Technical Report: GenoTyphi Implementation in Mykrobe

Danielle J. Ingle[1], Jane Hawkey[2], Martin Hunt[3,4], Zoe A. Dyson[2,5,6*], Kathryn E. Holt[2,5*]

1. Department of Microbiology and Immunology at the Peter Doherty Institute for Infection and Immunity, The University of Melbourne; Melbourne; Australia
2. Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004, Australia
3. European Molecular Biology Laboratory – European Bioinformatics Institute; Hinxton; UK
4. Nuffield Department of Medicine, University of Oxford; Oxford; UK
5. Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
6. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

* Indicates equal contribution.

Correspondence: kat.holt@lshtm.ac.uk, zoe.dyson@lshtm.ac.uk

December, 2022 • https://doi.org/10.5281/zenodo.7407985

Introduction

Salmonella enterica serovar Typhi ("Typhi") is the bacterial agent of typhoid fever, for which whole-genome sequencing (WGS) has become the standard assay for molecular characterisation in surveillance studies and public health investigations. GenoTyphi is a hierarchical genotyping scheme that divides the Typhi population into 4 major lineages, and >75 different clades and subclades, based on pre-defined canonical single nucleotide variants (SNVs). The scheme was developed using a dataset of nearly 2,000 Typhi genomes from 60 countries [1], and was introduced in 2016 [2]. GenoTyphi has since been widely adopted by the Typhi genomics community across both research and public health settings, and subsequent updates to the scheme, extending it to 82 genotypes, were summarised in 2021 [3]. Recently, three additional genotypes (sublineages of the 3.5.4 subclade originating in Samoa) were proposed [4].

The original code that was made available for assigning GenoTyphi genotypes to novel Typhi genomes [2], implemented in Python, required as input either WGS read alignments (BAM format) or variant call files (VCF format) generated by mapping WGS reads to the reference genome for Typhi CT18 [5]. Marker SNVs were identified from these inputs, and used to classify a genome sequence into a specific genotype. The requirement to pre-map raw WGS data to a specific reference sequence introduced unnecessary complexity for users who would not otherwise undertake mapping (e.g. if basing their other analyses on assemblies) or who prefer to use a different reference sequence; processing of BAM input files could be slow; and analysis of BAM or VCF files could sometimes fail if the read mapper used generated different headers from those expected.

We therefore sought a more efficient approach to detect canonical SNVs and calculate the hierarchical genotypes directly from WGS reads. The Mykrobe software, an open-source software platform first developed in 2015, uses a kmer-based approach to identify marker SNVs and assign hierarchical genotypes for *Mycobacterium tuberculosis* [6]. Probesets are also implemented for *Staphylococcus aureus* [7] and *Shigella sonnei* [8]. Mykrobe works directly from WGS sequencing reads (FASTQ files) and has been shown to work on noisy

long reads from Oxford Nanopore Technologies (ONT) instruments as well as high-accuracy short reads from Illumina platforms [7, 8].

Here, we describe the implementation of the GenoTyphi scheme in Mykrobe. We include all previously defined genotypes [3], including the recently proposed sublineages of 3.5.4 [4], and define two new genotypes that have been noted as being of epidemiological importance due to the emergence and persistence of clinically concerning antimicrobial resistance phenotypes. We validate the accuracy of the genotype calls using n=12,848 Typhi genomes by comparing Mykrobe derived calls to the original mapping-based genotyper as the gold standard.

New genotypes

3.5.4 sublineages from Samoa

Sikorski *et al* recently defined three new genotypes (3.5.4.1, 3.5.4.2, 3.5.4.3), based on WGS data from n=306 Typhi isolated in Samoa [4]. They reported that each genotype contained between 200 and 270 differentially-present genes distinguishing them from each other and the existing Typhi genotypes. Sikorksi *et al* proposed two marker SNVs each for 3.5.4.1 and 3.5.4.2, and one for 3.5.4.3. We selected the SNVs at CT18 positions 3960783 (G->A) and 1731534 (C->T) to serve as the canonical markers for 3.5.4.1 and 3.5.4.2, respectively, in the GenoTyphi scheme. The single marker proposed for 3.5.4.3, 295362 (C->T) was also included.

Drug-resistant 4.3.1.2.1 sublineages

A sublineage of Typhi 4.3.1.2 (also known as H58 lineage 2) carrying three fluoroquinolone resistance-associated mutations (GyrA-S83F, GyrA-D87N and ParC-S80I) was first reported in 2016 [9]. This variant was highly resistant to ciprofloxacin (minimum inhibitory concentration, MIC ≥24 mg/L), and was associated with clinical failure of antimicrobial chemotherapy with the fluoroquinolone gatifloxacin during a typhoid treatment trial in Nepal in 2013 [9]. Initial comparisons with public WGS data from other countries identified India as the likely origin of this fluoroquinolone-resistant (FQR) sublineage [9]. Subsequent phylogenomic studies of Typhi WGS data from South Asia showed the FQR sublineage to be common in India, with multiple transfers into neighbouring countries including Nepal [10–12], Pakistan [12] and Myanmar [13]. Bayesian spatiotemporal analyses estimated this lineage emerged in India circa 2008 [11]. Given the epidemiological importance of this FQR sublineage – including persistence of high-level resistance to a clinically important drug class, and geographical spread – we are designating a novel genotype, 4.3.1.2.1, to facilitate its identification and monitoring. We used a set of n=12,848 Illumina WGS readsets from globally distributed Typhi isolates to explore potential marker SNVs for 4.3.1.2.1, and identified a single SNV (A->G at CT18 position 1806478) that uniquely differentiates the FQR triple-mutant sublineage from its sister clades, which carry GyrA-S83F and ParC-S80I but lack GyrA-D87N (see **Figure 1**).

A variant of the FQR triple-mutant sublineage 4.3.1.2.1 has recently been identified as resistant to third-generation cephalosporins (3GCs), including cefixime and ceftriaxone. This phenotype is due to acquisition of a ~43 kbp IncX3 plasmid carrying the extended-spectrum beta-lactamase (ESBL) gene, $bla_{\text{SHV-12}}$ and the quinolone-resistance gene *qnrB7* [14]. This variant has been reported in multiple studies of Typhi isolated in Mumbai, India [14, 15] and also in travellers from India returning to Australia [16] and England [17]. Whilst this variant remains sensitive to chloramphenicol, co-trimoxazole and azithromycin, its co-resistance to ciprofloxacin and 3GCs is clinically concerning as cefixime and ceftriaxone are recommended first-line agents for treatment of typhoid therapy in South Asia (and

azithromycin is not recommended for treatment of complicated typhoid fever). We therefore designate this lineage 4.3.1.2.1.1, to facilitate identification and monitoring of this emerging public health threat. We identified two marker SNVs, one in core gene STY0130 (*leuC*) and the other in a hypothetical gene STY4446; we selected the SNV in *leuC* (CT18 position 131973, C->T) as this is more likely to be evolutionarily stable given its status as a conserved core gene in Typhi. Notably, both SNVs were present in the genome of SRR7049884, whose sequence matches that inferred for the most recent common ancestor of all bla_{SHV-12} -positive 4.3.1.2.1. SRR7049884 itself lacks the ESBL gene (see **Figure 1**), but is presumably very closely related to the ancestral strain of 4.3.1.2.1.1 that acquired bla_{SHV-12}.

With the addition of these five new genotypes, the GenoTyphi scheme now includes 87 genotypes and corresponding SNV markers (see https://github.com/katholt/genotyphi/blob/main/Genotype_specification.csv)

Implementation

Mykrobe probesets were created for all 87 marker SNVs using the `mykrobe variants make-probes` command using a kmer size of 21 and the Typhi CT18 (accession AL513382.1) reference genome. Additionally, we supplied Mykrobe with all SNVs detected against the Typhi CT18 reference for the set of test genomes to use as background SNVs when generating the probesets (in other words, to incorporate population SNVs which are within k bases of the marker SNVs into the probes).

We also created probesets to confirm input readsets as *Salmonella enterica* (based presence of marker gene *invA*) and serovar Typhi (based on multi-locus sequence typing using the 7-locus scheme for *S. enterica* [18], of which there are 26 known STs; see **Table 1**). Preliminary testing of probesets was conducted using a set of 91 genomes, with at least one per genotype (see **Table 2**).

These probesets, together with the hierarchical specification of genotypes, form the 'Typhi' typing panel for Mykrobe. The version described in this report is v20221207. The typing panel is distributed (along with panels for other species) with the Mykrobe code [6], available at https://github.com/Mykrobe-tools/mykrobe. A copy of the Typhi panel data is available in FigShare at https://doi.org/10.6084/m9.figshare.21695528.v1. Existing installations of Mykrobe can be updated to the latest version typing panels by running '`mykrobe panels update_metadata; mykrobe panels update_species all`'.

Mykrobe outputs a JSON file containing details of all marker SNVs identified, details of the kmer/read matching to probesets, and the calculated genotype. We developed a Python3 script, `parse_typhi_mykrobe.py`, available in the GenoTyphi repository, to batch-process these JSON files and generate a single, human-readable table (tab-delimited format) summarising the results. Key fields are `species` ('Typhi' if the *Salmonella* plus MLST probes confirm this, 'unknown' otherwise); `final_genotype`, `confidence` (strong, moderate, weak), and measures of support for each marker SNV detected (including Mykrobe support values and the number of reads supporting the marker and wildtype alleles). Full details of fields are given in the repository at https://github.com/katholt/genotyphi.

All code and instructions for genotyping isolates with GenoTyphi using Mykrobe is available in the GitHub repository at https://github.com/katholt/genotyphi. The version used in this report is v2.0, DOI: 10.5281/zenodo.7430538.

Accuracy of Mykrobe implementation

We validated the Mykrobe implementation of GenoTyphi using Illumina read sets for n=12,848 Typhi isolates. We ran Mykrobe v0.12.1 on all FASTQ files using the new Typhi panel (v20221207) and summarised results using `parse_typhi_mykrobe.py`.

To assess accuracy of Mykrobe genotype calls, we analysed the same read sets using the original mapping-based approach. Specifically, reads were mapped to the CT18 reference sequence (accession AL513382.1) using bwa-mem v0.7.17 via the CGPS mapping pipeline v1.2.2 (https://gitlab.com/cgps/ghru/pipelines/snp_phylogeny/), and the resulting BAM files passed to the `genotyphi.py` script to calculate genotypes (v2, updated to include the latest GenoTyphi v20221207 genotype set; code available in the GenoTyphi repository at https://github.com/katholt/genotyphi and DOI: 10.5281/zenodo.7430538).

Mykrobe confirmed all 12,848 readsets as Typhi, and generated confident genotypes calls for n=12,801 (99.63%; confident calls being defined as a quality of '1' reported by Mykrobe for all marker SNVs detected). Of the remaining 47 isolates, two had calls with moderate confidence and 45 with weak confidence.

All but one of the confident calls made by Mykrobe matched the genotype called using the original mapping-based approach, i.e. concordance of 99.99%. The discordant isolate, ERR5243665, was called as genotype 2 by Mykrobe and 0.1.3 by the mapping-based approach. The isolate clustered with other genotype 2 isolates in a distance-based phylogeny of genomes from the same study [19], suggesting the Mykrobe call is correct (see tree at https://pathogen.watch/collection/0z5knw9jic9b-guevara-et-al-2021).

We count as concordant six genomes for which mixtures were detected, which are reported differently between methods. For five genomes, Mykrobe gave a confident call of genotype 4.3.1.1 but with an additional marker detected ('2.1.2', supported by 54/54 reads); the mapping-based approach called these as '4.3.1.1,2.1.2'. A sixth genome was typed '4.3.1.1,4.3.1.3.Bdq' by the mapping-based approach but simply '4.3.1.1' by Mykrobe, with no additional markers. We also count as concordant 14 genomes belonging to nested subclades, which again are handled differently between methods. For eight genomes belonging to 3.5.3, which is nested within 3.5.4, Mykrobe gave a confident call of 3.5.3, with strong support for the 3.5.3 marker and the 3.5.4 marker; whereas the mapping-based implementation reported 3.5.4 as it reports the numerically last subclade. Similarly, for six genomes belonging to 2.3.3, which is nested within 2.3.2, Mykrobe gave a confident call of 2.3.3, with strong support for 2.3.2 and 2.3.3 markers.

Our Typhi Mykrobe parser script assigns 'moderate' confidence when 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. Information about poorly supported markers is provided to help users troubleshoot. In our test set, both genomes called with 'moderate' confidence were identified by Mykrobe as 4.3.1.2, but had poor support for the 4.3.1.2 marker SNV (n=46/65 and n=54/81 reads; compared with universal support for the other marker SNVs 1, 2, 3, 4 and 4.3.1). One also had some support for an additional marker of the sister lineage, 4.3.1.1 (n=24/67 reads), reported by Mykrobe. This genome was reported by the mapping-based approach as '4.3.1.1,4.3.1.2' with low support (0.27), and the other was reported as '4.3.1.1', also with low support (0.67). We conclude that, for these genomes, Mykrobe is making the 'right' calls and is accurately reflecting the level of evidence supported by the raw data, providing useful information for users to interpret the result (in these cases, a mix of 4.3.1 strain subtypes appears most likely).

Our Typhi Mykrobe parser script assigns ‘weak’ confidence when one or more marker SNVs are of low quality – defined as Mykrobe quality score of 0, or Mykrobe quality score of 0.5 and with minority support (i.e. <50% of reads) for the derived allele. For all 45 genomes called by Mykrobe with weak confidence, the genotype call was concordant with that of the mapping-based approach. In five cases, the mapping-based implementation called multiple genotypes and reported low support (≤0.2), in 12 cases it called a single genotype but with low support (≤0.52) and in 28 cases it called a single genotype with strong support (>0.95). Again, we conclude that Mykrobe is correctly classifying these genomes with final calls accurately reflects the level of support observed.

Conclusions

Mykrobe can reliably assign Typhi WGS Illumina read sets to genotypes according to the hierarchical GenoTyphi scheme. Compared with the original mapping-based approach, the Mykrobe implementation is simpler and faster to run (working direct from FASTQ files); equally accurate (agreeing in all cases except one, where the Mykrobe call shows better phylogenetic concordance than the mapping-based call); and provides richer information on the quality of evidence for each call, which can help users trouble-shoot and understand low-confidence calls.

Figure 1. Neighbour-joining tree showing representative isolates of Typhi 4.3.1.2, illustrating the position of sublineages for which genotypes have been designated.

Tips include n=47 representatives of 4.3.1.2.1 and n=4 of 4.3.1.2.1.1, and a random set of n=50 isolates of 4.3.1.2 (including single representatives of sublineages 4.3.1.2.EA2 and 4.3.1.2.E3) to show the phylogenetic structure. Tree was inferred using Pathogenwatch (https://pathogen.watch) [20]. Tips are coloured by genotype. Heatmap shows the drug resistance profile and presence of genetic determinants of resistance: CipR, ciprofloxacin resistant; CipNS, ciprofloxacin non-susceptible. An interactive version of this annotated tree is available at https://microreact.org/project/typhi-4312-subset; an interactive version with a tree including all 4.3.1.2 genomes available in Pathogenwatch is available at https://microreact.org/project/typhi-4312.

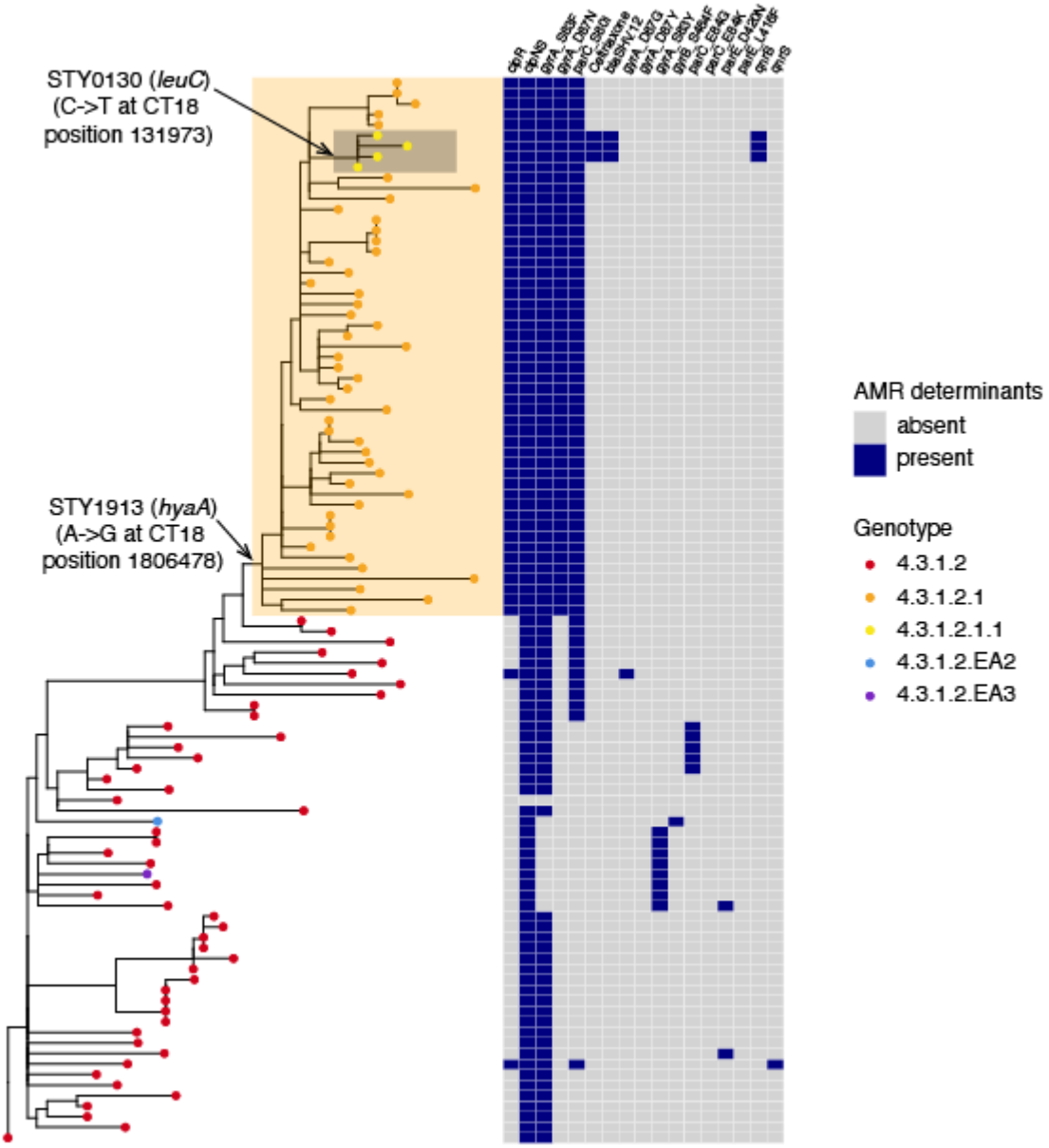

Table 1: Sequence types and allele profiles used to confirm *Salmonella enterica* serovar Typhi, according to the 7-locus multi-locus sequence typing scheme

Corresponding allele profiles can be downloaded from PubMLST at https://pubmlst.org/bigsdb?db=pubmlst_salmonella_seqdef

ST	*aroC*	*dnaN*	*hemD*	*hisD*	*purE*	*sucA*	*thrA*
1	1	1	1	1	1	1	5
2	1	1	2	1	1	1	5
8	1	1	2	3	1	1	5
2138	1	465	2	1	1	1	5
2160	1	1	385	3	1	1	5
2173	1	1	2	1	1	476	5
2209	1	1	2	1	484	1	5
2218	511	1	2	1	1	1	5
2231	512	1	2	1	1	1	5
2233	1	478	2	1	1	1	5
2244	1	1	2	616	1	1	5
2254	1	1	340	1	1	486	5
2267	518	133	356	94	488	226	4
2337	1	1	2	1	1	497	5
2338	1	492	1	1	1	1	5
2339	529	1	2	3	1	1	5
2341	1	1	1	1	532	1	5
2342	1	1	1	1	1	1	543
2350	531	1	1	1	1	1	5
2352	532	1	2	3	1	1	5
2353	1	1	2	1	1	1	544
2355	1	1	2	1	1	1	500
2356	1	1	2	635	1	1	5
2359	1	1	1	1	1	1	545
2360	1	1	1	1	1	498	5
2913	1	549	1	1	1	1	5

Table 2: Typhi read sets representative of unique genotypes, used for preliminary testing

Accession	Publication	Genotype
ERR314328	Baker et al, 2015 (ref 21)	0
ERR360655	Wong et al, 2016 (ref 2)	0.0.1
ERR352473	Wong et al, 2016 (ref 2)	0.0.2
ERR343337	Wong et al, 2016 (ref 2)	0.0.3
ERR360484	Wong et al, 2016 (ref 2)	0.1
ERR360505	Wong et al, 2016 (ref 2)	0.1.1
ERR360668	Wong et al, 2016 (ref 2)	0.1.2
ERR360486	Wong et al, 2016 (ref 2)	0.1.3
ERR360627	Wong et al, 2016 (ref 2)	1.1.1
ERR338008	Wong et al, 2016 (ref 2)	1.1.2
ERR360695	Wong et al, 2016 (ref 2)	1.1.3
ERR352453	Wong et al, 2016 (ref 2)	1.1.4
ERR998612	Park et al 2018 (ref 22)	1.2
ERR2663777	Tanmoy et al 2018 (ref 23)	1.2.1
ERR2663521	Tanmoy et al 2018 (ref 23)	2
ERR2663661	Tanmoy et al 2018 (ref 23)	2.0.1
ERR360492	Wong et al, 2016 (ref 2)	2.0.2
ERR357445	Wong et al, 2016 (ref 2)	2.1
ERR331214	Wong et al, 2016 (ref 2)	2.1.1
ERR343252	Wong et al, 2016 (ref 2)	2.1.2
ERR331225	Wong et al, 2016 (ref 2)	2.1.3
ERR343250	Wong et al, 2016 (ref 2)	2.1.4
ERR352302	Wong et al, 2016 (ref 2)	2.1.5
ERR331245	Wong et al, 2016 (ref 2)	2.1.6
ERR343289	Wong et al, 2016 (ref 2)	2.1.7
ERR352255	Wong et al, 2016 (ref 2)	2.1.7.1
ERR352309	Wong et al, 2016 (ref 2)	2.1.7.2
ERR331216	Wong et al, 2016 (ref 2)	2.1.8
ERR331217	Wong et al, 2016 (ref 2)	2.1.9

ERR2663529	Wong et al, 2016 (ref 2)	2.2
ERR352257	Wong et al, 2016 (ref 2)	2.2.1
ERR3804595	Wong et al, 2016 (ref 2)	2.2.2
ERR331310	Wong et al, 2016 (ref 2)	2.2.3
ERR343325	Wong et al, 2016 (ref 2)	2.2.4
ERR360494	Wong et al, 2016 (ref 2)	2.3.1
ERR360615	Wong et al, 2016 (ref 2)	2.3.2
ERR3290547	Rahman et al 2020 (ref 24)	2.3.3
ERR352269	Wong et al, 2016 (ref 2)	2.3.4
ERR352495	Wong et al, 2016 (ref 2)	2.3.5
ERR326664	Wong et al, 2016 (ref 2)	2.4
ERR360813	Wong et al, 2016 (ref 2)	2.4.1
ERR2663511	Tanmoy et al 2018 (ref 23)	2.5
ERR360646	Wong et al, 2016 (ref 2)	2.5.1
ERR360496	Wong et al, 2016 (ref 2)	2.5.2
ERR2663646	Tanmoy et al 2018 (ref 23)	3
ERR2663803	Tanmoy et al 2018 (ref 23)	3.0.1
ERR2663895	Tanmoy et al 2018 (ref 23)	3.0.2
ERR3804551	Britto et al 2020 (ref 9)	3.1
ERR360500	Wong et al, 2016 (ref 2)	3.1.1
ERR343291	Wong et al, 2016 (ref 2)	3.1.2
SRR5974898	Ingle et al 2019 (ref 25)	3.2
ERR2663559	Tanmoy et al 2018 (ref 23)	3.2.2
ERR2663868	Tanmoy et al 2018 (ref 23)	3.3
ERR3804596	Britto et al 2020 (ref 9)	3.3.1
ERR2663475	Tanmoy et al 2018 (ref 23)	3.3.2
ERR2663543	Tanmoy et al 2018 (ref 23)	3.3.2.Bd1
ERR2663600	Tanmoy et al 2018 (ref 23)	3.3.2.Bd2
ERR360747	Wong et al, 2016 (ref 2)	3.4
ERR352311	Wong et al, 2016 (ref 2)	3.5
ERR352501	Wong et al, 2016 (ref 2)	3.5.1
ERR331309	Wong et al, 2016 (ref 2)	3.5.2

ERR338070	Wong et al, 2016 (ref 2)	3.5.4
ERR343283	Sikorski et al 2022 (ref 4)	3.5.4.1
ERR338127	Sikorski et al 2022 (ref 4)	3.5.4.2
ERR357760	Sikorski et al 2022 (ref 4)	3.5.4.3
SRR5500443	Ingle et al 2019 (ref 25)	4
ERR2663643	Tanmoy et al 2018 (ref 23)	4.1
ERR357459	Wong et al, 2016 (ref 2)	4.2
ERR360689	Wong et al, 2016 (ref 2)	4.1.1
ERR331262	Wong et al, 2016 (ref 2)	4.2.1
ERR331260	Wong et al, 2016 (ref 2)	4.2.2
ERR352317	Wong et al, 2016 (ref 2)	4.2.3
ERR4042542	Duy et al 2020 (ref 26)	4.3.1
ERR360459	Wong et al, 2016 (ref 2)	4.3.1
ERR422757	Wong et al, 2016 (ref 2)	4.3.1
ERR2663470	Tanmoy et al 2018 (ref 23)	4.3.1
ERR349354	Wong et al, 2016 (ref 2)	4.3.1.1
ERR3477453	Hooda et al 2019 (ref 27)	4.3.1.1
ERR352442	Wong et al, 2016 (ref 2)	4.3.1.1.EA1
ERR3332551	Kariuki et al 2021 (ref 28)	4.3.1.1.EA1
SRR10918333	Rasheed et al 2020 (ref 29)	4.3.1.1.P1
ERR2663759	Tanmoy et al 2018 (ref 23)	4.3.1.2
ERR4790805	da Silva et al 2022 (ref 11)	4.3.1.2.1.1
ERR4992648	da Silva et al 2022 (ref 11)	4.3.1.2.1.1
ERR338134	Wong et al, 2016 (ref 2)	4.3.1.2.EA2
ERR3332773	Kariuki et al 2021 (ref 28)	4.3.1.2.EA2
ERR3332776	Kariuki et al 2021 (ref 28)	4.3.1.2.EA2
ERR3332782	Kariuki et al 2021 (ref 28)	4.3.1.2.EA2
SRR3049193	Ingle et al 2019 (ref 25)	4.3.1.2.EA3
ERR2663546	Tanmoy et al 2018 (ref 23)	4.3.1.3
ERR2933271	Rahman et al 2020 (ref 24)	4.3.1.3.Bdq

References

1. **Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ,** 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.
2. **Wong VK, Baker S, Connor TR, Pickard D, Page AJ,** et al. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. *Nat Commun* 2016;7:12827.
3. **Dyson ZA, Holt KE**. Five years of GenoTyphi: updates to the global *Salmonella* Typhi genotyping framework. *J Infect Dis* 2021;224:jiab414.
4. **Sikorski MJ, Hazen TH, Desai SN, Nimarota-Brown S, Tupua S,** et al. Persistence of Rare *Salmonella* Typhi Genotypes Susceptible to First-Line Antibiotics in the Remote Islands of Samoa. *Mbio* 2022;13:e01920-22.
5. **Parkhill J, Dougan G, James KD, Thomson NR, Pickard D,** et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 2001;413:848–852.
6. **Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M,** et al. Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res* 2019;4:191.
7. **Bradley P, Gordon NC, Walker TM, Dunn L, Heys S,** et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun* 2015;6:10063.
8. **Hawkey J, Paranagama K, Baker KS, Bengtsson RJ, Weill F-X,** et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, *Shigella sonnei*. *Nat Commun* 2021;12:2684.
9. **Thanh DP, Karkey A, Dongol S, Thi NH, Thompson CN,** et al. A novel ciprofloxacin-resistant subclade of H58 *Salmonella* Typhi is associated with fluoroquinolone treatment failure. *Elife* 2016;5:e14003.
10. **Britto CD, Dyson ZA, Mathias S, Bosco A, Dougan G,** et al. Persistent circulation of a fluoroquinolone-resistant *Salmonella enterica* Typhi clone in the Indian subcontinent. *J Antimicrob Chemoth* 2020;75:337–341.
11. **Britto CD, Dyson ZA, Duchene S, Carter MJ, Gurung M,** 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.
12. **da Silva KE, Tanmoy AM, Pragasam AK, Iqbal J, Sajib MSI,** et al. The international and intercontinental spread and expansion of antimicrobial-resistant *Salmonella* Typhi: a genomic epidemiology study. *Lancet Microbe* 2022;3:e567–e577.
13. **Oo KM, Myat TO, Htike WW, Biswas A, Hannaway RF,** et al. Molecular mechanisms of antimicrobial resistance and phylogenetic relationships of *Salmonella enterica* isolates from febrile patients in Yangon, Myanmar. *T Roy Soc Trop Med H* 2019;113:641–648.
14. **Jacob JJ, Pragasam AK, Vasudevan K, Veeraraghavan B, Kang G,** et al. *Salmonella* Typhi acquires diverse plasmids from other Enterobacteriaceae to develop cephalosporin resistance. *Genomics* 2021;113:2171–2176.
15. **Argimón S, Nagaraj G, Shamanna V, Darmavaram S, Vasanth AK,** et al. Circulation of third-generation cephalosporin resistant *Salmonella* Typhi in Mumbai, India. *Clin Infect Dis* 2015;74:ciab897-.
16. **Ingle DJ, Andersson P, Valcanis M, Wilmot M, Easton M,** et al. Genomic

Epidemiology and Antimicrobial Resistance Mechanisms of Imported Typhoid in Australia. *Antimicrob Agents Ch* 2021;65:e01200-21.
17. **Chattaway MA, Gentle A, Nair S, Tingley L, Day M,** et al. Phylogenomics and antimicrobial resistance of *Salmonella* Typhi and Paratyphi A, B and C in England, 2016–2019. *Microb Genom*;7. Epub ahead of print 2021. DOI: 10.1099/mgen.0.000633.
18. **Achtman M, Wain J, Weill F-X, Nair S, Zhou Z,** et al. Multilocus Sequence Typing as a Replacement for Serotyping in *Salmonella enterica*. *Plos Pathog* 2012;8:e1002776.
19. **Guevara PD, Maes M, Thanh DP, Duarte C, Rodriguez EC,** et al. A genomic snapshot of *Salmonella enterica* serovar Typhi in Colombia. *Plos Neglect Trop D* 2021;15:e0009755.
20. **Argimón S, Yeats CA, Goater RJ, Abudahab K, Taylor B,** et al. A global resource for genomic predictions of antimicrobial resistance and surveillance of *Salmonella* Typhi at pathogenwatch. *Nat Commun* 2021;12:2879.
21. **Baker KS, Burnett E, McGregor H, Deheer-Graham A, Boinett C,** et al. The Murray collection of pre-antibiotic era Enterobacteriacae: a unique research resource. *Genome Med* 2015;7:97
22. **Park SE, Pham DT, Boinett C, Wong VK, Pak GD**, et al. The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. *Nat Commun* 2018;9:5094
23. **Tanmoy AM, Westeel E, De Bruyne K, Goris J, Rajoharison**, et al*. Salmonella enterica* serovar Typhi in Bangladesh: exploration of genomic diversity and antimicrobial resistance. *mBio.* 2018;9:e02112-18
24. **Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE**, et al. Population structure and antimicrobial resistance patterns of *Salmonella* Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. *Plos Neglect Trop D* 2020;14:e0008036.
25. **Ingle DJ, Nair S, Hartman H, Ashton PM, Dyson ZA,** et al. Informal genomic surveillance of regional distribution of *Salmonella* Typhi genotypes and antimicrobial resistance via returning travellers. *Plos Neglect Trop D* 2019;9:e0007620
26. **Duy PT, Dongol S, Giri A, To NTN, Thanh HND**, et al. The emergence of azithromycin-resistant *Salmonella* Typhi in Nepal. *JAC Antimicrobial Resist* 2020;2:dlaa109
27. **Hooda Y, Sajib MSI, Rahman H, Luby SP, Bondy-Denomy J**, et al. Molecular mechanism of azithromycin resistance among typhoida*l Salmonella* strains in Bangladesh identified through passive pediatric surveillance. *Plos Neglect Trop D* 2019;13:e0007868
28. **Kariuki S, Revathi G, Kiiru J, Mengo DM, Mwituria J,** et al. Typhoid in Kenya is Associated with a Dominant Multidrug-Resistant *Salmonella enterica* Serovar Typhi Haplotype That Is Also Widespread in Southeast Asia. *J Clin Microbiol* 2021;48:2171-6.
29. **Rasheed F, Saeed M, Alikhan NF, Baker D, Khurshid M,** et al. Emergence of Resistance to Fluoroquinolones and Third-Generation Cephalosporins in *Salmonella* Typhi in Lahore, Pakistan. *Microorg* 2020;8:1336