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[2,3], Martin Hunt[4,5,6,7], Zamin Iqbal[4,8], Jacqueline A. Keane[9] Ayorinde O. Afolayan[10,11], Niyaz Ahmed[12], Saadia Andleeb[13], Philip M. Ashton[14,15], Isaac I. Bogoch[16], Megan E. Carey[17,18], Marie Anne Chattaway[19], John A. Crump[20], Paula Diaz Guevara[21], Benjamin P. Howden[1,22,23], Hidemasa Izumiya[24], Jobin John Jacob[25], Louise M. Judd[1,23], Arti Kapil[26], Karen H. Keddy[27], Justin Y. Kim[28], Myron M. Levine[29], Masatomo Morita[24], Satheesh Nair[19], Sophie Octavia[30], Iruka N. Okeke[10], Precious E. Osadebamwen[10], Sadia Isfat Ara Rahman[31], Assaf Rokney[32], David A. Rasko[29,33,34], Varun Shamanna[35], Michael J. Sikorski[29,33,], Anthony M. Smith[36,37], Gabriel T. Sunmonu[10], Kaitlin A. Tagg[28], Ryan R. Wick[1], Zoe A. Dyson[17,38]*, Kathryn E. Holt[2,17]* 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 Centre to Impact AMR, Monash University, Clayton, Victoria, Australia

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

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

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

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

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

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

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

34 University of Maryland School of Dentistry, Baltimore, Maryland, United States of America
35 Central Research Laboratory, KIMS, Bengaluru, India
36 Centre for Enteric Diseases, National Institute for Communicable Diseases, Johannesburg, South Africa
37 Department of Medical Microbiology, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa
38 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). 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 for isolates sequenced on both platforms (n =93 genomes). 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, although lower accuracy was observed for R9 ONT data. 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 [13,14,19–21].

151 The GenoTyphi scheme was originally defined based on nearly 2,000 WGS isolates from over
152 60 countries [18]. It uses a hierarchical framework based on the global phylogeny, stratifying the
153 Typhi population structure into four primary lineages, 16 clades, and 49 subclades (henceforth
154 referred to as genotypes), based on lineage-specific marker single nucleotide variants (SNVs).
155 A schematic mapping the hierarchical genotype labels to the Typhi phylogeny is shown in
156 **Supplementary Figure 1**. Under the GenoTyphi scheme, the lineage formerly known as 'H58'
157 was designated genotype 4.3.1, and is further divided into multiple higher-resolution genotypes
158 with the prefix 4.3.1 including 4.3.1.1, 4.3.1.2 (see **Supplementary Figure 1**). The H58 lineage,
159 or genotype 4.3.1 and associated sub-genotypes, has been associated with the spread of
160 multidrug resistance (MDR, resistant to amoxicillin/ampicillin, chloramphenicol, trimethoprim-
161 sulfamethoxazole) Typhi throughout Asia and into Eastern Africa [15]. The scheme was later
162 extended to provide further discriminatory power for identifying additional subpopulations of
163 epidemiological importance [21–27], such as the XDR Typhi lineage associated with outbreaks
164 in Pakistan (genotype 4.3.1.1.P1) [26,27]. The scheme currently includes 87 genotypes [28],
165 and its future maintenance and development will be managed by a working group within the
166 Global Typhoid Genomics Consortium (https://www.typhoidgenomics.org/).
167
168 The GenoTyphi scheme specification consists of a set of marker SNVs defining the hierarchical
169 genotypes, which in principle can be used by any bioinformatics tool to assign genotypes to new
170 genomes. The original GenoTyphi paper [18] was accompanied by a Python-based pipeline
171 (available at: https://github.com/typhoidgenomics/genotyphi) designed for typing Illumina WGS
172 data against the scheme. This implementation requires users to first map short reads or
173 assemblies to the reference sequence of Typhi strain CT18 (accession AL513382.1) [29], then
174 provide as input the resulting binary alignment map (BAM) or variant call file (VCF). Marker
175 SNVs were identified from these input files and used to assign genomes to genotypes.
176 Functionality was extended to include detection of mutations associated with reduced
177 susceptibility to ciprofloxacin or azithromycin [22]. This mapping-based approach had various
178 limitations, largely due to the need for users to pre-map their data to a specific reference, which
179 was cumbersome (e.g. there is wide variation in the upstream mapping tools used and some
180 users had issues with mapping tools creating incompatible output files, the BAM files are large
181 and inefficient to process), and AMR analysis was restricted to mutation detection. The
182 GenoTyphi scheme has also been implemented in other genotyping tools, including Typhi
183 PathogenWatch [13], which works by analysing assembled genomes, and the split-*k*-mer-based
184 BioHansel tool, which does not include AMR detection [30].

185

186 Here, we present a novel approach to rapid genotyping of both AMR and GenoTyphi lineages
187 from reads, implemented in the Mykrobe open-source software framework [31], dubbed Typhi
188 Mykrobe. Input into this new tool accepts raw WGS reads (FASTQ files), and has been tested
189 with reads generated by ONT and Illumina platforms. The tool performs rapid (i) lineage
190 genotyping using the GenoTyphi framework; (ii) detection of a broad range of clinically relevant
191 molecular determinants of AMR, including both acquired genes and mutations; and (iii)
192 detection of plasmid replicon markers common to Typhi. We validated GenoTyphi lineage
193 assignment from reads by comparison with the original mapping-based implementation, using
194 12,839 Typhi genomes [14]. We assessed AMR genotype calls by comparison with assembly-
195 based genotype calls from PathogenWatch, and with antimicrobial susceptibility phenotypes for
196 4,018 isolates. Finally, we assessed performance on long read data generated using Oxford
197 Nanopore Technologies (ONT) MinION devices, using isolates with matched ONT and Illumina
198 data.

199 Implementation

200 As outlined in the Introduction, GenoTyphi is a typing scheme specific to Typhi, which specifies
201 marker SNVs suitable to identify specific lineages. In principle GenoTyphi can be used with any
202 genotyping tool that is capable of searching for the marker SNVs in sequence data. GenoTyphi
203 provides the framework directing genotyping tools what SNVs to look for and how to interpret
204 them in terms of lineages to report. Mykrobe is a genotyping tool, which uses a *k*-mer-based
205 approach to identify marker SNVs in sequence reads. It was developed for typing
206 *Mycobacterium tuberculosis* and *Staphylococcus aureus* [32–34], but can in principle be used
207 with any genotyping scheme. Here we present a ready-to-run code base that allows users to
208 type genomes from raw reads, using Mykrobe, against the GenoTyphi framework to identify and
209 report known lineages – which we label “Typhi Mykrobe”. This tool additionally includes AMR
210 and plasmid typing frameworks, and generates predictions of resistance to the drugs
211 recommended for clinical treatment of typhoid. Details of this Implementation are described
212 below.
213

214 Genotyping targets

215 The current scope of the GenoTyphi scheme as discussed here (n=87 genotypes) is illustrated
216 in **Supplementary Figure 1**. This includes five new genotypes (3.5.4.1, 3.5.4.2, 3.5.4.3,
217 4.3.1.2.1, 4.3.1.2.1.1) recently described in a technical report [28]. The full scheme specification

218 is available in the GenoTyphi repository (https://github.com/typhoidgenomics/genotyphi) in the
219 file Genotype_specification.csv.
220
221 The AMR and plasmid targets included in Typhi Mykrobe are summarised in **Figure 1**. These
222 include sequence probes targeted to detect presence/absence of epidemiologically relevant
223 mobile AMR determinants [13,14], conferring resistance to the currently recommended drugs
224 ciprofloxacin, ceftriaxone, and azithromycin, as well as older drugs ampicillin, chloramphenicol,
225 trimethoprim-sulfamethoxazole, and tetracycline, as well as carbapenems. Fourteen SNVs
226 associated with resistance to ciprofloxacin and azithromycin are also included. In addition, Typhi
227 Mykrobe includes probes to detect the presence or absence of 13 plasmid replicons previously
228 reported in Typhi [14], including 11 associated with AMR plasmids, the cryptic plasmid pHCM2,
229 and linear plasmid pBSSB1, which carries the alternative flagellin z66 [35]. A marker SNV to
230 identify the plasmid sequence type 6 (PST6) lineage of the dominant IncHI1 plasmid is also
231 included [36]. The full list of AMR targets is available in the GenoTyphi repository
232 (https://github.com/typhoidgenomics/genotyphi) in the file AMR_genes_mutations_plasmids.csv.

233 **Implementation in Mykrobe**

234 The Mykrobe software, first developed in 2015 for *Mycobacterium tuberculosis* and
235 *Staphylococcus aureus* [32], uses a *k*-mer-based approach to identify marker SNVs and the
236 presence or absence of target genes using 'probe' sequences to screen against [31]. These
237 data are then used to assign hierarchical genotypes and report AMR determinants. Probe sets
238 have been previously developed and implemented for other bacterial pathogens including
239 *Mycobacterium tuberculosis* [32–34], *Staphylococcus aureus* [31], *Shigella sonnei* [37], and
240 *Salmonella* Paratyphi B [38]. Mykrobe is available at https://github.com/Mykrobe-tools/mykrobe.
241
242 Mykrobe probe sets were created for genotype and AMR marker SNVs using the *mykrobe*
243 *variants make-probes* command using a *k*-mer size of 21 and the Typhi CT18 reference
244 genome (accession: AL513382.1). The nested hierarchical relationships of the marker SNVs
245 were specified and Mykrobe used this information to identify the best supported genotype call
246 (e.g., a call of '4.3.1.1' is typically supported by detection of nested markers that demarcate
247 different levels of the hierarchy: 4, 4.3.1, 4.3.1.1). Briefly, probe sets for gene presence or
248 absence were created using a Python script that extracted the relevant AMR sequences. Details
249 of the specific commands and Python scripts are provided in
250 https://github.com/typhoidgenomics/genotyphi. Probe sets were also created to confirm input

251 read sets as *S. enterica* using the presence of marker gene *invA*, involved in invasion of
252 intestinal epithelium cells by *S. enterica*, and serovar Typhi based on the 26 known multi-locus
253 sequence types (STs) (see **Figure 1**). Further details of the preliminary testing and
254 implementation of the Typhi probe sets are available in the technical report
255 https://zenodo.org/doi/10.5281/zenodo.7407984. The Typhi Mykrobe probe sets and typing
256 panel are available at https://doi.org/10.6084/m9.figshare.27102334.

257 **Summarising Mykrobe outputs**

258 To support interpretation of the Mykrobe output, we developed a Python-based parser script
259 *parse_typhi_mykrobe.py,* which takes as input a set of JavaScript Object Notation (JSON) files
260 output by Mykrobe, and summarises them in a tab-delimited table with one row per genome,
261 and columns to indicate the genotype calls (see **Figure 1**). Also included are several quality-
262 control fields reported by Mykrobe, including read-level support for each marker SNV
263 contributing to the final genotype call ('node support'), and for any additional markers detected.
264 This information is helpful for identifying contaminated samples (see examples in Results). The
265 parser script also assigns a summary 'confidence' level in the final GenoTyphi lineage call,
266 based on the Mykrobe quality scores for each nested component marker that makes up the final
267 genotype call. In brief, a weak' confidence call is assigned when one or more marker SNVs are
268 of low quality, defined as Mykrobe quality score of 0 or Mykrobe quality score of 0.5 with
269 minority support (i.e., <50% of reads) for the derived allele. A 'moderate' confidence call is
270 assigned if one (and only one) marker SNV has a Mykrobe quality score of 0.5 but with majority
271 support (i.e. ≥50% of reads) for the derived allele. Otherwise, the genotype call is assigned
272 'strong' confidence.
273
274 The parser script summarises the AMR markers detected by Mykrobe in the form of a pseudo-
275 antibiogram with one column per drug. Each drug column contains either 'S' (indicating no
276 markers detected) or 'R' (detected markers). For ciprofloxacin, it is useful to distinguish between
277 low-level resistance (MIC 0.06-0.5 mg/L), which is associated with a single mutation or acquired
278 gene and high-level resistance (MIC >0.5 mg/L), which is associated with combinations of two
279 or more determinants [14,17], because while both groups are associated with increased fever
280 clearance times and probability of treatment failure compared with susceptible strains, high-level
281 resistance is associated with worse outcomes than low-level resistance [39]. Current guidelines
282 and terminology for these groups vary, but for simplicity, the Mykrobe parser reports these
283 groups as 'I' (for intermediate) and 'R', respectively. This matches the current CLSI guidelines

284 (https://clsi.org/standards/products/microbiology/documents/m100/). However, it should be
285 noted that (i) EUCAST defines the clinical breakpoint for 'R' in *Salmonella* spp. as MIC >0.06
286 mg/L (https://www.eucast.org/clinical_breakpoints), which includes both low-level and high-level
287 resistance; and (ii) previous literature has referred to the low-level range as 'decreased
288 ciprofloxacin susceptibility' rather than 'I' [40]. In addition to this antibiogram-style summary,
289 individual AMR and plasmid markers are reported in their own columns, coded as 1=present
290 and 0=absent (as seen in **Supplementary Table 1,** available at
291 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe).
292
293 All code and instructions for genotyping isolates with GenoTyphi using Typhi Mykrobe are freely
294 available in the GitHub repository at https://github.com/typhoidgenomics/genotyphi (DOI:
295 10.5281/zenodo.13859721).

296 **Validation data**

297 Validation on Illumina short-read data was conducted using the Global Typhoid Genomics
298 Consortium dataset described in Carey et al [14] . Reads were mapped to the CT18 reference
299 genome (accession: AL513382) [29] using bwa-mem v0.7.17 via the Centre for Genomic
300 Pathogen Surveillance (CGPS) mapping pipeline v1.2.2
301 (https://gitlab.com/cgps/ghru/pipelines/snp_phylogeny), to generate alignments in BAM format.
302 Assemblies were generated using the CGPS assembly pipeline v2.1.0
303 (https://gitlab.com/cgps/ghru/pipelines/dsl2/pipelines/assembly/) which utilises the SPAdes
304 assembler (v3.12.0) [41,42]. A total of 141 isolates were excluded due to excessive unclustered
305 heterozygous SNVs indicative of mixed cultures (n=92), assembly issues (n=46) or both (n=3),
306 resulting in 12,839 isolates for Mykrobe validation (listed in **Supplementary Table 2,** available
307 at https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). The genomes of
308 isolates called as 'not Typhi' were characterised for presence of the marker genes of *invA* and
309 the seven MLST genes using Blast in Bandage for *invA* [43] and MLST with the 'senterica'
310 database (https://github.com/tseemann/mlst). The dataset covered 85 of the 87 defined
311 genotypes (exceptions being genotypes 1 and 2.3, which represent internal nodes of the
312 hierarchy for which no genomes have been observed). Individual calls are given in
313 **Supplementary Table 1** and **Supplementary Table 2.**
314

315 **Validation of genotype calls**

316 Typhi Mykrobe v0.12.1 was run on all (n=12,839) Illumina FASTQ files using the new Typhi
317 panel (v20221208) and the results were summarised using *parse_typhi_mykrobe.py*. To
318 validate Typhi Mykrobe's GenoTyphi genotype calls, we compared them with those called by
319 the original mapping-based approach. To do this, we used BAM file read alignments generated
320 from the same n=12,839 FASTQs as described above as input to the Python script genotyphi.py
321 (v2.0; available at https://github.com/typhoidgenomics/genotyphi) [18,22,28]. To validate the
322 AMR and plasmid calls, we compared them to those called in corresponding genome
323 assemblies by Pathogenwatch (available for n=11,992 genomes, assemblies for the other
324 genomes were excluded as they did not meet Pathogenwatch quality-control criteria [13]). We
325 downloaded the GenoTyphi, AMR, and plasmid replicon typing reports generated by Basic
326 Local Alignment Search Tool (BLAST) searching assemblies for these genomes from
327 Pathogenwatch on 7 March 2023 (GenoTyphi v20221208 specification for lineage genotypes).
328 Categorical agreement (percentage of genomes where both tools agreed on the same call of
329 presence or absence) was calculated separately for each marker, and collectively by drug
330 (percentage of genomes where both tools agreed that at least one marker was present or that
331 no markers were present) (**Supplementary Table 3,** available at
332 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). Probes for plasmid-
333 borne carbapenemase genes circulating in Enterobacterales (*bla*NDM-5, *bla*KPC-2, *bla*VIM-1,
334 *bla*IMP-27, *bla*OXA-48) were added to GenoTyphi v20240407 (DOI: 10.5281/zenodo.13859721)
335 following the report of *bla*NDM-5 in Typhi from Pakistan [11], and validated using the genomes
336 of the two distinct morphologic variants from that study (accessions: SRR22801766 and
337 SRR22801806). Comparisons and statistical analyses were conducted in R v4.1.0 using
338 tidyverse v1.3.1 [44,45] .

339 **Benchmarking of run time for Mykrobe**

340 We further investigated the average run time for Typhi Mykrobe to demonstrate the rapid time to
341 a result. Here we used the short-read data from 100 Typhi genomes. These were randomly
342 selected but included at least one genome from each genotype and a diversity of AMR and
343 plasmid profiles. The `mykrobe predict` command was run for the 100 Typhi genomes on three
344 different computers where both the run time and RAM were recorded. The number of threads
345 was also varied from 1 to 64 threads for the two HPC and 1 to 16 for the Mac laptop.

346 **Accuracy of AMR phenotype prediction**

347 Antimicrobial susceptibility testing (AST) data were available for n=3,970 of the Illumina-
348 sequenced isolates, sourced from three separate datasets (**Supplementary Table 4,** available
349 at https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). Minimum
350 inhibitory concentration (MIC) data were available from the UK Health Security Agency
351 (UKHSA) for n=852 isolates following EUCAST standards and interpretive criteria (v10, 2020)
352 [46] and the US Centers for Disease Control and Prevention (US CDC) for n=720 isolates
353 following CLSI standards and interpretive criteria (M100, 2023) [47]. Disk diffusion data
354 following EUCAST standards and interpretive criteria (v8.0, 2018) were available for n=2,446
355 isolates collected and analysed from three countries as part of the Surveillance for Enteric Fever
356 in Asia Project (SEAP), with multiple source labs contributing AST data and isolates for
357 sequencing [6] . Typhi Mykrobe AMR calls following the S/I/R categorisation described above
358 were compared with the phenotypic antimicrobial susceptibility S/I/R categorisation using the
359 relevant CLSI or EUCAST standards for each data set, for the currently recommended drugs
360 azithromycin and ciprofloxacin and ceftriaxone, and the older drugs ampicillin, chloramphenicol,
361 trimethoprim-sulfamethoxazole, and tetracycline (note UKHSA tested amoxicillin rather than
362 ampicillin; and SEAP did not test tetracycline). To assess concordance between genotype and
363 phenotype, we followed the principles used for assessment of new AST devices by the US Food
364 and Drug Administration (FDA), which considers (i) categorical agreement, (ii) major error rate
365 (defined as the proportion of S isolates that test R with the new method [48,49]), and (iii) very
366 major error rate (defined as the proportion of R isolates that test S with the new method)
367 (**Supplementary Table 5,** available at
368 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe). The FDA standard
369 requires >90% overall categorical agreement (within one doubling dilution), <3% major errors for
370 each drug, and <1.5% very major errors for each drug; hence, we adopted these thresholds to
371 assess the accuracy of the genotype-based prediction of phenotype. For drugs with high error
372 rates, we investigated the relationship between individual resistance markers and phenotype,
373 which we visualised using upset plots generated in R v4.1.0 using the ComplexUpset package
374 v1.3.3 (10.5281/zenodo.3700590). For five isolates with unexplained trimethoprim-
375 sulfamethoxazole resistance, we used nucleotide blast search of the genome assemblies to
376 confirm the presence of wild-type chromosomal *folA* and *folP* (100% nucleotide match to
377 reference sequences from Typhi strain Ty2) and screened the assemblies for other AMR
378 determinants using CARD RGI v6.0.3 [50].

Matched ONT and Illumina validation data

For 98 isolates, long-read ONT data were available in addition to the Illumina data; these included 68 isolates from Carey et al 2023 [14] and 30 novel isolates from this study. Sample-level details including accessions and technical specifications such as library, flow cell, device, basecaller, and ONT processing (including demultiplexing, trimming and filtering), are available in **Supplementary Table 6,** available at https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe. These varied between isolates including different flow cells (e.g., R9.4.1 and R10.4.1) and base calling methods. Read depth for each ONT dataset was assessed by using seqtk v1.2-r94 (https://github.com/lh3/seqtk) to determine total number of bases per read set, which was divided by the length of the CT18 reference genome (4,809,037 bases). For five genomes with read depth <30×, Kraken2 was used for the initial read-level classification of ONT data [51]. These five genomes were also aligned to the CT18 reference genome using minimap2 [52,53] to determine the read coverage of the *invA* and the seven core genes in the *Salmonella enterica* MLST scheme [54]. ONT reads were assembled for these five genomes and in addition to four genomes to resolve discrepant plasmid calls. Assembly was done using Flye v2.9.1-b1780 [55] with default parameters and yielding genome sizes of the expected 4.8 Mbp. Due to the high read depth (784×), the ONT data for SRR17299234 was subsampled prior to assembly using Filtlong v0.2.1 (https://github.com/rrwick/Filtlong) to --target_bases 480000000 (resulting in 99.81× depth). The genomes of isolates called as ‘not Typhi’ were characterised for presence of the marker genes of *invA* and the seven MLST genes using Blast in Bandage for *invA* [43] and mlst with the ‘senterica’ database (https://github.com/tseemann/mlst).

Results

Validation of GenoTyphi genotype calling

To validate Typhi Mykrobe for assigning hierarchical GenoTyphi genotypes, we ran it on n=12,839 Illumina readsets (covering all existing types) and compared the resulting genotypes to those called using the original mapping-based implementation. Five genomes were typed as ‘not Typhi’ by Mykrobe **(Supplementary Table 1**)**.** Four of these genomes lacked the *invA* gene but MLST profiles were detected that were consistent with those on Pathogenwatch. In the remaining isolate, DRR071001, the *invA* gene was detected but no allele was reported for *hisD* in the MLST scheme, consistent with the MLST profiles from Pathogenwatch. Hence, these five isolates were reported as ‘not Typhi’ and not analysed further. Of the remaining n=12,834

411 genomes, Mykrobe called the same single genotype as the mapping-based approach for
412 n=12,807 (99.79%). For a further n=17 genomes, Mykrobe and the mapping-based pipeline
413 gave concordant reports of two genotypes being present, consistent with mixed samples (in the
414 Typhi Mykrobe output this is reported as one primary genotype plus additional markers, while in
415 the mapping-based pipeline the result is reported as a list of supported genotypes). Eight
416 genomes belonged to genotype 3.5.3, which was nested in 3.5.4, breaking the usual
417 hierarchical structure of the scheme. Typhi Mykrobe correctly reported these genomes as
418 genotype 3.5.3 with support for both markers 3.5.4 and 3.5.3, as it has this nested relationship
419 explicitly encoded in its hierarchy; whereas the mapping-based pipeline does not have this logic
420 and incorrectly reported these genomes as genotype 3.5.4.
421
422 There were two instances of genuine disagreement in the genotype calls generated by Typhi
423 Mykrobe as compared to the mapping-based approach. One isolate (ERR5243665) was called
424 as genotype 2 by Mykrobe with strong confidence (no reads support for any alternative alleles
425 or markers). It was genotyped as 0.1.3, albeit with low support, by the mapping-based
426 approach, but the genotype 2 marker was also detected. This isolate clustered with other
427 genotype 2 isolates in a distance-based phylogeny of genomes from the same study [6] (see
428 tree at https://pathogen.watch/collection/0z5knw9jic9b-guevara-et-al-2021), suggesting that the
429 Mykrobe call is correct. The other isolate (ERR2663783) was called as genotype 4.3.1.1 by
430 Typhi Mykrobe with strong confidence. It was reported as “4.3.1.1,4.3.1.3.Bdq” (with low
431 support) by the mapping approach; however, the support for the 4.3.1.3.Bdq marker was low.
432 This isolate clustered closely with other 4.3.1.1 isolates from the same study [25](see tree at
433 https://pathogen.watch/collection/b2k6dym3bdyq-tanmoy2018), which is consistent with the
434 Mykrobe call.
435
436 We demonstrate that the run time for Typhi Mykrobe’s `mykrobe predict` command on a modern
437 computer was <1minute to complete (**Supplementary Figure 2**). The `mykrobe predict`
438 command can be run with multiple threads using the `--threads`/`-t` option. Up to ~4 threads will
439 increase performance. Using multiple threads on a very fast CPU, the run time for each genome
440 was reduced to seconds (**Supplementary Figure 2**). Further, `mykrobe predict` is very
441 memory-efficient and will typically use less than 100 MB of RAM per genome.

442 **Validation of AMR genotyping**

443 To assess the accuracy of Mykrobe for detecting the presence of AMR genes and mutations,
444 we compared genotype calls from Mykrobe (called from *k*-mer analysis of reads) with those
445 from Pathogenwatch (called from BLAST analysis of genome assemblies of the same read
446 sets). We found very strong categorical agreement between the two methods at the individual
447 marker level, with 99.88% of the 299,004 marker-genome combinations yielding the same call
448 (summarised in **Table 2,** full details in **Supplementary Table 2**). High concordance was evident
449 for all genetic markers of AMR (98.56–100%, **Table 2**), as well as plasmid replicon markers
450 (99.74–100%, **Table 3**). The test dataset included all AMR and plasmid markers in GenoTyphi
451 v2.0 implemented Typhi Mykrobe v0.12.1. The *ereA*, *ermB* and carbapenemases are very rarely
452 reported in Typhi but were included in the Mykrobe panel in case they emerge.
453
454 Both Mykrobe and Pathogenwatch summarise AMR data in terms of a binary phenotype
455 categorization (S/R), assuming the presence of a marker associated with resistance to a drug
456 implies resistance 'R', with the exception of ciprofloxacin, where presence of any marker is
457 categorised as 'I' and specific combinations of markers are categorised as 'R', resulting in three
458 categories S/I/R (see Methods). Categorical agreement was very high between the drug-level
459 categorisations reported by both genotyping tools, with >99.7% agreement (n=2 to 33
460 discordant genomes) for each drug (**Table 2, Supplementary Table 3**), and for subsequent
461 categorisation as MDR (99.86%, n=17 discordant) or XDR (99.97%, n=4 discordant).

462 **Validation of AMR phenotype prediction**

463 To assess the accuracy of Mykrobe's genotype-based clinical resistance categorisations, we
464 compared them with AST data, which was available for a subset of n=4,018 isolates (see
465 Methods, **Supplementary Table 4**). These isolates were from three data sets, using different
466 methods (MIC with EUCAST standards, MIC with CLSI standards, disk diffusion with EUCAST
467 standards), hence we analyzed the results from each dataset separately. Categorical
468 agreement was high (>95%) for all drugs across all three datasets, and very high (>99%) for the
469 two sets of reference-laboratory MIC measures (**Figure 2, Supplementary Figure 5**).The only
470 exception was US CDC categorisation of low-level ciprofloxacin resistance (using the CLSI 'I'
471 threshold, ≥0.125 to 0.5 mg/L), which showed 97.36% agreement with Mykrobe calls of 'I' or 'R'.
472 This was mainly due to the presence of major errors (susceptible isolates with AMR
473 determinants detected, thus reported as R by Mykrobe, see Methods), whereby n=17 isolates
474 tested susceptible to ciprofloxacin (n=14 with MIC=0.06 mg/L, n=3 with MIC=0.015 mg/L) but

475 Mykrobe detected a known QRDR mutation (**Figure 3b**); categorical agreement within one
476 doubling dilution was 99.4%.
477
478 Major errors were infrequent (<3% prevalence) for all drugs in all datasets (**Figure 2**), except for
479 ciprofloxacin (**Supplementary Table 5**). Most of the ciprofloxacin major errors were in isolates
480 whose MIC or disk zone measurement were close to the breakpoint (**Figures 3, 4c**); all carried
481 well-characterized molecular resistance determinants (QRDR mutations and/or *qnr* genes, see
482 **Figures 3, 4c**), and in all but one case the same determinants were also detected in the
483 corresponding genome assemblies by Pathogenwatch (details below).
484
485 Very major errors (resistant isolates with no AMR determinants detected, thus reported as S by
486 Mykrobe) were also infrequent (<1.5% prevalence) for the reference-laboratory MIC datasets,
487 but errors were higher in the disk diffusion dataset (**Figure 2**). The only drugs showing major or
488 very major errors compared with the reference-laboratory MIC data were ciprofloxacin and
489 trimethoprim-sulfamethoxazole (**Figure 2**)**.** For ciprofloxacin, the US CDC categorisation of
490 high-level resistance (using the CLSI 'R' threshold, ≥1 mg/L) yielded four very major errors (R
491 isolate called as 'I' by Mykrobe). Two had a single mutation detected (GyrA-S83F) and two had
492 double QRDR mutations (GyrA-S83F plus GyrA-D87N or ParC-S80I) (see **Figure 3**);
493 Pathogenwatch calls agreed in all instances. Importantly, all four isolates had MIC=1 mg/L, i.e.
494 they agreed with the genotype-based call within one doubling dilution. UKHSA uses the
495 EUCAST standards, which have a single (low-level) breakpoint for ciprofloxacin (R >0.06 mg/L),
496 and this yielded just two major errors (2.9%) and two very major errors (0.26%) (see **Figure 3**).
497
498 For trimethoprim-sulfamethoxazole, Mykrobe analysis of the US CDC dataset yielded a single
499 error (R isolate called as S) but the MIC was 4 mg/L, i.e. within one doubling dilution, and the
500 isolate had no *sul* or *dfr* genes that were identified by Mykrobe or Pathogenwatch. Mykrobe
501 analysis of the UKHSA dataset yielded five errors for trimethoprim-sulfamethoxazole; all were R
502 isolates called as S by Mykrobe, which detected *sul* genes but no *dfrA* genes. Interestingly, all
503 five isolates tested susceptible to trimethoprim alone (MIC ≤0.5, well below the >4 mg/L
504 breakpoint), supporting the absence of *dfrA* genes; these were also not detected by
505 Pathogenwatch analysis of the corresponding genome assemblies, or by UKHSA's internal
506 pipeline [44]. Further investigation of these genome assemblies confirmed the presence of wild-
507 type *folA* and *folP*, and no additional determinants were reported by CARD RGI (v6.0.3).
508

509 Mykrobe analysis of the SEAP dataset showed lower overall categorical agreement with AST
510 phenotypes (95.3–98.2%). As the testing was done using disk diffusion, it was not possible to
511 assess the categorical agreement within one doubling dilution of a MIC. However, most of the
512 errors were within 2–4 mm of the breakpoint (see **Figure 4**), and in all but one instance, the
513 Mykrobe genotype result matched that called by Pathogenwatch from the corresponding
514 genome assemblies. The only exception was a ciprofloxacin resistant isolate (ERR4325960),
515 which Mykrobe reported as carrying GyrA-S83F with no other determinants, but Pathogenwatch
516 detected both GyrA-S83F and *qnrS* in the assembly.

517 **Validation on nanopore reads**

518 We explored the accuracy of Typhi Mykrobe for genotyping ONT long-read data by comparing
519 results from 98 pairs of ONT to Illumina reads generated from the same isolates. Data were
520 contributed by members of the Global Typhoid Genomics Consortium and generated using
521 different library preparation methods and sequencing devices (see **Supplementary Table 6**).
522 Mykrobe called one Illumina read set and five ONT read sets as non-Typhi (mean ONT read
523 depths estimated as 0.04×, 4.06×, 8.01×, 26.40×, 27.85×, all sequenced on R9 flowcells), these
524 were excluded from analysis of genotype accuracy as no genotype is called on isolates
525 identified as non-Typhi.
526
527 Among the n=92 Typhi genomes called from both ONT and Illumina reads, the Typhi Mykrobe
528 genotype lineage calls matched for n=89 (96.7%). For all three discrepant isolates, (all
529 sequenced on R.9 or R.9.4 flowcells), the set of SNV markers detected from ONT reads closely
530 matched those detected from Illumina reads, but with one or two markers missing from the
531 ONT-based calls and no unexpected markers reported resulting in final genotype calls of 4.3.1.1
532 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
533 variable ONT read depths, with ranges 3–9, 4–31 and 1–14, suggesting that the errors are likely
534 due to lack of informative ONT reads at the marker loci. Notably, the overall ONT read depths
535 for these were reasonable (mean 35×, 170× and 90× respectively), suggesting the lack of
536 informative reads for genotyping may be caused by contamination from non-Typhi reads, or
537 high error rates resulting in low rates of *k*-mer matching.
538
539 In this study, the five ONT genomes that had 'not Typhi' calls were further explored. These were
540 all classified as *Salmonella enterica*. However, Typhi was not the top serovar from read-level
541 classification, instead *Salmonella* Infantis was the top serovar for BZD62J, while *Salmonella*

542 Typhimurium was the top serovar for the remaining four. Further, the presence of the marker
543 genes of *invA* and the seven MLST genes were detected in the genomes at read depths
544 consistent with the genome mean read depth reported in **Supplementary Table 6**, with the
545 gene *invA* was detected in three of the five genomes from Blast analysis of the Flye assemblies.
546 The gene *invA* was not detected in the 611427 ONT assembly and only had a read depth of
547 0.43×, and BZD62J failed to assemble due to insufficient reads and was found to have no reads
548 mapping to *invA.* MLST analysis of the three complete genomes with *invA* detected reported,
549 were found to have partial matches to known alleles for all seven MLST genes and no reported
550 MLST profile. Of note, these five genomes had low read depth (0.04–28×), sequencing on older
551 R9 flowcells, and basecalling with older versions of guppy (**Supplementary Table 6**). The
552 failure of Typhi Mykrobe to identify these as Typhi is likely due to a combination of potential
553 contamination, low read depth and high error rates resulted in low rates of *k*-mer matching to
554 the probes. This highlights the importance of QC of the read data, as Typhi Mykrobe provides
555 some metrics for confidence of the calls, but does not itself perform detailed QC.
556
557 AMR marker detection from ONT reads was highly concordant with Illumina-based calls, with
558 98.77% overall agreement across the n=736 isolate-drug combinations (92 isolates across 8
559 drugs; see **Supplementary Table 7**). There were nine genomes with disagreement of AMR
560 calls; five of these had a marker detected in Illumina reads but not ONT reads (n=1 *catA1*, n=1
561 *qnr,* n=3 QRDR mutations), three had GyrB-464F detected in ONT reads but not Illumina, and
562 one had GyrB-S464F detected in Illumina reads and GyrA-S83F in ONT. In all nine cases, the
563 Typhi Mykrobe call from Illumina reads matched that from Pathogenwatch analysis of the
564 corresponding assemblies of Illumina reads. There was no overlap between ONT read sets with
565 AMR discrepancies and those with lineage genotype discrepancies. Overall, ONT-based
566 resistance prediction showed complete agreement with Illumina for six of the eight drugs, 98.9%
567 agreement for chloramphenicol resistance, and 91% for ciprofloxacin resistance.
568
569 Plasmid marker detection from ONT data was also highly concordant with Illumina-based calls,
570 with 99.67% overall agreement across the n=1,196 isolate-marker combinations (92 isolates
571 across 13 markers; see **Supplementary Table 8**). Four genomes had discrepant plasmid
572 marker calls, all with a marker detected from Illumina reads but not from ONT reads. One isolate
573 (DRR070993) had a typical IncHI1 MDR plasmid profile reported from Illumina reads (IncHI1A,
574 IncHI1BR27, IncHI1_ST6 indicating presence of IncHI1-ST6 plasmid, with *catA1*, *dfrA7*, bla_{TEM-1},
575 *sul1*, *sul2*), but the ONT read profile was missing the *IncHI1A* and *catA1* markers (and missing

576 *catA1* in the corresponding Illumina assembly and a mismatch for *IncHI1A*). The other three
577 isolates had pHCM2 detected in the Illumina reads and in the corresponding assemblies [14] but
578 not in ONT reads; for these isolates the ONT and Illumina sequencing was done on different
579 DNA extracts so this may reflect plasmid loss in the culture used for ONT sequencing.

580 **Discussion**

581 Typhi Mykrobe provides detection and genotyping of key features of clinical and public health
582 importance, including assigning lineages, AMR determinants, and plasmid replicons (**Figure 1**).
583 Compared with other tools for Typhi genotyping, Mykrobe has the advantage of not requiring
584 any pre-processing, aligning, or assembling of reads, but instead takes as inputs raw FASTQ
585 files. This simplifies bioinformatics workflows and reduces time-to-result as Typhi Mykrobe can
586 return a full genotype profile in typically <1 minute. Importantly, although Typhi Mykrobe is not a
587 QC tool *per se*, it provides useful data on the read-level support for individual genotype markers,
588 which can be used to help identify potentially mixed samples. As a command-line tool that is
589 easy to install locally using Bioconda, Typhi Mykrobe is suitable for research and public health
590 laboratory settings, facilitates standardized reporting of Typhi genotypes without the need to
591 share sequence data, and enables broader access to bioinformatics functionality across a
592 variety of settings.
593
594 Typhi Mykrobe showed very high genotyping accuracy, for GenoTyphi lineage assignment
595 (**Table 1**), and detection of AMR and plasmid markers (**Tables 2-3**), compared with existing
596 genotyping tools. Of note, the included IncY, pO111 and IncFIB for the cryptic pHCM2 replicon
597 targets have recently been shown to be sometimes carried on phage-plasmids [56,57], in
598 addition to conjugative plasmids such as the IncY plasmids associated with XDR Typhi [26].
599 Future work is required to identify additional targets to delineate the phage-plasmids, and the
600 Typhi Mykrobe software can be readily updated to include new markers.
601
602 In addition, we show that the panel of AMR markers currently included in Typhi Mykrobe can be
603 used to predict AMR phenotypes with high accuracy, demonstrating very high concordance with
604 S/I/R calls made from reference-laboratory MICs (**Figure 2**). Concordance was slightly lower
605 when compared with research laboratory disk diffusion data. However, as the phenotypic
606 resistance prevalence rates were similar across the three datasets, it is reasonable to assume
607 that the higher error rates in the disk diffusion dataset compared with the two reference-
608 laboratory MIC datasets most likely reflect comparatively lower accuracy of the disk diffusion

609 phenotype measurements, which are known to yield lower accuracy for some drugs such as
610 azithromycin [58,59] and in this case were completed across multiple laboratories. There are
611 currently no formal standards for assessing the accuracy of AMR phenotype predictions from
612 genomic data[60]; however, applying the standards for FDA licensing of new AST devices,
613 Typhi Mykrobe easily passed the target threshold for categorical agreement (>99% agreement
614 within one doubling dilution for reference-laboratory MIC data, >95% agreement for disk-
615 diffusion data; vs >90% target) (**Figure 2**).
616
617 Typhi Mykrobe also showed very low rates of major errors and very major errors compared with
618 the reference-laboratory MIC data, and for nearly all drugs these fell below the acceptable
619 thresholds for a new AST device (<3% and <1.5%, respectively) (**Figure 2**). The exception was
620 trimethoprim-sulfamethoxazole, for which five UKHSA isolates tested resistant but Typhi
621 Mykrobe predicted as susceptible due to a lack of *dfrA* genes, which are thought to be required
622 for resistance to trimethoprim-sulfamethoxazole along with *sul* genes, which were detected in
623 these genomes. Other genotyping tools (Typhi Pathogenwatch, CARD RGI) also could not
624 identify any known mechanisms of resistance to trimethoprim or trimethoprim-sulfamethoxazole
625 in these genomes, and the isolates tested sensitive to trimethoprim. Therefore, we hypothesise
626 these isolates may harbour efflux mutations or a novel mechanism of resistance to
627 trimethoprim-sulfamethoxazole that does not result in resistance to trimethoprim alone. Further
628 analyses to determine the exact resistance mechanism was beyond the scope of this study.
629 Ciprofloxacin showed very low error rates in the UKHSA MIC data, interpreted using the single
630 low-level resistance threshold (MIC >0.06 mg/L) recommended by EUCAST, but slightly higher
631 error rates for the US CDC MIC data, which was interpreted using low-level and high-level
632 resistance thresholds (although still below the acceptable thresholds within one doubling
633 dilution). Carbapenemase genes were added to the Typhi Mykrobe panel, but this resistance is
634 newly emerging and there was only a single carbapenem-resistant isolate sequenced at the
635 time of testing (in which *bla*NDM-5 was correctly identified by Typhi Mykrobe), therefore we
636 could not yet assess accuracy of phenotype prediction for carbapenems.
637
638 Finally, we demonstrated that Typhi Mykrobe can be used to genotype ONT reads, showing
639 high concordance with results obtained from the analysis of Illumina reads generated from the
640 same isolates (96.74% agreement on GenoTyphi lineage calls, 98.77% on AMR markers and
641 99.67% on plasmid markers). We did not have suitable data to directly assess ONT-based AMR
642 predictions with AST data. ONT-based and Illumina-based drug-level predictions of AST

643 phenotypes showed high agreement; however, future studies assessing ONT data compared to
644 reference-laboratory MIC data would be useful to confirm the accuracy of ONT data for
645 phenotype prediction. As one might expect, many of the errors occurred in isolates with lower
646 read depth. However, as the matched ONT-Illumina data available for testing were contributed
647 by different laboratories that used different ONT protocols and devices for library preparation,
648 sequencing, and base calling, we were unable to perform a detailed assessment of the impact
649 of read depth and quality on genotyping performance. The accuracy of ONT reads has greatly
650 increased in recent years and is approaching Illumina read accuracy of ~99%; hence modern
651 ONT data would be expected to perform similarly to Illumina data [61]
652 (doi.org/10.5281/zenodo.10397817). Future studies assessing the impact of read depth and
653 base calling methods on the performance of Typhi Mykrobe on ONT data would be desirable to
654 help guide laboratories on the most appropriate methods and target yield for ONT experiments
655 aiming to produce reliable genomic surveillance data for Typhi.

656 **Conclusions**

657 Typhi Mykrobe provides rapid and accurate genotyping of Typhi genomes directly from Illumina
658 reads, including (i) assignment of GenoTyphi lineage; (ii) detection of AMR determinants and
659 prediction of corresponding AMR phenotypes demonstrating >99% categorical agreement with
660 reference-laboratory MIC data; and (iii) detection of plasmid replicons. The software also
661 performs well on direct genotyping from ONT reads information, although further investigation is
662 needed to assess the accuracy of AMR phenotype prediction.
663

664 **Other sections**

665 **List of abbreviations**

666 AMR: antimicrobial resistance
667 AST: antimicrobial susceptibility testing
668 CARD: Comprehensive Antibiotic Resistance Database
669 CLSI: Clinical and Laboratory Standards Institute
670 EUCAST: The European Committee on Antimicrobial Susceptibility Testing
671 FDA: United States Food and Drug Administration
672 MDR: multidrug resistance
673 MIC: minimum inhibitory concentration

ONT: Oxford Nanopore Technologies
QRDR: quinolone resistance-determining regions
SEAP: Surveillance for Enteric Fever in Asia Project
SNV: single nucleotide variant
UKHSA: United Kingdom Health Security Agency
US CDC: United States of America Centers for Disease Control and Prevention
WGS: whole genome sequencing
XDR: extensively drug resistant

Availability and requirements

Project name: GenoTyphi
Project home page: https://github.com/typhoidgenomics/genotyphi
Operating system(s): Platform independent
Programming language: Python3 or higher
Other requirements: Mykrobe
License: GNU General Public License
Any restrictions to use by non-academics: None

Project name: Mykrobe
Project home page: https://github.com/Mykrobe-tools/mykrobe
Operating system(s): Platform independent
Programming language: Python3 or higher
Other requirements: anytree 2.8.0 or higher, Biopython 1.79 or higher, PyVCF3 1.0.3 or higher, requests 2.27.1 or higher, mongoengine 0.24.1 or higher, cython 0.29.28 or higher, numpy 1.22.0 or higher, pytest
License: MIT License
Any restrictions to use by non-academics: None.

Declarations

Ethics approval and consent to participate

Each contributing study or surveillance program obtained local ethical and governance approvals, as reported in the primary publication for each dataset. Inclusion of data that were not yet public domain by August 2021 was approved by the Observational / Interventions

705 Research Ethics Committee of the London School of Hygiene and Tropical Medicine (ref
706 #26408), on the basis of details provided on the local ethical approvals for sample and data
707 collection (listed in Carey et al, 2023, Supplementary file 1).

708 **Consent for publication**

709 Not applicable

710 **Availability of data and materials**

711 All code and instructions for genotyping isolates with GenoTyphi using Mykrobe is available in
712 the GitHub repository at https://github.com/typhoidgenomics/genotyphi. The version reported in
713 this paper is v2.1 (DOI: 10.5281/zenodo.13859721)
714
715 The datasets and analysis code used in this article are available in a GitHub repository at
716 https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe.
717

718 **Competing interests**

719 MML was a co-developer of a Trivalent Salmonella (Enteritidis/Typhimurium/Typhi Vi) conjugate
720 vaccine with Bharat Biotech International and the Wellcome Trust. MML has also received
721 payments from Pfizer for consultancy work. MML holds US patents for “Compositions and
722 Methods for Producing Bacterial Conjugate Vaccines”. MML was a member of a NIH DSMB that
723 oversaw US government-funded efficacy trials of COVID-19 vaccines. DSMB was disbanded
724 after several vaccines were given Emergency Use Authorization. MML was a member of the
725 Vaccines and Related Biological Products Advisory Committee of the FDA. IIB was a consultant
726 for the Weapons Threat Reduction Program, Global Affairs Canada. AJP has been involved an
727 Oxford University partnership with AZ for development of COVID-19 vaccines. AJP has received
728 payments for consultancy work from Shionogi. AJP is chair of DHSC’s Joint Committee on
729 Vaccination and Immunisation, is a chair of WHOs Salmonella TAG, and was a member of
730 WHOs SAGE. AJP received support from MRNA – Moderna. KLC has received payments from
731 Pfizer for presentations and travel support from BD. INO has received payments from the
732 Wellcome Trust for consultancy work and receives royalties for books or book chapters
733 published via Springer, Cornell University Press, and Oxford University Press. INO has received
734 travel support from BMGF, ESCMID, and the American ASM and has held leadership or
735 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 Leadership Fellowship 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 supported 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. RRW is supported by an ARC Discovery Early Career Researcher Award (DE250100677). JH was supported by an Emerging Leadership Fellowship from the NHMRC of Australia (GNT2034741).

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, JAK, 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, MEC, MAC, JAC, PDG, HI, JJJ, AK, KHK, MM, SN, SO, INO, SIAR, AR, DR, VS, MJS, AMS, GTS, KAT, RRW, ZAD and KEH - Formal analysis
DJI, JH, AOA, NA, SA, PMA, IIB, MC, MAC, JAC, PDG, BPH, HI, JJJ, LMJ, AK, KHK, JK, MML, MM, SN, SO, INO, PEO, SIAR, AR, DAR, VS, MJS, AMS, GTS, KAT, RRW, ZAD and KEH - Investigation,
DJI and KEH - Visualization,
DJI, ZAD and KEH - Writing – original draft,
DJI, JH, ZI, MH, JAK, AOA, NA, SA, PMA, IIB, MEC, MAC, JC, PDG, BPH, HI, JJJ, 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.
NA is grateful for support from ICMR India (IIRP-2023-2500).
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.
This work was supported by Monash eResearch capabilities, including the high performance computer M3 and Research Data Storage.

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

863 **References**

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

866 2. Collaborators G 2017 T and P, Stanaway JD, Reiner RC, Blacker BF, Goldberg EM, Khalil IA,
867 et al. The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global
868 Burden of Disease Study 2017. Lancet Infect Dis. 2019;19: 369–381. doi:10.1016/s1473-
869 3099(18)30685-6

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

872 4. Nabarro LE, McCann N, Herdman MT, Dugan C, Ladhani S, Patel D, et al. British infection
873 association guidelines for the diagnosis and management of enteric fever in England. J Infect.
874 2022;84: 469–489. doi:10.1016/j.jinf.2022.01.014

875 5. Collaborators G 2021 F, Vollset SE, Ababneh HS, Abate YH, Abbafati C, Abbasgholizadeh R,
876 et al. Burden of disease scenarios for 204 countries and territories, 2022–2050: a forecasting
877 analysis for the Global Burden of Disease Study 2021. Lancet. 2024;403: 2204–2256.
878 doi:10.1016/s0140-6736(24)00685-8

879 6. da Silva KE, Tanmoy AM, Pragasam AK, Iqbal J, Sajib MSI, Mutreja A, et al. The
880 international and intercontinental spread and expansion of antimicrobial-resistant *Salmonella*
881 Typhi: a genomic epidemiology study. Lancet Microbe. 2022;3: e567–e577. doi:10.1016/s2666-
882 5247(22)00093-3

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

886 8. Walker J, Chaguza C, Grubaugh ND, Carey M, Baker S, Khan K, et al. Assessing the global
887 risk of typhoid outbreaks caused by extensively drug resistant *Salmonella* Typhi. Nat Commun.
888 2023;14: 6502. doi:10.1038/s41467-023-42353-9

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

892 10. Wang Y, Lu D, Jin Y, Wang H, Lyu B, Zhang X, et al. Extensively Drug-Resistant (XDR)
893 *Salmonella* Typhi Outbreak by Waterborne Infection — Beijing Municipality, China, January–
894 February 2022. China CDC Wkly. 2022;4: 254–258. doi:10.46234/ccdcw2022.062

895 11. Nizamuddin S, Khan EA, Chattaway MA, Godbole G. Early Release - Case of Carbapenem-
896 Resistant *Salmonella* Typhi Infection, Pakistan, 2022 - Volume 29, Number 11—November
897 2023 - Emerging Infectious Diseases journal - CDC. Emerg Infect Dis. 2023;29: 2395–2397.
898 doi:10.3201/eid2911.230499

12. Dyson ZA, Klemm EJ, Palmer S, Dougan G. Antibiotic Resistance and Typhoid. Clin Infect Dis. 2019;68: S165–S170. doi: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. doi: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. doi: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: 632–639. doi: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. doi: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: 365–372. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.1371/journal.pntd.0006408

935 24. Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE, Chowdhury EK, et al. Population
936 structure and antimicrobial resistance patterns of Salmonella Typhi isolates in urban Dhaka,
937 Bangladesh from 2004 to 2016. Plos Neglect Trop D. 2020;14: e0008036.
938 doi:10.1371/journal.pntd.0008036

939 25. Tanmoy AM, Westeel E, Bruyne KD, Goris J, Rajoharison A, Sajib MSI, et al. *Salmonella*
940 *enterica* Serovar Typhi in Bangladesh: Exploration of Genomic Diversity and Antimicrobial
941 Resistance. Mbio. 2018;9: e02112-18. doi:10.1128/mbio.02112-18

942 26. Klemm EJ, Shakoor S, Page AJ, Qamar FN, Judge K, Saeed DK, et al. Emergence of an
943 Extensively Drug-Resistant *Salmonella enterica* Serovar Typhi Clone Harboring a Promiscuous
944 Plasmid Encoding Resistance to Fluoroquinolones and Third-Generation Cephalosporins. Mbio.
945 2018;9: e00105-18. doi:10.1128/mbio.00105-18

946 27. Kariuki S, Dyson ZA, Mbae C, Ngetich R, Kavai SM, Wairimu C, et al. Multiple introductions
947 of multidrug-resistant typhoid associated with acute infection and asymptomatic carriage,
948 Kenya. eLife. 2021;10: e67852. doi:10.7554/elife.67852

949 28. Ingle DJ, Hawkey J, Hunt M, Dyson ZA, Holt KE. Technical Report: GenoTyphi
950 Implementation in Mykrobe. Zenodo 2022. Available from: https://zenodo.org/records/7407985

951 29. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, et al. Complete genome
952 sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature.
953 2001;413: 848–852. doi:10.1038/35101607

954 30. Labbé G, Kruczkiewicz P, Robertson J, Mabon P, Schonfeld J, Kein D, et al. Rapid and
955 accurate SNP genotyping of clonal bacterial pathogens with BioHansel. Microb Genom. 2021;7:
956 000651. doi:10.1099/mgen.0.000651

957 31. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, et al. Antibiotic resistance
958 prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe.
959 Wellcome Open Res. 2019;4: 191. doi:10.12688/wellcomeopenres.15603.1

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

963 33. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, et al. The 2021 WHO
964 catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a
965 genotypic analysis. Lancet Microbe. 2022;3: e265–e273. doi:10.1016/s2666-5247(21)00301-3

966 34. Hall MB, Coin LJM. Assessment of the 2021 WHO *Mycobacterium tuberculosis* drug
967 resistance mutation catalogue on an independent dataset. Lancet Microbe. 2022;3: e645.
968 doi:10.1016/s2666-5247(22)00151-3

969 35. Baker S, Holt K, Whitehead S, Goodhead I, Perkins T, Stocker B, et al. A linear plasmid
970 truncation induces unidirectional flagellar phase change in H:z66 positive *Salmonella* Typhi. Mol
971 Microbiol. 2007;66: 1207–1218. doi:10.1111/j.1365-2958.2007.05995.x

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

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

978 38. Hawkey J, Frézal L, Dien AT, Zhukova A, Brown D, Chattaway MA, et al. Genomic
979 perspective on the bacillus causing paratyphoid B fever. Nat Commun. 2024;15: 10143.
980 doi:10.1038/s41467-024-54418-4

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

984 40. Crump JA, Kretsinger K, Gay K, Hoekstra RM, Vugia DJ, Hurd S, et al. Clinical Response
985 and Outcome of Infection with *Salmonella enterica* Serotype Typhi with Decreased
986 Susceptibility to Fluoroquinolones: a United States FoodNet Multicenter Retrospective Cohort
987 Study. Antimicrob Agents Chemother. 2008;52: 1278–1284. doi:10.1128/aac.01509-07

988 41. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New
989 Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol.
990 2012;19: 455–477. doi:10.1089/cmb.2012.0021

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

993 43. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo
994 genome assemblies. Bioinformatics. 2015;31: 3350–3352. doi:10.1093/bioinformatics/btv383

995 44. Wickham H. ggplot2. Wiley Interdiscip Rev: Comput Stat. 2011;3: 180–185.
996 doi:10.1002/wics.147

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

999 46. Chattaway MA, Gentle A, Nair S, Tingley L, Day M, Mohamed I, et al. Phylogenomics and
1000 antimicrobial resistance of Salmonella Typhi and Paratyphi A, B and C in England, 2016–2019.
1001 Microb Genom. 2021;7. doi:10.1099/mgen.0.000633

1002 47. CDC. National Antimicrobial Resistance Monitoring System (NARMS) Now: Human Data.
1003 Atlanta, Georgia: U.S. Department of Health and Human Services, CDC [Internet]. Available
1004 from: https://www.cdc.gov/narmsnow

1005 48. Reller LB, Weinstein M, Jorgensen JH, Ferraro MJ. Antimicrobial Susceptibility Testing: A
1006 Review of General Principles and Contemporary Practices. Clin Infect Dis. 2009;49: 1749–1755.
1007 doi:10.1086/647952

1008 49. U.S Food and Drug Administration. Antimicrobial Susceptibility Test (AST) Systems - Class
1009 II Special Controls Guidance for Industry and FDA. [Internet]. Available
1010 https://www.fda.gov/media/88069/download.

1011 50. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023:
1012 expanded curation, support for machine learning, and resistome prediction at the
1013 Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2022;51: D690–D699.
1014 doi:10.1093/nar/gkac920

1015 51. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol.
1016 2019;20: 257. doi:10.1186/s13059-019-1891-0

1017 52. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:
1018 3094–3100. doi:10.1093/bioinformatics/bty191

1019 53. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:
1020 4572–4574. doi:10.1093/bioinformatics/btab705

1021 54. Achtman M, Wain J, Weill F-X, Nair S, Zhou Z, Sangal V, et al. Multilocus Sequence Typing
1022 as a Replacement for Serotyping in *Salmonella enterica*. Plos Pathog. 2012;8: e1002776.
1023 doi:10.1371/journal.ppat.1002776

1024 55. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using
1025 repeat graphs. Nat Biotechnol. 2019;37: 540–546. doi:10.1038/s41587-019-0072-8

1026 56. Nair S, Barker CR, Bird M, Greig DR, Collins C, Painset A, et al. Presence of phage-
1027 plasmids in multiple serovars of *Salmonella enterica*. Microb Genom. 2024;10: 001247.
1028 doi:10.1099/mgen.0.001247

1029 57. Greig DR, Bird MT, Chattaway MA, Langridge GC, Waters EV, Ribeca P, et al.
1030 Characterization of a P1-bacteriophage-like plasmid (phage-plasmid) harbouring bla CTX-M-15
1031 in *Salmonella enterica* serovar Typhi. Microb Genom. 2022;8: mgen000913.
1032 doi:10.1099/mgen.0.000913

1033 58. Goldblatt J, Ward A, Yusuf M, Day M, Godbole G, Morris-Jones S. Azithromycin
1034 susceptibility testing for *Salmonella enterica* isolates: discordances in results using MIC gradient
1035 strips. J Antimicrob Chemother. 2020;75: 1820–1823. doi:10.1093/jac/dkaa097

1036 59. Jabeen K, Saleem S, Nizamuddin S, Arshad F, Jahan S, Hasnain F, et al. Reporting of
1037 Azithromycin Activity against Clinical Isolates of Extensively Drug-Resistant Salmonella enterica
1038 Serovar Typhi. Am J Trop Med Hyg. 2023;108: 942–947. doi:10.4269/ajtmh.22-0557

1039 60. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of
1040 whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the
1041 EUCAST Subcommittee. Clin Microbiol Infect. 2017;23: 2–22. doi:10.1016/j.cmi.2016.11.012

1042 61. Wick R. Yet another ONT accuracy test: Dorado v0.5.0. Available from:
1043 doi.org/10.5281/zenodo.10397817

1044 **Tables and Figures**

1045

1046 **Table 1. Validation of genotype calls**

1047 Summary of Typhi Mykrobe genotype calls (GenoTyphi scheme) for 12,839 Illumina readsets,
1048 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

1049

1050 **Table 2. Validation of AMR genotyping (vs assembly based)**

1051 Comparison of AMR genotyping by Typhi Mykrobe (*k*-mer-based analysis of reads) vs
1052 Pathogenwatch (based on nucleotide blast search of assemblies), for n=11,992 isolates.
1053 Complete details of the call comparisons are given in **Supplementary Table 3**. Categorical
1054 agreement at drug level is calculated based on agreement of the binary S/R categorisation of
1055 the 11,992 isolates by each method (i.e. 'Concordant' means Typhi Mykrobe and Typhi
1056 Pathogenwatch made the same call of S or R based on absence or presence, respectively, of
1057 known genetic determinants; 'Discordant' means one method identified determinants and called
1058 R, while the other did not and called S). *For ciprofloxacin, both tools categorise as S/I/R as
1059 defined in Methods, and categorical agreement was calculated across these three categories.
1060 **Results for *dfrA* genes are summarised together, and as some genomes have multiple *dfrA*
1061 alleles the total observations is greater than the number of isolates (see details in
1062 **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%**

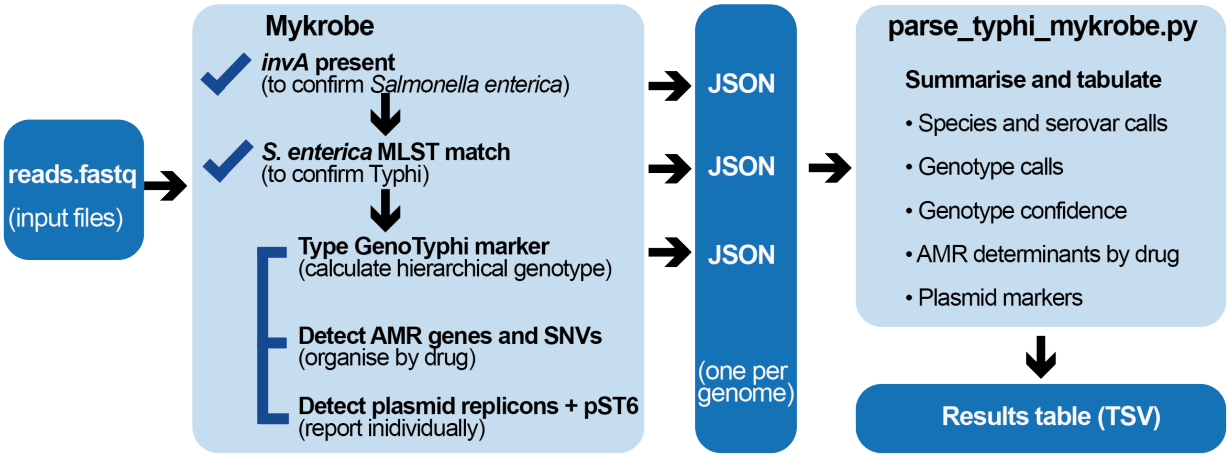
1063

1064 **Table 3 - Validation of plasmid replicon detection (vs assembly based).**
1065 Comparison of plasmid replicon markers identified in the genomes by Typhi Mykrobe (*k*-mer-
1066 based analysis of reads) vs Pathogenwatch (PW, based on nucleotide blast search of
1067 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%
IncX3	21	0	0	100.00%
IncY	666	0	0	100.00%

1068
1069

A. Typhi Mykrobe pipeline



B. AMR determinants included in Typhi Mykrobe

Drug	Acquired genes	Point mutations
ampicillin	*bla*TEM-1, *bla*OXA-7	
azithromycin	*ereA*, *ermB*, *mphA*	AcrB-717
carbapenems	*bla*KPC-2, *bla*NDM-5, *bla*VIM-1, *bla*IMP-27, *bla*OXA-48	
ceftriaxone	*ampC1*, *bla*CTX-M-15, *bla*OXA-134, *bla*SHV-12	
chloramphenicol	*catA1*, *cmlA1*	
ciprofloxacin	*qnrB1*, *qnrD1*, *qnrS1*	GyrA-83, GyrA-87, GyrB-464F, 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

1070
1071
1072 **Figure 1: Typhi Mykrobe functionality**

1073 A) A schematic overview of the Typhi Mykrobe pipeline. B) Table of epidemiologically important
1074 AMR determinants targeted for detection/typing by Typhi Mykrobe, grouped by drug. C) Table of
1075 epidemiologically important plasmid replicons targeted for detection by Typhi Mykrobe.

1076
1077

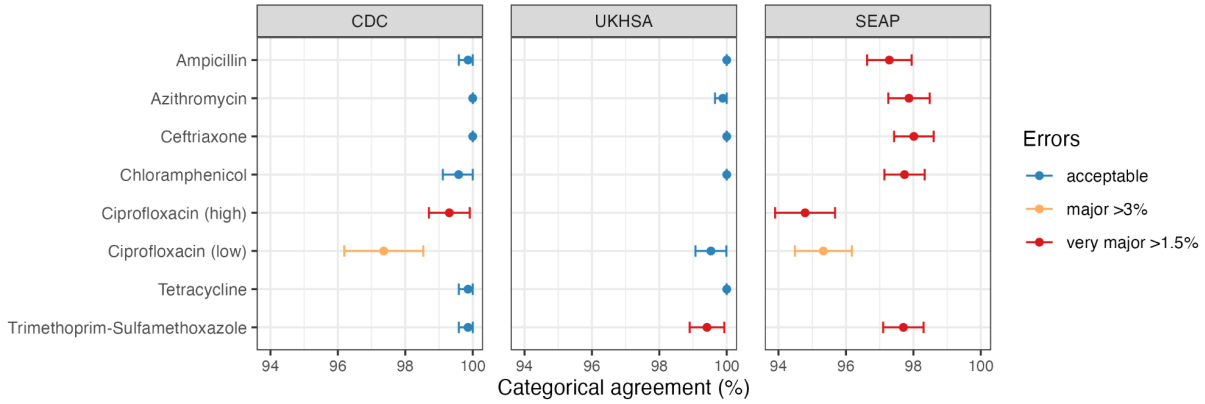


Figure 2: Validation of Typhi Mykrobe AMR predictions vs AMR phenotypes

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.

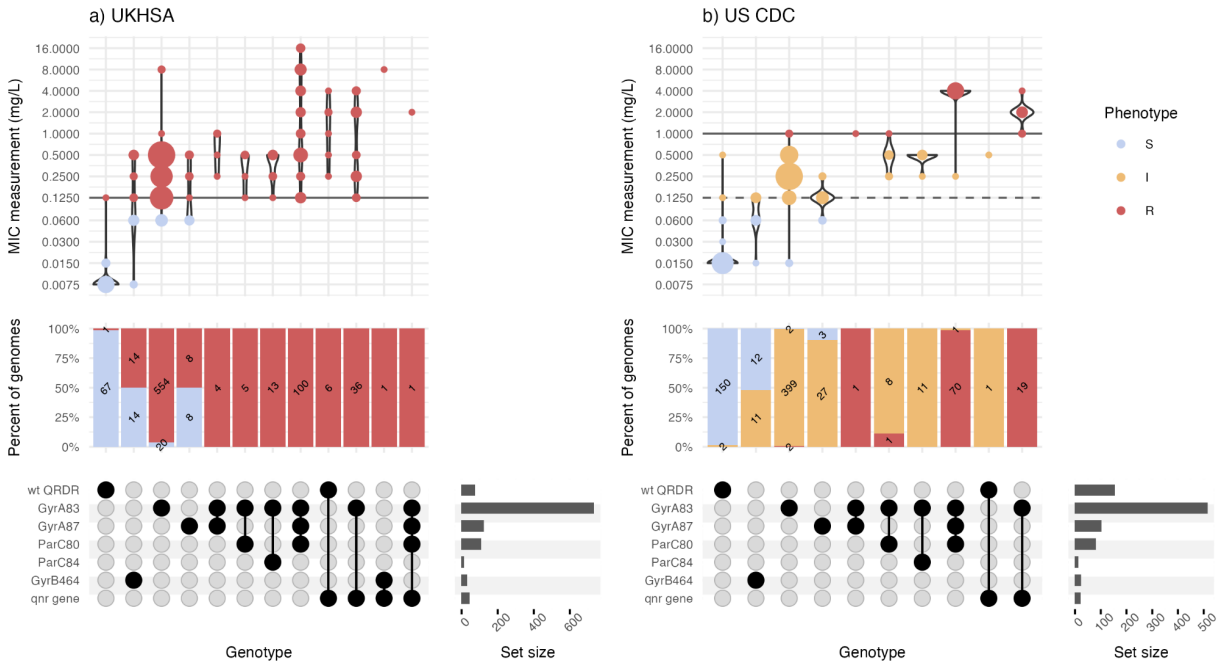


Figure 3: Typhi Mykrobe genotypes vs minimum inhibitory concentration from two reference laboratories, for ciprofloxacin.

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.

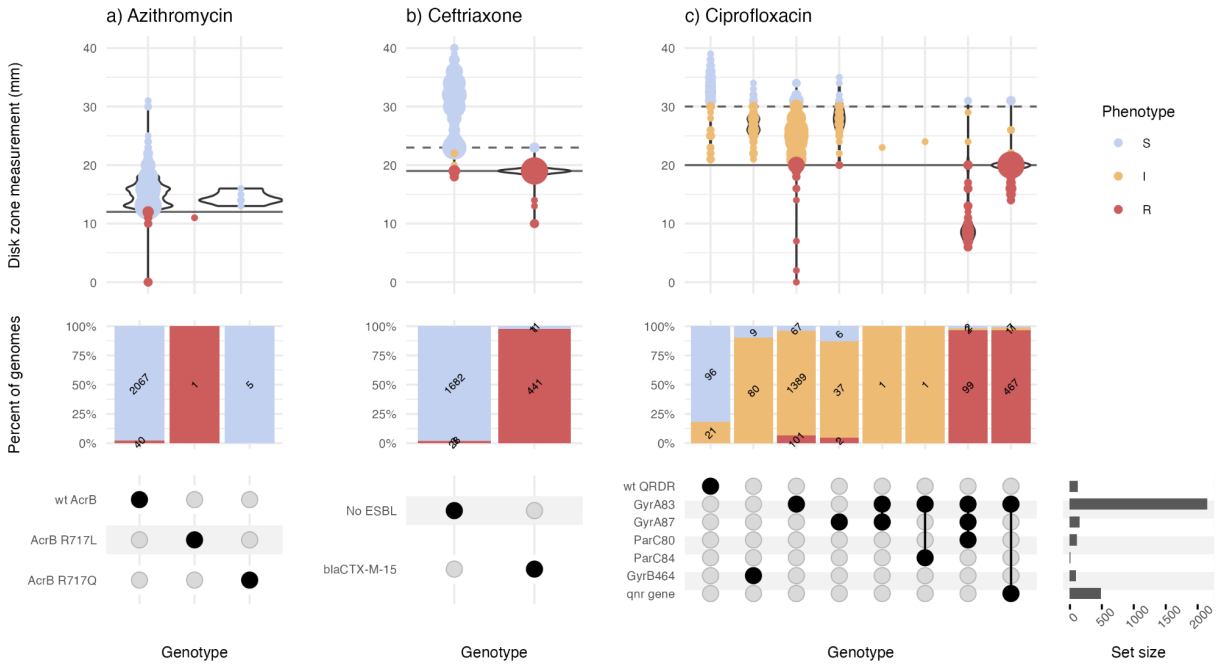


Figure 4: Typhi Mykrobe genotypes vs disk diffusion data from the SEAP study, for three drug classes

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, ceftriaxone 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.