v1.2.2
281 (https://gitlab.com/cgps/ghru/pipelines/snp_phylogeny), to generate alignments in BAM format.
282 Assemblies were generated using the CGPS assembly pipeline v2.1.0
283 (https://gitlab.com/cgps/ghru/pipelines/dsl2/pipelines/assembly/) which utilises the SPAdes
284 assembler (v3.12.0) [40,41]. A total of 141 isolates were excluded due to excessive unclustered
285 heterozygous SNVs indicative of mixed cultures (n=92), assembly issues (n=46) or both (n=3),
286 resulting in 12,839 isolates for Mykrobe validation (listed in **Supplementary Table 2,** available
287 at https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe).
288

289 **Validation of genotype calls**

290 Typhi Mykrobe v0.12.1 was run on all (n=12,839) Illumina FASTQ files using the new Typhi
291 panel (v20221208) and the results were summarised using *parse_typhi_mykrobe.py*. To
292 validate Typhi Mykrobe's GenoTyphi genotype calls, we compared them with those called by
293 the original mapping-based approach. To do this, we used BAM file read alignments generated
294 from the same n=12,839 FASTQs as described above as input to the Python script genotyphi.py
295 (v2.0; available at https://github.com/typhoidgenomics/genotyphi) [18,22,28]. To validate the
296 AMR and plasmid calls, we compared them to those called in corresponding genome
297 assemblies by Pathogenwatch (available for n=11,992 genomes, assemblies for the other
298 genomes were excluded as they did not meet Pathogenwatch quality-control criteria[13]). We
299 downloaded the GenoTyphi, AMR, and plasmid replicon typing reports generated by Basic
300 Local Alignment Search Tool (BLAST) searching assemblies for these genomes from
301 Pathogenwatch on 7 March 2023 (GenoTyphi v20221208 specification for lineage genotypes).
302 Categorical agreement (percentage of genomes where both tools agreed on the same call of
303 presence or absence) was calculated separately for each marker, and collectively by drug
304 (percentage of genomes where both tools agreed that at least one marker was present or that
305 no markers were present) (**Supplementary Table 3,** available at
306 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). Probes for plasmid-
307 borne carbapenemase genes circulating in Enterobacterales (*bla*NDM-5, *bla*KPC-2, *bla*VIM-1,
308 *bla*IMP-27, *bla*OXA-48) were added to GenoTyphi v20240407 (DOI: 10.5281/zenodo.13859721)
309 following the report of *bla*NDM-5 in Typhi from Pakistan [11], and validated using the genomes
310 of the two distinct morphologic variants from that study (accessions: SRR22801766 and

311 SRR22801806). Comparisons and statistical analyses were conducted in R v4.1.0 using
312 tidyverse v1.3.1 [42].

313 **Accuracy of AMR phenotype prediction**

314 Antimicrobial susceptibility testing (AST) data were available for n=3,970 of the Illumina-
315 sequenced isolates, sourced from three separate datasets (**Supplementary Table 4,** available
316 at https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). Minimum
317 inhibitory concentration (MIC) data were available from the UK Health Security Agency
318 (UKHSA) for n=852 isolates following EUCAST standards and interpretive criteria (v10, 2020)
319 [43] and the US Centers for Disease Control and Prevention (US CDC) for n=720 isolates
320 following CLSI standards and interpretive criteria (M100, 2023) [44]. Disk diffusion data
321 following EUCAST standards and interpretive criteria (v8.0, 2018) were available for n=2,446
322 isolates collected and analysed from three countries as part of the Surveillance for Enteric Fever
323 in Asia Project (SEAP), with multiple source labs contributing AST data and isolates for
324 sequencing [6]. Typhi Mykrobe AMR calls following the S/I/R categorisation described above
325 were compared with the phenotypic antimicrobial susceptibility S/I/R categorisation using the
326 relevant CLSI or EUCAST standards for each data set, for the currently recommended drugs
327 azithromycin and ciprofloxacin and ceftriaxone, and the older drugs ampicillin, chloramphenicol,
328 trimethoprim-sulfamethoxazole, and tetracycline (note UKHSA tested amoxicillin rather than
329 ampicillin; and SEAP did not test tetracycline). To assess concordance between genotype and
330 phenotype, we followed the principles used for assessment of new AST devices by the US Food
331 and Drug Administration (FDA), which considers (i) categorical agreement, (ii) major error rate
332 (defined as the proportion of S isolates that test R with the new method [45,46]), and (iii) very
333 major error rate (defined as the proportion of R isolates that test S with the new method)
334 (**Supplementary Table 5,** available at
335 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). The FDA standard
336 requires >90% overall categorical agreement (within one doubling dilution), <3% major errors for
337 each drug, and <1.5% very major errors for each drug; hence, we adopted these thresholds to
338 assess the accuracy of the genotype-based prediction of phenotype. For drugs with high error
339 rates, we investigated the relationship between individual resistance markers and phenotype,
340 which we visualised using upset plots generated in R v4.1.0 using the ComplexUpset package
341 v1.3.3 (10.5281/zenodo.3700590). For five isolates with unexplained trimethoprim-
342 sulfamethoxazole resistance, we used nucleotide blast search of the genome assemblies to
343 confirm the presence of wild-type chromosomal *folA* and *folP* (100% nucleotide match to

344 reference sequences from Typhi strain Ty2) and screened the assemblies for other AMR
345 determinants using CARD RGI v6.0.3 [47].

346 **Matched ONT and Illumina validation data**

347 For 93 isolates, long-read ONT data were available in addition to the Illumina data; these
348 included 68 isolates from Carey et al 2023 [14] and 30 novel isolates from this study. Sample-
349 level details including accessions and technical specifications such as library, flow cell, device,
350 basecaller, and ONT processing (including demultiplexing, trimming and filtering), are available
351 in **Supplementary Table 6,** available at
352 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe. These varied
353 between isolates including different flow cells (e.g., R9.4.1, n=30 and R10.4.1, n=24) and base
354 calling methods. Read depth for each ONT dataset was assessed by using seqtk v1.2-r94
355 (https://github.com/lh3/seqtk) to determine total number of bases per read set, which was
356 divided by the length of the CT18 reference genome (4,809,037 bases). For four genomes, ONT
357 reads were assembled to resolve discrepant plasmid calls, assembly was done using Flye
358 v2.9.1-b1780 [48] with default parameters and yielding genome sizes of the expected 4.8 Mbp.
359 Due to the high read depth (784×), the ONT data for SRR17299234 was subsampled prior to
360 assembly using Filtlong v0.2.1 (https://github.com/rrwick/Filtlong) to --target_bases 480000000
361 (resulting in 99.81× depth).

362 **Results**

363 **Validation of GenoTyphi genotype calling**

364 To validate Typhi Mykrobe for assigning hierarchical GenoTyphi genotypes, we ran it on
365 n=12,839 Illumina readsets and compared the resulting genotypes to those called using the
366 original mapping-based implementation. Five genomes were typed as ‘not Typhi’ by Mykrobe
367 and not analysed further. Of the remaining n=12,834 genomes, Mykrobe called the same single
368 genotype as the mapping-based approach for n=12,807 (99.79%). For a further n=17 genomes,
369 Mykrobe and the mapping-based pipeline gave concordant reports of two genotypes being
370 present, consistent with mixed samples (in the Typhi Mykrobe output this is reported as one
371 primary genotype plus additional markers, while in the mapping-based pipeline the result is
372 reported as a list of supported genotypes). Eight genomes belonged to genotype 3.5.3, which
373 was nested in 3.5.4, breaking the usual hierarchical structure of the scheme. Typhi Mykrobe
374 correctly reported these genomes as genotype 3.5.3 with support for both markers 3.5.4 and

375 3.5.3, as it has this nested relationship explicitly encoded in its hierarchy; whereas the mapping-
376 based pipeline does not have this logic and incorrectly reported these genomes as genotype
377 3.5.4.
378
379 There were two instances of genuine disagreement in the genotype calls generated by Typhi
380 Mykrobe as compared to the mapping-based approach. One isolate (ERR5243665) was called
381 as genotype 2 by Mykrobe with strong confidence (no reads support for any alternative alleles
382 or markers). It was genotyped as 0.1.3, albeit with low support, by the mapping-based
383 approach, but the genotype 2 marker was also detected. This isolate clustered with other
384 genotype 2 isolates in a distance-based phylogeny of genomes from the same study [6] (see
385 tree at https://pathogen.watch/collection/0z5knw9jic9b-guevara-et-al-2021), suggesting that the
386 Mykrobe call is correct. The other isolate (ERR2663783) was called as genotype 4.3.1.1 by
387 Typhi Mykrobe with strong confidence. It was reported as "4.3.1.1,4.3.1.3.Bdq" (with low
388 support) by the mapping approach; however, the support for the 4.3.1.3.Bdq marker was low.
389 This isolate clustered closely with other 4.3.1.1 isolates from the same study [25](see tree at
390 https://pathogen.watch/collection/b2k6dym3bdyq-tanmoy2018), which is consistent with the
391 Mykrobe call.

392 **Validation of AMR genotyping**

393 To assess the accuracy of Mykrobe for detecting the presence of AMR genes and mutations,
394 we compared genotype calls from Mykrobe (called from *k*-mer analysis of reads) with those
395 from Pathogenwatch (called from BLAST analysis of genome assemblies of the same read
396 sets). We found very strong categorical agreement between the two methods at the individual
397 marker level, with 99.88% of the 299,004 marker-genome combinations yielding the same call
398 (summarised in **Table 2,** full details in **Supplementary Table 2**). High concordance was evident
399 for all genetic markers of AMR (98.56–100%, **Table 2**), as well as plasmid replicon markers
400 (99.74–100%, **Table 3**).
401
402 Both Mykrobe and Pathogenwatch summarise AMR data in terms of a binary phenotype
403 categorization (S/R), assuming the presence of a marker associated with resistance to a drug
404 implies resistance 'R', with the exception of ciprofloxacin, where presence of any marker is
405 categorised as 'I' and specific combinations of markers are categorised as 'R', resulting in three
406 categories S/I/R (see Methods). Categorical agreement was very high between the drug-level
407 categorisations reported by both genotyping tools, with >99.7% agreement (n=2 to 33

408 discordant genomes) for each drug (**Table 2, Supplementary Table 3**), and for subsequent
409 categorisation as MDR (99.86%, n=17 discordant) or XDR (99.97%, n=4 discordant).

410 **Validation of AMR phenotype prediction**

411 To assess the accuracy of Mykrobe's genotype-based clinical resistance categorisations, we
412 compared them with AST data, which was available for a subset of n=4,018 isolates (see
413 Methods, **Supplementary Table 4**). These isolates were from three data sets, using different
414 methods (MIC with EUCAST standards, MIC with CLSI standards, disk diffusion with EUCAST
415 standards), hence we analyzed the results from each dataset separately. Categorical
416 agreement was high (>95%) for all drugs across all three datasets, and very high (>99%) for the
417 two sets of reference-laboratory MIC measures (**Figure 2, Supplementary Figure 5**). The only
418 exception was US CDC categorisation of low-level ciprofloxacin resistance (using the CLSI 'I'
419 threshold, ≥0.125 to 0.5 mg/L), which showed 97.36% agreement with Mykrobe calls of 'I' or 'R'.
420 This was mainly due to the presence of major errors (susceptible isolates with AMR
421 determinants detected, thus reported as R by Mykrobe, see Methods), whereby n=17 isolates
422 tested susceptible to ciprofloxacin (n=14 with MIC=0.06 mg/L, n=3 with MIC=0.015 mg/L) but
423 Mykrobe detected a known QRDR mutation (**Figure 3b**); categorical agreement within one
424 doubling dilution was 99.4%.
425
426 Major errors were infrequent (<3% prevalence) for all drugs in all datasets (**Figure 2**), except for
427 ciprofloxacin (**Supplementary Table 5**). Most of the ciprofloxacin major errors were in isolates
428 whose MIC or disk zone measurement were close to the breakpoint (**Figures 3, 4c**); all carried
429 well-characterized molecular resistance determinants (QRDR mutations and/or *qnr* genes, see
430 **Figures 3, 4c**), and in all but one case the same determinants were also detected in the
431 corresponding genome assemblies by Pathogenwatch (details below).
432
433 Very major errors (resistant isolates with no AMR determinants detected, thus reported as S by
434 Mykrobe) were also infrequent (<1.5% prevalence) for the reference-laboratory MIC datasets,
435 but errors were higher in the disk diffusion dataset (**Figure 2**). The only drugs showing major or
436 very major errors compared with the reference-laboratory MIC data were ciprofloxacin and
437 trimethoprim-sulfamethoxazole (**Figure 2**). For ciprofloxacin, the US CDC categorisation of
438 high-level resistance (using the CLSI 'R' threshold, ≥1 mg/L) yielded four very major errors (R
439 isolate called as 'I' by Mykrobe). Two had a single mutation detected (*gyrA*-S83F) and two had
440 double QRDR mutations (*gyrA*-S83F plus *gyrA*-D87N or *parC*-S80I) (see **Figure 3**);

441 Pathogenwatch calls agreed in all instances. Importantly, all four isolates had MIC=1 mg/L, i.e.
442 they agreed with the genotype-based call within one doubling dilution. UKHSA uses the
443 EUCAST standards, which have a single (low-level) breakpoint for ciprofloxacin (R >0.06 mg/L),
444 and this yielded just two major errors (2.9%) and two very major errors (0.26%) (see **Figure 3**).
445
446 For trimethoprim-sulfamethoxazole, Mykrobe analysis of the US CDC dataset yielded a single
447 error (R isolate called as S) but the MIC was 4 mg/L, i.e. within one doubling dilution, and the
448 isolate had no *sul* or *dfr* genes that were identified by Mykrobe or Pathogenwatch. Mykrobe
449 analysis of the UKHSA dataset yielded five errors for trimethoprim-sulfamethoxazole; all were R
450 isolates called as S by Mykrobe, which detected *sul* genes but no *dfrA* genes. Interestingly, all
451 five isolates tested susceptible to trimethoprim alone (MIC ≤0.5, well below the >4 mg/L
452 breakpoint), supporting the absence of *dfrA* genes; these were also not detected by
453 Pathogenwatch analysis of the corresponding genome assemblies, or by UKHSA's internal
454 pipeline [44]. Further investigation of these genome assemblies confirmed the presence of wild-
455 type *folA* and *folP*, and no additional determinants were reported by CARD RGI (v6.0.3).
456
457 Mykrobe analysis of the SEAP dataset showed lower overall categorical agreement with AST
458 phenotypes (95.3–98.2%). As the testing was done using disk diffusion, it was not possible to
459 assess the categorical agreement within one doubling dilution of a MIC. However, most of the
460 errors were within 2–4 mm of the breakpoint (see **Figure 4**), and in all but one instance, the
461 Mykrobe genotype result matched that called by Pathogenwatch from the corresponding
462 genome assemblies. The only exception was a ciprofloxacin resistant isolate (ERR4325960),
463 which Mykrobe reported as carrying *gyrA*-S83F with no other determinants, but Pathogenwatch
464 detected both *gyrA*-S83F and *qnrS* in the assembly.

465 **Validation on nanopore reads**

466 We explored the accuracy of Typhi Mykrobe for genotyping ONT long-read data by comparing
467 results from 98 pairs of ONT to Illumina reads generated from the same isolates. Data were
468 contributed by members of the Global Typhoid Genomics Consortium and generated using
469 different library preparation methods and sequencing devices (see **Supplementary Table 6**).
470 Mykrobe called one Illumina read set and five ONT read sets as non-Typhi (mean ONT read
471 depths estimated as 0.04×, 4.06×, 8.01×, 26.40×, 27.85×), these were excluded from further
472 comparison as Mykrobe does not further genotype once an isolate is called as non-Typhi.
473

474 Among the n=92 Typhi genomes called from both ONT and Illumina reads, the Typhi Mykrobe
475 genotype lineage calls matched for n=89 (96.7%). For all three discrepant isolates, (all
476 sequenced on R.9 or R.9.4 flowcells), the set of SNV markers detected from ONT reads closely
477 matched those detected from Illumina reads, but with one or two markers missing from the
478 ONT-based calls and no unexpected markers reported resulting in final genotype calls of 4.3.1.1
479 vs 4.3.1, 4.3.1.2 vs 4.3.1, 2.1.7.2 vs 2.1.7. For the missing markers, Mykrobe reported low and
480 variable ONT read depths, with ranges 3–9, 4–31 and 1–14, suggesting that the errors are likely
481 due to lack of informative ONT reads at the marker loci. Notably, the overall ONT read depths
482 for these were reasonable (mean 35×, 170× and 90× respectively), suggesting the lack of
483 informative reads for genotyping may be caused by contamination from non-Typhi reads, or
484 high error rates resulting in low rates of *k*-mer matching.
485
486 AMR marker detection from ONT reads was highly concordant with Illumina-based calls, with
487 98.77% overall agreement across the n=736 isolate-drug combinations (92 isolates across 8
488 drugs; see **Supplementary Table 7**). There were nine genomes with disagreement of AMR
489 calls; five of these had a marker detected in Illumina reads but not ONT reads (n=1 *catA1*, n=1
490 *qnr,* n=3 QRDR mutations), three had *gyrB*-464F detected in ONT reads but not Illumina, and
491 one had *gyrA*-S464F detected in Illumina reads and *gyrA*-S83F in ONT. In all nine cases, the
492 Typhi Mykrobe call from Illumina reads matched that from Pathogenwatch analysis of the
493 corresponding assemblies of Illumina reads. There was no overlap between ONT read sets with
494 AMR discrepancies and those with lineage genotype discrepancies. Overall, ONT-based
495 resistance prediction showed complete agreement with Illumina for six of the eight drugs, 98.9%
496 agreement for chloramphenicol resistance, and 91% for ciprofloxacin resistance.
497
498 Plasmid marker detection from ONT data was also highly concordant with Illumina-based calls,
499 with 99.67% overall agreement across the n=1,196 isolate-marker combinations (92 isolates
500 across 13 markers; see **Supplementary Table 8**). Four genomes had discrepant plasmid
501 marker calls, all with a marker detected from Illumina reads but not from ONT reads. One isolate
502 (DRR070993) had a typical IncHI1 MDR plasmid profile reported from Illumina reads (IncHI1A,
503 IncHI1BR27, IncHI1_ST6 indicating presence of IncHI1-ST6 plasmid, with *catA1*, *dfrA7*, $bla_{\text{TEM-1}}$,
504 *sul1*, *sul2*), but the ONT read profile was missing the *IncHI1A* and *catA1* markers (and missing
505 *catA1* in the corresponding Illumina assembly and a mismatch for *IncHI1A*). The other three
506 isolates had pHCM2 detected in the Illumina reads and in the corresponding assemblies [14] but

507 not in ONT reads; for these isolates the ONT and Illumina sequencing was done on different
508 DNA extracts so this may reflect plasmid loss in the culture used for ONT sequencing.

509 **Discussion**

510 Typhi Mykrobe provides detection and genotyping of key features of clinical and public health
511 importance, including assigning lineages, AMR determinants, and plasmid replicons (**Figure 1**).
512 Compared with other tools for Typhi genotyping, Mykrobe has the advantage of not requiring
513 any pre-processing, aligning, or assembling of reads, but instead takes as inputs raw FASTQ
514 files. This simplifies bioinformatics workflows and reduces time-to-result as Typhi Mykrobe can
515 return a full genotype profile in typically <1 minute. Importantly, although Typhi Mykrobe is not a
516 QC tool *per se*, it provides useful data on the read-level support for individual genotype markers,
517 which can be used to help identify potentially mixed samples. As a command-line tool that is
518 easy to install locally using Bioconda, Typhi Mykrobe is suitable for research and public health
519 laboratory settings, facilitates standardized reporting of Typhi genotypes without the need to
520 share sequence data, and enables broader access to bioinformatics functionality across a
521 variety of settings.
522
523 Typhi Mykrobe showed very high genotyping accuracy, for GenoTyphi lineage assignment
524 (**Table 1**), and detection of AMR and plasmid markers (**Tables 2-3**), compared with existing
525 genotyping tools. Of note, the included IncY, pO111 and IncFIB for the cryptic pHCM2 replicon
526 targets have recently been shown to be sometimes carried on phage-plasmids [49,50], in
527 addition to conjugative plasmids such as the IncY plasmids associated with XDR Typhi [26].
528 Future work is required to identify additional targets to delineate the phage-plasmids, and the
529 Typhi Mykrobe software can be readily updated to include new markers.
530
531 In addition, we show that the panel of AMR markers currently included in Typhi Mykrobe can be
532 used to predict AMR phenotypes with high accuracy, demonstrating very high concordance with
533 S/I/R calls made from reference-laboratory MICs (**Figure 2**). Concordance was slightly lower
534 when compared with research laboratory disk diffusion data. However, as the phenotypic
535 resistance prevalence rates were similar across the three datasets, it is reasonable to assume
536 that the higher error rates in the disk diffusion dataset compared with the two reference-
537 laboratory MIC datasets most likely reflect comparatively lower accuracy of the disk diffusion
538 phenotype measurements, which are known to yield lower accuracy for some drugs such as
539 azithromycin [51,52] and in this case were completed across multiple laboratories. There are

540 currently no formal standards for assessing the accuracy of AMR phenotype predictions from
541 genomic data [53]; however, applying the standards for FDA licensing of new AST devices,
542 Typhi Mykrobe easily passed the target threshold for categorical agreement (>99% agreement
543 within one doubling dilution for reference-laboratory MIC data, >95% agreement for disk-
544 diffusion data; vs >90% target) (**Figure 2**).
545
546 Typhi Mykrobe also showed very low rates of major errors and very major errors compared with
547 the reference-laboratory MIC data, and for nearly all drugs these fell below the acceptable
548 thresholds for a new AST device (<3% and <1.5%, respectively) (**Figure 2**). The exception was
549 trimethoprim-sulfamethoxazole, for which five UKHSA isolates tested resistant but Typhi
550 Mykrobe predicted as susceptible due to a lack of *dfrA* genes, which are thought to be required
551 for resistance to trimethoprim-sulfamethoxazole along with *sul* genes, which were detected in
552 these genomes. Other genotyping tools (Typhi Pathogenwatch, CARD RGI) also could not
553 identify any known mechanisms of resistance to trimethoprim or trimethoprim-sulfamethoxazole
554 in these genomes, and the isolates tested sensitive to trimethoprim. Therefore, we hypothesise
555 these isolates may harbour efflux mutations or a novel mechanism of resistance to
556 trimethoprim-sulfamethoxazole that does not result in resistance to trimethoprim alone. Further
557 analyses to determine the exact resistance mechanism was beyond the scope of this study.
558 Ciprofloxacin showed very low error rates in the UKHSA MIC data, interpreted using the single
559 low-level resistance threshold (MIC >0.06 mg/L) recommended by EUCAST, but slightly higher
560 error rates for the US CDC MIC data, which was interpreted using low-level and high-level
561 resistance thresholds (although still below the acceptable thresholds within one doubling
562 dilution). Carbapenemase genes were added to the Typhi Mykrobe panel, but this resistance is
563 newly emerging and there was only a single carbapenem-resistant isolate sequenced at the
564 time of testing (in which *bla*NDM-5 was correctly identified by Typhi Mykrobe), therefore we
565 could not yet assess accuracy of phenotype prediction for carbapenems.
566
567 Finally, we demonstrated that Typhi Mykrobe can be used to genotype ONT reads, showing
568 high concordance with results obtained from the analysis of Illumina reads generated from the
569 same isolates (96.74% agreement on GenoTyphi lineage calls, 98.77% on AMR markers and
570 99.67% on plasmid markers). We did not have suitable data to directly assess ONT-based AMR
571 predictions with AST data. ONT-based and Illumina-based drug-level predictions of AST
572 phenotypes showed high agreement; however, future studies assessing ONT data compared to
573 reference-laboratory MIC data would be useful to confirm the accuracy of ONT data for

574 phenotype prediction. As one might expect, many of the errors occurred in isolates with lower
575 read depth. However, as the matched ONT-Illumina data available for testing were contributed
576 by different laboratories that used different ONT protocols and devices for library preparation,
577 sequencing, and base calling, we were unable to perform a detailed assessment of the impact
578 of read depth and quality on genotyping performance. The accuracy of ONT reads has greatly
579 increased in recent years and are approaching Illumina read accuracy of ~99%; hence modern
580 ONT data would be expected to perform similarly to Illumina data [54]
581 (doi.org/10.5281/zenodo.10397817). Future studies assessing the impact of read depth and
582 base calling methods on the performance of Typhi Mykrobe on ONT data would be desirable to
583 help guide laboratories on the most appropriate methods and target yield for ONT experiments
584 aiming to produce reliable genomic surveillance data for Typhi.

585 **Conclusions**

586 Typhi Mykrobe provides rapid and accurate genotyping of Typhi genomes directly from Illumina
587 reads, including (i) assignment of GenoTyphi lineage; (ii) detection of AMR determinants and
588 prediction of corresponding AMR phenotypes demonstrating >99% categorical agreement with
589 reference-laboratory MIC data; and (iii) detection of plasmid replicons. The software also
590 performs well on direct genotyping from ONT reads information, although further investigation is
591 needed to assess the accuracy of AMR phenotype prediction.
592

593 **Other sections**

594 **List of abbreviations**

595 AMR: antimicrobial resistance
596 AST: antimicrobial susceptibility testing
597 CARD: Comprehensive Antibiotic Resistance Database
598 CLSI: Clinical & Laboratory Standards Institute
599 EUCAST: The European Committee on Antimicrobial Susceptibility Testing
600 FDA: United States of America Food and Drug Administration
601 MDR: multidrug resistance
602 MIC: minimum inhibitory concentration
603 ONT: Oxford Nanopore Technologies

604 QRDR: quinolone resistance-determining regions
605 SEAP: Surveillance for Enteric Fever in Asia Project
606 SNV: single nucleotide variants
607 UKHSA: United Kingdom Health Security Agency
608 US CDC: United States of America Centres for Disease Control and Prevention
609 WGS: whole genome sequencing
610 XDR: extensively drug resistant

611 **Availability and requirements**

612 Project name: GenoTyphi
613 Project home page: https://github.com/typhoidgenomics/genotyphi
614 Operating system(s): Platform independent
615 Programming language: Python3 or higher
616 Other requirements: Mykrobe
617 License: GNU General Public License
618 Any restrictions to use by non-academics: None
619
620 Project name: Mykrobe
621 Project home page: https://github.com/Mykrobe-tools/mykrobe
622 Operating system(s): Platform independent
623 Programming language: Python3 or higher
624 Other requirements: anytree 2.8.0 or higher, Biopython 1.79 or higher, PyVCF3 1.0.3 or higher,
625 requests 2.27.1 or higher, mongoengine 0.24.1 or higher, cython 0.29.28 or higher, numpy
626 1.22.0 or higher, pytest
627 License: MIT License
628 Any restrictions to use by non-academics: None.

629 **Declarations**

630 **Ethics approval and consent to participate**

631 Each contributing study or surveillance program obtained local ethical and governance
632 approvals, as reported in the primary publication for each dataset. Inclusion of data that were
633 not yet public domain by August 2021 was approved by the Observational / Interventions
634 Research Ethics Committee of the London School of Hygiene and Tropical Medicine (ref

#26408), on the basis of details provided on the local ethical approvals for sample and data collection (listed in Carey et al, 2023, Supplementary file 1).

Consent for publication

Not applicable

Availability of data and materials

All code and instructions for genotyping isolates with GenoTyphi using Mykrobe is available in the GitHub repository at https://github.com/typhoidgenomics/genotyphi. The version reported in this paper is v2.1 (DOI: 10.5281/zenodo.13859721)

The datasets and analysis code used in this article are available in a GitHub repository at https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe.

Competing interests

MML was a co-developer of a Trivalent Salmonella (Enteritidis/Typhimurium/Typhi Vi) conjugate vaccine with Bharat Biotech International and the Wellcome Trust. MML has also received payments from Pfizer for consultancy work. MML holds US patents for “Compositions and Methods for Producing Bacterial Conjugate Vaccines”. MML was a member of a NIH DSMB that oversaw US government-funded efficacy trials of COVID-19 vaccines. DSMB was disbanded after several vaccines were given Emergency Use Authorization. MML was a member of the Vaccines and Related Biological Products Advisory Committee of the FDA. IIB was a consultant for the Weapons Threat Reduction Program, Global Affairs Canada. AJP has been involved an Oxford University partnership with AZ for development of COVI19 vaccines. AJP has received payments for consultancy work from Shionogi. AJP is chair of DHSC’s Joint Committee on Vaccination and Immunisation, is a chair of WHOs Salmonella TAG, and was a member of WHOs SAGE. AJP received support from MRNA – Moderna. KLC has received payments from Pfizer for presentations and travel support from BD. INO has received payments from the Wellcome Trust for consultancy work and receives royalties for books or book chapters published via Springer, Cornell University Press, and Oxford University Press. INO has received travel support from BMGF, ESCMID, and the American ASM and has held leadership or advisory roles for Wellcome SEDRIC, the BMGF surveillance advisory group, the Thomas

Bassir Biomedical Foundation, and International Centre for Antimicrobial Resistance Solutions (ICARS) Technical Advisory Forum. ZI has received travel support from ETH Zurich.

Funding

DJI was supported by an Emerging Leadership Fellowship from the National Health and Medical Research Council (NHMRC) of Australia (GNT1195210). ZAD received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie (grant agreement No 845681). KEH received funding from the Wellcome Trust (grant 226432/Z/22/Z). INO was supported by Official Development Assistance (ODA) funding from the NIHR (grant number #16/136/111) and a Calestous Juma Science LeadershipFellowship from the BMGF (INV-036234). AMS is supported by the SEQAFRICA project, funded by the Department of Health and Social Care's Fleming Fund using UK aid. MJS was supported by BMGF (INV-000049, OPP1161058) and National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (F30AI156973). JAC was supposed by the Bill & Melinda Gates Foundation Typhoid Vaccine Acceleration Consortium (TyVAC, INV-030857). BPH was supported by a NHMRC Leadership Fellowship (GNT1196103). KAT and this work are supported by the Centers for Disease Control and Prevention. MML was supported by BMGF (OPP1194582, INV-000049, and INV-029806). MH was supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with the UK Health Security Agency (UKHSA) (NIHR200915), and supported by the NIHR Biomedical Research Centre, Oxford. DAR is supported by the NIH (U19 AI110802-01, NIH T32AI162579). JJJ was supported by grants from Bill and Melinda Gates Foundation, USA (Investment ID INV-009497 OPP1159351) for the Project, National Surveillance System for Enteric Fever in India.

This research was funded in part by the Wellcome Trust. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission

Authors' Contributions

DJI, JH, ZAD and KEH – Conceptualization.

DJI, JK, ZAD and KEH - Data curation.

JH, MH, ZI, ZAD and KEH – Software

DJI, JH, RRW, ZAD and KEH - Methodology
DJI, JK, AOA, NA, SA, PMA, IIB, MC, MAC, JC, PDG, HI, JJ, AK, KHK, MM, SN, SO, INO, SIAR, AR, DR, VS, MJS, AMS, GTS, KT, RRW, ZAD and KEH - Formal analysis
DJI, JK, AOA, NA, SA, PMA, IIB, MC, MAC, JC, PDG, BPH, HI, JJ, LMJ, AK, KHK, JK, ML, MM, SN, SO, INO, PEO, SIAR, AR, DR, VS, MJS, AMS, GTS, KT, RRW, ZAD and KEH - Investigation,
DJI and KEH - Visualization,
DJI, ZAD and KEH - Writing – original draft,
DJI, JK, ZI, MH, JHK, AOA, NA, SA, PMA, IIB, MEC, MAC, JC, PDG, BPH, HI, JJ, LMJ, AK, KHK, JK, ML, MM, SN, SO, INO, PEO, SIAR, AR, DR, VS, MJS, AMS, GTS, KT, RRW, ZAD and KEH - Writing – review and editing
All authors read and approved the final manuscript.

Acknowledgements

MAC and SN are affiliated to the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Genomics and Enabling Data at University of Warwick in partnership with the UK Health Security Agency (UKHSA). MAC and SN are based at UKHSA. MH is affiliated to the NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University. The views expressed are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care or the UK Health Security Agency.

SIAR from the icddr,b is grateful to the Governments of Bangladesh, and Canada for providing unrestricted support.

KAT thanks the epidemiology and laboratory partners in state and local health departments. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Names of specific vendors, manufacturers, or products are included for public health and informational purposes; inclusion does not imply endorsement of the vendors, manufacturers, or products by the Centers for Disease Control and Prevention or the US Department of Health and Human Services.

AMS thanks all participants of the GERMS-SA Laboratory Surveillance Network, South Africa, for submission of clinical isolates of Salmonella species to the National Institute for Communicable Diseases; the surveillance network includes laboratories belonging to the Department of Health (NHLS laboratories) and laboratories that form part of the private sector.

We thank all the lab members involved in SEFI reference lab activities at CMC Vellore implicated in culture maintenance and ONT data generation.

Global Typhoid Genomics Consortium Group Authorship

Ali H. Abbas[1], Safina Abdul Razzak[2], Mabel K. Aworh[3], Stephen Baker[4], Buddha Basnyat[5], Ka Lip Chew[6], Thomas C. Darton[7], Andrew R. Greenhill[8], Madhu Gupta[9], Mochammad Hatta[10], Aamer Ikram[11], Dasaratha Ramaiah Jinka[12], A. C. Lauer[13], Octavie Lunguya[14], Jaspreet Mahindroo[15], Tapfumanei Mashe[16], Elli Mylona[17], Maria Pardos de la Gandara[18], Andrew J. Pollard[19], Saikt Rahman[20], Elrashdy M Redwan[21,22,23], Jean Pierre Rutanga[24], Jivan Shakya[25], Arif Tanmoy[26], Mathew S. Thomas[27], Sandra Van Puyvelde[28], Jacqueline M Wright[29]

1 Department of Microbiology, Faculty of Veterinary Medicine, University of Kufa, Kufa- Najaf Street, Najaf, Iraq

2 Department of Paediatrics and Child Health, Aga Khan University, P.O. Box 3500, National Stadium Rd, Karachi, Sindh, Pakistan.

3 North Carolina State University, 1060 William Moore Drive, Raleigh, NC 27607, USA

4 University of Cambridge, JCBC, Puddicome way, Cambridge, CB2 0AW, United Kingdom

5 OUCRU-Nepal, Jhamshikhel, Lalitpur, Nepal

6 National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074, Republic of Singapore

7 Clinical Infection Research Group, School of Medicine and Population Health, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom

8 Federation University Australia, Gippsland Campus, Churchill, Victoria, Australia

9 Postgraduate Institute of Medical Education and Research, sector 12, Chandigarh. CMC Vellore, India

10 Department of Molecular Biology and Immunology, Faculty of Medicine, Hasanuddin University, Makassar, 90245, Indonesia

11 Fauji Foundation, Rawal Road, Rawalpindi, Pakistan

12 RDT Hospital, Bathalapalli, Srisathya Sai District, 515661, India

13 Centers for Disease Control and Prevention, Atlanta, United States of America

14 Department of Microbiology, Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo

15 MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, W12 0BZ, United Kingdom

767 16 World Health Organisation, Zimbabwe
768 17 Cambridge Institute of Therapeutic Immunology and Infectious Disease (CITIID), University
769 of Cambridge, Cambridge CB1 0AW, United Kingdom
770 18 Institut Pasteur, France
771 19 Oxford Vaccine Group, Department of Paediatrics, University of Oxford and the NIHR Oxford
772 Biomedical Research Centre, Oxford, UK
773 20 RDT Hospital, Bathalapalli, Srisathya Sai District, 515661, India
774 21 Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah,
775 Saudi Arabia
776 22 Centre of Excellence in Bionanoscience Research, King Abdulaziz University, Jeddah
777 21589, Saudi Arabia
778 23 Therapeutic and Protective Proteins Laboratory, Protein Research Department, Genetic
779 Engineering and Biotechnology Research Institute, City of Scientific Research and
780 Technological Applications, New Borg EL-Arab 21934, Alexandria, Egypt.
781 24 University of Rwanda, KK 737 street, Kigali, B.P 3900, Rwanda
782 25 Institute for Research in Science and Technology, Sachal-2, Lalitpur, Nepal
783 26 Child Health Research Foundation, Dhaka, Bangladesh
784 27 CMAI, ICMDA
785 28 University of Antwerp, Antwerp, Belgium
786 29 Institute of Environmental Sciences and Research 27 Creyke Road, Christchurch, 8041, New
787 Zealand

References

1. Meiring JE, Khanam F, Basnyat B, Charles RC, Crump JA, Debellut F, et al. Typhoid fever. Nat Rev Dis Prim. 2023;9:71. https://doi.org/10.1038/s41572-023-00480-z

2. Collaborators G 2017 T and P, Stanaway JD, Reiner RC, Blacker BF, Goldberg EM, Khalil IA, et al. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect Dis. 2019;19:369–81. https://doi.org/10.1016/S1473-3099(18)30685-6

3. The WHO AWaRe (Access, Watch, Reserve) antibiotic book. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

4. Nabarro LE, McCann N, Herdman MT, Dugan C, Ladhani S, Patel D, et al. British infection association guidelines for the diagnosis and management of enteric fever in England. J Infect. 2022;84:469–89. https://doi.org/10.1016/j.jinf.2022.01.014

5. GBD 2021 Forecasting Collaborators, Vollset SE, Ababneh HS, Abate YH, Abbafati C, Abbasgholizadeh R, et al. Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403:2204–56. https://doi.org/10.1016/S0140-6736(24)00685-8

6. da Silva KE, Tanmoy AM, Pragasam AK, Iqbal J, Sajib MSI, Mutreja A, et al. The international and intercontinental spread and expansion of antimicrobial-resistant *Salmonella* Typhi: a genomic epidemiology study. Lancet Microbe. 2022;3:e567–77. https://doi.org/10.1016/S2666-5247(22)00093-3

7. Hughes MJ, Birhane MG, Dorough L, Reynolds JL, Caidi H, Tagg KA, et al. Extensively Drug-Resistant Typhoid Fever in the United States. Open Forum Infect Dis. 2021;8:ofab572. https://doi.org/10.1093/ofid/ofab572

8. Walker J, Chaguza C, Grubaugh ND, Carey M, Baker S, Khan K, et al. Assessing the global risk of typhoid outbreaks caused by extensively drug resistant *Salmonella* Typhi. Nat Commun. 2023;14:6502. https://doi.org/10.1038/s41467-023-42353-9

9. Posen HJ, Wong W, Farrar DS, Campigotto A, Chan T, Barker KR, et al. Travel-associated extensively drug-resistant typhoid fever: a case series to inform management in non-endemic regions. J Travel Med. 2022;30:taac086. https://doi.org/10.1093/jtm/taac086

10. Wang Y, Lu D, Jin Y, Wang H, Lyu B, Zhang X, et al. Extensively Drug-Resistant (XDR) *Salmonella* Typhi Outbreak by Waterborne Infection — Beijing Municipality, China, China CDC Wkly. 2022;4:254–8. https://doi.org/10.46234/ccdcw2022.062

11. Nizamuddin S, Khan EA, Chattaway MA, Godbole G. Case of Carbapenem-Resistant *Salmonella* Typhi Infection, Pakistan, 2022. Emerg Infect Dis. 2023;29:2395–7. https://doi.org/10.3201/eid2911.230499

12. Dyson ZA, Klemm EJ, Palmer S, Dougan G. Antibiotic Resistance and Typhoid. Clin Infect Dis. 2019;68:S165–70. https://doi.org/10.1093/cid/ciy1111

13. Argimón S, Yeats CA, Goater RJ, Abudahab K, Taylor B, Underwood A, et al. A global resource for genomic predictions of antimicrobial resistance and surveillance of *Salmonella* Typhi at pathogenwatch. Nat Commun. 2021;12:2879. https://doi.org/10.1038/s41467-021-23091-2

14. Carey ME, Dyson ZA, Ingle DJ, Amir A, Aworh MK, Chattaway MA, et al. Global diversity and antimicrobial resistance of typhoid fever pathogens: Insights from a meta-analysis of 13,000 *Salmonella* Typhi genomes. eLife. 2023;12:e85867. https://doi.org/10.7554/eLife.85867

15. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. Nat Genet. 2015;47:632–9. https://doi.org/10.1038/ng.3281

16. Hooda Y, Sajib MSI, Rahman H, Luby SP, Bondy-Denomy J, Santosham M, et al. Molecular mechanism of azithromycin resistance among typhoidal *Salmonella* strains in Bangladesh identified through passive pediatric surveillance. Plos Neglect Trop D. 2019;13:e0007868. https://doi.org/10.1371/journal.pntd.0007868
17. Day MR, Doumith M, Nascimento VD, Nair S, Ashton PM, Jenkins C, et al. Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Salmonella enterica* serovars Typhi and Paratyphi. J Antimicrob Chemoth. 2017;73:365–72. https://doi.org/10.1093/jac/dkx379
18. Wong VK, Baker S, Connor TR, Pickard D, Page AJ, Dave J, et al. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. Nat Commun. 2016;7:12827. https://doi.org/10.1038/ncomms12827
19. Ingle DJ, Andersson P, Valcanis M, Wilmot M, Easton M, Lane C, et al. Genomic Epidemiology and Antimicrobial Resistance Mechanisms of Imported Typhoid in Australia. Antimicrob Agents Ch. 2021;65:e01200-21. https://doi.org/10.1128/aac.01200-21
20. Ikhimiukor OO, Oaikhena AO, Afolayan AO, Fadeyi A, Kehinde A, Ogunleye VO, et al. Genomic characterization of invasive typhoidal and non-typhoidal *Salmonella* in southwestern Nigeria. PLoS Neglected Trop Dis. 2022;16:e0010716. https://doi.org/10.1371/journal.pntd.0010716
21. Sikorski MJ, Hazen TH, Desai SN, Nimarota-Brown S, Tupua S, Sialeipata M, et al. Persistence of Rare *Salmonella* Typhi Genotypes Susceptible to First-Line Antibiotics in the Remote Islands of Samoa. Mbio. 2022;13:e01920-22. https://doi.org/10.1128/mbio.01920-22
22. Dyson ZA, Holt KE. Five years of GenoTyphi: updates to the global *Salmonella* Typhi genotyping framework. J Infect Dis. 2021;224:jiab414. https://doi.org/10.1093/infdis/jiab414
23. Britto CD, Dyson ZA, Duchene S, Carter MJ, Gurung M, Kelly DF, et al. Laboratory and molecular surveillance of paediatric typhoidal *Salmonella* in Nepal: Antimicrobial resistance and implications for vaccine policy. Plos Neglect Trop D. 2018;12:e0006408. https://doi.org/10.1371/journal.pntd.0006408
24. Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE, Chowdhury EK, et al. Population structure and antimicrobial resistance patterns of *Salmonella* Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. Plos Neglect Trop D. 2020;14:e0008036. https://doi.org/10.1371/journal.pntd.0008036
25. Tanmoy AM, Westeel E, Bruyne KD, Goris J, Rajoharison A, Sajib MSI, et al. *Salmonella enterica* Serovar Typhi in Bangladesh: Exploration of Genomic Diversity and Antimicrobial Resistance. mBio. 2018;9:e02112-18. https://doi.org/10.1128/mbio.02112-18
26. Klemm EJ, Shakoor S, Page AJ, Qamar FN, Judge K, Saeed DK, et al. Emergence of an Extensively Drug-Resistant *Salmonella enterica* Serovar Typhi Clone Harboring a Promiscuous Plasmid Encoding Resistance to Fluoroquinolones and Third-Generation Cephalosporins. mBio. 2018;9:e00105-18. https://doi.org/10.1128/mbio.00105-18
27. Kariuki S, Dyson ZA, Mbae C, Ngetich R, Kavai SM, Wairimu C, et al. Multiple introductions of multidrug-resistant typhoid associated with acute infection and asymptomatic carriage, Kenya. eLife. 2021;10:e67852. https://doi.org/10.7554/eLife.67852
28. Ingle DJ, Hawkey J, Hunt M, Dyson ZA, Holt KE. Technical Report: GenoTyphi Implementation in Mykrobe. Zenodo. 2022. Available from: https://zenodo.org/doi/10.5281/zenodo.7407984
29. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature. 2001;413:848–52. https://doi.org/10.1038/35101607
30. Labbé G, Kruczkiewicz P, Robertson J, Mabon P, Schonfeld J, Kein D, et al. Rapid and accurate SNP genotyping of clonal bacterial pathogens with BioHansel. Microb Genom. 2021;7:000651. https://doi.org/10.1099/mgen.0.000651
31. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, et al. Antibiotic resistance

prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. Wellcome Open Res. 2019;4:191. https://doi.org/10.12688/wellcomeopenres.15603.1

32. Baker S, Holt K, Whitehead S, Goodhead I, Perkins T, Stocker B, et al. A linear plasmid truncation induces unidirectional flagellar phase change in H:z66 positive *Salmonella* Typhi. Mol Microbiol. 2007;66:1207–18. https://doi.org/10.1111/j.1365-2958.2007.05995.x

33. Holt KE, Phan MD, Baker S, Duy PT, Nga TVT, Nair S, et al. Emergence of a Globally Dominant IncHI1 Plasmid Type Associated with Multiple Drug Resistant Typhoid. Plos Neglect Trop D. 2011;5:e1245. https://doi.org/10.1371/journal.pntd.0001245

34. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and *Mycobacterium tuberculosis*. Nat Commun. 2015;6:10063. https://doi.org/10.1038/ncomms10063

35. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, et al. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. Lancet Microbe. 2022;3:e265–73. https://doi.org/10.1016/S2666-5247(21)00301-3

36. Hall MB, Coin LJM. Assessment of the 2021 WHO *Mycobacterium tuberculosis* drug resistance mutation catalogue on an independent dataset. Lancet Microbe. 2022;3:e645. https://doi.org/10.1016/S2666-5247(22)00151-3

37. Hawkey J, Paranagama K, Baker KS, Bengtsson RJ, Weill F-X, Thomson NR, et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, *Shigella sonnei*. Nat Commun. 2021;12:2684. https://doi.org/10.1038/s41467-021-22700-4

38. Thanh DP, Karkey A, Dongol S, Thi NH, Thompson CN, Rabaa MA, et al. A novel ciprofloxacin-resistant subclade of H58 *Salmonella* Typhi is associated with fluoroquinolone treatment failure. Elife. 2016;5:e14003. https://doi.org/10.7554/eLife.14003

39. Crump JA, Kretsinger K, Gay K, Hoekstra RM, Vugia DJ, Hurd S, et al. Clinical Response and Outcome of Infection with *Salmonella enterica* Serotype Typhi with Decreased Susceptibility to Fluoroquinolones: a United States FoodNet Multicenter Retrospective Cohort Study. Antimicrob Agents Chemother. 2008;52:1278–84. https://doi.org/10.1128/aac.01509-07

40. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021

41. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinform. 2020;70:e102. https://doi.org/10.1002/cpbi.102

42. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4:1686. https://doi.org/10.21105/joss.01686

43. Chattaway MA, Gentle A, Nair S, Tingley L, Day M, Mohamed I, et al. Phylogenomics and antimicrobial resistance of *Salmonella* Typhi and Paratyphi A, B and C in England, 2016–2019. Microb Genom. 2021;7. https://doi.org/10.1099/mgen.0.000633

44. CDC. National Antimicrobial Resistance Monitoring System (NARMS) Now: Human Data. Atlanta, Georgia: U.S. Department of Health and Human Services, CDC [Internet]. Available from: https://www.cdc.gov/narmsnow

45. Reller LB, Weinstein M, Jorgensen JH, Ferraro MJ. Antimicrobial Susceptibility Testing: A Review of General Principles and Contemporary Practices. Clin Infect Dis. 2009;49:1749–55. https://doi.org/10.1086/647952

46. U.S Food and Drug Administration. Antimicrobial Susceptibility Test (AST) Systems - Class II Special Controls Guidance for Industry and FDA. [Internet]. Available https://www.fda.gov/media/88069/download.

47. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the

Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2022;51:D690–9. https://doi.org/10.1093/nar/gkac920

48. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6. https://doi.org/10.1038/s41587-019-0072-8

49. Greig DR, Bird MT, Chattaway MA, Langridge GC, Waters EV, Ribeca P, et al. Characterization of a P1-bacteriophage-like plasmid (phage-plasmid) harbouring bla CTX-M-15 in *Salmonella enterica* serovar Typhi. Microb Genom. 2022;8:mgen000913. https://doi.org/10.1099/mgen.0.000913

50. Nair S, Barker CR, Bird M, Greig DR, Collins C, Painset A, et al. Presence of phage-plasmids in multiple serovars of *Salmonella enterica*. Microb Genom. 2024;10:001247. https://doi.org/10.1099/mgen.0.001247

51. Goldblatt J, Ward A, Yusuf M, Day M, Godbole G, Morris-Jones S. Azithromycin susceptibility testing for *Salmonella enterica* isolates: discordances in results using MIC gradient strips. J Antimicrob Chemother. 2020;75:1820–3. https://doi.org/10.1093/jac/dkaa097

52. Jabeen K, Saleem S, Nizamuddin S, Arshad F, Jahan S, Hasnain F, et al. Reporting of Azithromycin Activity against Clinical Isolates of Extensively Drug-Resistant *Salmonella enterica* Serovar Typhi. Am J Trop Med Hyg. 2023;108:942–7. https://doi.org/10.4269/ajtmh.22-0557

53. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect. 2017;23:2–22. https://doi.org/10.1016/j.cmi.2016.11.012

54. Wick R. Yet another ONT accuracy test: Dorado v0.5.0 2023. Zenodo. Available from: https://zenodo.org/records/10397818

965 **Tables and Figures**

966

967 **Table 1. Validation of genotype calls**

968 Summary of Typhi Mykrobe genotype calls (GenoTyphi scheme) for 12,839 Illumina readsets,
969 compared with mapping-based genotype calls.

Typhi Mykrobe vs mapping-based genotype calls	**Genomes (N)**
invA-negative, thus designated non-Typhi and not genotyped by Mykrobe	5
Single genotype, exact match	12,807
Two genotypes, exact match (reported differently)	17
Genotype 3.5.3, mismatch (reported incorrectly by mapping pipeline as 3.5.4 due to breaking hierarchy assumption)	8
Discordant result (low support reported by mapping-based pipeline, phylogenetic tree supports Mykrobe call)	2

970

971 **Table 2. Validation of AMR genotyping (vs assembly based)**

972 Comparison of AMR genotyping by Typhi Mykrobe (kmer-based analysis of reads) vs
973 Pathogenwatch (based on nucleotide blast search of assemblies), for n=11,992 isolates.
974 Complete details of the call comparisons are given in **Supplementary Table 3**. Categorical
975 agreement at drug level is calculated based on agreement of the binary S/R categorisation of
976 the 11,992 isolates by each method (i.e. 'Concordant' means Typhi Mykrobe and Typhi
977 Pathogenwatch made the same call of S or R based on absence or presence, respectively, of
978 known genetic determinants; 'Discordant' means one method identified determinants and called
979 R, while the other did not and called S). *For ciprofloxacin, both tools categorise as S/I/R as
980 defined in Methods, and categorical agreement was calculated across these three categories.
981 **Results for *dfrA* genes are summarised together, and as some genomes have multiple *dfrA*
982 alleles the total observations is greater than the number of isolates (see details in
983 **Supplementary Table 3**).

Drug	Agreement (drug level)	Genetic marker	Concordant	Discordant	Agreement (marker level)
Ampicillin	99.72%	*bla*TEM-1	11,958	34	99.72%
		*bla*OXA-7	11,991	1	99.99%
Azithromycin	99.98%	*acrB*_717	11,991	1	99.99%
		mphA	11,992	0	100.00%
Ceftriaxone	99.97%	*bla*CTX-M	11,990	3	99.97%
		*bla*SHV-12	11,991	1	99.99%
		*bla*OXA-134	11,992	0	100.00%
		ampC	11,992	0	100.00%
Chloramphenicol	99.89%	*catA1*	11,979	13	99.89%
		cmlA1	11,992	0	100.00%
Ciprofloxacin*	99.72%	*qnrB*	11,979	13	99.89%
		qnrD	11,992	0	100.00%
		qnrS	11,987	5	99.96%
		*gyrA*_83	11,986	6	99.95%
		*gyrA*_87	11,986	10	99.92%
		*gyrB*_464	11,493	8	99.93%
		*parC*_80	11,976	17	99.86%
		*parC*_84	11,992	0	100.00%
Sulfonamides	99.81%	*sul1*	11,971	21	99.82%
		sul2	11,819	173	98.56%
Trimethoprim	99.83%	*dfrA* genes**	12,049	22	99.82%
Tetracycline	99.85%	*tetA(A)*	11,986	6	99.95%
		tetA(B)	11,980	12	99.90%
		tetA(C)	11,970	0	100.00%
		tetA(D)	11,970	0	100.00%
-	-	**Total**	**299,004**	**346**	**99.88%**

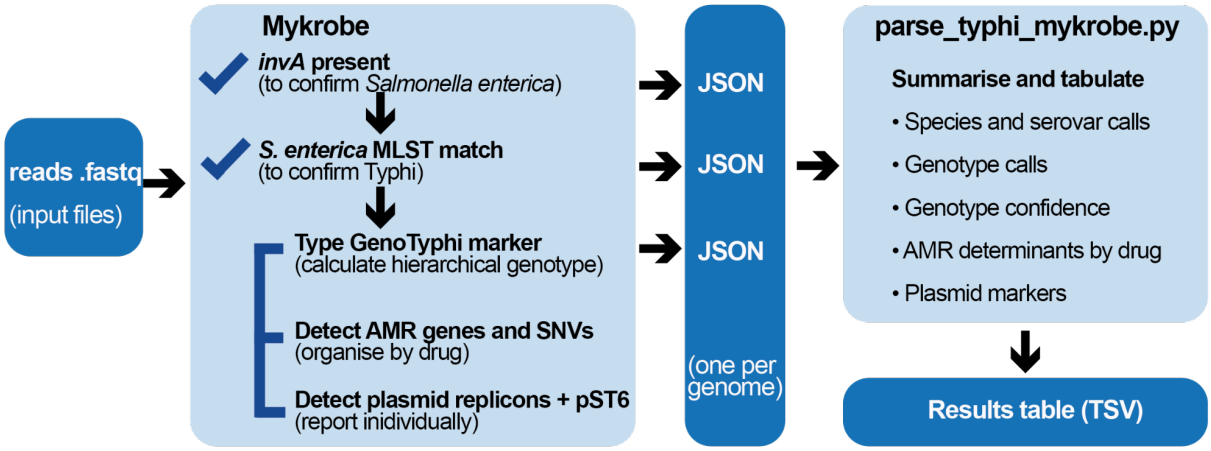
984

985 **Table 3 - Validation of plasmid replicon detection (vs assembly based).**
986 Comparison of plasmid replicon markers identified in the genomes by Typhi Mykrobe (kmer-
987 based analysis of reads) vs Pathogenwatch (PW, based on nucleotide blast search of
988 assemblies), for n=11,992 isolates.

Plasmid replicon marker	**Both**	**Mykrobe only**	**PW only**	**Agreement**
IncFIAHI1	251	0	0	100%
IncFIB(K)	175	0	1	99.99%
IncFIB(pHCM2)	1406	26	5	99.74%
IncHI1A	872	9	3	99.90%
IncHI1B	874	7	1	99.93%
IncHI2A	3	0	0	100.00%
IncI1	5	0	0	100.00%
IncL/M	1	0	0	100.00%
IncN	100	2	0	99.98%
IncX1				
IncX3	21	0	0	100.00%
IncY	666	0	0	100.00%
p0111				

989
990

A. Typhi Mykrobe pipeline



B. AMR determinants included in Typhi Mykrobe

Drug	Acquired genes	Point mutations
ampicillin	*blaTEM-1, OXA-7*	
azithromycin	*ereA, ermB, mphA*	*acrB-717*
carbapenems	*blaKPC-2, blaNDM-5, blaVIM-1, blaIMP-27, blaOXA-48*	
ceftriaxone	*ampC1, CTX-M-15, OXA-134, SHV-12*	
chloramphenicol	*catA1, cmlA1*	
ciprofloxacin	*qnrB1, qnrD1, qnrS1*	*gyrA-83, gyrA-87, parC-80, parC-84*
sulfonamides	*sul1, sul2*	
trimethoprim	*dfrA1, dfrA5, dfrA7, dfrA14, dfrA15, dfrA17, dfrA18*	
tetracycline	*tetA(A), tetA(B), tetA(C), tetA(D)*	

C. Plasmids included in Typhi Mykrobe

Group	Plasmid probe target
AMR-linked replicons	IncFIBK, IncHI2A, IncI1, IncL/M, IncN, IncX1, IncX3, IncY
IncHI1 targets	IncHI1A, IncHI1BR27, IncFIAHI1 (rep markers); C19241A (marker of PST6 lineage)
Non-AMR plasmids	IncFIB_pHCM2 (cryptic), pBSSB1 (z66 flagellin), pO111

991
992
993 **Figure 1: Typhi Mykrobe functionality**

994 A) A schematic overview of the Typhi Mykrobe pipeline. B) Table of epidemiologically important
995 AMR determinants targeted for detection/typing by Typhi Mykrobe, grouped by drug. C) Table of
996 epidemiologically important plasmid replicons targeted for detection by Typhi Mykrobe.

997
998

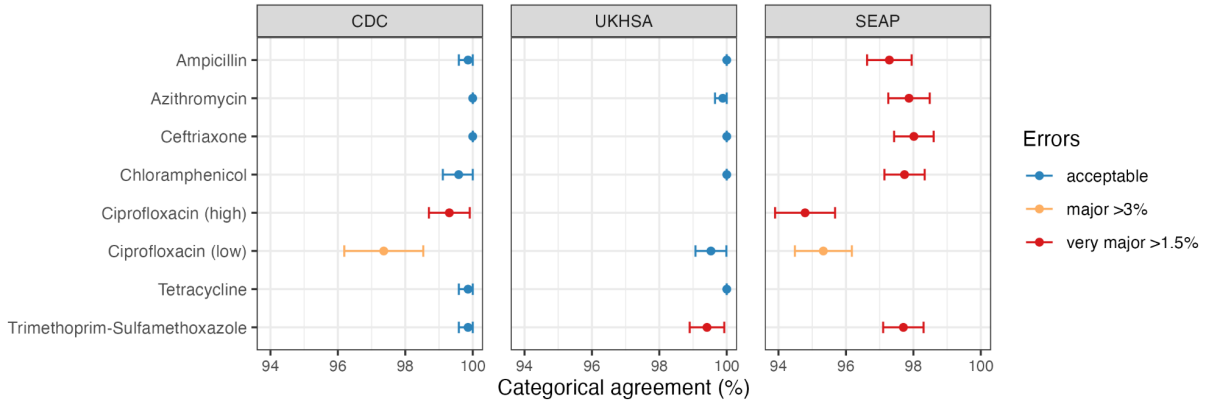


Figure 2: Validation of Typhi Mykrobe AMR predictions vs AMR phenotypes

Summary of Typhi Mykrobe's genotype-based clinical resistance categorisations compared with susceptibility phenotypes, for 4,018 isolates with publicly available matched genome and phenotype data. Comparisons are summarised as categorical agreement, and calculated separately for the three different source datasets, which used different phenotyping methods and interpretive standards (US CDC: MIC, CLSI; UKHSA: MIC, EUCAST; SEAP: disk diffusion, EUCAST; see **Methods**). Major errors were defined as susceptible isolates with AMR determinants detected (reported as R by Mykrobe). Very major errors were defined as resistant isolates with no AMR determinants detected (reported as S by Mykrobe). Error bars are shown that represent 95% confidence interval of proportion. Note UKHSA tested amoxicillin rather than ampicillin, the result is reported in the ampicillin row.

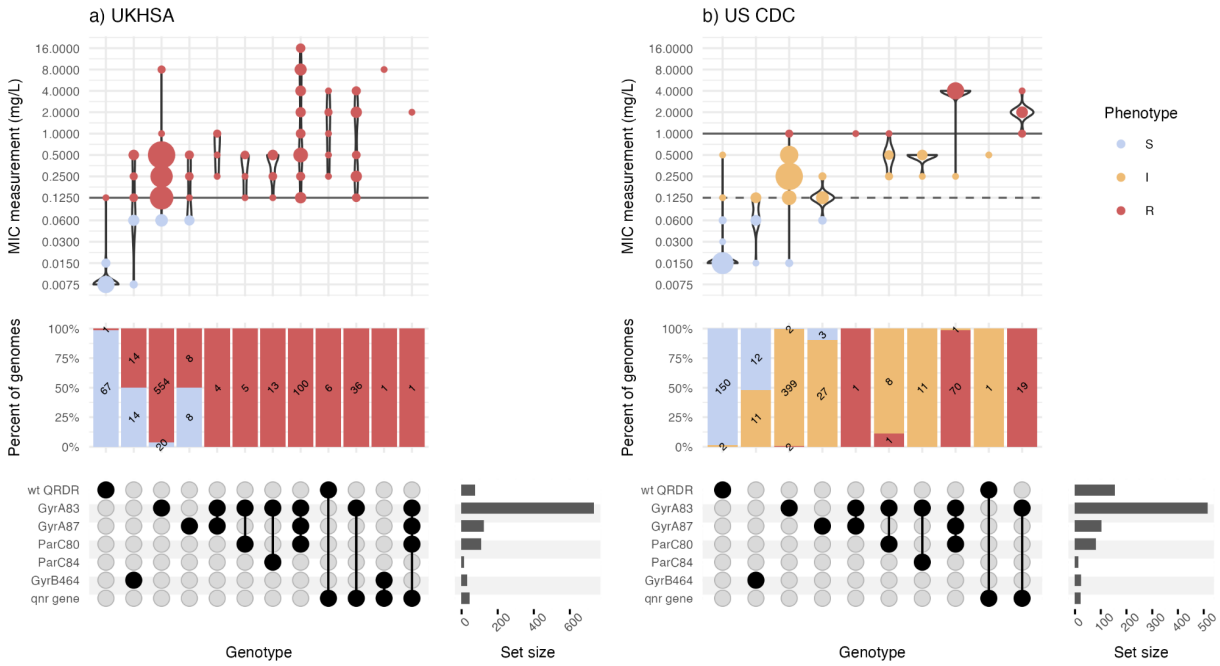


Figure 3: Typhi Mykrobe genotypes vs minimum inhibitory concentration from two reference laboratories, for ciprofloxacin.

a) UKHSA dataset, using EUCAST. **b)** US CDC dataset, using CLSI. Each column represents the set of isolates in which Typhi Mykrobe identified a unique combination of genetic determinants, indicated in the panel at the bottom. For each column, the violin plots show the distribution of ciprofloxacin MIC values, and stacked barplots show the proportion of genomes called as S, I, or R (coloured as per inset legend, and labelled with counts). The solid horizontal lines on the violin plots mark the R breakpoint, the dashed line marks the CLSI I breakpoint used for the US CDC dataset.

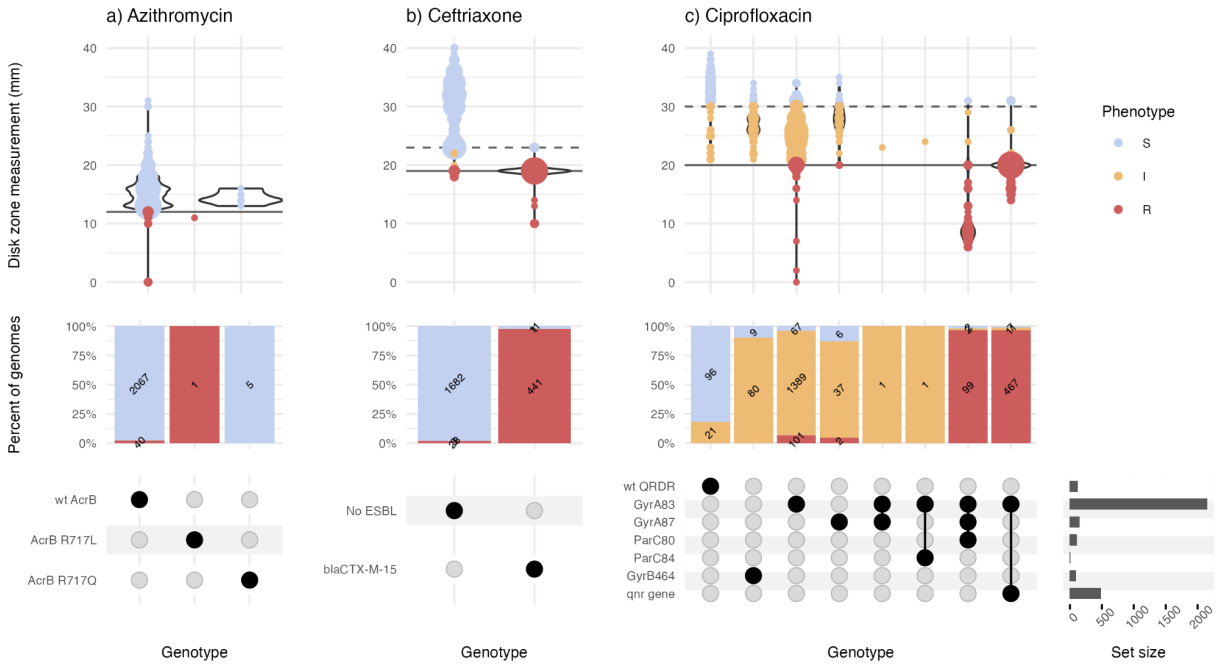


Figure 4: Typhi Mykrobe genotypes vs disk diffusion data from the SEAP study, for three drug classes

Data for three drugs are shown in panels **a)** Azithromycin, **b)** Ceftriaxone, and **c)** Ciprofloxacin. Each column represents the set of isolates in which Typhi Mykrobe identified a unique combination of genetic determinants, indicated in the panel at the bottom. For each column, the violin plots show the distribution of disk diffusion values for azithromycin, ceftriaxonne or ciprofloxacin. The stacked barplots show the proportion of genomes called as S, I, or R (coloured as per inset legend, and labelled with counts). The solid horizontal lines on the violin plots mark the R EUCAST breakpoint used for the SEAP disk diffusion data. The dashed line marks the EUCAST breakpoint for I.