Author/s:
Ingle, DJ;Hawkey, J;Hunt, M;Iqbal, Z;Keane, JA;Afolayan, AO;Ahmed, N;Andleeb, S;Ashton, PM;Bogoch, II;Carey, ME;Chattaway, MA;Crump, JA;Diaz Guevara, P;Howden, BP;Izumiya, H;Jacob, JJ;Judd, LM;Kapil, A;Keddy, KH;Kim, JY;Levine, MM;Morita, M;Nair, S;Octavia, S;Okeke, IN;Osadebamwen, PE;Rahman, SIA;Rokney, A;Rasko, DA;Shamanna, V;Sikorski, MJ;Smith, AM;Sunmonu, GT;Tagg, KA;Wick, RR;Dyson, ZA;Holt, KE;Abbas, AH

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

Date:
2025-12-01

Citation:
Ingle, D. J., Hawkey, J., Hunt, M., Iqbal, Z., Keane, J. A., Afolayan, A. O., Ahmed, N., Andleeb, S., Ashton, P. M., Bogoch, I. I., Carey, M. E., Chattaway, M. A., Crump, J. A., Diaz Guevara, P., Howden, B. P., Izumiya, H., Jacob, J. J., Judd, L. M., Kapil, A. ,... Abbas, A. H. (2025). Typhi Mykrobe: fast and accurate lineage identification and antimicrobial resistance genotyping directly from sequence reads for the typhoid fever agent Salmonella Typhi. Genome Medicine, 17 (1), pp.130-. https://doi.org/10.1186/s13073-025-01551-4.

Persistent Link:
https://hdl.handle.net/11343/368016

Ingle *et al. Genome Medicine* (2025) 17:130
https://doi.org/10.1186/s13073-025-01551-4

Genome Medicine

SOFTWARE

Open Access

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

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

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

[†]Zoe A. Dyson and Kathryn E. Holt contributed equally to this work.

*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
Full list of author information is available at the end of the article

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 categorization (S/I/R) for eight drug classes, and we show strong agreement with phenotypic categorizations generated from reference laboratory minimum inhibitory concentration (MIC) data for $n = 1572$ 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.

Keywords Typhoid fever, *Salmonella* Typhi, Antimicrobial resistance (AMR), Whole genome sequencing (WGS), Pathogen sequencing, Genotyping, Lineage

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].

The GenoTyphi scheme was originally defined based on nearly 2000 WGS isolates from over 60 countries [18]. It uses a hierarchical framework based on the global phylogeny, stratifying the Typhi population structure into four primary lineages, 16 clades, and 49 subclades (henceforth referred to as genotypes), based on lineage-specific marker single-nucleotide variants (SNVs). A schematic mapping the hierarchical genotype labels to the Typhi phylogeny is shown in Additional File 1:Fig. S1. Under the GenoTyphi scheme, the lineage formerly known as 'H58' was designated genotype 4.3.1, and is further divided into multiple higher-resolution genotypes with the prefix 4.3.1 including 4.3.1.1, 4.3.1.2 (see Additional File 1:Fig. S1). The H58 lineage, or genotype 4.3.1 and associated sub-genotypes, has been associated with the spread of multidrug resistance (MDR, resistant to amoxicillin/ampicillin, chloramphenicol, trimethoprim-sulfamethoxazole) Typhi throughout Asia and

into Eastern Africa [15]. The scheme was later extended to provide further discriminatory power for identifying additional subpopulations of epidemiological importance [21–27], such as the XDR Typhi lineage associated with outbreaks in Pakistan (genotype 4.3.1.1.P1) [26, 27]. The scheme currently includes 87 genotypes [28], and its future maintenance and development will be managed by a working group within the Global Typhoid Genomics Consortium [29].

The GenoTyphi scheme specification consists of a set of marker SNVs defining the hierarchical genotypes, which in principle can be used by any bioinformatics tool to assign genotypes to new genomes. The original GenoTyphi paper [18] was accompanied by a Python-based pipeline (available at: https://github.com/typhoidgenomics/genotyphi) [30] designed for typing Illumina WGS data against the scheme. This implementation requires users to first map short reads or assemblies to the reference sequence of Typhi strain CT18 (accession AL513382.1) [31], then provide as input the resulting binary alignment map (BAM) or variant call file (VCF). Marker SNVs were identified from these input files and used to assign genomes to genotypes. Functionality was extended to include detection of mutations associated with reduced susceptibility to ciprofloxacin or azithromycin [22]. This mapping-based approach had various limitations, largely due to the need for users to pre-map their data to a specific reference, which was cumbersome (e.g., there is wide variation in the upstream mapping tools used and some users had issues with mapping tools creating incompatible output files, the BAM files are large and inefficient to process), and AMR analysis was restricted to mutation detection. The GenoTyphi scheme has also been implemented in other genotyping tools, including Typhi PathogenWatch [13], which works by analyzing assembled genomes, and the split-*k*-mer-based BioHansel tool, which does not include AMR detection [32].

Here, we present a novel approach to rapid genotyping of both AMR and GenoTyphi lineages from reads, implemented in the Mykrobe open-source software framework [33], dubbed Typhi Mykrobe. Input into this new tool accepts raw WGS reads (FASTQ files), and has been tested with reads generated by Oxford Nanopore Technologies (ONT) and Illumina platforms. The tool performs rapid (i) lineage genotyping using the GenoTyphi framework; (ii) detection of a broad range of clinically relevant molecular determinants of AMR, including both acquired genes and mutations; and (iii) detection of plasmid replicon markers common to Typhi. We validated GenoTyphi lineage assignment from reads by comparison with the original mapping-based implementation, using 12,839 Typhi genomes [14]. We assessed AMR genotype calls by comparison with assembly-based genotype calls from PathogenWatch, and with antimicrobial susceptibility phenotypes for 4018 isolates. Finally, we assessed performance on long read data generated using Oxford Nanopore Technologies (ONT) MinION devices, using isolates with matched ONT and Illumina data.

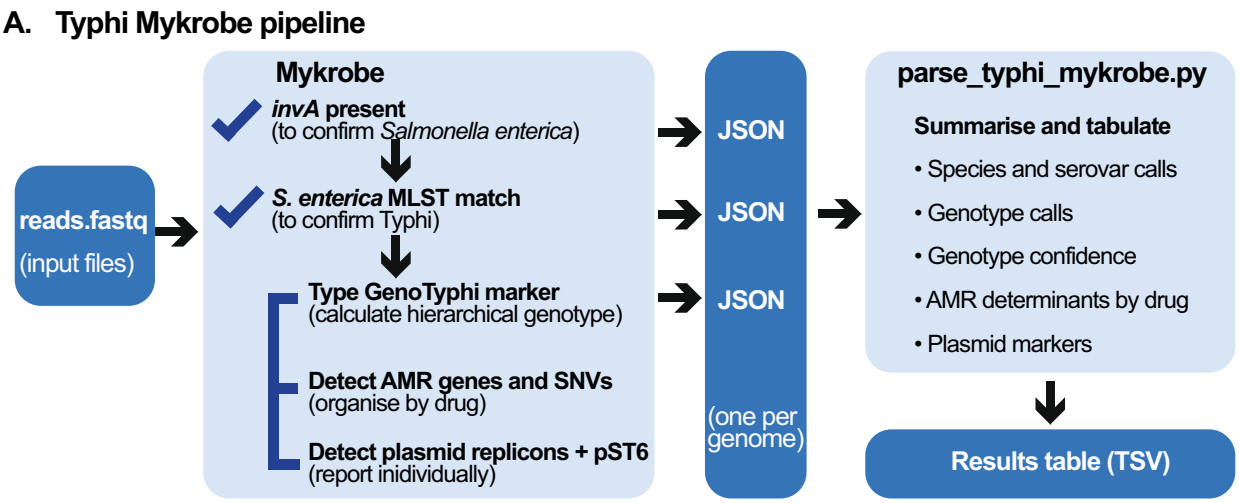
Implementation

As outlined in the Introduction, GenoTyphi is a typing scheme specific to Typhi, which specifies marker SNVs suitable to identify specific lineages. In principle, GenoTyphi can be used with any genotyping tool that is capable of searching for the marker SNVs in sequence data. GenoTyphi provides the framework directing genotyping tools what SNVs to look for and how to interpret them in terms of lineages to report. Mykrobe is a genotyping tool, which uses a *k*-mer-based approach to identify marker SNVs in sequence reads. It was developed for typing *Mycobacterium tuberculosis* and *Staphylococcus aureus* [34–36], but can in principle be used with any genotyping scheme. Here, we present a ready-to-run code base that allows users to type genomes from raw reads, using Mykrobe, against the GenoTyphi framework to identify and report known lineages—which we label "Typhi Mykrobe." This tool additionally includes AMR and plasmid typing frameworks, and generates predictions of resistance to the drugs recommended for clinical treatment of typhoid. Details of this Implementation are described below.

Genotyping targets

The current scope of the GenoTyphi scheme as discussed here ($n = 87$ genotypes) is illustrated in Additional File 1:Fig. S1. This includes five new genotypes (3.5.4.1, 3.5.4.2, 3.5.4.3, 4.3.1.2.1, 4.3.1.2.1.1) recently described in a technical report [28]. The full scheme specification is available in the GenoTyphi repository (https://github.com/typhoidgenomics/genotyphi) [30] in the file Genotype_specification.csv.

The AMR and plasmid targets included in Typhi Mykrobe are summarized in Fig. 1. These include sequence probes targeted to detect presence/absence of epidemiologically relevant mobile AMR determinants [13, 14], conferring resistance to the currently recommended drugs ciprofloxacin, ceftriaxone, and azithromycin, as well as older drugs ampicillin, chloramphenicol, trimethoprim-sulfamethoxazole, and tetracycline, as well as carbapenems. Fourteen SNVs associated with resistance to ciprofloxacin and azithromycin are also included. In addition, Typhi Mykrobe includes probes to detect the presence or absence of 13 plasmid replicons previously reported in Typhi [14], including 11 associated with AMR plasmids, the cryptic plasmid pHCM2, and linear plasmid pBSSB1, which carries the alternative flagellin



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

Fig. 1 Typhi Mykrobe functionality. **A** A schematic overview of the Typhi Mykrobe pipeline. **B** Table of epidemiologically important AMR determinants targeted for detection/typing by Typhi Mykrobe, grouped by drug.**C** Table of epidemiologically important plasmid replicons targeted for detection by Typhi Mykrobe

z66 [37]. A marker SNV to identify the plasmid sequence type 6 (PST6) lineage of the dominant IncHI1 plasmid is also included [38]. The full list of AMR targets is available in the GenoTyphi repository [30] in the file AMR_genes_mutations_plasmids.csv.

Implementation in Mykrobe

The Mykrobe software, first developed in 2015 for *Mycobacterium tuberculosis* and *Staphylococcus aureus* [34], uses a *k*-mer-based approach to identify marker SNVs and the presence or absence of target genes using "probe" sequences to screen against [33]. These data are then used to assign hierarchical genotypes and report AMR determinants. Probe sets have been previously developed and implemented for other bacterial pathogens including *Mycobacterium tuberculosis* [34–36], *Staphylococcus aureus* [33], *Shigella sonnei* [39], and *Salmonella* Paratyphi B [40]. Mykrobe is available at https://github.com/Mykrobe-tools/mykrobe [41].

Mykrobe probe sets were created for genotype and AMR marker SNVs using the *mykrobe variants make-probes* command using a *k*-mer size of 21 and the Typhi

CT18 reference genome (accession: AL513382.1). The nested hierarchical relationships of the marker SNVs were specified and Mykrobe used this information to identify the best supported genotype call (e.g., a call of "4.3.1.1" is typically supported by detection of nested markers that demarcate different levels of the hierarchy: 4, 4.3.1, 4.3.1.1). Briefly, probe sets for gene presence or absence were created using a Python script that extracted the relevant AMR sequences. Details of the specific commands and Python scripts are provided in the GenoTyphi GitHub [30]. Probe sets were also created to confirm input read sets as *S. enterica* using the presence of marker gene *invA*, involved in invasion of intestinal epithelium cells by *S. enterica*, and serovar Typhi based on the 26 known multi-locus sequence types (STs) (see Fig. 1). Further details of the preliminary testing and implementation of the Typhi probe sets are available in the technical report [28].The Typhi Mykrobe probe sets and typing panel are available at https://doi.org/10.6084/m9.figshare.27102334 [42].

Summarizing Mykrobe outputs

To support interpretation of the Mykrobe output, we developed a Python-based parser script *parse_typhi_mykrobe.py*, which takes as input a set of JavaScript Object Notation (JSON) files output by Mykrobe, and summarizes them in a tab-delimited table with one row per genome, and columns to indicate the genotype calls (see Fig. 1). Also included are several quality-control fields reported by Mykrobe, including read-level support for each marker SNV contributing to the final genotype call ("node support"), and for any additional markers detected. This information is helpful for identifying contaminated samples (see examples in " Results" section). The parser script also assigns a summary "confidence" level in the final GenoTyphi lineage call, based on the Mykrobe quality scores for each nested component marker that makes up the final genotype call. In brief, a "weak" confidence call is assigned when one or more marker SNVs are of low quality, defined as Mykrobe quality score of 0 or Mykrobe quality score of 0.5 with minority support (i.e., < 50% of reads) for the derived allele. A "moderate" confidence call is assigned if one (and only one) marker SNV has a Mykrobe quality score of 0.5 but with majority support (i.e., ≥ 50% of reads) for the derived allele. Otherwise, the genotype call is assigned "strong" confidence.

The parser script summarizes the AMR markers detected by Mykrobe in the form of a pseudo-antibiogram with one column per drug. Each drug column contains either "S" (indicating no markers detected) or "R" (detected markers). For ciprofloxacin, it is useful to distinguish between low-level resistance (MIC 0.06–0.5 mg/L), which is associated with a single mutation or acquired gene and high-level resistance (MIC > 0.5 mg/L), which is associated with combinations of two or more determinants [14, 17], because while both groups are associated with increased fever clearance times and probability of treatment failure compared with susceptible strains, high-level resistance is associated with worse outcomes than low-level resistance [43]. Current guidelines and terminology for these groups vary, but for simplicity, the Mykrobe parser reports these groups as "I" (for intermediate) and "R," respectively. This matches the current Clinical and Laboratory Standards Institute (CLSI) guidelines [44]. However, it should be noted that (i) The European Committee on Antimicrobial Susceptibility Testing (EUCAST) defines the clinical breakpoint for "R" in *Salmonella* spp. as MIC > 0.06 mg/L [45], which includes both low-level and high-level resistance; and (ii) previous literature has referred to the low-level range as "decreased ciprofloxacin susceptibility" rather than "I" [46]. In addition to this antibiogram-style summary, individual AMR and plasmid markers are reported in their own columns, coded as 1 = present and 0 = absent (as seen in Additional File 2:TableS1, available at the Typhi Mykrobe GitHub [47].

Validation data

Validation on Illumina short-read data was conducted using the Global Typhoid Genomics Consortium dataset described in Carey et al. [14]. Reads were mapped to the CT18 reference genome (accession: AL513382) [31] using bwa-mem v0.7.17 via the Centre for Genomic Pathogen Surveillance (CGPS) mapping pipeline v1.2.2 (https://gitlab.com/cgps/ghru/pipelines/snp_phylogeny) [48], to generate alignments in BAM format. Assemblies were generated using the CGPS assembly pipeline v2.1.0 (https://gitlab.com/cgps/ghru/pipelines/dsl2/pipelines/assembly/) [49] which utilizes the SPAdes assembler (v3.12.0) [50, 51]. A total of 141 isolates were excluded due to excessive unclustered heterozygous SNVs indicative of mixed cultures ($n = 92$), assembly issues ($n = 46$) or both ($n = 3$), resulting in 12,839 isolates for Mykrobe validation (listed in Additional File 3:TableS2, available at the Typhi Mykrobe GitHub [47]. The genomes of isolates called as "not Typhi" were characterized for presence of the marker genes of *invA* and the seven MLST genes using Blast in Bandage for *invA* [52] and MLST with the "senterica" database (https://github.com/tseemann/mlst) [53]. The dataset covered 85 of the 87 defined genotypes (exceptions being genotypes 1 and 2.3, which represent internal nodes of the hierarchy for which no genomes have been observed). Individual calls are given in Additional File 2:TableS1 and Additional File 3:TableS2.

Validation of genotype calls

Typhi Mykrobe v0.12.1 was run on all ($n=12{,}839$) Illumina FASTQ files using the new Typhi panel (v20221208) and the results were summarized using *parse_typhi_mykrobe.py*. To validate Typhi Mykrobe's GenoTyphi genotype calls, we compared them with those called by the original mapping-based approach. To do this, we used BAM file read alignments generated from the same $n=12{,}839$ FASTQs as described above as input to the Python script genotyphi.py (v2.0; available at GenoTyphi GitHub [18, 22, 28, 30]. To validate the AMR and plasmid calls, we compared them to those called in corresponding genome assemblies by Pathogenwatch (available for $n=11{,}992$ genomes, assemblies for the other genomes were excluded as they did not meet Pathogenwatch quality-control criteria [13]). We downloaded the GenoTyphi, AMR, and plasmid replicon typing reports generated by Basic Local Alignment Search Tool (BLAST) searching assemblies for these genomes from Pathogenwatch on 7 March 2023 (GenoTyphi v20221208 specification for lineage genotypes). Categorical agreement (percentage of genomes where both tools agreed on the same call of presence or absence) was calculated separately for each marker, and collectively by drug (percentage of genomes where both tools agreed that at least one marker was present or that no markers were present) (Additional File 4:TableS3, available at Typhi Mykrobe GitHub [47]. Probes for plasmid-borne carbapenemase genes circulating in Enterobacterales (*bla*NDM-5, *bla*KPC-2, *bla*VIM-1, *bla*IMP-27, *bla*OXA-48) were added to GenoTyphi v20240407 [28] following the report of *bla*NDM-5 in Typhi from Pakistan [11], and validated using the genomes of the two distinct morphologic variants from that study (accessions: SRR22801766 and SRR22801806). Comparisons and statistical analyses were conducted in R v4.1.0 using tidyverse v1.3.1 [54, 55].

Benchmarking of run time for Mykrobe

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

Accuracy of AMR phenotype prediction

Antimicrobial susceptibility testing (AST) data were available for $n=3970$ of the Illumina-sequenced isolates, sourced from three separate datasets (Additional File 5:TableS4, available at Typhi Mykrobe GitHub [47]. Minimum inhibitory concentration (MIC) data were available from the UK Health Security Agency (UKHSA) for $n=852$ isolates following EUCAST standards and interpretive criteria (v10, 2020) [56] and the US Centers for Disease Control and Prevention (US CDC) for $n=720$ isolates following CLSI standards and interpretive criteria (M100, 2023) [57]. Disk diffusion data following EUCAST standards and interpretive criteria (v8.0, 2018) were available for $n=2446$ isolates collected and analyzed from three countries as part of the Surveillance for Enteric Fever in Asia Project (SEAP), with multiple source labs contributing AST data and isolates for sequencing [6]. Typhi Mykrobe AMR calls following the S/I/R categorization described above were compared with the phenotypic antimicrobial susceptibility S/I/R categorization using the relevant CLSI or EUCAST standards for each data set, for the currently recommended drugs azithromycin and ciprofloxacin and ceftriaxone, and the older drugs ampicillin, chloramphenicol, trimethoprim-sulfamethoxazole, and tetracycline (note UKHSA tested amoxicillin rather than ampicillin; and SEAP did not test tetracycline). To assess concordance between genotype and phenotype, we followed the principles used for assessment of new AST devices by the US Food and Drug Administration (FDA), which considers (i) categorical agreement, (ii) major error rate (defined as the proportion of S isolates that test R with the new method [58, 59]), and (iii) very major error rate (defined as the proportion of R isolates that test S with the new method) (Additional File 6:TableS5, available at Typhi Mykrobe GitHub [47].The FDA standard requires > 90% overall categorical agreement (within one doubling dilution), < 3% major errors for each drug, and < 1.5% very major errors for each drug; hence, we adopted these thresholds to assess the accuracy of the genotype-based prediction of phenotype. For drugs with high error rates, we investigated the relationship between individual resistance markers and phenotype, which we visualized using upset plots generated in R v4.1.0 using the ComplexUpset package v1.3.3 (https://doi.org/10.5281/zenodo.3700590) [60]. For five isolates with unexplained trimethoprim-sulfamethoxazole resistance, we used nucleotide blast search of the genome assemblies to confirm the presence of wild-type chromosomal *folA* and *folP* (100% nucleotide match to reference sequences from Typhi strain Ty2) and screened the assemblies for other AMR determinants using Comprehensive Antibiotic Resistance Database (CARD) RGI v6.0.3 [61].

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 Additional File 7:TableS6, available at Typhi Mykrobe GitHub [47]. 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) [62] 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 [63]. These five genomes were also aligned to the CT18 reference genome using minimap2 [64, 65] to determine the read coverage of the *invA* and the seven core genes in the *Salmonella enterica* MLST scheme [66]. 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 [67] 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 (available at the Filtlong GitHub [68]) to –target_bases 480,000,000 (resulting in 99.81×depth). The genomes of isolates called as "not Typhi" were characterized for presence of the marker genes of *invA* and the seven MLST genes using Blast in Bandage for *invA* [52] and mlst with the "senterica" database [53].

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 (Table 1,Additional File 2:TableS1). 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 analyzed further. Of the remaining $n = 12{,}834$ genomes, Mykrobe called the same single genotype as the mapping-based approach for $n = 12{,}807$ (99.79%) (Table 1). For a further $n = 17$ genomes, Mykrobe and the mapping-based pipeline gave concordant reports of two genotypes being present, consistent with mixed samples (in the Typhi Mykrobe output this is reported as one primary genotype plus additional markers, while in the mapping-based pipeline the result is reported as a list of supported genotypes). Eight genomes belonged to genotype 3.5.3, which was nested in 3.5.4, breaking the usual hierarchical structure of the scheme. Typhi Mykrobe correctly reported these genomes as genotype 3.5.3 with support for both markers 3.5.4 and 3.5.3, as it has this nested relationship explicitly encoded in its hierarchy; whereas the mapping-based pipeline does not have this logic and incorrectly reported these genomes as genotype 3.5.4.

There were two instances of genuine disagreement in the genotype calls generated by Typhi Mykrobe as compared to the mapping-based approach. One isolate (ERR5243665) was called as genotype 2 by Mykrobe with strong confidence (no reads support for any alternative alleles or markers). It was genotyped as 0.1.3, albeit with low support, by the mapping-based approach, but the genotype 2 marker was also detected. This isolate clustered with other genotype 2 isolates in a distance-based phylogeny of genomes from the same study [6] (see tree at Pathogenwatch [69]), suggesting that the Mykrobe call is correct. The other isolate (ERR2663783) was called as genotype 4.3.1.1 by Typhi Mykrobe with strong confidence. It was reported as "4.3.1.1,4.3.1.3.Bdq" (with low support) by the mapping approach; however, the support for the 4.3.1.3.Bdq marker was low. This isolate clustered closely with other 4.3.1.1 isolates from the same study [25] (see tree at Pathogenwatch [70]), which is consistent with the Mykrobe call.

Table 1 Validation of genotype calls. Summary of Typhi Mykrobe genotype calls (GenoTyphi scheme) for 12,839 Illumina readsets, 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

We demonstrate that the run time for Typhi Mykrobe's "mykrobe predict" command on a modern computer was < 1 min to complete (Additional file1:Fig. S2). The "mykrobe predict" command can be run with multiple threads using the "–threads"/"-t" option. Up to ~ "4 threads will increase performance. Using multiple threads on a very fast CPU, the run time for each genome was reduced to seconds (Additional file1:Fig. S2). Further, "mykrobe predict" is very memory-efficient and will typically use less than 100 MB of RAM per genome.

Validation of AMR genotyping

To assess the accuracy of Mykrobe for detecting the presence of AMR genes and mutations, we compared genotype calls from Mykrobe (called from *k*-mer analysis of reads) with those from Pathogenwatch (called from BLAST analysis of genome assemblies of the same read sets). We found very strong categorical agreement between the two methods at the individual marker level, with 99.88% of the 299,004 marker-genome combinations yielding the same call (summarized in Table 2, full details in Additional file 3:TableS2). High concordance was evident for all genetic markers of AMR (98.56–100%,

Table 2 Validation of AMR genotyping (vs assembly based). Comparison of AMR genotyping by Typhi Mykrobe (*k*-mer-based analysis of reads) vs Pathogenwatch (based on nucleotide blast search of assemblies), for n=11,992 isolates. Complete details of the call comparisons are given in Additional file 4: Table S3. Categorical agreement at drug level is calculated based on agreement of the binary S/R categorisation of the 11,992 isolates by each method (i.e. 'Concordant' means Typhi Mykrobe and Typhi Pathogenwatch made the same call of S or R based on absence or presence, respectively, of known genetic determinants; 'Discordant' means one method identified determinants and called R, while the other did not and called S).*For ciprofloxacin, both tools categorise as S/I/R as defined in Methods, and categorical agreement was calculated across these three categories. **Results for *dfrA* genes are summarised together, and as some genomes have multiple *dfrA* alleles the total observations is greater than the number of isolates (see details in Additional file 4: Table S3)

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	11,982	10	99.92%
		GyrB-464	11,984	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%

Table 2), as well as plasmid replicon markers (99.74–100%, Table 3). The test dataset included all AMR and plasmid markers in GenoTyphi v2.0 implemented Typhi Mykrobe v0.12.1. The *ereA*, *ermB*, and carbapenemases are very rarely reported in Typhi but were included in the Mykrobe panel in case they emerge.

Both Mykrobe and Pathogenwatch summarize AMR data in terms of a binary phenotype categorization (S/R), assuming the presence of a marker associated with resistance to a drug implies resistance "R," with the exception of ciprofloxacin, where presence of any marker is categorized as "I" and specific combinations of markers are categorized as "R," resulting in three categories S/I/R (see "Methods" section). Categorical agreement was very high between the drug-level categorizations reported by both genotyping tools, with > 99.7% agreement ($n = 2$ to 33 discordant genomes) for each drug (Table 2,Additional file 4:TableS3), and for subsequent categorization as MDR (99.86%, $n = 17$ discordant) or XDR (99.97%, $n = 4$ discordant).

Table 3 Validation of plasmid replicon detection (vs assembly based). Comparison of plasmid replicon markers identified in the genomes by Typhi Mykrobe (*k*-mer-based analysis of reads) vs Pathogenwatch (PW, based on nucleotide blast search of 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%

Validation of AMR phenotype prediction

To assess the accuracy of Mykrobe's genotype-based clinical resistance categorizations, we compared them with AST data, which was available for a subset of $n = 4018$ isolates (see "Methods" section, Additional file 5:TableS4). These isolates were from three data sets, using different methods (MIC with EUCAST standards, MIC with CLSI standards, disk diffusion with EUCAST standards), hence we analyzed the results from each dataset separately. Categorical agreement was high (> 95%) for all drugs across all three datasets, and very high (> 99%) for the two sets of reference-laboratory MIC measures (Fig. 2,Additional file 6:Table S5). The only exception was US CDC categorization of low-level ciprofloxacin resistance (using the CLSI "I" threshold, ≥ 0.125 to 0.5 mg/L), which showed 97.36% agreement with Mykrobe calls of "I" or "R." This was mainly due to the presence of major errors (susceptible isolates with AMR determinants detected, thus reported as R by Mykrobe, see Methods), whereby $n = 17$ isolates tested susceptible

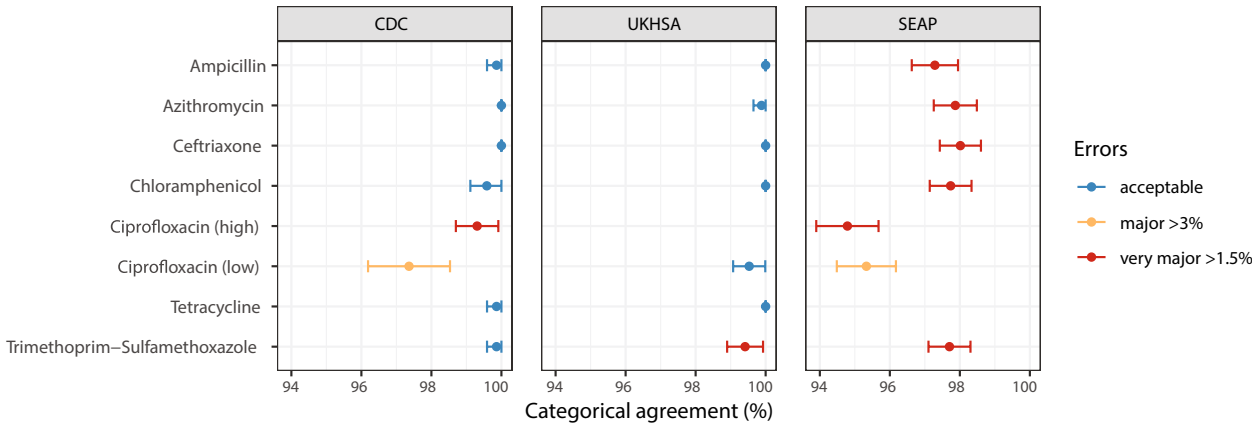


Fig. 2 Validation of Typhi Mykrobe AMR predictions vs AMR phenotypes. Summary of Typhi Mykrobe's genotype-based clinical resistance categorisations compared with susceptibility phenotypes, for 4,018 isolates with publicly available matched genome and phenotype data. Comparisons are summarised as categorical 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

to ciprofloxacin ($n=14$ with MIC = 0.06 mg/L, $n=3$ with MIC = 0.015 mg/L) but Mykrobe detected a known QRDR mutation (Fig. 3b); categorical agreement within one doubling dilution was 99.4%.

Major errors were infrequent (< 3% prevalence) for all drugs in all datasets (Fig. 2), except for ciprofloxacin (Additional file 6:TableS5). Most of the ciprofloxacin major errors were in isolates whose MIC or disk zone measurement were close to the breakpoint (Figs. 3 and 4c); all carried well-characterized molecular resistance determinants (QRDR mutations and/or *qnr* genes, see Figs. 3 and 4c), and in all but one case the same determinants were also detected in the corresponding genome assemblies by Pathogenwatch (details below).

Very major errors (resistant isolates with no AMR determinants detected, thus reported as S by Mykrobe) were also infrequent (< 1.5% prevalence) for the reference-laboratory MIC datasets, but errors were higher in the disk diffusion dataset (Fig. 2). The only drugs showing major or very major errors compared with the reference-laboratory MIC data were ciprofloxacin and trimethoprim-sulfamethoxazole (Fig. 2). For ciprofloxacin, the US CDC categorization of high-level resistance (using the CLSI "R" threshold, ≥ 1 mg/L) yielded four very major errors (R isolate called as "I" by Mykrobe). Two had a single mutation detected (GyrA-S83F) and two had double QRDR mutations (GyrA-S83F plus GyrA-D87N or ParC-S80I) (see Fig. 3); Pathogenwatch calls agreed in all instances. Importantly, all four isolates had MIC = 1 mg/L, i.e., they agreed with the genotype-based call within one doubling dilution. UKHSA uses the EUCAST standards, which have a single (low-level) breakpoint for ciprofloxacin (R > 0.06 mg/L), and this yielded just two major errors (2.9%) and two very major errors (0.26%) (see Fig. 3).

For trimethoprim-sulfamethoxazole, Mykrobe analysis of the US CDC dataset yielded a single error (R isolate called as S) but the MIC was 4 mg/L, i.e., within one doubling dilution, and the isolate had no *sul* or *dfr* genes that were identified by Mykrobe or Pathogenwatch. Mykrobe analysis of the UKHSA dataset yielded five errors for trimethoprim-sulfamethoxazole; all were R isolates called as S by Mykrobe, which detected *sul* genes but no *dfrA* genes. Interestingly, all five isolates tested susceptible to trimethoprim alone (MIC ≤ 0.5, well below the > 4 mg/L breakpoint), supporting the absence of *dfrA* genes; these were also not detected by Pathogenwatch analysis of the corresponding genome assemblies, or by UKHSA's internal pipeline [44]. Further investigation of these genome assemblies confirmed the presence of wild-type *folA* and

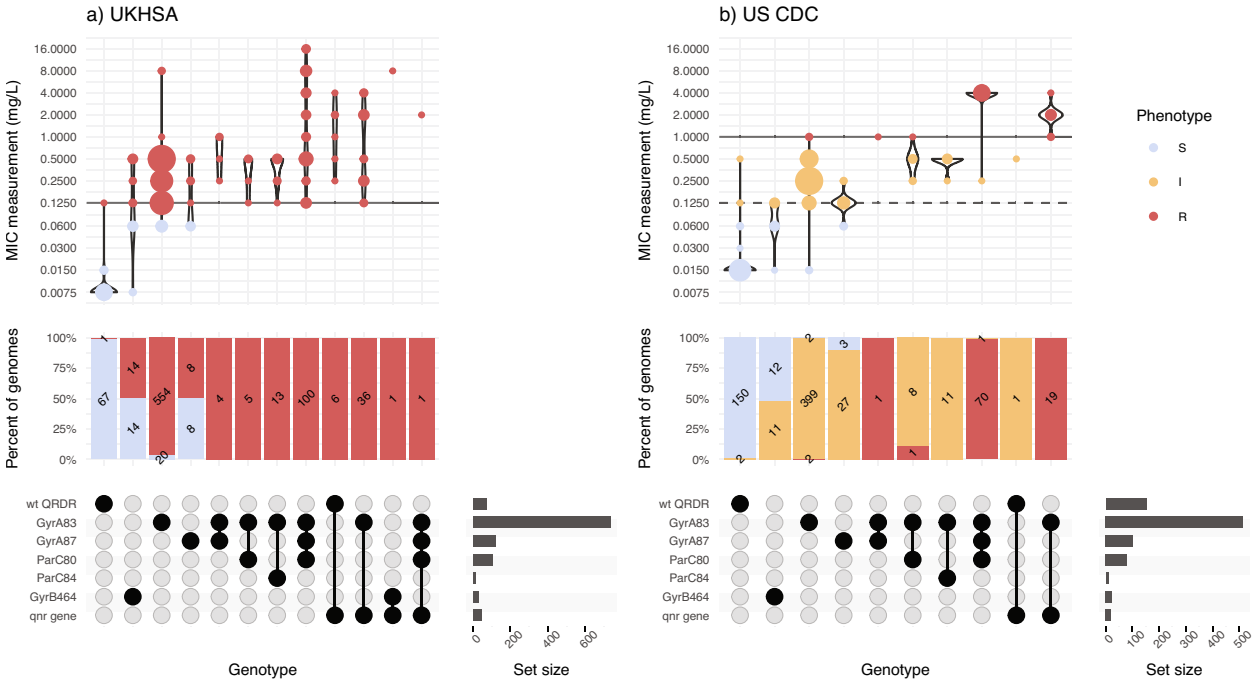


Fig. 3 Typhi Mykrobe genotypes vs minimum inhibitory concentration from two reference laboratories, for ciprofloxacin. **a** UKHSA dataset, using EUCAST. **b** US CDC dataset, using CLSI. Each column represents the set of isolates in which Typhi Mykrobe identified a unique combination of genetic determinants, 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

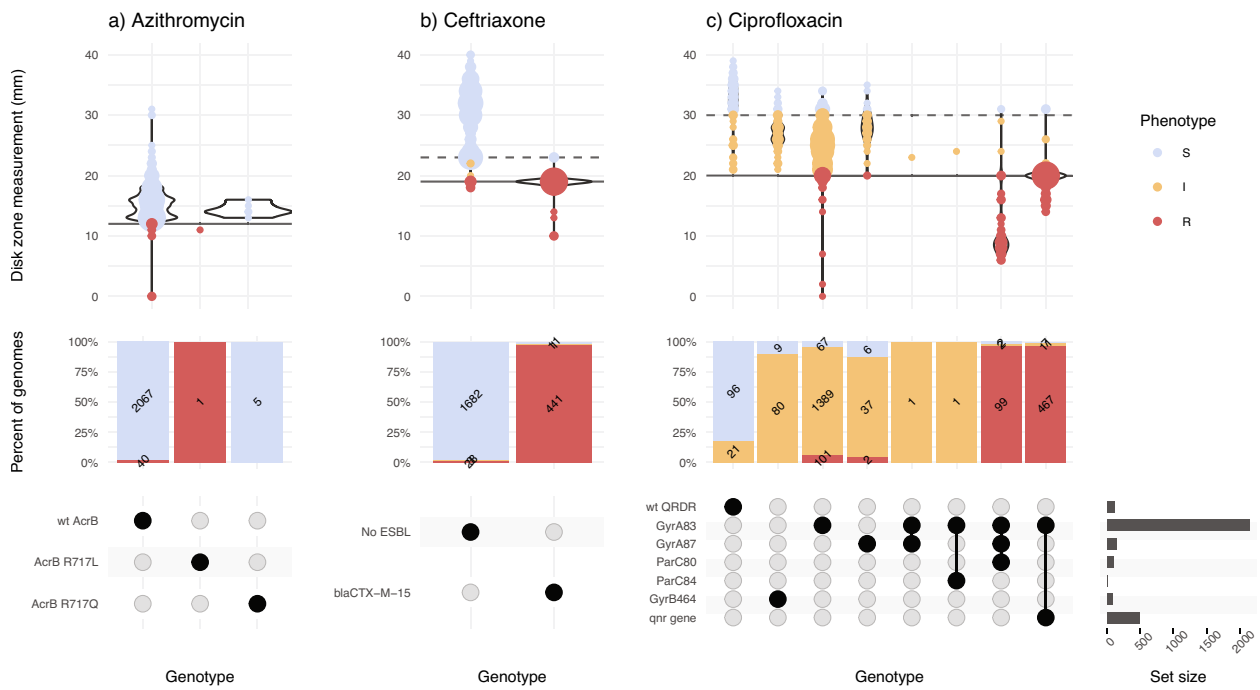


Fig. 4 Typhi Mykrobe genotypes vs disk diffusion data from the SEAP study, for three drug classes. Data for three drugs are shown in panels **a)** Azithromycin,**b** Ceftriaxone, and **c)** Ciprofloxacin. Each column represents the set of isolates in which Typhi Mykrobe identified a unique combination 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

folP, and no additional determinants were reported by CARD RGI (v6.0.3).

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

Validation on nanopore reads

We explored the accuracy of Typhi Mykrobe for genotyping ONT long-read data by comparing results from 98 pairs of ONT to Illumina reads generated from the same isolates. Data were contributed by members of the Global Typhoid Genomics Consortium and generated using different library preparation methods and sequencing devices (see Additional file 7:TableS6). Mykrobe called one Illumina read set and five ONT read sets as non-Typhi (mean ONT read depths estimated as 0.04×, 4.06×, 8.01×, 26.40×, 27.85×, all sequenced on R9 flowcells); these were excluded from analysis of genotype accuracy as no genotype is called on isolates identified as non-Typhi.

Among the $n=92$ Typhi genomes called from both ONT and Illumina reads, the Typhi Mykrobe genotype lineage calls matched for $n=89$ (96.7%). For all three discrepant isolates (all sequenced on R.9 or R.9.4 flowcells), the set of SNV markers detected from ONT reads closely matched those detected from Illumina reads, but with one or two markers missing from the ONT-based calls and no unexpected markers reported resulting in final genotype calls of 4.3.1.1 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 variable ONT read depths, with ranges 3–9, 4–31 and 1–14, suggesting that the errors are likely due to lack of informative ONT reads at the marker loci. Notably, the overall ONT read depths for these were reasonable (mean 35×, 170×, and 90× respectively), suggesting the lack of informative reads for genotyping may be caused by contamination from non-Typhi reads, or high error rates resulting in low rates of *k*-mer matching.

In this study, the five ONT genomes that had "not Typhi" calls were further explored. These were all classified as *Salmonella enterica*. However, Typhi was not the top serovar from read-level classification, instead *Salmonella* Infantis was the top serovar for BZD62J, while *Salmonella* Typhimurium was the top serovar for the remaining four. Further, the presence of the marker genes of *invA* and the seven MLST genes were detected in the genomes at read depths consistent with the genome mean read depth reported in Additional file 7:TableS6, with the gene *invA* was detected in three of the five genomes from Blast analysis of the Flye assemblies. The gene *invA* was not detected in the 611,427 ONT assembly and only had a read depth of 0.43×, and BZD62J failed to assemble due to insufficient reads and was found to have no reads mapping to *invA.* MLST analysis of the three complete genomes with *invA* detected reported, were found to have partial matches to known alleles for all seven MLST genes and no reported MLST profile. Of note, these five genomes had low read depth (0.04–28×), sequencing on older R9 flowcells, and basecalling with older versions of guppy (Additional file 7:TableS6). The failure of Typhi Mykrobe to identify these as Typhi is likely due to a combination of potential contamination, low read depth, and high error rates resulted in low rates of *k*-mer matching to the probes. This highlights the importance of QC of the read data, as Typhi Mykrobe provides some metrics for confidence of the calls, but does not itself perform detailed QC.

AMR marker detection from ONT reads was highly concordant with Illumina-based calls, with 98.77% overall agreement across the $n=736$ isolate-drug combinations (92 isolates across 8 drugs; see Additional file1:TableS7). There were nine genomes with disagreement of AMR calls; five of these had a marker detected in Illumina reads but not ONT reads ($n=1$ *catA1*, $n=1$ *qnr*, $n=3$ QRDR mutations), three had GyrB-464F detected in ONT reads but not Illumina, and one had GyrB-S464F detected in Illumina reads and GyrA-S83F in ONT. In all nine cases, the Typhi Mykrobe call from Illumina reads matched that from Pathogenwatch analysis of the corresponding assemblies of Illumina reads. There was no overlap between ONT read sets with AMR discrepancies and those with lineage genotype discrepancies. Overall, ONT-based resistance prediction showed complete agreement with Illumina for six of the eight drugs, 98.9% agreement for chloramphenicol resistance, and 91% for ciprofloxacin resistance.

Plasmid marker detection from ONT data was also highly concordant with Illumina-based calls, with 99.67% overall agreement across the $n=1196$ isolate-marker combinations (92 isolates across 13 markers; see Additional file 1:TableS8). Four genomes had discrepant plasmid marker calls, all with a marker detected from Illumina reads but not from ONT reads. One isolate (DRR070993) had a typical IncHI1 MDR plasmid profile reported from Illumina reads (IncHI1A, IncHI1BR27, IncHI1_ST6 indicating presence of IncHI1-ST6 plasmid, with *catA1*, *dfrA7*, $bla_{\text{TEM-1}}$, *sul1*, *sul2*), but the ONT read profile was missing the *IncHI1A* and *catA1* markers (and missing *catA1* in the corresponding Illumina assembly and a mismatch for *IncHI1A*). The other three isolates had pHCM2 detected in the Illumina reads and in the corresponding assemblies [14] but not in ONT reads; for these isolates the ONT and Illumina sequencing was done on different DNA extracts so this may reflect plasmid loss in the culture used for ONT sequencing.

Discussion

Typhi Mykrobe provides detection and genotyping of key features of clinical and public health importance, including assigning lineages, AMR determinants, and plasmid replicons (Fig. 1). Compared with other tools for Typhi genotyping, Mykrobe has the advantage of not requiring any pre-processing, aligning, or assembling of reads, but instead takes as inputs raw FASTQ files. This simplifies bioinformatics workflows and reduces time-to-result as Typhi Mykrobe can return a full genotype profile in typically < 1 min. Importantly, although Typhi Mykrobe is not a QC tool per se, it provides useful data on the read-level support for individual genotype markers, which can be used to help identify potentially mixed samples. As a command-line tool that is easy to install locally using Bioconda, Typhi Mykrobe is suitable for research and public health laboratory settings, facilitates standardized reporting of Typhi genotypes without the need to share sequence data, and enables broader access to bioinformatics functionality across a variety of settings.

Typhi Mykrobe showed very high genotyping accuracy, for GenoTyphi lineage assignment (Table 1), and detection of AMR and plasmid markers (Tables 2 and 3), compared with existing genotyping tools. Of note, the included IncY, pO111, and IncFIB for the cryptic pHCM2 replicon targets have recently been shown to be sometimes carried on phage-plasmids [71, 72], in addition to conjugative plasmids such as the IncY plasmids associated with XDR Typhi [26]. Future work is required to identify additional targets to delineate the phage-plasmids, and the Typhi Mykrobe software can be readily updated to include new markers.

In addition, we show that the panel of AMR markers currently included in Typhi Mykrobe can be used to predict AMR phenotypes with high accuracy, demonstrating very high concordance with S/I/R calls made from reference-laboratory MICs (Fig. 2). Concordance was slightly lower when compared with research laboratory

disk diffusion data. However, as the phenotypic resistance prevalence rates were similar across the three datasets, it is reasonable to assume that the higher error rates in the disk diffusion dataset compared with the two reference-laboratory MIC datasets most likely reflect comparatively lower accuracy of the disk diffusion phenotype measurements, which are known to yield lower accuracy for some drugs such as azithromycin [73, 74] and in this case were completed across multiple laboratories. There are currently no formal standards for assessing the accuracy of AMR phenotype predictions from genomic data [75]; however, applying the standards for FDA licensing of new AST devices, Typhi Mykrobe easily passed the target threshold for categorical agreement (> 99% agreement within one doubling dilution for reference-laboratory MIC data, > 95% agreement for disk-diffusion data; vs. > 90% target) (Fig. 2).

Typhi Mykrobe also showed very low rates of major errors and very major errors compared with the reference-laboratory MIC data, and for nearly all drugs these fell below the acceptable thresholds for a new AST device (< 3% and < 1.5%, respectively) (Fig. 2). The exception was trimethoprim-sulfamethoxazole, for which five UKHSA isolates tested resistant but Typhi Mykrobe predicted as susceptible due to a lack of *dfrA* genes, which are thought to be required for resistance to trimethoprim-sulfamethoxazole along with *sul* genes, which were detected in these genomes. Other genotyping tools (Typhi Pathogenwatch, CARD RGI) also could not identify any known mechanisms of resistance to trimethoprim or trimethoprim-sulfamethoxazole in these genomes, and the isolates tested sensitive to trimethoprim. Therefore, we hypothesize these isolates may harbor efflux mutations or a novel mechanism of resistance to trimethoprim-sulfamethoxazole that does not result in resistance to trimethoprim alone. Further analyses to determine the exact resistance mechanism was beyond the scope of this study. Ciprofloxacin showed very low error rates in the UKHSA MIC data, interpreted using the single low-level resistance threshold (MIC > 0.06 mg/L) recommended by EUCAST, but slightly higher error rates for the US CDC MIC data, which was interpreted using low-level and high-level resistance thresholds (although still below the acceptable thresholds within one doubling dilution). Carbapenemase genes were added to the Typhi Mykrobe panel, but this resistance is newly emerging and there was only a single carbapenem-resistant isolate sequenced at the time of testing (in which *bla*NDM-5 was correctly identified by Typhi Mykrobe), therefore we could not yet assess accuracy of phenotype prediction for carbapenems.

Finally, we demonstrated that Typhi Mykrobe can be used to genotype ONT reads, showing high concordance with results obtained from the analysis of Illumina reads generated from the same isolates (96.74% agreement on GenoTyphi lineage calls, 98.77% on AMR markers and 99.67% on plasmid markers). We did not have suitable data to directly assess ONT-based AMR predictions with AST data. ONT-based and Illumina-based drug-level predictions of AST phenotypes showed high agreement; however, future studies assessing ONT data compared to reference-laboratory MIC data would be useful to confirm the accuracy of ONT data for phenotype prediction. As one might expect, many of the errors occurred in isolates with lower read depth. However, as the matched ONT-Illumina data available for testing were contributed by different laboratories that used different ONT protocols and devices for library preparation, sequencing, and base calling, we were unable to perform a detailed assessment of the impact of read depth and quality on genotyping performance. The accuracy of ONT reads has greatly increased in recent years and is approaching Illumina read accuracy of ~ 99%; hence modern ONT data would be expected to perform similarly to Illumina data [76]. Future studies assessing the impact of read depth and base calling methods on the performance of Typhi Mykrobe on ONT data would be desirable to help guide laboratories on the most appropriate methods and target yield for ONT experiments aiming to produce reliable genomic surveillance data for Typhi.

Conclusions

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

Abbreviations

AMR	Antimicrobial resistance
AST	Antimicrobial susceptibility testing
CARD	Comprehensive Antibiotic Resistance Database
CLSI	Clinical and Laboratory Standards Institute
EUCAST	The European Committee on Antimicrobial Susceptibility Testing
FDA	United States Food and Drug Administration
MDR	Multidrug resistance
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

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-025-01551-4.

Additional file 1: Includes Fig S1 and S2 and Tables S7 and S8 (pdf). Table S1. Tabulated Typhi Mykrobe output table for all genomes included in validation analyses (available in GitHub). Table S2: Genome data used for validation (available in GitHub). Table S3. Details of AMR genotype calls comparison (available in GitHub). Table S4. Genome data for all isolates with publicly available antimicrobial susceptibility testing (AST) (available in GitHub). Table S5. Comparison and error rates for AMR genotype and phenotype data (available in GitHub). Table S6. Details of validation of typing from nanopore reads (available in GitHub).

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, Iraq2 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, USA4 University of Cambridge, JCBC, Puddicome way, Cambridge, CB2 0AW, United Kingdom5 OUCRU-Nepal, Jhamshikhel, Lalitpur, Nepal6 National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074, Republic of Singapore7 Clinical Infection Research Group, School of Medicine and Population Health, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom8 Federation University Australia, Gippsland Campus, Churchill, Victoria, Australia9 Postgraduate Institute of Medical Education and Research, sector 12, Chandigarh. CMC Vellore, India10 Department of Molecular Biology and Immunology, Faculty of Medicine, Hasanuddin University, Makassar, 90245, Indonesia11 Fauji Foundation, Rawal Road, Rawalpindi, Pakistan12 RDT Hospital, Bathalapalli, Srisathya Sai District, 515661, India13 Centers for Disease Control and Prevention, Atlanta, United States of America14 Department of Microbiology, Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of the Congo15 MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, W12 0BZ, United Kingdom 16 World Health Organisation, Zimbabwe17 Cambridge Institute of Therapeutic Immunology and Infectious Disease (CITIID), University of Cambridge, Cambridge CB1 0AW, United Kingdom18 Institut Pasteur, France19 Oxford Vaccine Group, Department of Paediatrics, University of Oxford and the NIHR Oxford Biomedical Research Centre, Oxford, UK20 RDT Hospital, Bathalapalli, Srisathya Sai District, 515661, India21 Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia22 Centre of Excellence in Bionanoscience Research, King Abdulaziz University, Jeddah 21589, Saudi Arabia23 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, Rwanda25 Institute for Research in Science and Technology, Sachal-2, Lalitpur, Nepal26 Child Health Research Foundation, Dhaka, Bangladesh27 CMAI, ICMDA28 University of Antwerp, Antwerp, Belgium29 Institute of Environmental Sciences and Research 27 Creyke Road, Christchurch, 8041, New Zealand.

Availability and requirements
Project name: GenoTyphi. Project home page: https://github.com/typhoid-genomics/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.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.

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.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.

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.

Data availability

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: [https:/doi.org/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]. The version reported in this paper is v1 (DOI = https:/doi.org/10.5281/zenodo.17117923).

Declarations

Ethics approval and consent to participate

All research has been conducted in accordance with the Declaration of Helsinki. 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 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.

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 COVID-19 vaccines. AJP has received payments for consultancy work from Shionogi. AJP is chair of DHSC's Joint Committee on Vaccination and Immunization, 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. The remaining authors decare that they have no competing interests.

Author details

[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, VIC 3004, Australia. [3]Centre to Impact AMR, Monash University, Clayton, VIC, Australia. [4]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK. [5]Nuffield Department of Medicine, University of Oxford, Oxford, UK. [6]National Institute of Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Headley Way, Oxford, UK. [7]Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, UK. [8]Milner Centre for Evolution, University of Bath, Bath, UK. [9]Department of Medicine, Institute of Therapeutic Immunology and Infectious Disease, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge, Cambridge Biomedical Campus, Cambridge, UK. [10]Faculty of Pharmacy and College of Medicine, University of Ibadan, Ibadan, Nigeria. [11]Institute for Infection Prevention and Control, Medical Centre, University of Freiburg, Freiburg, Germany. [12]Pathogen Biology Laboratory, Department of Biotechnology and Bioinformatics, University of Hyderabad, Hyderabad, India. [13]Atta-Ur-Rahman School of Applied Biosciences, National University of Sciences and Technology, Rawalpindi, Pakistan. [14]Malawi Liverpool Wellcome Programme, Blantyre, Malawi. [15]Institute of Infection, Ecology and Veterinary Sciences, University of Liverpool, Liverpool, UK. [16]Department of Medicine, University of Toronto, London, Canada. [17]Department Infection Biology, London School of Hygiene and Tropical Medicine, London, UK. [18]IAVI, Chelsea & Westminster Hospital, London, UK. [19]Gastrointestinal Bacteria Reference Unit, United Kingdom Health Security Agency, London, UK. [20]Centre for International Health, University of Otago, Dunedin, New Zealand. [21]Grupo de Microbiologia, Instituto Nacional de Salud, Bogota, Colombia. [22]Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology & Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia. [23]Innovation Hub, Centre for Pathogen Genomics, University of Melbourne (Doherty Institute), Melbourne, Australia. [24]Department of Bacteriology I, National Institute of Infectious Diseases, Tokyo, Japan. [25]Department of Clinical Microbiology, Christian Medical College, Vellore, India. [26]Institute of Medical Sciences, New Delhi, India. [27]Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Pretoria, South Africa. [28]Centers for Disease Control and Prevention, Atlanta, USA. [29]Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA. [30]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia. [31]International Centre for Diarrhoeal Disease Research Bangladesh (Icddr,B), Dhaka, Bangladesh. [32]Ministry of Health, Jerusalem, Israel. [33]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA. [34]University of Maryland School of Dentistry, Baltimore, Maryland, USA. [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, UK.

Received: 4 October 2024 Accepted: 15 September 2025
Published online: 24 October 2025

References

1. Meiring JE, Khanam F, Basnyat B, Charles RC, Crump JA, Debellut F, et al. Typhoid fever. Nat Rev Dis Primers. 2023;9:71.
2. GBD 2017 Typhoid and Paratyphoid Collaborators, 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.

3. World Health Organization. The WHO AWaRe (Access, Watch, Reserve) antibiotic book. Geneva; 2022.
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(4):469–89.
5. Collaborators G 2021 F, 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.
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.
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.
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.
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.
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, January–February 2022. China CDC Wkly. 2022;4:254–8.
11. Nizamuddin S, Khan EA, Chattaway MA, Godbole G. Early release—case of carbapenem-resistant *Salmonella typhi* infection, Pakistan, 2022 - Volume 29, Number 11—November 2023 - Emerging Infectious Diseases journal - CDC. Emerg Infect Dis. 2023;29:2395–7.
12. Dyson ZA, Klemm EJ, Palmer S, Dougan G. Antibiotic resistance and typhoid. Clin Infect Dis. 2019;68:S165–70.
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.
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.
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.
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 Negl Trop Dis. 2019;13:e0007868.
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 Chemother. 2017;73:365–72.
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.
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 Chemother. 2021;65:e01200–21.
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 Negl Trop Dis. 2022;16:e0010716.
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.
22. Dyson ZA, Holt KE. Five years of GenoTyphi: updates to the global *Salmonella* Typhi genotyping framework. J Infect Dis. 2021;224: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 Negl Trop Dis. 2018;12:e0006408.
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 Negl Trop Dis. 2020;14:e0008036.
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-e2118.
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-e118.
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.
28. Ingle DJ, Hawkey J, Hunt M, Dyson ZA, Holt KE. Technical Report: GenoTyphi Implementation in Mykrobe. Zenodo. 2022. Available from: https://zenodo.org/records/7407985.
29. Global Typhoid Genomics Consortium. 2025. Available from: https://www.typhoidgenomics.org/.
30. Dyson ZA et al. Genotyping *Salmonella* Typhi. GitHub. 2025. Available from: https://github.com/typhoidgenomics/genotyphi.
31. 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.
32. 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.
33. 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.
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.
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.
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.
37. 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.
38. 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 Negl Trop Dis. 2011;5:e1245.
39. 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.
40. Hawkey J, Frézal L, Dien AT, Zhukova A, Brown D, Chattaway MA, et al. Genomic perspective on the bacillus causing paratyphoid B fever. Nat Commun. 2024;15:10143.
41. Hunt M. Mykrobe. GitHub. 2023. Available from: https://github.com/Mykrobe-tools/mykrobe.
42. Hunt M. Mykrobe panel typhi version 20240407. Figshare. 2024. Available from: https://doi.org/10.6084/m9.figshare.27102334.
43. 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.
44. Clinical and Laboratory Standards Institute. CLSI M100 Performance Standards for Antimicrobial Susceptibility Testing. 2024. Available from: https://clsi.org/standards/products/microbiology/documents/m100/).
45. European Committee on Antimicrobial Susceptibility Testing. Clinical breakpoints—breakpoints and guidance. 2024. Available from: https://www.eucast.org/clinical_breakpoints.

46. 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.
47. Ingle DJ et al. Data and code supporting typhi mykrobe development and validation. GitHub. 2025. DOI = https://doi.org/10.5281/zenodo.17117923 Available from: https://github.com/typhoidgenomics/TyphoidGenomicsConsortiumMykrobe.
48. The Centre for Genomic Pathogen Surveillance (CGPS). snp_phylogeny. GitHub. 2021. Available from: https://gitlab.com/cgps/ghru/pipelines/snp_phylogeny.
49. The Centre for Genomic Pathogen Surveillance (CGPS). assembly. GitHub. 2021. Available from: https://gitlab.com/cgps/ghru/pipelines/dsl2/pipelines/assembly/.
50. 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.
51. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinform. 2020;70:e102.
52. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.
53. Seemann T. mlst. GitHub. 2022. Available from: https://github.com/tseemann/mlst.
54. Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;3:180–5.
55. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4(43):1686.
56. 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, Microb Genom. 2021;7:2016–2019.
57. CDC. National Antimicrobial Resistance Monitoring System (NARMS) Now: Human Data. Atlanta, Georgia: U.S. Department of Health and Human Services, CDC . Available from: https://www.cdc.gov/narmsnow.
58. 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.
59. U.S Food and Drug Administration. Antimicrobial Susceptibility Test (AST) Systems - Class II Special Controls Guidance for Industry and FDA. 2009. Available https://www.fda.gov/media/88069/download.
60. Krassowski M. ComplexUpset. Zenodo. 2022. Available from: https://doi.org/10.5281/zenodo.3700590.
61. 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.
62. Li H. seqtk. GitHub. 2024. Available from: https://github.com/lh3/seqtk.
63. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257.
64. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
65. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4.
66. Achtman M, Wain J, Weill F-X, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog. 2012;8:e1002776.
67. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6.
68. Wick RR. Filtlong. GitHub. 2024. Available from: https://github.com/rrwick/Filtlong.
69. Guevara et al. Pathogenwatch Genotype 2. Pathogenwatch. 2021. Available from: https://pathogen.watch/collection/0z5knw9jic9b-guevara-et-al-2021.
70. Tanmoy et al. Pathogenwatch Genotype 4.3.1.1 Pathogenwatch. Available from: https://pathogen.watch/collection/b2k6dym3bdyq-tanmoy2018.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.
76. Wick R. Yet another ONT accuracy test: Dorado v0.5.0. Zenodo. 2023. Available from: https://doi.org/10.5281/zenodo.10397817.

Publisher's Note