Genomewide analysis of *Salmonella* Typhi and Paratyphi A, B circulating in Bangladesh to understand the population structure, transmission pattern and evolutionary dynamics using whole genome sequencing

A DISSERTATION SUBMITTED TO THE UNIVERSITY OF DHAKA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOCHEMISTRY AND MOLECULAR BIOLOGY

SADIA ISFAT ARA RAHMAN

Registration No- 14/2017-18

January, 2022

Department of Biochemistry and Molecular Biology
Faculty of Biological Sciences
University of Dhaka
Bangladesh

DECLARATION

This is to certify that the research presented in this dissertation entitled "Genomewide analysis of *Salmonella* Typhi and Paratyphi A, B circulating in Bangladesh to understand the population structure, transmission pattern and evolutionary dynamics using whole genome sequencing" submitted by Sadia Isfat Ara Rahman, has been carried out under our direct supervision and guidance. The research work described in this dissertation is original and it has not been submitted, in whole or in part, in any previous application for other degree.

Academic supervisor
Dr. Yearul Kabir
Chairman, Professor, Department of Biochemistry and Molecular Biology
University of Dhaka, Bangladesh

Co-supervisor
Dr. Emran Kabir Chowdhury
Professor, Department of Biochemistry and Molecular Biology
University of Dhaka (Deputation), Bangladesh
Vice Chancellor, Comilla University, Bangladesh

Co-supervisor
Dr. Firdausi Qadri
Senior Scientist, Infectious Disease Division,
Head, Mucosal Immunology and Vaccinology Laboratory, icddr,b

Dedicated

To

My Beloved Parents

ACKNOWLEDGMENTS

All praise to the sustainer of the almighty Allah for blessing me with courage, mental power, patience and good health to accomplish this dissertation. The work described in this thesis arose as a result of four years of insistent hard work that I have been carried out in collaboration with icddr,b, Wellcome Sanger Institute and University of Cambridge. During the journey of my PhD, I have worked with a number of people whose contribution to be specially mentioned and would like to convey my gratitude to all of them for their continuous support.

I am deeply indebted to my supervisor **Professor Dr. Yearul Kabir**, Chairman, Department of Biochemistry and Molecular Biology, University of Dhaka for his academic support and encouragement. I would like to thank him from the bottom of my heart to take the responsibilities of supervising my PhD in such a crisis moment. I would like to convey my special gratitude to my co-supervisor **Dr. Emran Kabir Chowdhury,** Department of Biochemistry and Molecular Biology, University of Dhaka and Vice Chancellor of Comilla University for his endless inspiration, affectionate guidance and cordial supervision. I am thankful to **all of my honourable teachers** of Department of Biochemistry and Molecular Biology for scholastic guidance.

I am overwhelmed to express my respect and sincere gratitude to my mentor and co-supervisor **Dr. Firdausi Qadri**, Senior Scientist and Head of the Mucosal Immunology and Vaccinology Laboratory, icddr,b for giving me the opportunity to conduct a collaborative PhD on genomics. Her strong personality and impressive scientific knowledge always inspires me to push myself outside the confines of my comfort zone. I could not think of this PhD happen without her support, inspiration, guidance, positive spirit towards me. I would have lost in my personal problems if she was not there to hold my hand and to take care of me as her own child. I am really grateful to her for nourishing my intellectual maturity during my hard times.

I convey special acknowledgement to my co-supervisors **Professor Dr. Gordon Dougan**, University of Cambridge, United Kingdom and **Professor Dr. Nicholas Robert Thomson**, Wellcome Sanger Institute, Cambridge for creative advice, vital contribution, scholastic guidance and most importantly for giving me the opportunity to join their team as Visiting Researcher. I am thankful for introducing me to the new world of bioinformatics, for listening my plans, giving their valuable suggestions to reshape my ideas, supporting me in genomic data analysis, manuscript writing. I sincerely extend my gratitude to **Professor Stephen Baker, To Nguyen Thi Nguyen** for helping me to conduct *S.* Paratyphi A genomics study and **Professor Dr. Ankur Mutreja, Dereck Pickard, Fahad Khokar** for guiding me to optimize a genome based conventional single-plex PCR assay at icddr,b.

I owe a debt of gratitude to **Dr. Elizabeth Klemm, Dr. Kathryn Holt, Dr. Zoe Dyson, Dr. Alyce Taylor-Brown, Dr. Gal Horesh** for teaching me from the very basic Linux command to different bioinformatics pipelines, R programs. As my research career started with immunological based studies at icddr,b, it was not an easy task for me to gain bioinformatics skills overnight. It would have never been possible without their contribution. My special thanks to **Zoe and Alyce** for giving their precious time to solve all command line errors and discuss research ideas during manuscript writing when I was at Cambridge and even also through skype virtually in Bangladesh. I would like to express special thanks to review my PhD thesis.

It was a pleasure to work for a short period in the Pathogen Genomics Group at Wellcome Sanger Institute and University of Cambridge. I also thank the members of **Pathogen Informatics Team** for configuring Sanger Unix account in my laptop and teaching me the basic commands to run different sequence data analysis pipeline. Special thanks to **the Core Sequencing Team** for generating whole genome sequence data of the collection *S.* Typhi and Paratyphi strains used in this PhD work. I am grateful to **Siva,** my Landlord for her nice hospitality and delicious food during my stay at Cambridge.

My deepest gratitude goes to **Dr. Farhana Khanam**, **Dr. Taufiqur Rahman Bhuiyan, Dr. Yasmin Ara Begum, Dr. Ashraful Islam Khan** for valuable suggestions, incredible support, guidance, cordial help throughout the journey of my PhD. My heartfelt thanks to **Fatema Khatun, Nasrin Parvin** for blessing me with good advices, moral support when I felt low. Special thanks to **Marjahan Akhter** for sharing her PhD experience and for giving me practical suggestions from the very beginning of PhD enrolment to thesis submission. She is my inspiration from whom I got the motivation of doing a collaborative PhD. I am grateful to **Hassan Afrad** for being there to solve any sort of scientific, non-scientific technical problems and helping me in PhD synopsis writing.

I gratefully acknowledge the support of all laboratory team members, field workers at the icddr,b and the funding sources. The work was supported by the grants from Wellcome Trust, Bill and Melinda Gates Foundation (grant no. OPP1135223, OPP50419), The National Institutes of Health (NIH), including the National Institute of Allergy and Infectious Diseases (AI100023, AI058935), SIDA fund (54100020, 51060029).

Finally, I like to express my outmost gratitude to my beloved **parents and family members** for their endless moral support, kind prayers during these years and for always being proud of me no matter what happens.

ABSTRACT

Enteric fever, caused by *Salmonella enterica* serovar Typhi (*S*. Typhi), *Salmonella enterica* Paratyphi serovars (*S*. Paratyphi A, B, and C), remains a major public health concern in low and middle income countries (LMIC) with poor hygiene and sanitation system. However, there is still limited understanding on the genomic epidemiology and population structure of these pathogens in low endemic settings including Bangladesh. The rapid emergence of multidrug resistance (MDR) and lack of credible systemic surveillance data representing the true disease burden emphasize the importance of genome based surveillance studies. In this thesis, a collection of whole genome sequence (WGS) data of 202 *S.* Typhi, 67 *S.* Paratyphi A and 79 *S.* Paratyphi B strain collected from different already existing enteric disease surveillance sites across Bangladesh was utilized with the aim to study population structure, antimicrobial resistance (AMR) dynamics and placing these sequence data in global genome context to investigate both global and regional transmission patterns.

Phylogenetic analyses showed diverse *S.* Typhi population structure harboring nine distinct genotypes compared to *S.* Paratyphi A comprising only three lineages with the evidence of circulation of these Bangladeshi pathogens throughout neighboring South Asian countries. In addition, WGS analysis further classified previously serotyped Bangladeshi *S.* Paratyphi B strains as biotype Java, which were associated with diarrheal symptoms. Moreover, genotypic distribution and AMR patterns were observed among the *S.* Typhi Bangladeshi population during the study time period of 2004 to 2006. The high prevalence of reduced fluoroquinolone susceptibility with absence of MDR genes was identified among *S.* Typhi non-H58 genotypes 2.3.3, 3.2.2, 3.3.2 (from median 22.2% per year in 2004-2010 to 65.2% per year in 2011-2016) emphasizing the shift in treatment practice towards third generation cephalosporin and azithromycin. With the rising trend of fluoroquinolone resistant *S.* Typhi and *S.* Paratyphi A strain described in this dissertation, single azithromycin resistant *S.* Paratyphi A strain isolated in 2018 due to acquisition of R717L mutation in *acrB* gene has narrowed down the overall treatment option for enteric fever. Despite of having increasing reports of antibiotic usage without prescription in Bangladesh, there was limited evidence for widespread extrinsic resistance gene and plasmid acquisition in Bangladeshi *S.* Java strains. Notably, a pHCM2-like cryptic

plasmid carrying *S.* Paratyphi A strains in Bangladesh was reported first time here that has not previously detected in other global regions. A comparative Basic Local Alignment Search Tool (BLAST) analysis and pHCM2 phylogenetic analysis including plasmid sequences from the 17 *S.* Paratyphi A, 334 *S.* Typhi, and one *S.* Java straind from Bangladesh showed high similarity between typhoidal *S.* Typhi and *S.* Paratyphi A that were different from nontyphoidal *S.* Paratyphi B biotype Java. Comparative pan-genome analysis was also conducted to investigate genomic composition of core and accessory genes which revealed closed pan-genome for *S.* Paratyphi A but open pan-genome for *S.* Java. Lastly, the performance of a new WGS based diagnostic approach was evaluated by targeting unique genes for *S.* Typhi including globally dominant H58 lineage and *S.* Paratyphi A that has been considered as rapid, reliable, cost-effective compared to other currently available diagnostic methods for enteric fever detection.

The findings in this dissertation will improve the understanding of genomic diversity of these pathogens to facilitate the identification of epidemiologically important new genotypes and to monitor the ongoing evolution of AMR in Bangladesh. The assimilation of WGS data will provide a framework for future genomic epidemiology studies in low-middle income countries as well as inform public health information to implement innovative diagnostic approaches, treatment regimens and enteric fever control strategies.

LIST OF PUBLICATIONS

Based on this thesis work, the following three papers have been published in peer reviewed journals:

1. Population structure and antimicrobial resistance patterns of *Salmonella* Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. PLOS Neglected Tropical Disease.

Sadia Isfat Ara Rahman, Zoe A Dyson, Elizabeth J Klemm, Farhana Khanam, Kathryn E Holt, Emran Kabir Chowdhury, Gordon Dougan, Firdausi Qadri.

PLOS Neglected Tropical Diseases 2020

DOI: 10.1371/journal.pntd.0008036

2. Genetic diversity of *Salmonella* Paratyphi A isolated from enteric fever patients in Bangladesh from 2008 to 2018

Sadia Isfat Ara Rahman, To Nguyen Thi Nguyen, Farhana Khanam, Nicholas R. Thomson, Zoe A. Dyson, Alyce Taylor-Brown, Emran Kabir Chowdhury, Gordon Dougan, Stephen Baker, Firdausi Qadri

PLOS Neglected Tropical Diseases 2021

DOI: 10.1371/journal.pntd.0009748

3. Genome-wide analysis provides a deeper understanding of the population structure of the *Salmonella enterica* serotype Paratyphi B complex in Bangladesh

Sadia Isfat Ara Rahman, Alyce Taylor-Brown, Farhana Khanam, Ashraful Islam Khan, Gal Horesh, Zoe A. Dyson, Yasmin Ara Begum, Emran Kabir Chowdhury, Firdausi Qadri, Gordon Dougan, Nicholas R. Thomson

Microbial Genomics 2021 Vol. 7 Issue 9

DOI: 10.1099/mgen.0.000617

TABLE OF CONTENTS

TITLE	PAGE
CHAPTER 1 INTRODUCTION	**1**
1.1 Classification of the genus *Salmonella*	1
1.2 General overview: *S.* Typhi and *S.* Paratyphi A, B	3
1.2.1 Antigenic and biochemical characteristics	3
1.2.2 Global epidemiology	4
1.2.3 Pathogenesis and clinical features	6
1.2.4 Diagnostic approaches	6
1.2.4.1 Microbiological culture based diagnosis	7
1.2.4.2 Serological diagnosis	7
1.2.4.3 Molecular diagnosis	7
1.2.5 Antibiotic treatment and resistance	8
1.2.6 Vaccination	9
1.3 Next generation sequencing (NGS) approaches	10
1.3.1 An introduction to the bioinformatics approaches	12
1.3.2 Archiving WGS data	14
1.3.3 *S.* Typhi and *S.* Paratyphi reference genome sequences	14
1.4 WGS application for *S.* Typhi and *S.* Paratyphi genomics	15
1.4.1 Phylogeography and population genomics	16
1.4.2 Understanding evolutionary changes using genomics	19
1.4.3 Comparative genomics and transmission dynamics	21
1.4.4 Molecular basis of AMR	22
1.4.4.1 Genomic architecture of MDR *S.* Typhi	22
1.4.4.2 AMR profile of *S.* Paratyphi A, B	24
1.4.5 Implementation of WGS data to design PCR for diagnosis	24
1.5 Thesis outline	25
1.6 Hypothesis	26
1.7 Aims and objectives of the dissertation	26

CHAPTER 2 METHODS AND MATERIALS 27

2.1 Study settings and sample collection 27

2.2 Bacterial isolation 28

2.3 Genomic DNA extraction 29

2.4 Whole genome sequencing by Illumina platform 30

2.5 WGS data analysis 30

2.5.1 Mapping and SNP calling 31

2.5.2 SNP based genotyping analysis 31

2.5.3 Phylogenetic and population genetic analysis 32

2.5.4 *In silico* AMR gene, plasmid, virulence factor and *S.* Paratyphi B biotype identification 32

2.5.5 Comparative pHCM2 plasmid analysis 33

2.5.6 *Denovo* genome assembly and annotation 34

2.5.7 Comparative pan-genome analysis 34

2.6 Statistical analysis 35

2.7 Novel WGS approach based diagnostic assay 35

2.7.1 Bacterial strains and DNA extraction 35

2.7.2 *S.* Typhi, H58 *S.* Typhi and *S.* Paratyphi A primer design 35

2.7.3 Optimization of conventional single-plex, multi-plex PCR 36

CHAPTER 3 RESULTS 38

3.1 Study I: *S.* Typhi genomics study 38

3.1.1 Genotypic diversity of *S.* Typhi in Bangladesh 38

3.1.2 Transmission pattern of global and regional *S.* Typhi strains 40

3.1.3 AMR and plasmid profile among *S.* Typhi population 41

3.1.4 Correlation of genotypic distribution and AMR pattern over time period between 2004-2016 43

3.2 Study II: *S.* Paratyphi A genomics study 44

3.2.1 Population structure of *S.* Paratyphi A in Bangladesh 44

3.2.2 AMR genes and plasmid profile of *S.* Paratyphi A 44

3.2.3 Evolutionary history of pHCM2 plasmid in Bangladesh 46
3.2.4 Comparative pan-genome analysis 47
3.3 Study III: *S.* Paratyphi B genomics study 47
3.3.1 *In silico S.* Paratyphi B biotype identification 47
3.3.2 Population structure *S.* Paratyphi B complex in Bangladesh 47
3.3.3 AMR genes and plasmid replicons of *S.* Paratyphi B complex 50
3.3.4 Comparative pan-genome analysis 50
3.4 Study IV: Novel WGS approach based diagnostic assay 51

CHAPTER 4 DISCUSSION 54
4.1 General discussion 54
4.2 Future work and conclusion 61

REFERENCES 63

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 1.1	Kaufmann-White antigenic classification of *Salmonella* serovars	3
Table 2.1	Biochemical tests and serotyping for identification of *S.* Typhi and *S.* Paratyphi A, B	29
Table 2.2	List of primers targeting *S.* Typhi, *S.* Paratyphi A and H58 *S.* Typhi genes	36
Table 2.3	The thermal cycle of PCR reaction	36
Table 3.1	List of genetic determinants of AMR in 202 *S.* Typhi strains	42

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 1.1	Classification of the genus *Salmonella* and its disease.	2
Figure 1.2	Estimated incidence rates of enteric fever per 100,000 populations by country, according to GBD study in 2017	5
Figure 1.3	Timeline of the use of antibiotic and development of resistance in enteric fever from 1950 to 2018	9
Figure 1.4	Three different generations of whole genome sequencing platforms	11
Figure 1.5	Basic workflow for WGS data analysis.	14
Figure 1.6	Global population structure of *S.* Typhi based on genome-wide SNP.	17
Figure 1.7	Global population structure for *S.* Paratyphi A.	18
Figure 1.8	Population structure of *S.* Paratyphi B complex.	19
Figure 1.9	Pseudogene formation during the evolution of *S.* Paratyphi A and *S.* Typhi.	20
Figure 1.10	The genetic structure of MDR and XDR *S.* Typhi.	23
Figure 3.1	Spatial analysis of 202 Bangladeshi *S.* Typhi in Dhaka city with detailed distribution of genotypes	39
Figure 3.2	Global population structure of non-H58 and H58 Bangladeshi *S.* Typhi genotypes.	40
Figure 3.3	Timeline of urban Dhaka *S.* Typhi genotypes and AMR profiles from 2004-2016.	43
Figure 3.4	Global population structure of Bangladesh *S.* Paratyphi A with AMR genes and plasmid replicons.	45
Figure 3.5	Comparative pHCM2 plasmid analysis.	46
Figure 3.6	Mid-point rooted maximum-likelihood phylogeny of *S.* Java BAPS cluster 1.1.	49
Figure 3.7	Pan-genome dynamics of the *S.* Paratyphi B complex.	51
Figure 3.8	Single-plex PCR for the diagnosis of typhoidal pathogens.	52
Figure 3.9	Multi-plex PCR for the diagnosis of typhoidal pathogens.	53

LIST OF ABBREVIATIONS

ACT	Artemis Comparison Tool
AMR	Antimicrobial resistance
ARIBA	Antimicrobial Resistance Identification By Assembly
BAM	Binary Alignment Map
BAPS	Bayesian Analysis of Population Structure
BEAST	Bayesian Evolutionary Analysis Sampling Tree
BHI	Brain heart infusion
BiPs	Biallelic polymorphisms
BITID	Bangladesh Institute of Tropical and Infectious Diseases
BLAST	Basic Local Alignment Search Tool
BWA	Burrows-Wheeler Alignment
CARD	Comprehensive Antibiotic Resistance Database
CDS	Coding Sequence
CRISPR	Clustered regularly interspaced short palindromic repeats
DGHS	Directorate General of Health Services
dNTPs	Di-deoxynucleotides
dTa	Dextrorotatory tartrate
ESBL	Extended Spectrum Beta Lactamase
EBI	European Bioinformatics Institute
GATK	Genome Analysis Tool Kit
icddr,b	International Centre for Diarrhoeal Disease Research, Bangladesh
IEDCR	Institute of Epidemiology, Disease Control and Research
IncH1	Incompatibility group H subgroup 1
InDel	Insertions and Deletions
KIA	Kligler's Iron Agar
LMICs	Low and middle income countries
LPS	Lipopolysacharide
MCC	Maximum Clade Credibility
mcr-1	Mobilized Colistin Resistance-1
MDR	Multidrug resistance
MIC	Minimal inhibitory concentration
MIVU	Mucosal Immunology and Vaccinology Team
MIU	Motile Indole Urea
ML	Maximum likelihood
MLST	Multi locus sequence typing
MRCA	Most recent common ancestor
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing

NTS	Non-typhoidal *Salmonella*
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PCR	Polymerase Chain Reaction
PFGE	Pulse-field gel electrophoresis
PG	Phylogroups
pMLST	Plasmid multi-locus sequence typing
PMQR	Plasmid mediated quinolone resistance
QC	Quality Control
QRDR	Quinolone Resistance Determining Region
RAxML	Randomized Accelerated Maximum Likelihood
SAM	Sequence Alignment/Map
SGI-1	*Salmonella* genomic island 1
SISTR	Salmonella *In Silico* Typing Resource
SNP	Single nucleotide polymorphisms
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
SPI-1	Salmonella Pathogenicity Island-1
SRST2	Short Read Sequence Typing
ST	Sequence Type
STRATAA	Strategic Typhoid Alliance across Africa and Asia
T3SS	Type three secretion system
TIS	Typhoid Immunization Surveillance study
TyVac	Typhoid Vaccine Acceleration Consortium
VCF	Variant call format
VFDB	Virulence factor database
Vi-PS	Vi Polysaccharide
Vi-TCV	Vi-Typhoid Conjugate Vaccine
WASH	Water sanitation and hygiene
WGS	Whole genome sequence
WSI	Wellcome Sanger Institute
WHO	World Health Organization
XDR	Extensively drug resistant

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Reagents and Preparation of Different Media	75
Appendix B	Published Paper I	77
Appendix C	Published Paper II	78
Appendix D	Published Paper III	79

CHAPTER 1

INTRODUCTION

Salmonella enterica is a leading cause of bacterial foodborne illness in many low- and middle-income countries (LMICs) [1]. *Salmonella enterica* serovars Typhi, Paratyphi A, Paratyphi B (d-tartrate-negative) and Paratyphi C are referred to collectively as typhoidal *Salmonella*, whereas other serovars are grouped as non-typhoidal *Salmonella* (NTS). Typhoidal *Salmonella* strains are human host-restricted bacteria that cause typhoid fever and paratyphoid fever, together referred to as enteric fever. Enteric fever has a wide variety of presentations that may range from systemic, febrile illness to relatively minor cases of diarrhea accompanied with low-grade fever [2]. The emergence of antibiotic resistance in *S*. Typhi and *S*. Paratyphi A has resulted in loss of the value of traditional first-line drugs and fluoroquinolones over time which poses a major problem for disease control [3]. Recent advances in genome sequencing offer insights into the evolutionary mechanism of *S*. Typhi and *S*. Paratyphi A, B which will be explored in this dissertation to reveal the population structure, transmission dynamics and antimicrobial resistance (AMR) pattern in Bangladesh.

1.1 Classification of the genus *Salmonella*

Salmonella was first discovered in 1884 by the American pathologist Daniel Elmer Salmon who isolated *Salmonella choleraesuis* from porcine intestine. *Salmonella* is a Gram-negative, facultative, lactose non-fermenting and rod-shaped bacterium. The genus *Salmonella* belongs to the family *Enterobacteriaceae* and is divided into two species, *Salmonella enterica* and *Salmonella bongori. Salmonella enterica* is further divided into six subspecies, which contains more than 2,500 serovars or serotypes [4]. Each *Salmonella* serovar has been classified and named according to the disease, the geographical region, the animal from which the organism was isolated. Approximately 99% of *Salmonella* serovars belonging to *Salmonella enterica* subspecies *enterica* cause human and animal infections which are then divided into typhoidal and non-typhoidal serovars **(Figure 1.1)** [5].

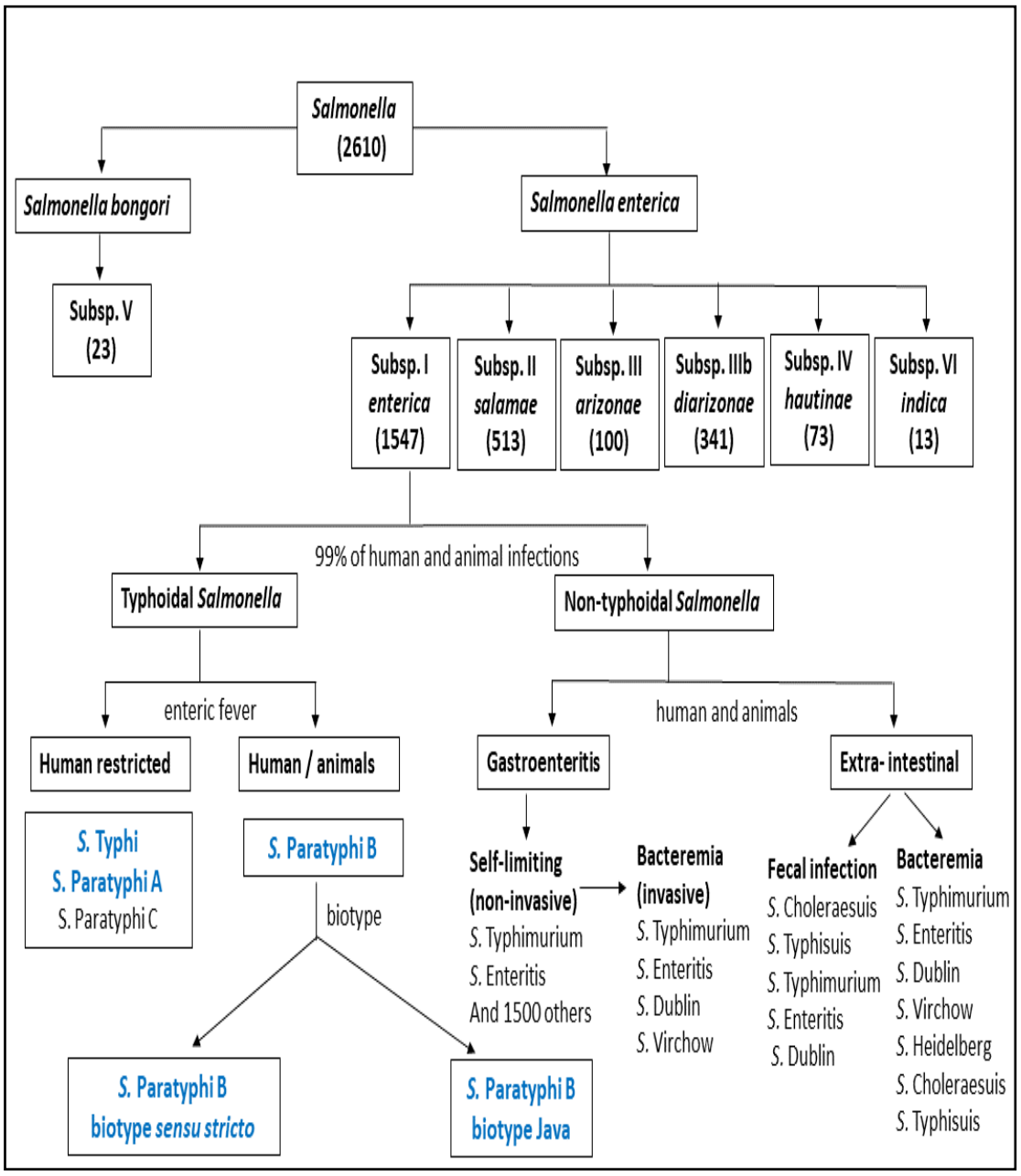

Figure 1.1 Classification of the genus *Salmonella* and its disease.
Numbers in the brackets indicate the total number of serovars in each subspecies. Blue highlighted serovars/biotypes are described in this dissertation (Adapted from: Akyala and Alsam. 2015 [6])

According to the White-Kauffmann-Le Minor scheme, the serovars are differentiated by antigenic structure based on the O (somatic or lipopolysaccharide; LPS) antigen, H (flagellar) antigens (phase 1, phase 2) and less frequently by the capsular antigens (Vi) [4, 5]. Each *Salmonella* serovar is recognized by its unique combination of these antigens which is known as its antigenic formula **(Table 1.1)**.

Table 1.1 Kaufmann-White antigenic classification of *Salmonella* serovars (Adapted from Selander *et al.* 1990 [7])

Group	Serovar	Somatic "O" antigen	Flagellar (H) antigens		Antigenic formula
			Phase 1	Phase 2	
A	*S.* Paratyphi A	1, 2, 12	a	-	1,2,12:a:-
B	*S.* Paratyphi B	1, 4, 5, 12	b	1, 2	1,4,[5],12:b:1,2
	S. Typhimurium	1, 4, 5, 12	i	1, 2	1,4,[5],12:i:1,2
	S. Gloucester	1, 4, 12, 27	i	i, w	1,4,[5],12:i:i,w
C1	*S.* Paratyphi C	6, 7, Vi	c	1, 5	6,7,[Vi]:c:1,5
	S. Cholerasuis	6, 7	c	1, 5	6,7:c:1,5
C2	*S.* Newport	6, 8	e, h	1, 2	6,8:e,h:1,2
D	*S.* Typhi	9, 12, Vi	d	-	9,12,[Vi]:d:-
	S. Enteridis	1, 9, 12	g, m	-	1,9,12:g,m:-
E	*S.* Anatum	3, 10	e, h	1, 6	3,10:e,h:1,6
F	*S.* Aberdeen	11	i	1, 2	11:i:1,2
G	*S.* Poona	13, 22	z	1, 6	13,22:z:1,6

1.2 General overview: *S.* Typhi and *S.* Paratyphi A, B

1.2.1 Antigenic and biochemical characteristics

Both *S.* Typhi and *S.* Paratyphi A cause enteric fever with similar clinical symptoms, but they are variable on the basis of antigenic property. For example: serotyping defines for *S.* Typhi strain by the presence of the somatic antigen O9, 12 (group D), flagella antigen Hd and Vi polysaccharide capsular antigen with antigenic formula written as: 9,12[Vi]: d: -, whereas the antigenic formula for *S.* Paratyphi A is 1,2,12: a: - **(Table 1.1)**.

Conversely, *Salmonella enterica* serotype Paratyphi B is differentiated into two distinct biotypes based on dextrorotatory tartrate (dTa) fermentation and slime wall formation: *Salmonella enterica* serotype Paratyphi B biotype *sensu stricto* (*S.* Paratyphi B *sensu stricto*; dTa^-, slime wall positive) and *Salmonella enterica* serotype Paratyphi B biotype Java (*S.* Java; $d Ta^+$; slime wall negative), although both typhoidal and non-typhoidal biotypes share the antigenic formula 1,4,[5],12:b:1,2. Collectively,

these two biotypes comprise the *S.* Paratyphi B complex. These biotypical differences are still used to differentiate between invasive *S.* Paratyphi B *sensu stricto* and non-invasive *S.* Java which are responsible for human-restricted paratyphoid fever and gastroenteritis respectively, however this method may be unreliable for identifying these two biotypes in terms of pathogenic mechanisms [8, 9]. Recently whole genome sequencing (WGS) approach has resolved the confusion around the *S.* Paratyphi B complex by defining detailed population structure and diversity of invasive and non-invasive lineages of *S.* Paratyphi B biotypes [8].

1.2.2 Global epidemiology

Enteric fever remain a major public health concern in LMICs having low socio-economic conditions with limited access of safe water supply and inadequate sanitation system [10, 11]. Recent global estimates implicate that typhoid fever causes around 12 million cases with 129,000 deaths whereas paratyphoid fever causes 3.4 million cases with 19,100 deaths each year globally [1, 12-15]. According to the Global Burden of Disease study (GBD) in 2017, South Asia experienced the highest incidence rate (549 episodes per 100,000 person-years, accounting for 71.8% of global case) and mortality rate (94,700 deaths, accounting for 69.6% of global cases) of enteric fever **(Figure 1.2)** [12].

The Dhaka Hospital of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), the world's largest diarrheal disease hospital, was reported previously the longer prodrome of illness including diarrhoeal symptoms among the majority of the hospitalized enteric fever patients [16]. Moreover, mostly communities living close to the rivers Buriganga, Turag, and Balu have an elevated risk of enteric fever infection compared to communities in other locations [17]. Recently, the enteric fever burden has assessed in a multicenter population based study as a part of Strategic Typhoid Alliance across Africa and Asia (STRAATA). The overall incidence of blood culture confirmed *S.* Typhi and *S.* Paratyphi A was high in Dhaka, Bangladesh (161 cases and 42 cases per 100,000 person-years, respectively), followed by Kathmandu, Nepal (74 cases and 6 cases per 100,000 person-years, respectively) and Blantyre, Africa (58 cases and 0 cases per 100,000 person-years, respectively) [18]. However, the incidence of typhoid fever has decreased dramatically in the high income developed countries due to improvement of hygiene and introduction of an

effective nationwide vaccination program. But still the developed countries are facing high transmission of enteric fever cases because of frequent travelling across the countries where enteric fever is endemic. In the United States, 74% of typhoid fever cases reported to the United States Centers for Disease Control and Prevention each year due to travel to Mexico, Philippines, Haiti and the Indian subcontinent [19].

Recently, *S.* Paratyphi A has accounted for a growing proportion of paratyphoid cases in Asia, causing up to 35% of all enteric fever episodes in India and Nepal and more than 60% in China [20]. However, substantial numbers of invasive paratyphoid fever cases caused by *S.* Paratyphi B biotype *sensu stricto* were reported in Chile in the 1980s [9, 21]. Of note, an increasing incidence of non-typhoidal *S.* Paratyphi B biotype Java infection, especially from poultry sources have been observed in Germany, Netherlands, Belgium since 1990 [22], in the United Kingdom since 2010 [23] and also reported in other countries such as Saudi Arabia [24] and, Bangladesh [25].

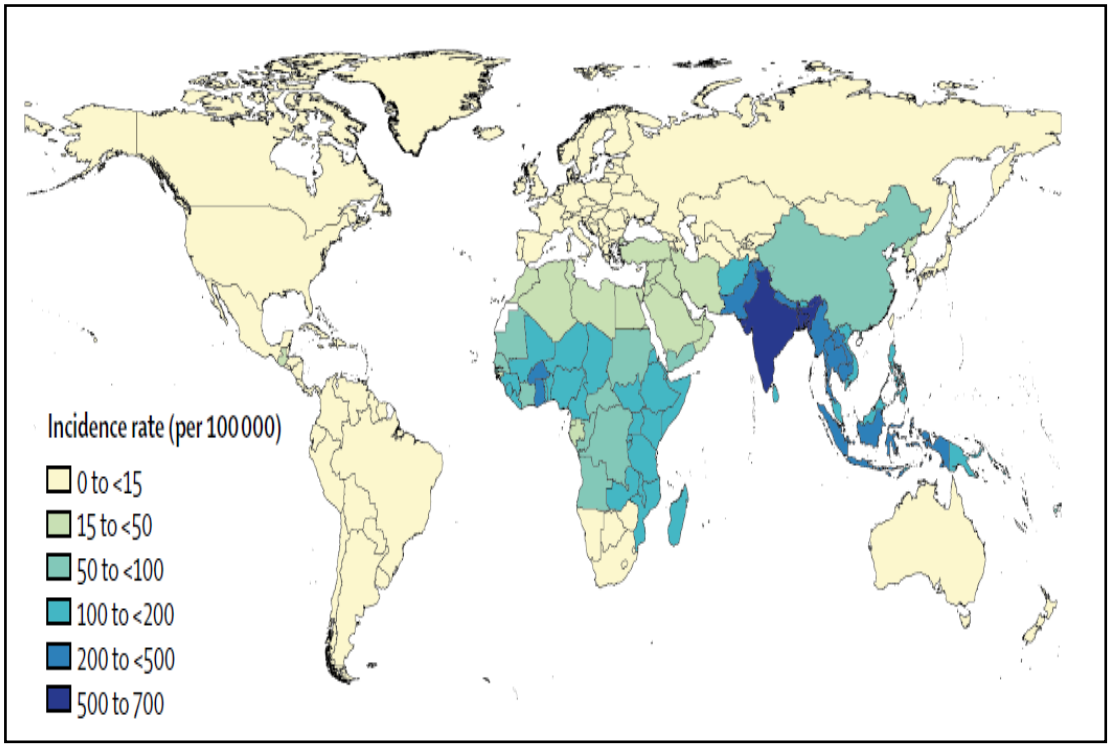

Figure 1.2 Estimated incidence rates of enteric fever per 100,000 populations by country, according to GBD study in 2017 (Source: Stanaway. *et al.* 2019 [12])

1.2.3 Pathogenesis and clinical features

Salmonella infection in human occur through the ingestion of contaminated food and water containing the infectious dose of this organism. Following ingestion, NTS invades the intestinal epithelial cells by delivering effector proteins through *Salmonella* Pathogenicity Island-1 (SPI-1) encoded type three secretion system (T3SS) to cause gastroenteritis [26]. In case of systemic infection, typhoidal *S.* Typhi and *S.* Paratyphi A colonizes the small intestine through M cells of the Peyer's patches to invade the mucosal surface, subsequently migrating into the mesenteric lymph nodes where multiplication occurs. The bacteria then reaches to the circulation via thoracic duct resulting in primary bacteremia [27, 28]. Afterwards, the bacterial load is removed from blood by macrophages which spread throughout the body via the reticuloendothelial system including the liver, spleen and bone marrow. Following the incubation period of 7 to 14 days, there is onset of clinical symptoms such as fever and malaise accompanied by chills, headache, anorexia, nausea, abdominal discomfort, dry cough, myalgia, coated tongue, tender abdomen, hepatomegaly, and splenomegaly [27, 29].

Approximately 2-5% patients fail to recover from the acute infection due to colonization of *S.* Typhi in the gallbladder and shed high numbers of *S.* Typhi in stool for more than one year after acute infection, leading to long term chronic fecal shedding [30-32]. Most of the severe complications such as: intestinal hemorrhage and perforation, gastrointestinal bleeding, toxic myocarditis, bronchitis, encephalopathy, pneumonia develop in the third and fourth weeks of the infection in untreated patients [33]. However, the complications and occurrence of stool shedding in convalescent stage may be lower due to early antibiotic usage [31]

1.2.4 Diagnostic approaches

The diagnosis of enteric fever has become difficult due to the fact that typhoid and paratyphoid infection present as nonspecific clinical syndromes which can be difficult to differentiate from other febrile illnesses caused by pathogenic organisms.

1.2.4.1 Microbiological culture based diagnosis

A confirmed diagnosis of enteric fever requires culturing of the organism either from blood, stool or bone marrow which are not only considered as gold standards for detection of the bacteria but also offer antimicrobial susceptibility profile to prescribe appropriate antibiotics for early patient management [34]. Both *S.* Typhi and *S.* Paratyphi can be detected from the bloodstream in around 90% of untreated patients in the first week of illness but the isolation rate decreases to 50% at the third week of illness [35-37]. Stool culture can give positive result only after the first week of infection and is useful for chronic carrier detection because of bacterial shedding through the gallbladder even after acute illness [38]. However, bone marrow culture is more accurate (80-95% sensitive) compared to blood and stool culture [39]. The major drawback of these bacterial isolation methods is that these are time consuming and are not feasible for routine use in many endemic settings where proper microbiology laboratory facilities are inadequate.

1.2.4.2 Serological diagnosis

Various serological tests such as Widal test [40], Tubex [41], Typhidot [42] are widely used in low resource settings which detect serum antibodies against the somatic O and flagellar H antigens. These serological tests are rapid, cheap and simple to perform, but it may give false positive result because of cross reactivity of antigens with other *Salmonella* serovars [43].

1.2.4.3 Molecular diagnosis

Molecular diagnostic methods such as polymerase chain reaction (PCR), nested PCR, microarray, next generation sequencing (NGS) have considered to be highly sensitive, specific and rapid compared to conventional culture-based and serological methods. PCR based approach can amplify *Salmonella* serovar specific target genes even when very low numbers of bacteria are present in blood or stool specimen [44-46]. Several target genes such as O somatic antigen gene, H flagellar antigen gene, Vi capsular antigen gene, 16s rRNA genes have previously used for both *S.* Typhi and *S.* Paratyphi detection which are also found in other *Salmonella* serovars [45, 47]. Recently, computational approaches such as nucleic acid sequence alignment tools, primer designing tools have used to identify unique, serovar specific target genes

which can be used as primer to design novel multiplex PCR at wet lab for detection [45, 48] .

1.2.5 Antibiotic treatment and resistance

Over the past 70 years, traditional first line antimicrobials (chloramphenicol, ampicillin and trimethoprim-sulfamethoxazole) have been used to treat enteric fever and reduced the overall mortality rate to less than 1% [37]. However, resistance to chloramphenicol was reported in 1972 and subsequently resistance to all first line antibiotics, known as multidrug resistance (MDR), have emerged by the end of 1980 **(Figure 1.3)** [49]. Over the past two decades, a high burden of MDR typhoid has been previously reported in many endemic settings throughout Asia [50-53], Africa [54-58] and the Middle East [59] which led to changes in empirical treatment regimens. The emergence of MDR exacerbates the widespread use of second generation fluoroquinolones (ciprofloxacin, ofloxacin) [60, 61]. However, fluoroquinolone was not commonly used in children due to reported side effects associated with joint or cartilage damage. Different studies conducted in Vietnam and Nepal have reported full nalidixic acid and intermediate levels of ciprofloxacin resistant *S.* Typhi that emerged within a couple of years of ciprofloxacin introduction [61, 62]. Moreover, *S.* Paratyphi A became dominant in recent years because of high frequency of fluoroquinolone non-susceptibility, particularly in Asia. A recent study in Nepal reported decreased fluoroquinolone susceptibility with no evidence of MDR among *S.* Paratyphi A strain [53]. Due to indiscriminate and increasing over-the-counter sale of fluoroquinolone over the last decade, World Health Organization (WHO) recommended third generation cephalosporin (ceftriaxone, cefixime) and azithromycin for treatment [63]. However, sporadic reports of ceftriaxone resistant *S.* Typhi and azithromycin resistant *S.* Typhi, *S.* Paratyphi A in Bangladesh are of great concern as only these antibiotics are currently available for treatment [64, 65]. In addition, the first major outbreak of extensively drug resistant (XDR) *S.* Typhi (i.e. defined as resistant to first-line antibiotics, fluoroquinolones and third-generation cephalosporin) was observed in Pakistan from 2016 onwards [66].

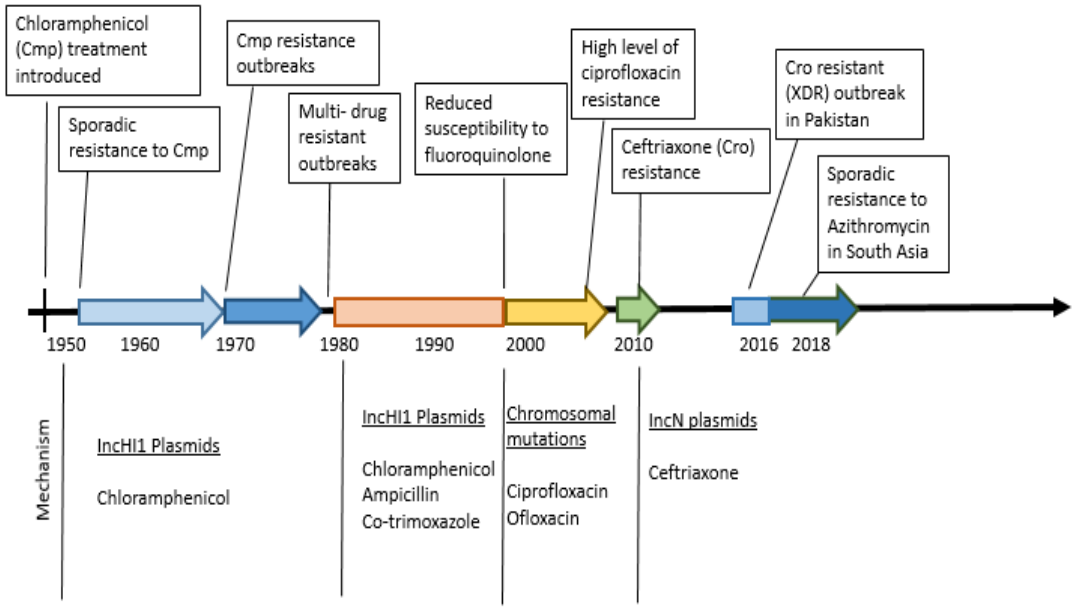

Figure 1.3 Timeline of the use of antibiotic and development of resistance in enteric fever from 1950 to 2018 (Adapted from Akram *et al.* 2020 [67], Andrews *et al.* 2018 [68])

1.2.6 Vaccination

Water sanitation and hygiene (WASH) interventions are generally considered as major preventive measures for enteric fever. However, a significant proportion of individuals from LMICs are living under inadequate infrastructure with limited access to clean water and sanitation facilities [39]. The introduction of nationwide vaccination program against *S.* Typhi targeting the high risk populations in endemic countries could be a beneficial control strategy for enteric fever. Currently two licensed typhoid vaccines have been used: live attenuated oral vaccine Ty21a (chemically mutated Ty2 *S.* Typhi parental strain) and intramuscular injectable Vi polysaccharide vaccine (purified Vi capsular polysaccharide of Ty2 strain). But these vaccines are not widely implemented due to moderate effectiveness among young children (efficacy of 58% and 59%, respectively) [37, 69]. As Vi polysaccharide vaccine alone could not stimulate T cell mediated immune response, the strategy of conjugating polysaccharide to a carrier protein have considered to overcome the limitation. In January 2018, WHO prequalified a single dose of Vi-TT typhoid conjugate vaccine (Vi polysaccharide conjugated to tetanus toxoid protein) manufactured by Bharat Biotech International, Hyderabad, India as safe, immunogenic in all age groups. Recently, the Typhoid Vaccine Acceleration Consortium (TyVAC) has carried out a

large scale, cluster randomized, controlled, clinical Vi-TT vaccine trial in Bangladesh, Nepal and Malawi which proven to be safe, immunogenic and showed 81% efficacy among Bangladeshi children at ages under two years [70, 71]. Vi-polysaccharide conjugate vaccine confers no protection against *S.* Paratyphi A, B, as they lack the Vi polysaccharide antigen. However, Ty21a vaccine reportedly provides some cross-protection against infection with *S.* Paratyphi A and B, which share the O12 antigen with *S.* Typhi [9, 72]. With the rapid advances in NGS technology, a number of promising O-antigen specific live attenuated and conjugate *S.* Paratyphi A vaccines are in the early phases of clinical development [72, 73].

1.3 Next generation sequencing (NGS) approaches

The advancement of NGS technology allows the possibility to capture information of whole genome at once, whereas earlier only a few genes could be studied at a time through targeted approaches. In 1975, Frederick Sanger initiated the path to WGS technology with Sanger sequencing which dominated for about 25 years. As it was expensive and time consuming for routine use, second generation sequencing has taken the place of first generation Sanger sequencing method. Later on in between 2005-2007, the second generation sequencing technologies: Roche-454 (array-based pyrosequencing approach), Illumina/Solexa (sequencing-by-synthesis method) and Applied biosystem SOliD (Sequencing by Oligonucleotide Ligation and Detection) has been widely used **(Figure 1.4)** [74]. The basic concept of all these platforms involves random fragmentation of genomic DNA, adapter ligation to both fragment ends, amplification on the surface (bead/chip) and record incorporation of each nucleotide thorough camera optics. From costs per run and costs per gigabyte of data perspective, Illumina Mi-Seq and Hi-Seq platforms are currently dominating the sequencing platform [75] that enable to generate up to 15 gigabyte of output with 25 million of sequencing reads [76]. Moreover, third generation of long-read sequencing platforms including Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) enable to produce 10-100 kilo base read lengths although these methods have relatively high error rate (approximately 10% compared with 0.1% for Illumina) [77, 78].

In this dissertation, all the *S.* Typhi and *S.* Paratyphi A, B strains were sequenced using the Illumina HiSeq2500 platform. This technology works on a sequencing by synthesis approach which involves binding of the four differently fluorescent labeled reversible terminator di-deoxynucleotides (dNTPs) with the DNA templates. The cluster of DNA fragments is clonally amplified to generate millions of copies of single stranded DNA in a process called cluster generation. After each cycle, the Illumina software performs base calling and assigns A, C, T or G to each cluster based on the fluorescent signals [79].

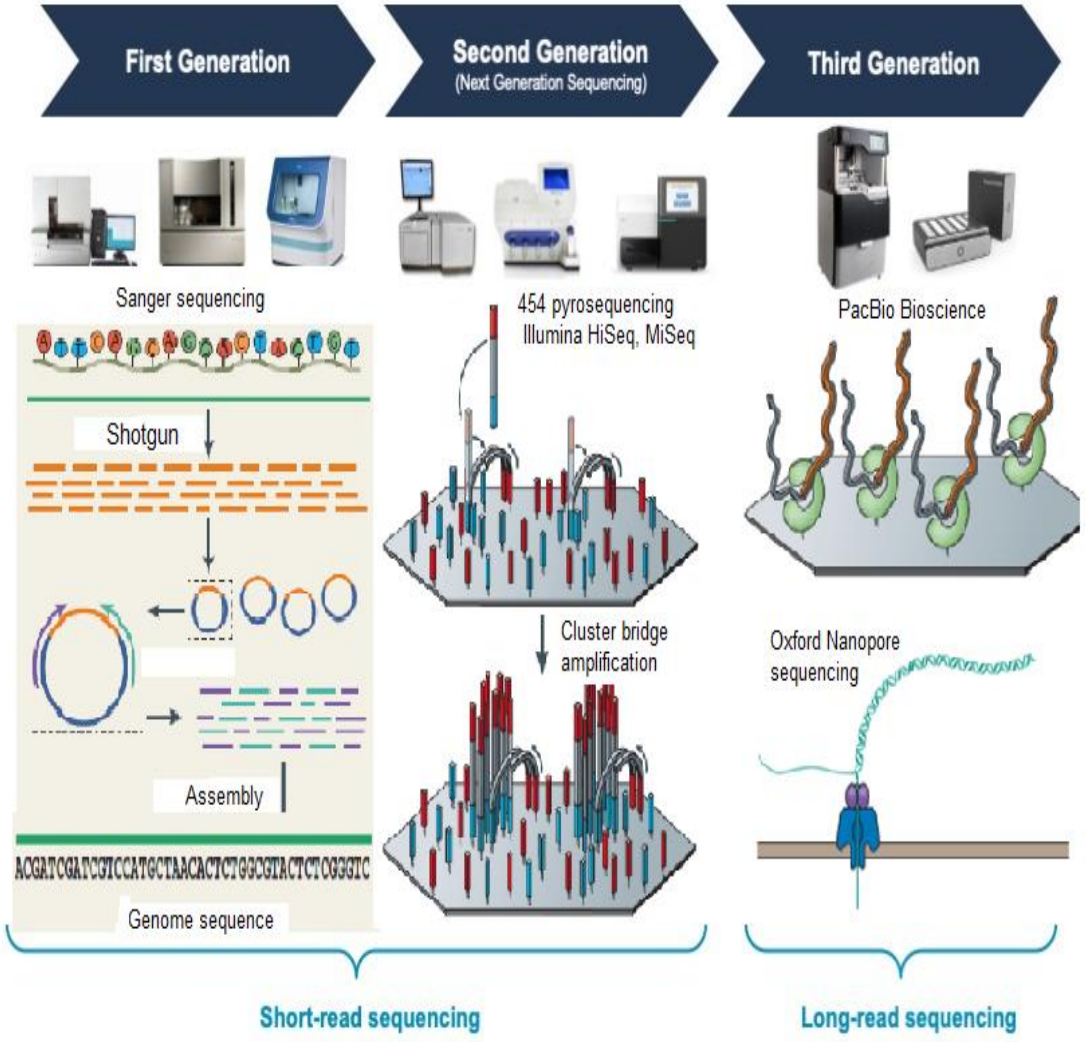

Figure 1.4 Three different generations of whole genome sequencing platforms
First generation (Whole shotgun sequencing, Sanger sequencing), Second generation (454 pyrosequencing, Illumina platform), Third generation (Pac-Bio Bioscience, Oxford Nanopore sequencing). Adapted from Loman *et al.* 2015 [80]

1.3.1 An introduction to the bioinformatics approaches

High-throughput sequencer platforms sequence each base multiple times to minimize the false positive base calls. Dealing with a bulk amount of sequence data is challenging to perform meaningful downstream analysis and also for answering biological relevant questions. Different bioinformatics pipelines together with computational algorithms are needed to process, analyze, compare, interpret, identify and visualize the complex sequence data along with structural variants which are explained in the following steps **(Figure 1.4)**:

Step I: Quality Control (QC) of raw reads

Raw sequence data in fastq format often contain low-quality, contaminant sequence reads with adapter sequences due to improper trimming which affect assembly procedure and might lead to false conclusions. FastQC, Trimmomatic tools are used to filter out low quality reads or N bases and trim adapters by observing Phred or Q score, GC content distribution, per base sequence content, read length distribution, and sequence duplication level [81, 82].

Step II: Genome assembly/alignment

De novo assembly of short reads is performed by identifying overlaps in short reads and joining them to form a contig without using reference genome, whereas reference assembly maps short sequence reads against reference genome to produce whole assembled genome in Sequence Alignment Map (SAM) format [83, 84]. Sometimes, alignment tools mis-assemble low coverage or repetitive regions that can be resolved by re-aligning the original reads. Moreover, read duplication error may occur due to PCR step in sequencing library preparation process that cause uneven amplification of DNA fragments. To reduce these harmful effects prior to variant calling, Genome Analysis Tool Kit (GATK) that realigns small insertions, deletions regions against a reference genome [85] and Picard Mark Duplicates tool (http://broadinstitute.github.io/picard/) that removes duplicated mapped reads are used. Additionally, assembly algorithms in different alignment tools calculate the probability for the correctness of the alignment for the whole read and also validate the performance of the assembly program optimally with a given distribution of library sizes [84]. For example, Burrows-Wheeler Alignment tool (BWA) utilizes

three algorithms: BWA backtrack, BWA-SW and BWA-MEM from which BWA backtrack is designed for Illumina sequence reads up to 100bp, while the rest of the two are designed for longer sequence reads ranged from 70bp to 1Mbp [86, 87].

Step III: Variant/Single nucleotide polymorphism (SNP) calling

Variant calling is important for comparative genomics to identify nucleotide-level differences by looking up every position in the genome with polymorphism relative to a reference genome. SAMtools [88], in combination of BCFtools [89] are commonly used to identify sequence variants such as: SNP, insertions and deletions (InDel), structural variants in coding, non-coding or intragenic regions. SNP calling is important to understand structural, functional and evolutionary changes as well as species diversification.

Step IV: Visualization

Large-scale sequencing datasets produced by NGS platform are visualized in an interactive, intuitive and graphical format to understand the features, functions, structure and evolutions of DNA and protein. Different gene visualization and annotation tools are able to display a range of sequence features in the context of genome annotation [90].

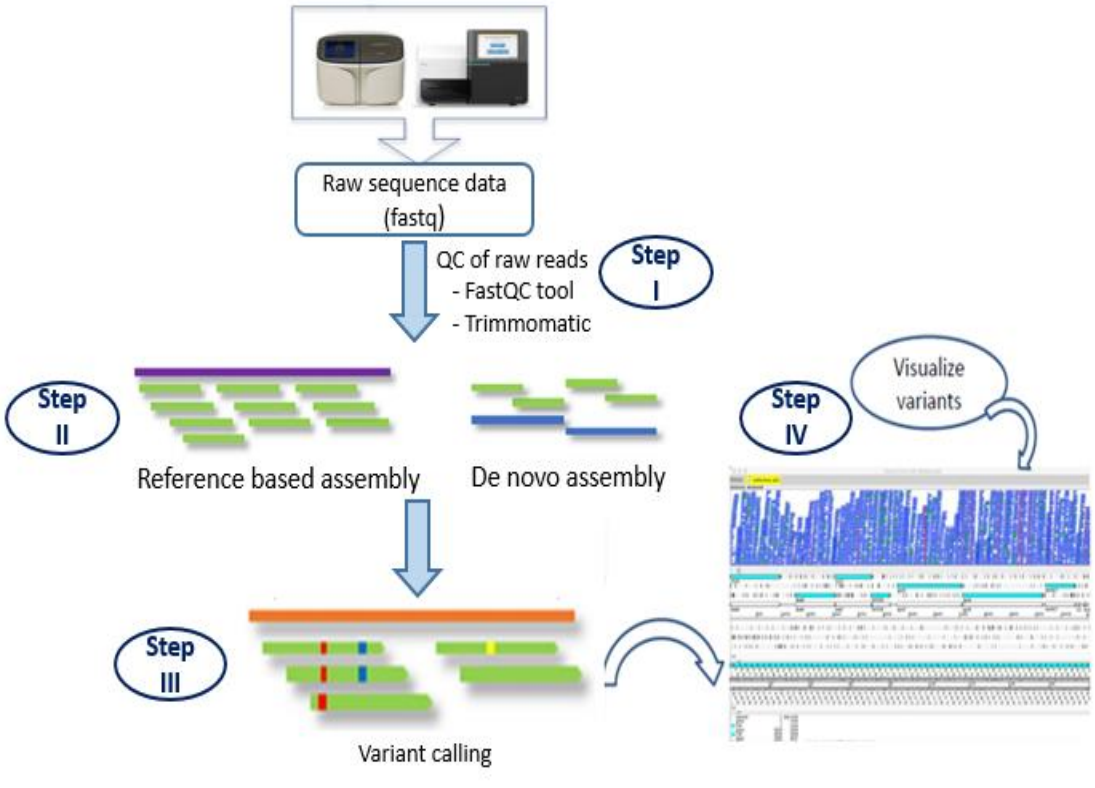

Figure 1.5 Basic workflow for WGS data analysis.
After generation of the raw sequence data from sequencer, the analysis of sequence data has undergone through multiple steps (Step I to step IV highlighted in blue circle)

1.3.2 Archiving WGS data

To manage a bulk number of sequence data under large scale genome sequencing projects, a data repository system is required which can archive, share and disseminate raw sequence reads. Genome sequence archive platforms such as National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI) provide comprehensively record of sequence assembly information, metadata of the released sequences and also allow free access to publicly available sequence data by browsing, downloading, sharing sequence data files with worldwide scientific communities [91].

1.3.3 *S.* Typhi and *S.* Paratyphi reference genome sequences

WGS era of *S.* Typhi initiated in 2001 with the release of the first complete genome of MDR *S.* Typhi CT18 (clinical strain isolated from typhoid patient in Vietnam) followed by the publication of P-stx-12 (clinical strain obtained from a typhoid carrier in India), Ty2 (laboratory pathogenic strain) and Ty21a (attenuated vaccine strain

from parental Ty2 strain) complete genome [92-96]. The genome size of the *S.* Typhi CT18 is 4,809,037 bp and harbours 106,516-bp cryptic plasmid (pHCM2), which shows recent common ancestry with a virulence plasmid of *Yersinia pestis* and 218,150 bp MDR associated IncH1 plasmid (pHCM1) [94].

The first *S.* Paratyphi A genome ATCC 9150 (4,585,229 bp genome in size) was sequenced in 2004, [97]. It was followed by publication of second complete genome *S.* Paratyphi A strain AKU_12601 which facilitated genome comparison between *S.* Typhi and *S.* Paratyphi A [97, 98]. *S.* Paratyphi A AKU_12601 genome consists of a 4,581,797 bp circular chromosome, encoding 4,285 coding sequence (CDS), and a 212,711 bp MDR IncHI1 like pAKU_1 plasmid [98, 99].

S. Paratyphi B complete genome SPB7 (NCBI Taxonomy ID: 1016998) was isolated from the stool of a female patient in Penang Malaysia and was annotated in 2017. The genome size is 4,858,887 bp and contains 4,718 genes of which 4,596 genes are CDS. Later in 2015, the complete genome sequence of *S.* Paratyphi B biotype Java strain harboring the mobilized colistin resistance gene (*mcr-1*) was published from the collection of the German National Reference Laboratory [100].

1.4 WGS application for *S.* Typhi and *S.* Paratyphi genomics

The wide advent of low cost, high-throughput NGS technologies and progressive increase in the number of genome sequence data have made a vast transformation in the field of bacterial genomics that offer new approaches in molecular microbiology, epidemiology, diagnostics, novel treatment strategy and new vaccine development. With the release of publicly available complete annotated *S.* Typhi and *S.* Paratyphi reference genomes, WGS successfully captured the fine genomic variations, evolutionary changes, population structure which play a vital role to identify genomic markers of AMR, resistance mechanism, virulence factors, emerging dominant variants during outbreak investigation [93]. To understand the disease dynamics and implement control strategies in poor endemic regions, WGS based enteric fever surveillance approach could contribute by collaborating and sharing genome surveillance data within countries. Recently, STRATAA consortium has combined traditional epidemiological approach with WGS approach for assessing disease burden

and investigating pathogen population structure, host genetics, AMR determinants, transmission dynamics [11, 18].

1.4.1 Phylogeography and population genomics

Both *S.* Typhi and *S.* Paratyphi A are slowly evolving monomorphic pathogens that undergo limited genetic variation due to accumulation of small numbers of SNP over time (approximately 1 substitution/year/genome) [101, 102]. Thus, it is often difficult to discriminate individual sequence variation of these highly clonal pathogens by phage typing [103], pulse-field gel electrophoresis (PFGE) [104] and, multi locus sequence typing (MLST) [105]. The advent of high-throughput WGS approach has overcome the limitations of these traditional typing approaches and provide greater discriminatory power with improved phylogenetically informative markers [101].

Roumagnac haplotyping scheme provided the first phylogenetic framework for epidemiological studies by classifying the *S.* Typhi population into 85 haplotypes based on 88 biallelic polymorphisms (BiPs) within 200 gene fragments (500 base pairs each) of a mixture of housekeeping, cell surface, regulatory, pathogenicity-associated genes, pseudogenes [106].

In 2016, a comprehensive robust SNP based genotyping scheme named "GenoTyphi" was developed to define phylogenetically informative lineages for *S.* Typhi by utilizing a set of 68 phylogenetic informative SNPs based on Bayesian Analysis of Population Structure (BAPS) which has recently updated with additional epidemiologically important and geographically clustered *S.* Typhi subpopulation [101, 107] (**Figure 1.5)**. Under this GenoTyphi nomenclature, the globally disseminated MDR clone haplotype H58 named as genotype 4.3.1 which was further subdivided into 4.3.1.1 (lineage I), 4.3.1.2 (lineage II) [51, 52] and 4.3.1.3 (formerly named "lineage bd") [108, 109]. Genotype 4.3.1 was initially emerged in South Asia, then subsequently entered into several countries of East Africa but completely absent in West Africa where genotype 3.1.1 (haplotype H56) found to be the dominating lineage [58, 110, 111]. During WGS study conducted in the Dhaka Shishu (Children) hospital, Dhaka, "Lineage bd" was originally reported which then formed a Bangladesh specific monophyletic sublineage (named Bdq) containing with high level ciprofloxacin resistance associated *qnr* genes [107, 108]. This scheme illustrates a

framework to establish genome based surveillance studies in poor endemic settings to track the spread of epidemiologically important genotypes to support public health measures and also has the potential to accurately predict origins at the country level [101]. However, the current genome based surveillance effort has added great value in monitoring impact of large scale immunization programme on pathogen population. The introduction of nationwide vaccination program in Thailand has successfully eradicated the endemic *S.* Typhi population by limiting disease outbreak and dissemination of lineages between different location [111] .

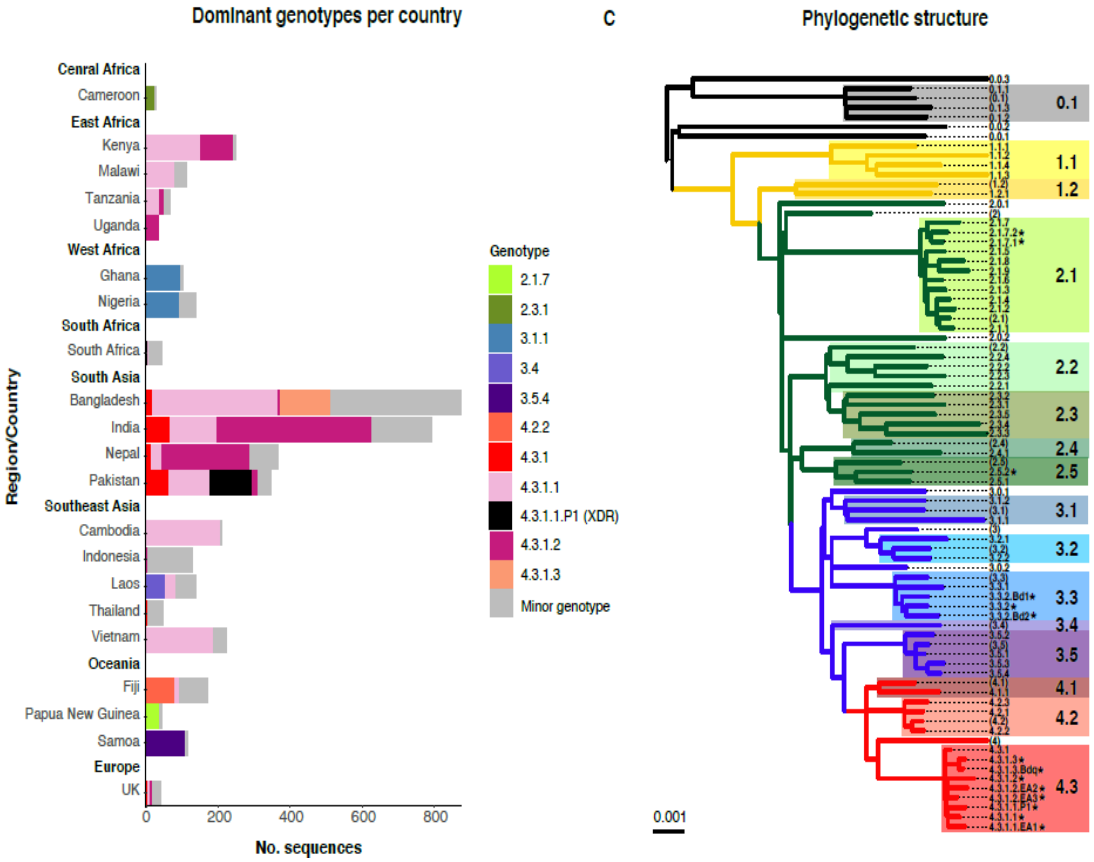

Figure 1.6 Global population structure of *S.* Typhi based on genome-wide SNP.

(a) Dominant and minor genotypes were shown for each country and coloured as per the inset legends (b) Whole genome SNP based phylogenetic tree backbone showing 16 *S.* Typhi clades which further differentiated into 63 subclades/sublineages (Source: Dyson *et al.* 2021 [107])

The population structure of *S.* Paratyphi A was first described by Bayesian Evolutionary Analysis Sampling Tree (BEAST) and Maximum Clade Credibility (MCC) phylogenetic tree approaches which identified seven lineages (designated as lineage A to G) among 149 global *S.* Paratyphi A genomes on the basis of 4,584 SNPs

in the non-repetitive, non-recombinant core genome (**Figure 1.6A)** [112]. Among the seven lineages, lineage A and lineage C were reported as the most globally dominant which further subdivided into sublineage A1-A2 and sublineage C1-C5, respectively during the outbreak in Cambodia and China [112-114]. Recently, a SNP based genotyping scheme named "Paratype" has developed using 1379 *S.* Paratyphi A global genome data to divide the total population into three primary, seven secondary and 18 subclades/genotypes **(Figure 1.6B)** [115].

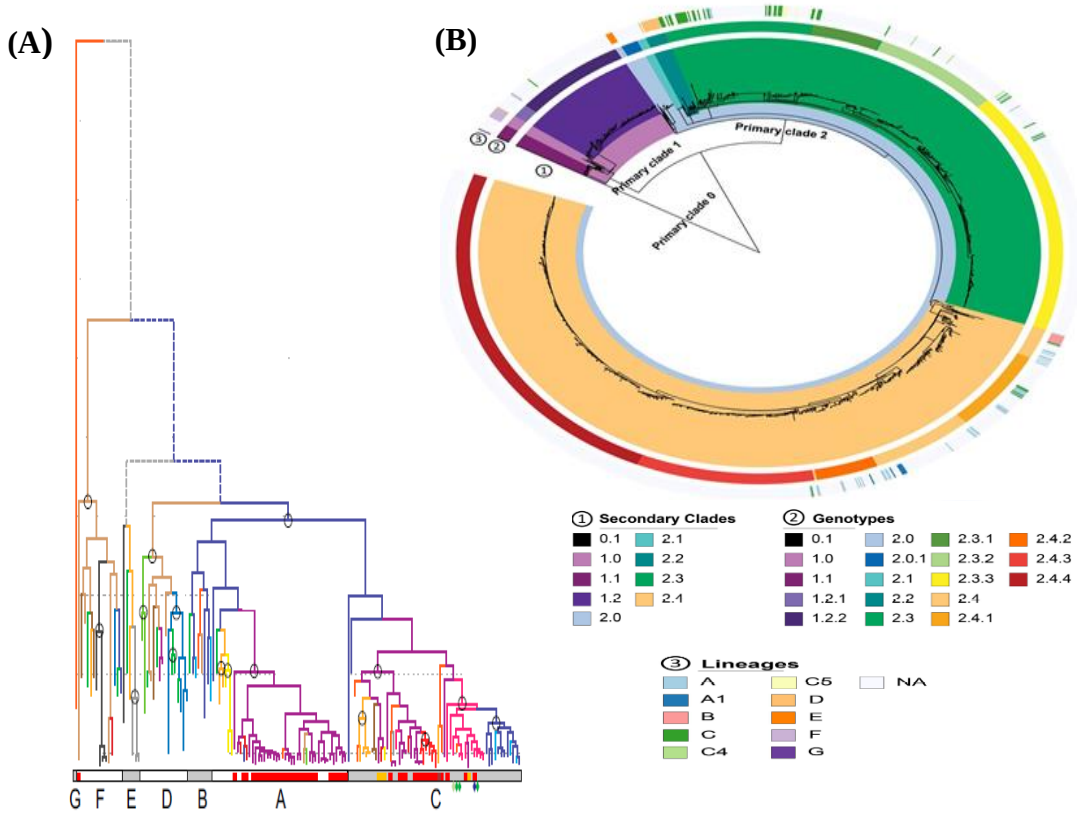

Figure 1.7 Global population structure for *S.* Paratyphi A.
(A) MCC tree of 149 global *S.* Paratyphi A strain defining seven lineages (lineage A-G) at the base (B) Whole genome SNP based phylogenetic tree of 1379 global *S.* Paratyphi A strain showing primary, secondary clades and subclades/genotypes (in the inner rings 1,2) and the previously proposed lineages in the outer ring 3. (Source: (A) Zhou *et al.* 2014 [112] and (B) Tanmoy *et al.* 2021 [115])

The biotypes of *S.* Paratyphi B complex sharing same antigenic formula 1,4,[5],12:b:1,2 was previously distinguished by Phage typing [116], IS*200* profiling [117], MLST) [118], clustered regularly interspaced short palindromic repeats (CRISPR) typing [119]. Later, the WGS analysis utilized the BAPS approach to subdivide the polyphyletic population structure of *S.* Paratyphi B complex into 10

distinct phylogroups (PG1 to PG10) in which the invasive biotype *S.* Paratyphi B *sensu stricto* (*d*Ta^{-}) strain grouped into PG1, while the remaining PGs (PG2 to PG10) comprised strain of biotype *S.* Java (*d*Ta^{+}) [8].

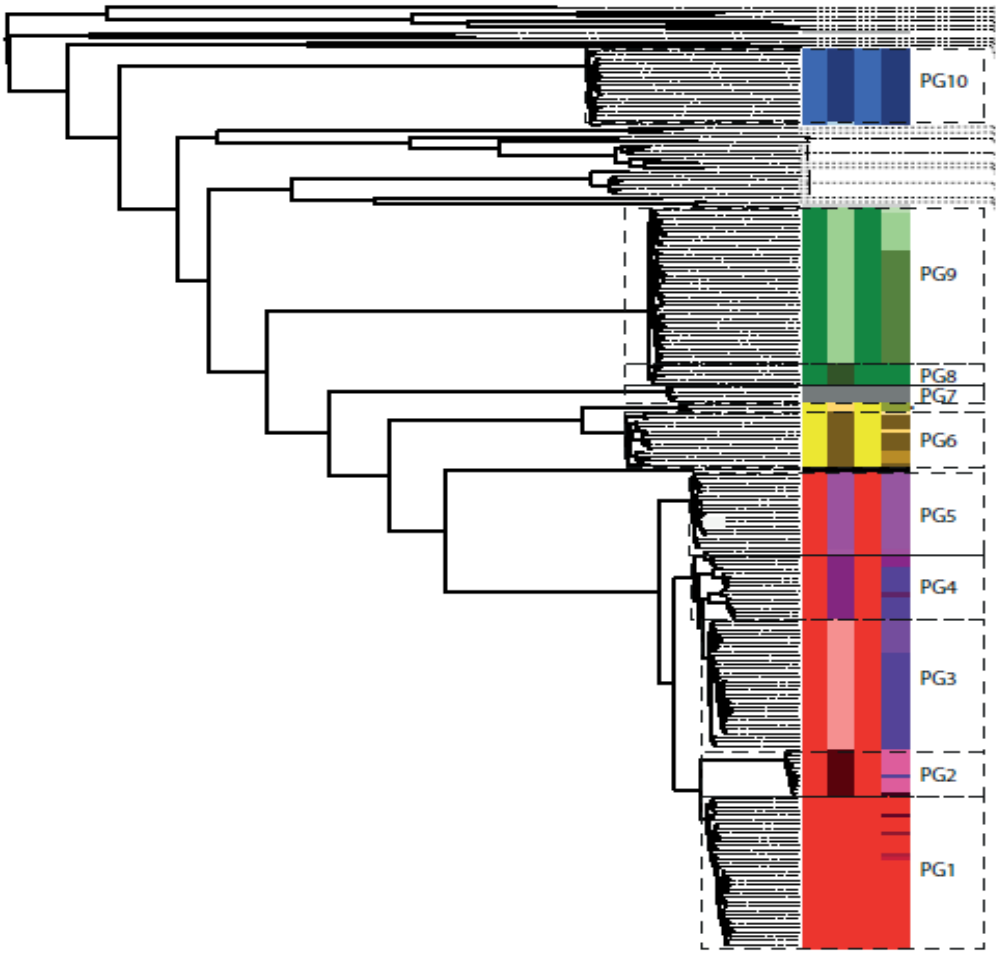

Figure 1.8 Population structure of *S.* Paratyphi B complex.
Whole-genome maximum likelihood phylogeny tree is constructed based on alignment of core genome SNPs. Coloured blocks indicate ten PGs (PG1-10) from where PG1belonged to *S.* Paratyphi B biotype *sensu stricto* and the remaining PGs were assigned to *S.* Paratyphi B biotype Java (Source: Connor *et al.* 2016 [8])

1.4.2 Understanding evolutionary changes using genomics

Salmonella primarily coevolved for millions of years since diverging from common ancestor *Escherichia. coli* and then undergone either neutral or detrimental mutations to facilitate host evolution of *Salmonella* serovars from broad host range to host restricted during infection. The most recent common ancestor (MRCA) implied that typhoidal *Salmonella* serovars Typhi and Paratyphi A had emerged from around 50,000 and 450 years ago, respectively, suggesting more recent evolutionary origin

compared to *S. enterica* and *E. coli* (dated 120-160 million years ago) [112, 120-122]. Horizontal gene transfers (HGTs) via exchange of plasmids, transposons, phages; acquisition of several *Salmonella* pathogenicity islands (SPIs), various virulence genes and extensive loss of gene function due to accumulation of pseudogenes are linked to the degree of invasiveness of the bacteria and have role in host specificity[123-125]. *S.* Typhi and *S.* Paratyphi A are host-restricted pathogens and have the ability to cause human specific systemic enteric fever by different paths. Both genome independently accumulated 210 and 173 pseudogenes, respectively that is around 5% of total 4400 protein coding genes. On the other hand, host generalist non-typhoidal *Salmonella* serovar such as Typhimurium accumulate only 0.9% pseudogene of total genome, facilitating to grow in wide variety animals, human and cause gastroenteritis [97, 98]. Ancestral pseudogenes inherited from a common ancestor arose slowly in the initial period of divergence prior to recombination. While majority of the shared pseudogenes containing different inactivating mutations accumulated rapidly after the recombination between a quarter of their genomes and became fixed as strain specific pseudogenes in each population. This might result host adaptation by disseminating to systemic niche to cause systemic invasive infection only in human and less capable of surviving in other niches (**Figure 1.7)** [98, 123, 125-127].

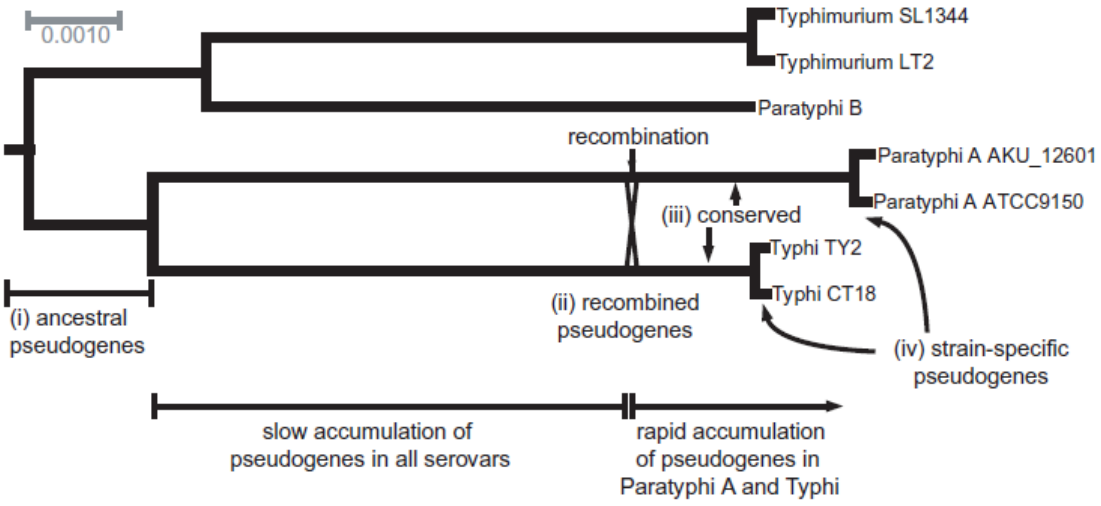

Figure 1.9 Pseudogene formation during the evolution of *S.* Paratyphi A and *S.* Typhi.

Phylogenetic tree was constructed based on multiple alignments of all non-recombined genes and rooted using *S.* bongori and *E.* coli as outgroups. Scale bar represents nucleotide divergence (Source: Holt *et al.* 2009 [98])

1.4.3 Comparative genomics and transmission dynamics

Genomic comparison between *S.* Typhi and *S.* Paratyphi A show low nucleotide divergence (mean 0.8%) and share similar gene content across 23% of their genomes via recombination indicating that they are more closely related each other at DNA level compared to multiple lineages of clinically important *Salmonella* serovars such as *S.* Paratyphi B complex. The remaining 77% genome were more nucleotide divergent (mean 1.2%) like other *Salmonella* serovars [98]. The *S.* Paratyphi A genome size is smaller than *S.* Typhi mainly due to the lack of SPI7 which is involved in Vi polysaccharide synthesis in *S.* Typhi [97]. On the other hand, the genetic variability in the polyphyletic *S.* Paratyphi B complex have occurred due to recombination at the flagellum loci that confer a wide range of host adaptation mechanisms and ability to colonize multiple hosts [8]. Host adaptation is accompanied by loss of bacterial fitness for inter-host transmission and apparent convergence in pathogenesis [128, 129]. Pseudogenes, fimbrial operons, virulence plasmids, lysogenic phages and genes belonging to SPI are involved in conferring host specificity and restricting the host range [130]. For example: *sopE* phage encoding an invasion associated secreted protein is present in invasive typhoidal *S.* Typhi, *S.* Paratyphi A, *S.* Paratyphi B biotype *sensu stricto* genome but absent in non-invasive, non-typhoidal *S.* Paratyphi B biotype Java which might confer a broad host range from human to animal [8].

Comparative genomic analysis has revealed vast differences in gene content between members of a single bacterial species, leading to the definition of the pan-genome. The pan-genome is the total number of genes including the core genome (genes present in ≥ 95% of total strains) and accessory genome (set of genes share by some but not all strains i.e < 95% of total strains) [131]. Pan-genome can be classified as open or close pan-genome. The size of pan-genome increases with addition of new genomes and thus additional strains yield novel genes. On the other hand, close pan-genomes are characterized by large core genome and small accessory genome indicating that addition of new sequenced genomes will not lead to acquire new genes. The higher functional pan-genome size in *S.* Java concluded open pan-genome due to importation of new functional genes while *S.* Typhi and *S.* Paratyphi A pan-genome are closed indicating limited potential for horizontal gene acquisition [8, 124, 132].

An open pan-genome of *S.* Java means it can colonize within multiple environments and exchange genetic material in multiple ways such as duplication of existing sequences, diversification, importation of new genetic material from other species [131, 133].

1.4.4 Molecular basis of AMR

Antimicrobials mainly have four mechanisms to inhibit the growth and even cause bacterial cell death: (1) inhibition of double DNA formation by inhibiting the DNA gyrase (eg. fluoroquinolones); (2) inhibition of RNA synthesis (eg. rifamycins); (3) inhibition of protein synthesis by inhibiting the 50S or 30S ribosome (eg. macrolides, aminoglycosides, tetracyclines); (4) inhibition of cell wall synthesis (eg. beta-lactams, glycopeptides, lipopeptides)[134] .

1.4.4.1 Genomic architecture of MDR *S.* Typhi

Emergence of MDR has become a major threat for typhoid treatment that may influenced reshaping of current *S.* Typhi population. The acquisition of MDR genes *blaTEM-1*, *catA1*, *sul1*, *sul2*, *dfrA*, *strA*, *strB*, *tetA* which confer resistance to ampicillin, chloramphenicol, sulfamethoxazole, trimethoprim, streptomycin, tetracycline by *S.* Typhi was historically associated with self-transmissible IncHI1 plasmids [51, 135, 136] . The global burden of MDR is largely driven by the dissemination of highly clonal, expanding haplotype H58 (genotype 4.3.1) [137]. These genes are encoded within a Tn*2670*-like complex transposable element comprising transposon Tn*6029*, which carries *blaTEM-1*, *strAB* and *sul2*, inserted into transposon Tn*2670* and itself comprises Tn*21* carrying a class I integron (including *sul1*, with *dfrA7* in the gene cassette) inserted into Tn*9* carrying *catA1*. Integration of the MDR composite transposon between the chromosomal genes STY3618 and STY3619, near *cyaA* gene may facilitate loss of the IncHI1 plasmid and thus reduce the potential fitness cost associated with MDR phenotype maintenance [49, 52, 53, 109]. Interestingly in West Africa, the evolution of MDR in *S.* Typhi have become established as IncHI1 plasmid associated H56 haplotype (genotype 3.1.1) instead of globally dominant H58 haplotype [58].

The high frequency of reduced fluoroquinolone susceptibility has occurred due to the accumulation of chromosomal mutations in the quinolone resistance determining

region (QRDR) of DNA gyrase genes *gyrA*, *gyrB* and DNA topoisomerase IV genes p*arC* and *parE*. [50, 60, 110]. The *S.* Typhi H58 triple mutant clade carrying three QRDR mutations (*gyrA-S83F*, *gyrA-D87N*, *parC-S80I*) have been observed in Nepal, India and also in Bangladesh acquiring a different sets of QRDR triple mutant (*gyrA-S83F*, *gyrA-D87G*, *parC-E84K*) associated with reduced fluoroquinolone efficacy [53, 108, 138]. Moreover, sporadic case of H58 sublineage bdq reported in Bangladesh which carried *qnrS* gene, *gyrA*-S83Y mutation associated with high median ciprofloxacin minimal inhibitory concentration (MIC 4 µg/mL) [108].

Recently, MDR H58 lineage I *S.* Typhi populations have evolved into an XDR clone by acquiring an additional novel IncY plasmid encoding a *bla*CTX-M-15 extended-spectrum β-lactamase gene potentially from an *E. coli* strain or another enteric bacterial donor which then underwent a clonal expansion in Pakistan (genotype 4.3.1.1.P1) [49, 66]. Due to rising trends of fluoroquinolone resistance and ongoing sporadic outbreak of XDR *S.* Typhi, azithromycin has become the last orally administered drug for typhoid treatment across South Asia. However, the increasing use of azithromycin places selective pressure for the emergence of azithromycin-resistant *S.* Typhi via the acquisition of non-synonymous point mutation in *acrB* gene at codon 717 which has been sporadically reported in Bangladesh, Nepal, India and Pakistan [64, 139-141].

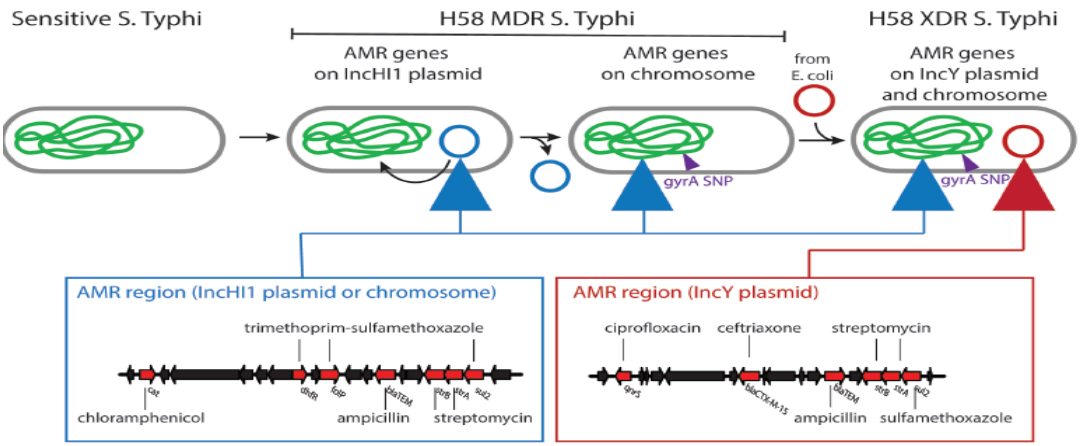

Figure 1.10 The genetic structure of MDR and XDR *S.* Typhi.
Oval outlines represent individual bacteria harboring chromosome (green) and plasmid (red and blue circles) elements. Red and blue triangles represent acquired resistance loci. Small purple triangles represent SNPs in the chromosome associated with resistance (Source: Dyson *et al.* 2019 [49])

1.4.4.2 AMR profile of *S*. Paratyphi A, B

The genetic basis of MDR in *S.* Paratyphi A was revealed by the sequencing of plasmid pAKU_1 which shared a common IncHI1 sequence backbone with the *S.* Typhi plasmid pHCM1 and an *S. enterica* serovar Typhimurium plasmid pR27. Thus, these plasmids appear to have acquired similar antibiotic resistance genes independently via the horizontal transfer of mobile DNA elements [99]. Apart from MDR prevalence, intermediate susceptibility to ciprofloxacin and full resistance to fluoroquinolones are very common in *S.* Paratyphi A [3, 53, 142, 143]. Furthermore, introduction of sporadic cases of cephalosporin and azithromycin resistance due to acquisition of extended-spectrum beta-lactamase (ESBL) genes and mutations in *acrB* gene, respectively are of great concern [144, 145]. However, MDR *S.* Paratyphi B biotype Java strains associated with poultry, poultry products have been observed since the late 1990s, [146]. Recently, WGS based study has reported a chromosomally located class 2 integron carrying *aadA24, dfrA, sul1, sul2, gyrA* genes that confer aminoglycoside, trimethoprim-sulphamethoxazole, fluoroquinolone resistance, respectively among poultry source associated PG10 lineage of *S.* Paratyphi B complex [8].

1.4.5 Implementation of WGS data to design PCR for diagnosis

Recent advances in WGS analysis of *S.* Typhi CT18, *S.* Paratyphi A AKU 12607 reference genome have contributed to identify serovar specific unique conserved regions for *S.* Paratyphi A and *S.* Typhi. In addition to detection of these serovars, the SNP based genotyping approach has added a new dimension to identify lineage specific unique SNP against the reference genome *S.* Typhi CT18 for the detection of MDR associated H58 Typhi lineages and emerging AMR clones [48]. A rapid, cost-effective and highly sensitive single-plex or multi-plex PCR can be designed utilizing serovar specific unique markers as primer sets for typhoidal pathogen diagnosis directly from either blood or stool samples. Coupling of *in silico* computational genome analysis with wet lab based PCR approach may contribute to implement innovative genome based surveillance for routine enteric fever diagnosis and also has the potential to inform disease control measures for implementing treatment strategies in LMICs [45, 147].

1.5 Thesis outline

In this dissertation, the population structure, transmission patterns and evolutionary dynamics of *S.* Typhi, *S.* Paratyphi A and *S.* Paratyphi B in Bangladesh were investigated using whole genome sequence analysis. In **'Study I',** a total of 202 Bangladeshi blood culture positive *S.* Typhi strains were assigned according to the extended *S.* Typhi SNP based genotyping framework to investigate genotypic distribution. Moreover, temporal trends in AMR genes and genotypes over time period between 2004-2016 were examined. In **'Study II'**, WGS-based approach was employed to assess genomic variation, AMR determinants and auxiliary plasmid profiles of 67 Bangladeshi *S.* Paratyphi A strains isolated from enteric fever patients in the last 10 years until 2018. The evolutionary history of the cryptic pHCM2-like plasmid identified in Bangladeshi *S.* Typhi, *S.* Paratyphi A, *S.* Paratyphi B populations were also investigated to reveal important aspects of the evolution of the plasmids and their movement between pathogen populations. In addition, comparative genomic analysis was carried out between Bangladeshi and globally circulating *S.* Typhi and *S.* Paratyphi A strains to better understand their transmission dynamics in **'Study I and II'.** A detailed genomic variation of *S.* Paratyphi B complex causing wide range of diseases, from gastroenteritis to paratyphoid fever, depending on the biotypes Java and *sensu stricto,* was examined in **'Study III'**. This is the first WGS study on Bangladeshi *S.* Paratyphi B in which Paratyphi B serotyped strains were taken from a large, longitudinal, nationwide multi-pathogen, enteric disease surveillance study where patients infected with a variety of enteric pathogens such as *Vibrio cholerae, Shigella spp,* enterotoxigenic *E. coli* (ETEC), typhoidal and non-typhoidal *Salmonella.* WGS analysis classified all these serotyped Paratyphi B strains as biotype Java associated with diarrhoeal symptoms which concluded the dominance of gastroenteritis caused by *S.* Paratyphi B biotype Java with no presence of paratyphoid causing *S.* Paratyphi B biotype *sensu stricto* in Bangladesh. Furthermore, comparative pan-genomic analysis was also conducted to investigate the composition of core and accessory genes among Bangladeshi *S.* Paratyphi A and *S.* Paratyphi B strains. Finally, **'Study IV'** describes the potential implementation of genomic data to evaluate an in-house built genome based diagnostic assay at icddr,b in collaboration with the University of Cambridge to improve better diagnostic accuracy which was validated by comparing the sensitivity with standard blood culture method. This assay

has been originally developed at the University of Cambridge by utilizing their *in silico* designed primer sets targeting unique genes for *S.* Typhi, H58 Typhi and *S.* Paratyphi A detection.

1.6 Hypothesis

The hypotheses of this research that population genomics and phylogeny clustering will define typhoidal *S.* Typhi, *S.* Paratyphi A and non-typhoidal *S.* Paratyphi B lineages in context of disease phenotype. Further, the study of these genomes will reveal genetic determinants and mechanisms (either chromosomal or plasmid mediated) of antimicrobial resistance which may play a vital role in implementation of serovar or biotype specific control and treatment strategies in endemic settings like Bangladesh.

1.7 Aims and objectives of the dissertation

The aim of this PhD dissertation was to explore genomic variation, evolutionary changes, and transmission dynamics of *S.* Typhi and *S.* Paratyphi A, B pathogen populations in Bangladesh utilizing high resolution whole genome sequence data in order to:

I. Understand the genotype distribution, AMR patterns and transmission dynamics of *S.* Typhi strains in urban Dhaka – **Study I (Paper I)**
II. Comprehensively assess the population structure, pan-genomic variation, AMR determinants and pHCM2 plasmid evolutionary history among *S.* Paratyphi A strains – **Study II (Paper II)**
III. Characterize the genetic diversity of diarrhoeal disease causing *S.* Paratyphi B biotype Java (initially serotyped as Paratyphi B), pan-genomic variation and AMR determinants of *S.* Paratyphi B complex – **Study III (Paper III)**
IV. Evaluate WGS based novel diagnostic assay as simple, rapid, cost effective and highly sensitive for *S.* Typhi including globally dominant H58 lineage and *S.* Paratyphi A detection– **Study IV**

CHAPTER 2

METHODS AND MATERIALS

2.1 Study settings and sample collection

International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) is an international health research organization located at the Mohakhali area in Dhaka which conducted various enteric disease surveillance studies in all over eight divisions across Bangladesh. All the genomic studies of this dissertation were carried out by the stored collection of *S.* Typhi and *S.* Paratyphi A, B strains from the previously established enteric disease surveillance studies of icddr,b.

'Study I' was conducted in three study sites of Typhoid Immunization Surveillance study (TIS study) [35, 36, 148] inside Dhaka city: icddr,b Dhaka hospital, Mohakhali; icddr,b Kamalapur field site and icddr,b Mirpur field site, in between 2004 and 2016. Kamalapur is situated in the southeast part of Dhaka, whereas Mirpur is located in the northeast part of Dhaka metropolitan area. Blood samples of suspected typhoid fever patients were enrolled from these three study sites based on the criteria of fever at least 38°C with a minimum duration of 3 days.

'Study II' was carried out in eight different study sites across Bangladesh from two already established enteric disease surveillance studies. Nationwide enteric disease surveillance study conducted in collaboration with the Institute of Epidemiology, Disease Control and Research (IEDCR) and icddr,b at 22 sentinel surveillance sites located in 21 different districts across seven divisions of Bangladesh [149]. Among them, *S.* Paratyphi A strains were taken from five hospitals (Site 1: Naogaon Sadar Hospital, Site 2: Potuakhali General Hospital, Site 3: Bangladesh Institute of Tropical and Infectious Diseases (BITID), Chittagong, Site 4: Uttara Adhunik Hospital, Site 7: Dhaka Medical College), in between 2014 and 2018. TIS study conducted within three urban areas in Dhaka city from which *S.* Paratyphi A strains were taken from all of these three sites (Site 5: icddr,b Mirpur field site, Site 6: icddr,b Mohakhali Hospital and Site 8: icddr,b Kamalapur field site) in between 2008 and 2016 [35, 36,

148]. Blood samples were collected from enteric fever suspected patients based on enrollment criteria as described above.

'Study III' was conducted in ten hospital sites of nationwide enteric disease surveillance study [149]. The surveillance sites were selected based on reports of acute watery diarrhoea according to the national District Health Information Software v2 Database from Directorate General of Health Services (DGHS) [150]. It is a large, longitudinal, multi-pathogen surveillance study which included patients infected with a variety of enteric pathogens: *Vibrio cholerae*, ETEC, *Shigella spp*, typhoidal and non-typhoidal *Salmonella*. Patients were enrolled into the nationwide enteric disease surveillance study if they were over two months old, and attended hospital with either a) three times or more loose or liquid stools; b) less than three times loose or liquid stools causing dehydration; or c) at least one bloody loose stool in the previous 24 hours [149]. As stool was the main sample source collected from the enrolled patients in the nationwide surveillance study, stool samples were included based on the above enrollment criteria in '**Study III'**.

Lastly, **'Study IV'** was designed to evaluate WGS based novel diagnostic approach by utilizing a subset of strain collection of study I and II (detail mentioned above)[35, 36, 148, 149].

2.2 Bacterial isolation

For '**Studies I, II and IV'**, blood samples (3mL for children <5 years of age and 5mL for others) were inoculated into blood culture bottle and were cultured using BacT/Alert automated system [151, 152]. All blood culture positive bottles were sub-cultured onto MacConkey agar plates. For '**Study III'**, stool samples were cultured by streaking on MacConkey agar and *Salmonella-Shigella* (SS) agar plates. After overnight incubation at 37°C, suspected non-lactose fermenting colonies from agar plates were inoculated into Kligler's Iron Agar (KIA), Motile Indole Urea (MIU) and citrate tubes and kept for overnight at 37°C. Following growth on enriched medium, *S.* Typhi, *S.* Paratyphi A and *S.* Paratyphi B were identified based on standard biochemical reactions which were further confirmed by slide agglutination test with *Salmonella*-specific O and flagellar H antiserum **(Table 2.1)** and stored with 20% glycerol at -70°C for further use [35, 36, 148]. On the basis of blood and stool culture

confirmation result, all the available stored 202 *S.* Typhi strains for **'Study I'**, 67 *S.* Paratyphi A strains for **'Study II'**, 79 *S.* Paratyphi B strains for **'Study III'** were included. For **'Study IV',** total 62 strains (20 *S.* Typhi, 20 H58 *S.* Typhi and 22 *S.* Paratyphi A) were selected from the collection of study I and study II.

Table 2.1 Biochemical tests and serotyping for identification of *S.* Typhi and *S.* Paratyphi A, B

Biochemical test						Serotyping		Organism
KIA test		MIU test			Citrate test	Somatic "O" Ag	Flagellar "H" Ag	
Glucose	H_2S	Motile	Indole	Urea	Citrate			
Acid	+ ve	+ ve	- ve	- ve	- ve	O9, vi	H_d	*S.* Typhi
Acid + Gas	- ve	+ ve	- ve	- ve	- ve	O2	H_a	*S.* Paratyphi A
Acid + Gas	+ ve	+ ve	- ve	- ve	- ve	O4	H_b	*S.* Paratyphi B

2.3 Genomic DNA extraction

A loop of bacterial colony from MacConkey agar plate was inoculated into brain heart infusion (BHI) broth and was grown in shaker with vigorous shaking (250 rpm) for overnight at 37°C. Bacterial cells were pelleted by centrifuging the overnight bacterial culture. Genomic DNA extraction from the bacterial cells was performed by myself with the support of Mucosal Immunology and Vaccinology Team (MIVU) at icddr,b using commercially available the Wizard Genomic DNA Kit [153]. Briefly, cells were lysed by nuclei lysis solution, incubated for 5 minutes at 80°C and then cool to room temperature. RNase solution mix was then added to the cell lysate, incubated for 15-60 minutes at 37°C and then cool to room temperature. After RNase treatment, protein precipitation solution was mixed properly with the cell lysate by vortexing vigorously and proteins were then removed by centrifugation. Supernatant containing the DNA was transferred to new 1.5 mL microcentrifuge tube which was purified by washing with isopropanol and then with 70% ethanol. Purified DNA pellet was resuspended in rehydration solution and stored at 2-4°C for further use. DNA purity and concentration of genomic DNA was estimated by measuring absorbance of ultraviolet light at 260 nm and 280 nm wavelength using Nanodrop spectrophotometer.

The ratio of the absorbance at 260 and 280 nm ($A_{260/280}$) 1.8 is generally accepted as pure DNA [154].

2.4 Whole genome sequencing by Illumina platform

All the extracted genomic DNA from *S.* Typhi, *S.* Paratyphi A, *S.* Paratyphi B strains used in **'Studies I, II and IV'** were shipped to the Wellcome Sanger Institute (WSI). Core sequencing team at WSI was performed WGS using Illumina Hiseq 2500 platform (Illumina, San Diego, CA, USA) [155]. Briefly, DNA was fragmented and tagged for multiplexing with NexteraXT DNA Sample Preparation Kit. Index-tagged paired end Illumina sequencing libraries with an insert size of 500 bp were prepared for each sample to generate 150 bp paired end reads. Sequence quality was assessed on the basis of adapter contamination, average base read quality and any unusual GC bias using FastQC tool. Sequence data having median Phred score greater than 30 (99.9% accuracy) was considered for downstream genomic analysis.

Raw sequence data in fastq format for 202 *S.* Typhi, 67 *S.* Paratyphi A, 79 *S.* Paratyphi B generated in this thesis work was submitted to the European Nucleotide Archive (ENA) with accession numbers for each strain. Additionally, previously published sequence data of 88 Bangladeshi *S.* Typhi from Wong *et al*. 2016, and 528 from Tanmoy *et al.* 2018 were also included to provide detailed genomic context (raw sequence data are available in ENA under study accessions ERP001718 and PRJEB27394, respectively) in **'Study I'**. To provide global context, previously published 1,560 *S.* Typhi genomes in **'Study I'**; 242 *S.* Paratyphi A genome in **'Study II'**; 180 *S.* Paratyphi B genomes sharing the antigenic formula 1,4,[5],12:b:[1,2] and 12 *S.* Paratyphi B biotype *sensu stricto* genomes from patients presenting with enteric fever symptoms in **'Study III'** were included and subjected to further mapping, SNP calling and phylogenetic analysis [8, 9, 52, 53, 66, 101].

2.5 WGS data analysis

All of the genomic data analysis using different bioinformatics tools for this thesis work was performed by using access of Sanger computing cluster and done by myself with the support of pathogen informatics team and Nick Thomoson's team at WSI; Gordon Dougan and Steve Baker's team at University of Cambridge.

2.5.1 Mapping and SNP calling

Sequence reads produced by the Illumina sequencing platform were mapped against the indexed reference genome [88]. PCR duplicate reads were identified using Picard (http://broadinstitute.github.io/picard/) and flagged as duplicates in binary alignment map (BAM) file. In **'Study I'**, *S.* Typhi paired-end reads were mapped to the reference genome CT18 (accession no. AL513382) by RedDog mapping pipeline which used Bowtie alignment tool [156]. In **'Study II' and 'Study III'**, *S.* Paratyphi A and *S.* Paratyphi B sequence reads were mapped against the reference genome of *S.* Paratyphi A AKU_12607 (accession number FM200053) and *S.* Paratyphi B SPB7 (accession number CP000886) respectively using mapping tool SMALT [98, 157].

SNP calling was performed using samtools mpileup with parameters "-d 1000 -DSugBf" to identify SNPs having a phred quality score above 30 and removed low confidence alleles with consensus base quality ≤20 (represented with a gap character), read depth ≤5 (represent putative repeated sequences) or heterozygous base calls [88]. SNPs located in prophage regions, repetitive sequences or recombinant regions as detected by Gubbins [158] were also excluded. SNPs that passes these criteria were extracted from the multi-FASTA alignment using SNP-sites [159] and resulted 14852, 5419 and 132593 chromosomal SNPs for 2378 *S.* Typhi, (study I), 309 *S.* Paratyphi A (study II) and 271 *S.* Paratyphi B strains (study III), respectively. Pairwise SNP distances between strains (number of core genome SNP loci at which pairs of strain had discordant alleles) were calculated from each alignment using the dna.dist() function in the Analysis of Phylogenetics and Evolution (ape) R package (study I, II, III) [160].

2.5.2 SNP based genotyping analysis

In study I, short read alignment BAM files output from the RedDog mapping pipeline were used to assign genotypes for each read set according to an extended *S.* Typhi genotyping framework [101] using a custom Python script genotyphi.py (available at https://github.com/katholt/genotyphi). The resulting variant call format (VCF) file is then processed to identify the presence of cluster-, clade and/or subclade-defining SNP alleles that pass a minimum quality threshold (default consensus base phred score ≥ 20) and assigned the read set to a cluster, clade and subclade. To define novel

genotype 3.3.2 including 3.3.2.Bd1 and 3.3.2.Bd2 sublineage, unique SNPs which were absent in other *S.* Typhi genotypes were detected by extracting non-synonymous SNPs from SNP allele table [109].

2.5.3 Phylogenetic and population genetic analysis

Maximum likelihood (ML) phylogenetic trees were constructed from the aforementioned chromosomal SNP alignments using RAxML (study I, II, III) [161]. A generalized time-reversible model and a gamma distribution was used to model site-specific rate variation (GTR+ Γ substitution model; GTRGAMMA in RAxML) with 100 bootstrap pseudoreplicates to assess branch support for the ML phylogeny. SNP alleles from *S.* Paratyphi A AKU1_12601 in study I, *S.* Paratyphi A lineage G in study II and poultry originated *S.* Java PG10 in study III were included as outgroup to root the phylogenetic tree. In study III, hierarchical BAPS implemented in RhierBAPS R package was performed to re-define the *S.* Paratyphi B subpopulation structure [162]. A mid-point rooted phylogeny was constructed with all the *S.* Java strains belonging to BAPS cluster 1.1 (PGs 3, 4). All the resulting phylogenies were visualized and annotated using FigTree (available at http://tree.bio.ed.ac.uk/software/figtree), Microreact [163], R package ggtree [164], phandango [165] and iTOL [166].

2.5.4 *In silico* AMR gene, plasmid, virulence factor and *S.* Paratyphi B biotype identification

In '**Study I'**, the mapping based allele typer SRST2 [167] was used in conjunction with the ARG-Annot [168] and PlasmidFinder [169] database to detect acquired AMR genes and plasmids, respectively. Plasmid multi-locus sequence typing (pMLST) scheme was used to detect IncHI1 plasmid sequence type (PST) by using slurm_srst2.py python script and specifying the MLST definitions parameters [135]. ISMapper was performed was used to screen for both IS1 and IS26 insertion sites using CT18 reference genome to identify the location of these insertions into the chromosome of each Bangladesh *S.* Typhi strain [170]. Genomes with less common AMR profiles and without evidence of known resistance plasmid were assembled using Unicycler [171] and examined visually using the assembly graph viewer tool Bandage to inspect the composition and insertion sites of resistance-associated composite transposons [172]. QRDR mutations in *gyrA* and *parC* genes associated

with reduced susceptibility to fluoroquinolones were determined using GenoTyphi from RedDog output binary alignment map (BAM) files.

In *S.* Paratyphi A and *S.* Paratyphi B genomic studies ('**Study II, III'**), ARIBA (Antimicrobial Resistance Identification By Assembly) [173] in conjunction with the comprehensive antibiotic resistance database (CARD) [174] and PlasmidFinder database [169] were used to detect AMR genes and plasmid replicons, respectively. The same approach was used to detect virulence factors using the virulence factor database (VFDB) [175] among *S.* Paratyphi B genomes ('**Study III'**). In addition, AMR mutations in *S*. Paratyphi A genomes (study II), including point mutations in the QRDR of *gyrA, parC* genes associated with reduced fluoroquinolone susceptibility and mutation in *acrB* gene associated with azithromycin resistance were identified using the genoparatyphi python script (https://github.com/zadyson/genoparatyphi/).

In '**Study III'**, The Salmonella *In Silico* Typing Resource (SISTR) [176] implemented in PathogenWatch [177] was used to classify the *S.* Paratyphi B biotype *sensu stricto* and *S.* Paratyphi B biotype Java. Moreover, SRST2 was used to assign sequence types (STs) to each *S.* Paratyphi B genome according to the *S. enterica* MLST database [167].

2.5.5 Comparative pHCM2 plasmid analysis

To investigate the genetic diversity, evolution, and circulation of the pHCM2 plasmid in Bangladesh, pHCM2 plasmid phylogeny was constructed with 334 *S*. Typhi, 17 *S*. Paratyphi A, and single *S.* Paratyphi B variant Java genomes from collection of the study I, II and III respectively. Briefly, all the raw pHCM2 plasmid sequencing reads from these organisms were mapped against the reference *S.* Typhi CT18 pHCM2 plasmid (accession number: AL513384.1) using Bowtie2 [156]. Total 1788 core SNPs were identified by SAMTools [88] which were then used to construct a ML plasmid phylogeny using RAxML [161]. In addition, the comparison between annotated pHCM2 plasmid sequence of *S*. Paratyphi A and *S.* Paratyphi B biotype Java with the reference pHCM2 plasmid was performed by Basic Local Alignment Search Tool (BLAST) [178], Artemis Comparison Tool (ACT) [179] and visualized these comparisons using Easyfig [180].

2.5.6 *Denovo* genome assembly and annotation

In '**Study II, III'**, raw sequence reads were assembled *de novo* using Unicycler [171] and annotation was performed by prokka [181]. The quality of genome assemblies with the assembly metrics i.e. number of contigs (≥1,000bp), total assembly length (contigs ≥1,000bp) were assessed using genome assembly quality assessment tool QUAST [182]. Due to fragmentation of genome assemblies and low-quality of few global *S.* Paratyphi A sequence reads, genome assemblies that meet the threshold less than 100 contigs by using the formula (mean + 0.5 standard deviation) was considered for comparative *S.* Paratyphi A pan-genome analysis (study II).

2.5.7 Comparative pan-genome analysis

In **'Study II and III'**, roary was performed on annotated assembled genomes to identify the pan-genome by using a blastp percentage sequence identity of 95% and a core definition 95% of the included strains [183]. The Heaps function within the Micropan R package [184] was used to the curve fit constant according to Heaps law[185]: $n = k*N^{-\alpha}$, where n is pan-genome size, N is the number of genomes, k is the curve-specific constant. The exponential term, α determines whether the pan-genome of a bacterial variant is closed ($\alpha>1$) or open ($\alpha<1$) [124].

Comparative pan-genome *S.* Paratyphi B analysis was performed to identify biotype specific genes (*S.* Paratyphi B biotypes *sensu stricto* and Java) and PG specific (PG3/PG4 clades relative to rest of the PGs) defined as core genes present in more than 95% of the specific biotype/clade and absent from more than 95% of the strain of the other biotype/clade using custom R program and python scripts. To investigate the synteny in the loci containing the biotype- or clade-specific genes, a synteny graph was constructed in the regions of interest [186] and were visualized by Phandango [187] and Cytoscape [188]. In addition, BLAST [178], InterProScan [189] and EffectiveDB [190] were used to identify the distribution of group-specific genes throughout the species, to predict the function of the group-specific hypothetical proteins and to predict secreted proteins, respectively.

2.6 Statistical analysis

In '**Study I'**, base R function chisq.test [191] was used to calculate p value of age groups and *S.* Typhi genotypes using Chi-Squared test. Shannon diversity was calculated using the diversity function in the R package vegan [192]. In '**Study II and III'**, p value was calculated using fisher's exact test implemented in an integrated statistical software package STATA to determine statistical differences of lineage, ST distribution among the study sites [193].

2.7 Novel WGS approach based diagnostic assay

2.7.1 Bacterial strains and DNA extraction

In '**Study IV'**, a total 62 blood culture positive bacterial strains of enteric fever patients including 20 *S.* Typhi, 20 H58 *S.* Typhi and 22 *S.* Paratyphi A were used for PCR assay from the strain collection of study I, II (detailed sample collection and bacterial isolation specified above in method section 2.1 and 2.2). All of these 62 known blood culture positive bacterial strains were chosen to validate the accuracy of the PCR assay with the blood culture result. For DNA extraction, one loop of bacteria from MacConkey agar plate was suspended in 300 µL of nuclease free water (NFW). The bacterial suspension was heated at 95°C on water bath for 10 minutes and centrifuged at 14,000 rpm for 5 minutes. The supernatant was collected in a new 1.5 mL microcentrifuge tube which were used as DNA template for PCR reaction.

2.7.2 *S.* Typhi, H58 *S.* Typhi and *S.* Paratyphi A primer design

Whole genome comparison of *S.* Typhi CT18 and *S.* Paratyphi A AKU_12601 reference genome was performed to identify conserved unique regions to *S.* Typhi and *S.* Paratyphi A serovars by using BLASTn [94, 98, 178]. In **'Study IV'**, *S.* Typhi and *S.* Paratyphi A specific primers targeting conserved region STY0307 and SSPA2308, respectively were prioritized as these targets genes are not found in other *Salmonella* serovars [45, 194] . On the other hand, H58 *S.* Typhi specific primer was designed based on SNP based genotyping scheme (available at https://github.com/katholt/genotyphi) [101]. H58 *S.* Typhi specific SNP was defined by the synonymous mutation T349T in the STY2513 gene at nucleotide position 2348902 in *S.* Typhi CT18 encoding anaerobic glycerol-3-phosphate dehydrogenase A

(*glpA*) gene [48]. All these primers were designed and synthesized by Professor Ankur Mutreja's team at University of Cambridge using Primer-BLAST [195] and Integrated DNA Technologies (IDT, USA), respectively which then shipped to MIVU laboratory of icddr,b for setting up conventional single-plex, multi-plex PCR.

Table 2.2 List of primers targeting *S.* Typhi, *S.* Paratyphi A and H58 *S.* Typhi genes

Target gene	Primer name	Primer sequence (5'-3')	Amplicon length (bp)
STY0307	ST_227F	GGCAGATATACTTTCGCAGGCA	227
	ST_227R	CCCAGAACCAAATTTGCTTACA	
SSPA2308	SPA_305F	AGGGATGAGAATTTTCAGACGT	305
	SPA_305R	ACCCCAGCTCTGAGAGATATCT	
STY2513	H58_509F	GGGCTTGATGGCTTCATTAGT	509
	H58_509R	ACAGGTTGTACGCCTTTCCA	

2.7.3 Optimization of conventional single-plex, multi-plex PCR

Firstly, conventional single-plex PCR assay was carried out in a total reaction volume of 25 µL containing 2 µL template DNA, 12.5 µL 2x PCR master mix (Thermo scientific), 0.5 µL of each set of primers (mentioned above) and 9.5 µL of NFW for total 30 cycles as per suggested final version of standard operating procedure (SOP) generated by Ankur Mutreja's team at University of Cambridge. Finally, multi-plex PCR has planned to optimize with the same thermal condition followed for single-plex PCR except increasing the elongation time from 45 seconds to 60 seconds at 68°C.

Table 2.3 The thermal cycle of PCR reaction

Stage	Temperature	Time	Cycles
Preheating step	95 °C	120 seconds	1
Denaturation	95 °C	30 seconds	30
Annealing	60°C	45 seconds	30
Elongation	68°C	45 seconds (60 seconds for multi-plex PCR)	30
Final extension step	68°C	600 seconds	1
Final hold	4 °C	Indefinite	1

The amplified PCR products were analyzed by agarose gel electrophoresis on 2.0% (w/v) agarose gel containing 4 µL of fluorescent nucleic acid dye GelRed (Biotium). 4.5 µL of amplified PCR product was mixed with 3 µL loading dye to load into the gel and run the electrophoresis at 150 volt for 30 minutes. The band was visible on an automated gel documentation system (Gel Doc, Bio-Rad, USA) under ultraviolet light.

CHAPTER 3

RESULTS

3.1 Study I: *S.* Typhi genomics study

3.1.1 Genotypic diversity of *S.* Typhi in Bangladesh

The WGS of 202 *S.* Typhi strains isolated from three urban areas in Dhaka city between 2004-2016 revealed the presence nine distinct *S.* Typhi genotypes 2.0.0 (n=2, 0.99%), 2.0.1 (n=10, 4.95%), 2.1.7 (n=2, 0.99%), 2.3.3 (n=30, 14.85%), 3.0.1 (n=1, 0.50%), 3.1.2 (n=2, 0.99%), 3.2.2 (n=44, 21.80%), 3.3.0 (n=28, 13.86%) and 4.3.1 (n=83, 41.1%). Our phylogenetic data revealed that genotypes 2.3.3, 3.3.2, 3.2.2, 4.3.1.1 and 4.3.1.3 were present in all three study sites (**Figure 3.1A)** and were intermingled in the phylogeny (**Figure 3.1B)**. In addition, there was no significant difference in genotypes (p value 0.344 using Chi-squared test) among different age groups (young children under 5 years of age, older children from 5-15 years of age and adults above 15 years of age) studied in this collection.

The majority of the strains belonged to H58 (genotype 4.3.1) with H58 lineage I (genotype 4.3.1.1) being dominant (n=63, 31.2%) followed by H58 lineage II (genotype 4.3.1.2) (n=1, 0.50%). Moreover, the remaining 19 undifferentiated H58 strains were matched with H58 lineage Bd (genotype 4.3.1.Bd) reported in earlier study (Tanmoy *et al.* 2018) which was defined in this present study as genotype 4.3.1.3 by possessing characteristics SNPs STY0054-C711A (561056 nucleotide position in CT18) [108]. Additionally, a novel genotype 3.3.2 (previously typed as genotype 3.3.0) was identified on the basis of a SNP STY3641-A224G (3498544 nucleotide position in CT18). Further examination of this novel genotype revealed two Bangladeshi specific sublineages 3.3.2.Bd1 (SNPs STY2588-G378A, position 2424394 in CT18) and 3.3.2.Bd2 (STY2441-G439A, position 2272144 in CT18), carrying the *gyrA*-S83F and *gyrA*-D87N mutations, respectively.

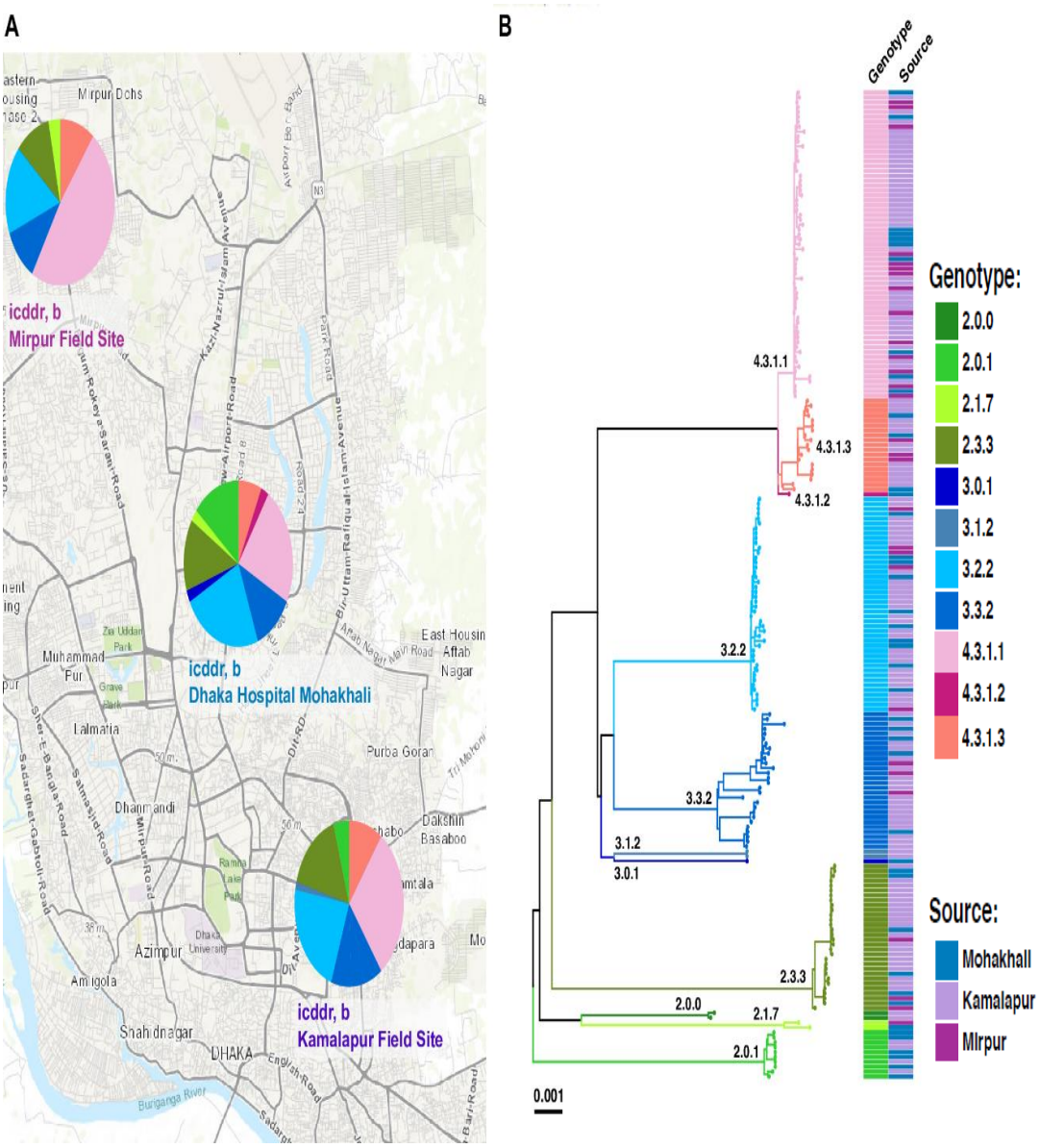

Figure 3.1 Spatial analysis of 202 Bangladeshi *S.* Typhi in Dhaka city with detailed distribution of genotypes (A)

Map of Dhaka indicating the prevalence of *S.* Typhi genotypes present at each of the three study sites of icddr,b. Map sourced from LandsatLook Viewer: https://landsatlook.usgs.gov/ and **(B)** Maximum likelihood tree of 202 urban Dhaka *S.* Typhi strains from the three study sites. Online interactive phylogeny available via Microreact https://microreact.org/project/sP2Uwk_DI. Branch colors indicate the genotype (as labelled) and the colored heatmap (on the right) shows for each strain, genotype, and source of the strain as per the inset legend.

3.1.2 Transmission pattern of global and regional *S.* Typhi strains

To better understand the global transmission patterns, a phylogeny including 1560 *S.* Typhi strains from previously published global WGS collection was constructed (interactive phylogeny available at https://microreact.org/project/5GzJ7Umoz) and showed that our Bangladeshi *S.* Typhi strains were closely related to strains from Nepal, India and Pakistan. Notably, a single introduction of South African *S.* Typhi strain among genotype 4.3.1.3 strain (approximately 2 SNPs differences) were also observed. In addition, multiple localized monophyletic lineages were observed within genotype 2.3.3, 3.2.2, 3.3.2, 4.3.1.1 and 4.3.1.3 indicating ongoing clonal transmission within Bangladesh (**Figure 3.2)**.

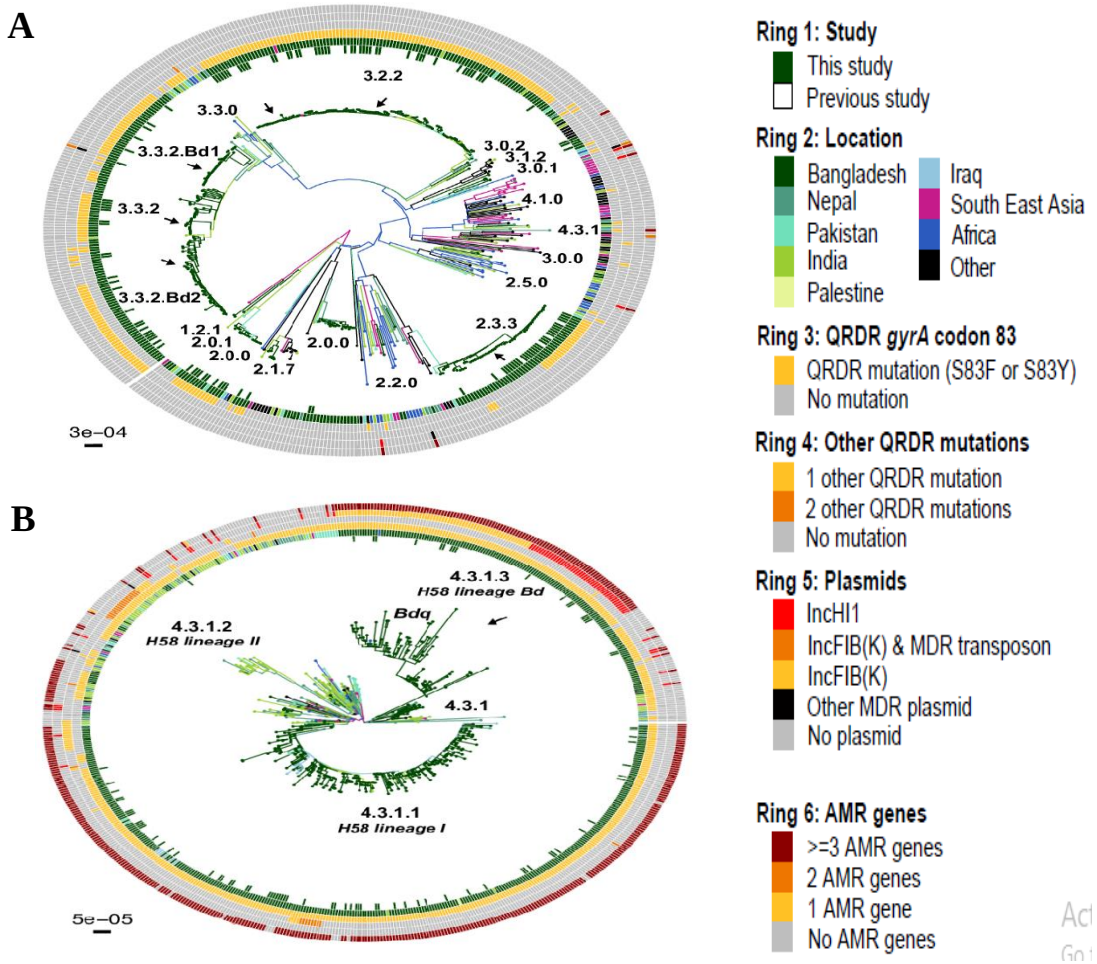

Figure 3.2 Global population structure of non-H58 and H58 Bangladeshi *S.* Typhi genotypes.
Maximum likelihood phylogenies of non-H58 (A) and H58 (B) global *S.* Typhi strains were outgroup rooted with *S.* Paratyphi A strain AKU_12601. All branches and rings are colored as per the inset legend. Arrows indicate localized lineages of Bangladeshi strains

3.1.3 AMR and plasmid profile among *S.* Typhi population

The acquisition of five AMR genes *bla$_{TEM-1}$, catA1, dfrA7, sul1, sul2* resistant to ampicillin, chloramphenicol, cotrimoxazole (trimethoprim and sulphamethoxazole collectively) respectively were detected in Bangladeshi *S.* Typhi strains (n=74, 36.63%) (**Table 3.1, Figure 3.2)**. Among the 202 *S.* Typhi analyzed, H58 strains (n=57, 28.22%) were strongly associated with MDR than non-H58 strain (n=0, 0.0%). Moreover, H58 lineage I strains acquired AMR genes without evidence of the IncHI1 plasmid indicating that these have been acquired via chromosomal translocation of MDR transposable element. Among the H58 lineage I population, there appear to be two variants of this MDR loci; a composite transposon encoding seven AMR genes (*catA1, dfrA7, sul1, blaTEM-1, strB, strA, sul2)* and a typical transposon encoding three AMR genes (*catA1, dfrA7, sul1).* GenoplotR analysis revealed that the integration of both transposons were mediated by IS*1* transposition with both inverted repeat and 8 bp target site duplication GGTTTAGA.

Moreover, the newly defined H58 genotype 4.3.1.3 strains revealed two different plasmid mediated AMR patterns clustering within two distinct clades. The first cluster (n=8) carried IncHI1 plasmids of plasmid sequence type 6 (IncHI1-PST6) encoding eight AMR genes *bla$_{TEM-1}$, catA1, dfrA7, sul1, sul2, strAB,* and *tetB* (resistance to ampicillin, chloramphenicol, trimethoprim, sulfonamides, streptomycin, tetracycline, respectively) and the second cluster (n=9) (termed H58 sublineage Bdq previously in Tanmoy *et al*.2018) carried a IncFIB(K) plasmid encoding the four AMR genes *bla$_{TEM-1}$, sul2, qnrS1, tetA* (resistance to ampicillin, sulfonamides, quinolone, respectively).

An alarming acquisition rate of QRDR mutations associated with reduced fluoroquinolone susceptibility observed in both H58 strains (n=83, 41.1%) and non-H58 strains (n=102, 50.5%). QRDR double mutant (*gyrA*-S83F and *parC*-E84K) was detected in only one H58 lineage I strain, but no QRDR triple mutant observed in our collection. Only 17 strains of total 202 *S.* Typhi collection were susceptible to antibiotics and did not carry any known genetic determinants of AMR. In addition, 14 strains belonging to genotype 3.2.2 carried IncX1 plasmid without any evidence of AMR genes and 77 strains from the most prevalent genotypes 4.3.1.1, 2.3.3, 3.2.2, 3.3.2.Bd1, 3.3.2.Bd2 carried cryptic plasmid pHCM2.

Table 3.1 List of genetic determinants of AMR in 202 *S.* Typhi strains

Resistance patterns	**H58 strain (n=83)**	**Non-H58 strain(n=119)**	**Total strain (n=202)**
Acquired AMR genes / QRDR mutations	**74 (89.16%)**	**0 (0.0%)**	**74 (36.63%)**
bla$_{TEM-1}$, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB* / *gyrA*-S83F	47 (56.63%)	0 (0.00%)	47 (23.27%)
bla$_{TEM-1}$, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB*, *tetB* / *gyrA*-S83Y	8 (9.64%)	0 (0.00%)	8 (3.96%)
bla$_{TEM-1}$, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB* / *gyrA*-S83F, *parC*-E84K	1 (1.20%)	0 (0.00%)	1 (0.50%)
bla$_{TEM-1}$, *catA1*, *dfrA7*, *sul1* / *gyrA*-S83F	1 (1.20%)	0 (0.00%)	1 (0.50%)
catA1, *dfrA7*, *sul1* / *gyrA*-S83F	8 (9.64%)	0 (0.00%)	8 (3.96%)
bla$_{TEM-1}$, *sul2*, *qnrS*, *tetA* / *gyrA*-S83Y	9 (10.84%)	0 (0.00%)	9 (4.46%)
QRDR mutations only	**83 (100%)**	**102 (85.71%)**	**185 (91.6%)**
gyrA-S83F	62 (74.70%)	85 (83.33%)	147 (72.77%)
gyrA-S83Y	19 (22.90%)	3 (2.94%)	22 (10.89%)
gyrA-D87G	1 (1.20%)	0 (0.00%)	1 (0.50%)
gyrA-D87N	0 (0.00%)	12 (11.76%)	12 (5.94%)
gyrA-D87Y	0 (0.00%)	2 (1.96%)	2 (0.99%)
gyrA-S83F, *parC*-E84K	1 (1.20%)	0 (0.00%)	1 (0.50%)
Susceptible to all antibiotics	**0 (0.00%)**	**17 (16.67%)**	**17 (8.42%)**

3.1.4 Correlation of genotypic distribution and AMR pattern over time period between 2004-2016

The MDR associated H58 genotype 4.3.1.1 and 4.3.1.3 appeared to be dominant in Bangladesh across the three sites sampled prior to 2011, but from 2011 onwards a relative decrease in H58 strains (median 66.67% per year in 2004-2010 versus 28.03% in 2011-2016) with increasing trends of non-H58 genotypes 2.3.3, 3.2.2, 3.3.2 (median 22.2% per year in 2004-2010 versus 65.2% in 2011-2016) **(Figure 3.3)**. An overall Shannon diversity increased from 1.34 in 2004-2010 to 1.83 in 2011-2016. The recent spread of these three new non-H58 genotypes indicates an alarming situation as it had started acquisition of QRDR mutation, leading to shift the treatment practice towards third generation cephalosporin.

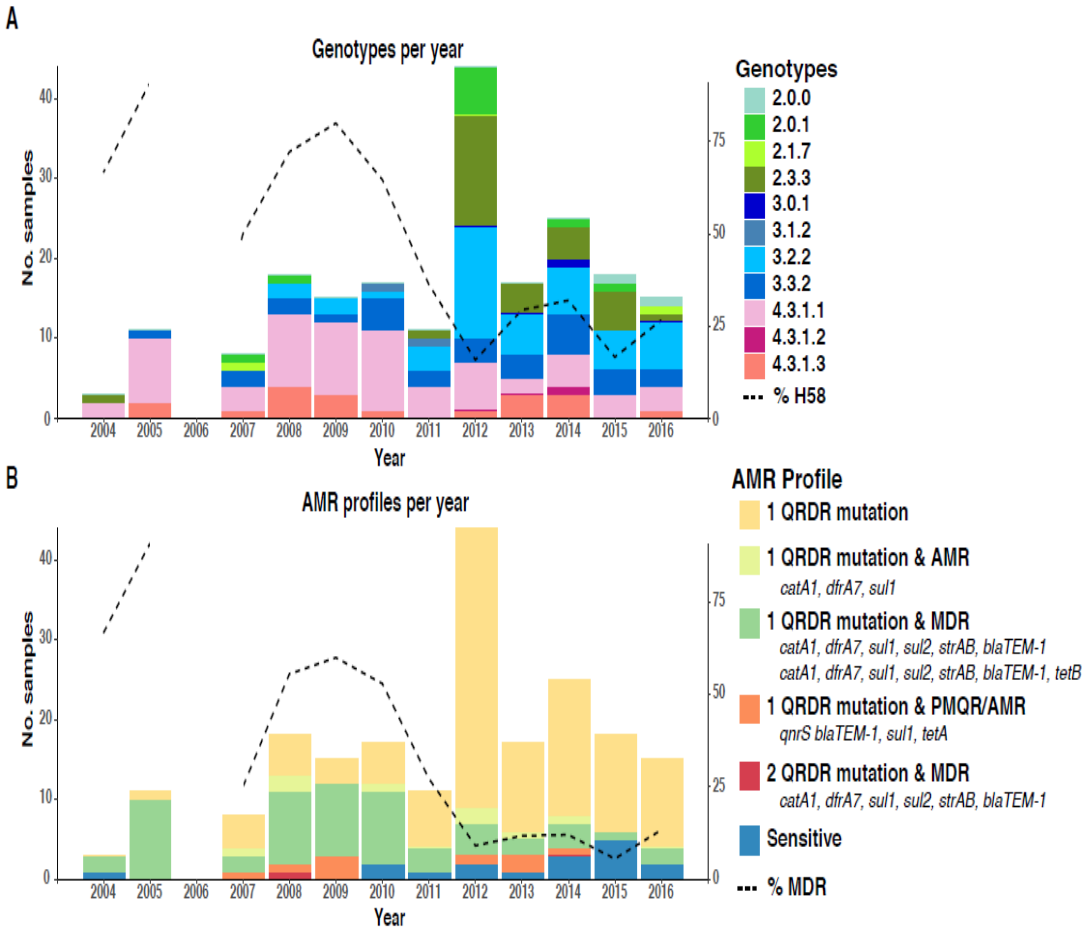

Figure 3.3 Timeline of urban Dhaka *S.* Typhi genotypes and AMR profiles from 2004-2016.

(A) Bar plots showing distribution of genotypes per year. Overlaid line indicates the percentage of H58 strain per year. (B) Bar plots showing AMR trends per year. Overlaid line indicates the percentage of MDR strain per year. Genotypes and AMR profiles are colored as per inset legend.

3.2 Study II: *S.* Paratyphi A genomics study

3.2.1 Population structure of *S.* Paratyphi A in Bangladesh

Phylogenetic analysis of 67 Bangladeshi *S.* Paratyphi A sequence data together with 242 previously sequenced global collection of *S.* Paratyphi A strains showed Bangladeshi *S.* Paratyphi A strains belonged to only lineages A (n=45, 67.2%), C (n=13, 19.4%), F (n=9, 13.4%) and were closely related to strains from Nepal, India, Pakistan, and Myanmar (**Figure 3.4)**. Notably, the majority of the Bangladeshi *S.* Paratyphi A strains (n=42, 62.6%) formed a monophyletic sublineage within lineage A which is here defined as sublineage A3 (median distance of ~70 SNPs away from the sublineage A1 found in Nepal). Moreover, distribution of lineages A, C, F among 67 strains collected from eight study sites covering four divisions within Bangladesh was also observed. There was no significant difference (p = 0.15 using Fisher's exact test) of the lineage distributions between study sites inside Dhaka (site 4, 5, 6, 7; *n*=52) and outside Dhaka (site 1, 2, 3; *n*= 15).

3.2.2 AMR genes and plasmid profile of *S.* Paratyphi A

All 67 Bangladeshi *S.* Paratyphi A strains were susceptible to traditional first line antibiotics due to absence of *catA, dfrA7, sul1, sul2, strAB,* bla_{TEM-1} genes but carried single QRDR mutation of *gyrA* gene at either codon 83 *gyrA*-S83F (n=57, 85.07%), *gyrA*-S83Y (n=6, 8.95%) or codon 87 *gyrA*-D87G (n=4, 5.97%) associated with reduced susceptibility to fluoroquinolones (**Figure 3.4)**. In addition, only one *S.* Paratyphi A isolated from Dhaka Medical College in 2018 (site 7) carried *acrB*-R717L mutation conferring azithromycin resistance together with *gyrA*-S83F QRDR mutation. Notably, non-AMR associated plasmids IncX1 plasmid (n=12,17.9%) and cryptic IncFIB-pHCM2 plasmid (n= 17,25.3%) were detected among this collection.

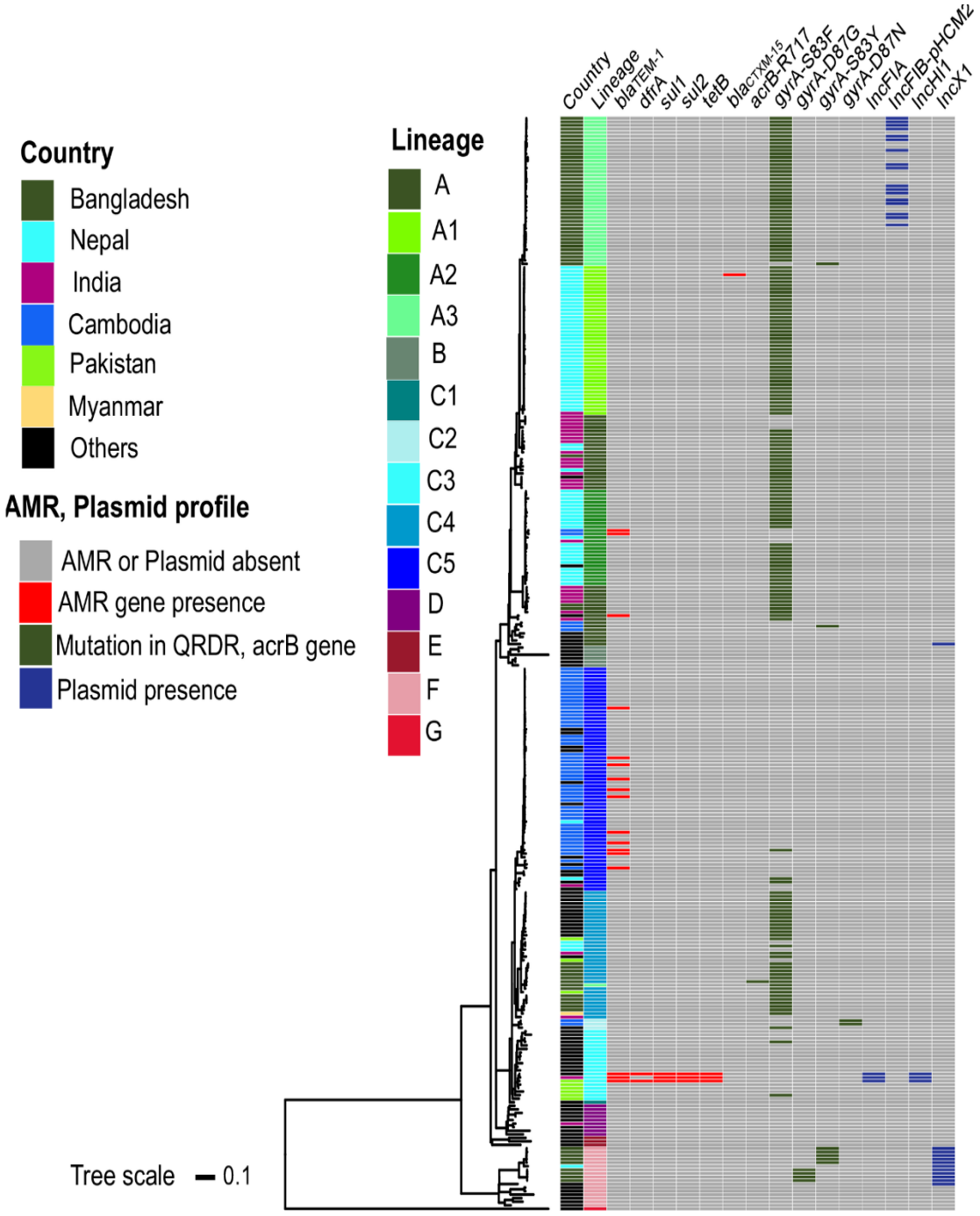

Figure 3.4 Global population structure of Bangladesh *S.* Paratyphi A with AMR genes and plasmid replicons.

Maximum-likelihood phylogenetic tree of 309 *S.* Paratyphi A genome sequences from the global collection including Bangladeshi strains. The coloured heatmap shows country, lineages, AMR genes, point mutations in the *gyrA* and *acrB* genes, plasmid replicons for each strain as per the inset legend.

3.2.3 Evolutionary history of pHCM2 plasmid in Bangladesh

The pHCM2 plasmid phylogeny including *S.* Typhi, *S.* Paratyphi A and *S.* Paratyphi B variant Java strains from Bangladesh (described details in methods) showed highly genetic similarity between typhoidal *S.* Typhi and *S.* Paratyphi A pHCM2 plasmid (99.98% nucleotide identity and average 19 SNPs difference) but different from non-typhoidal *S.* Paratyphi B biotype Java (98.8% nucleotide identity and 1697 SNPs away from *S.* Typhi). Moreover, pHCM2 plasmid length in *S.* Java was slightly smaller than that in *S.* Typhi and *S.* Paratyphi A (107,362 kb vs. 106,706 kb). Comparative pHCM2 plasmid analysis showed 12 non-identical regions including three insertions and one deletion in *S.* Java pHCM2 against reference *S.* Typhi CT18 **(Figure 3.5)**. In addition, comparative pan-genome analyses identified 113 hypothetical proteins and 14 annotated proteins which were present in only 17 pHCM2 plasmids harboring Bangladeshi *S.* Paratyphi A strains but absent in other global *S.* Paratyphi A strains. Furthermore, phylogenetic analysis showed that all the 17 pHCM2 carrying *S.* Paratyphi A strains were from same sublineage A3 and isolated in between 2011-2018 suggesting ancestral origins of pHCM2 in the *S.* Paratyphi A strain might have been derived from *S.* Typhi around 2011.

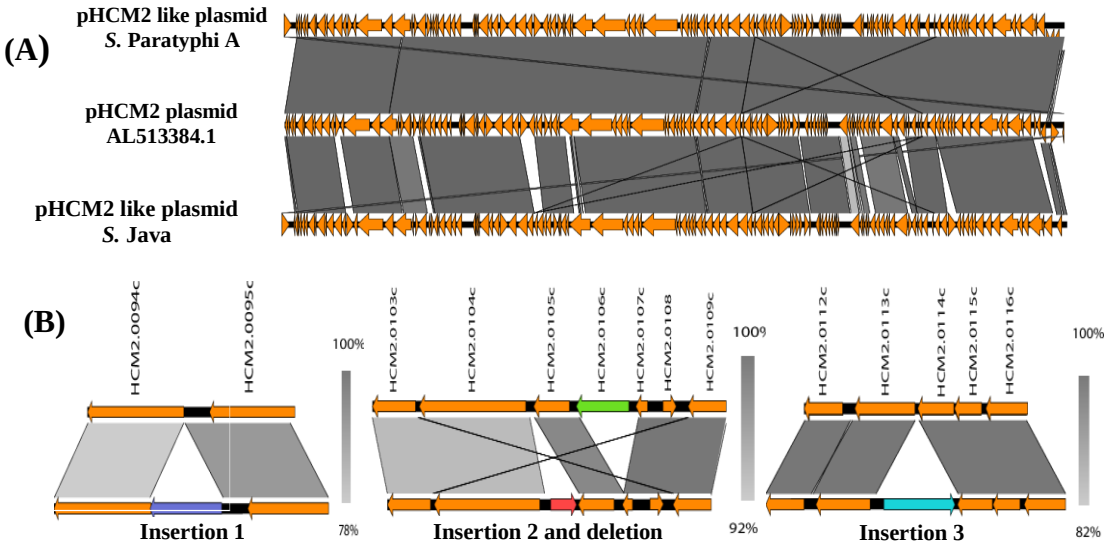

Figure 3.5 Comparative pHCM2 plasmid analysis.
(A) Pairwise comparison of pHCM2 plasmid reference genome (AL513383) with pHCM2-like plasmid of *S.* Paratyphi A and *S.* Java. including **(B)** three insertions and one deletion event against pHCM2 plasmid reference

3.2.4 Comparative pan-genome analysis

The pan-genome analysis with 295 quality-filtered *S.* Paratyphi A genome assemblies (cut-off determination explained detail in method) revealed a total of 9,093 genes comprised the pan-genome, with 4,125 core genes (present in ≥ 95% of all genomes) and 4,968 accessory genes (present in < 95% of all genomes). The gene accumulation curve for *S.* Paratyphi A was slowly flattened with the addition of new genomes and pan-genome curve parameter α value of 1.000015 indicated that *S.* Paratyphi A pan-genome was closed.

3.3 Study III: *S.* Paratyphi B genomics study

3.3.1 *In silico S.* Paratyphi B biotype identification

Among the 29,537 diarrhoeal patients enrolled from ten sentinel surveillance sites between June 2014 and June 2018, 107 patients were confirmed as *Salmonella enterica* serotype Paratyphi B by serotyping, with the antigenic formula O1,4,5,12:Hb:1,2. The proportion of *S.* Paratyphi B positive cases in each surveillance sites were comparatively low, ranging from 0.06% to 1.12%, compared to other enteric pathogens such as 1.01-18.3% for *Vibrio cholerae*. Among the total identified *S.* Paratyphi B positive cases, 79 available *S.* Paratyphi B strains were included for further genomic study. WGS data showed that all these 79 serotyped Bangladeshi *S.* Paratyphi B strains were further classified as biotype Java. This genomic data correlated with the clinical data, in which patients exhibited the non-invasive diarrhoeal symptoms (i.e. loose or rice watery diarrhea with a duration of 1-5 days, dehydration, vomiting, abdominal cramp) rather than invasive paratyphoid fever caused by *S.* Paratyphi B biotype *sensu stricto*.

3.3.2 Population structure *S.* Paratyphi B complex in Bangladesh

The global phylogenetic analysis including 192 contextual *S.* Paratyphi B complex genomes originating from over 20 countries and 79 Bangladeshi *S.* Java genomes showed that all the Bangladeshi *S.* Java strains clustered within previously-described *S.* Java lineages, either PG3 (n=2) or PG4 (n=77). Interestingly, the addition of 79 Bangladeshi *S.* Java strains to the global phylogeny (previously reported in Connor *et*

al) resulted in some intermingling of PG3 and PG4 clades into a single cluster at BAPS level 1 (defined as BAPS cluster 1.1) which was then resolved by Bayesian hierarchical clustering. BAPS cluster 1 was then further divided the into seven sub-clusters at BAPS level two (BAPS sub-clusters 2.1 to 2.7). In the updated phylogeny, PG3 clade grouped into sub-clusters 2.2, 2.3, 2.7 and PG4 clade grouped into 2.1, 2.4, 2.5, 2.6.

However, Bangladeshi *S.* Java strains contributed to the clonal expansion of PG4 (separated by 4,709 SNPs from PG3) residing into three distinct BAPS sub-clusters 2.4, 2.5, 2.6 which corresponded with STs 2113 (n = 19, 24.1%), 1577 (n = 7, 8.9%), 43 (n = 53, 67.1%), respectively (**Figure 3.5)**. STs 2113 and 1577 have only detected in the UK and Bangladesh and were strongly associated with patients having rice watery stool (RWS) (100% and 72% respectively) rather than loose watery stool (LWS). While globally distributed ST43 was associated with both RWS and LWS at similar frequencies (n=22 and n=28, respectively). The distribution of virulence factors (VFs) among the Bangladeshi *S.* Java population was also examined in context of BAPS sub-clusters. BAPS sub-cluster 2.6 acquired *tcfABCD* genes encoding *S.* Typhi colonization factor which are also present in other NTS serovars but lacked fimbrial *stfACDEFG* gene clusters.

Phylogeographical data showed no significant difference of ST distributions (p = 0.121, Fisher's exact test) between strains from inside Dhaka sites (site 3,4,5) and the rest of the strains from outside Dhaka sites across Bangladesh (site 1,2,6,7,8,9). Furthermore, there was no significant effect of age groups (children, adult) or gender (male, female) on the distribution of *S.* Java STs (p = 0.361 and p = 0.469, respectively using Fisher's exact test).

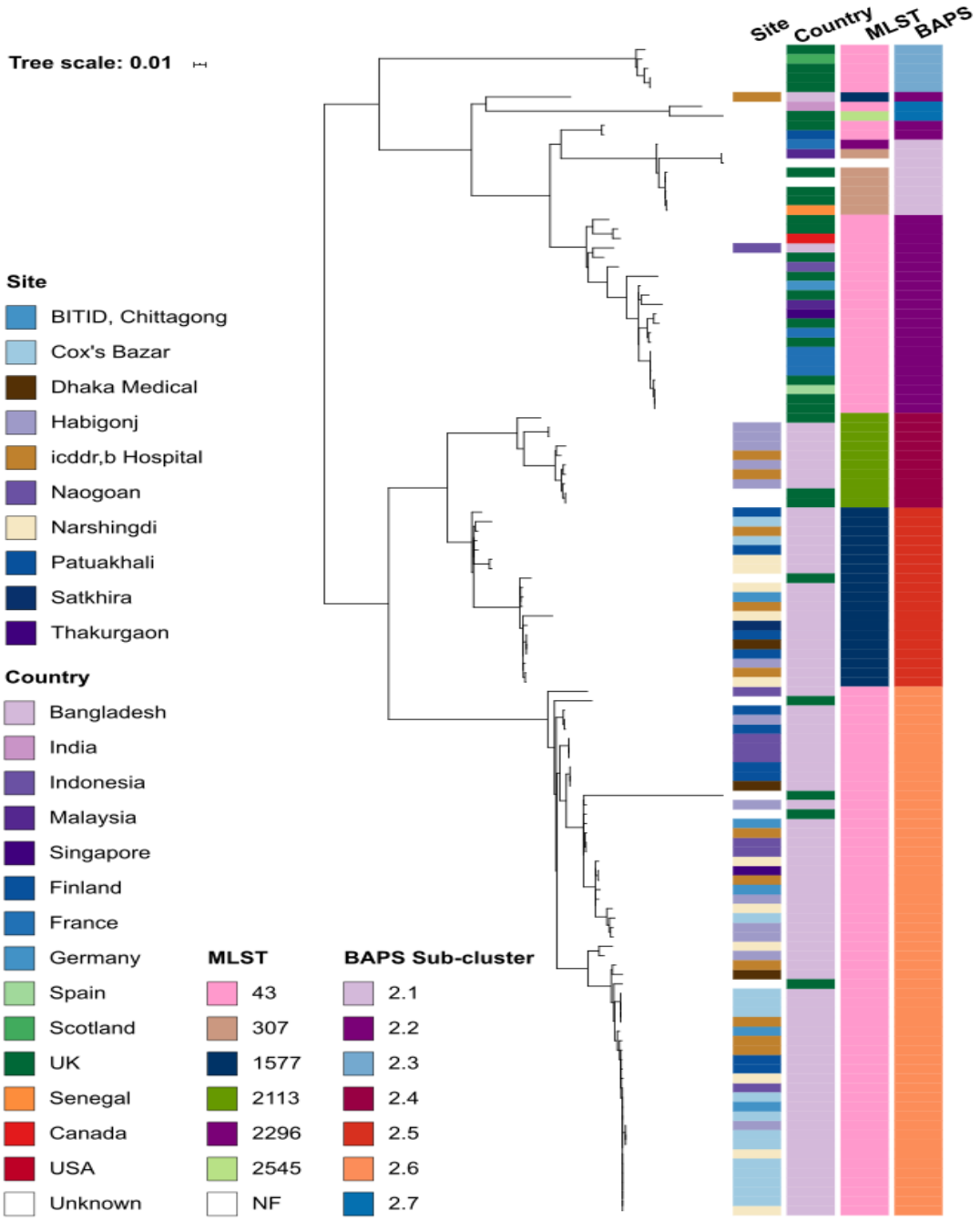

Figure 3.6 Mid-point rooted maximum-likelihood phylogeny of *S.* Java BAPS cluster 1.1.

Whole genome SNP tree 123 *S.* Java strains including Bangladeshi *S.* Java strains isolated from ten surveillance sites across Bangladesh. The coloured strips as onset legend alongside the tree show the surveillance site, country or region of isolation, MLST, and BAPS sub-cluster for each strain.

3.3.3 AMR genes and plasmid replicons of *S.* Paratyphi B complex

Majority of the Bangladeshi *S.* Java strains lacked known determinants of AMR genes where only three strains acquired *mphA, qnrB,* $bla_{\text{DHA-7}}$, and *sul1* AMR genes (confer macrolide, fluoroquinolone, beta-lactam, and sulfamethoxazole resistance respectively) and also carried both IncFIA and IncFIB plasmids. Additionally, two of these strains also carried *mphE* and *msrE,* encoding resistance to macrolide or erythromycin and streptogramin, respectively. Interestingly, a different set of AMR genes bla_{CARB}, *aadA, floR* , *sul1* genes, which confer resistance to beta-lactam, aminoglycoside, chloramphenicol and sulfonamide respectively was observed among global *S.* Paratyphi B strains (n=10, 3.7%). Furthermore, high frequency of AMR genes (*aadA24, dfrA, sul1, sul2, gyrA* genes conferring integron-encoded aminoglycoside, trimethoprim-sulphamethoxazole, fluoroquinolone resistance, respectively) were previously reported in Connor *et al* among poultry originated *S.* Java strains (PG10) which was absent in human originated *S.* Java strains (PG2-9).

3.3.4 Comparative pan-genome analysis

A total 271 *S.* Paratyphi B complex pan-genomes comprised 11,929 genes including 3,706 core genes (present in ≥ 95% of total genomes) and 8,223 accessory genes (present in < 95% of total genomes). Among them, 4,141 core and 789 accessory genes were for 46 *S.* Paratyphi B *sensu stricto* pan-genomes whereas 3,787 core and 7838 accessory genes were for 225 *S.* Java pan-genomes. The curve fitting parameter values of $\gamma = 0.40$ and $\gamma = 0.41$ for *S.* Paratyphi B complex and *S.* Java indicated open pan-genome but accumulation curve of *S.* Paratyphi B *sensu stricto* was closed ($\gamma = -0.05$) with less evidence of importation of new genes **(Figure 3.7)**.

Next, clade-wise comparisons were performed to identify *S.* Paratyphi B biotypes *sensu stricto* and Java specific genes. Comparative pan-genome analysis identified a set of 20 and 30 core genes specific to *S.* Paratyphi B *sensu stricto* and *S.* Java, respectively. *S.* Paratyphi B *sensu stricto*-specific genes were at same locus containing approximately 59 genes that encoded bacteriophage Mu F-like protein, SPI-2 secreted effector protein (*sopE*). *S.* Java-specific genes were clustered at three genetic loci within a 181 CDS-long chromosomal region. These loci carried genes encoded (1) hypothetical protein with a carboxymuconolactone decarboxylase; (2) ABC

transporters for amino acid transport; (3) numerous hydrolase and oxidoreductase enzymes for amino acids, carbohydrates and nitric oxide metabolism.

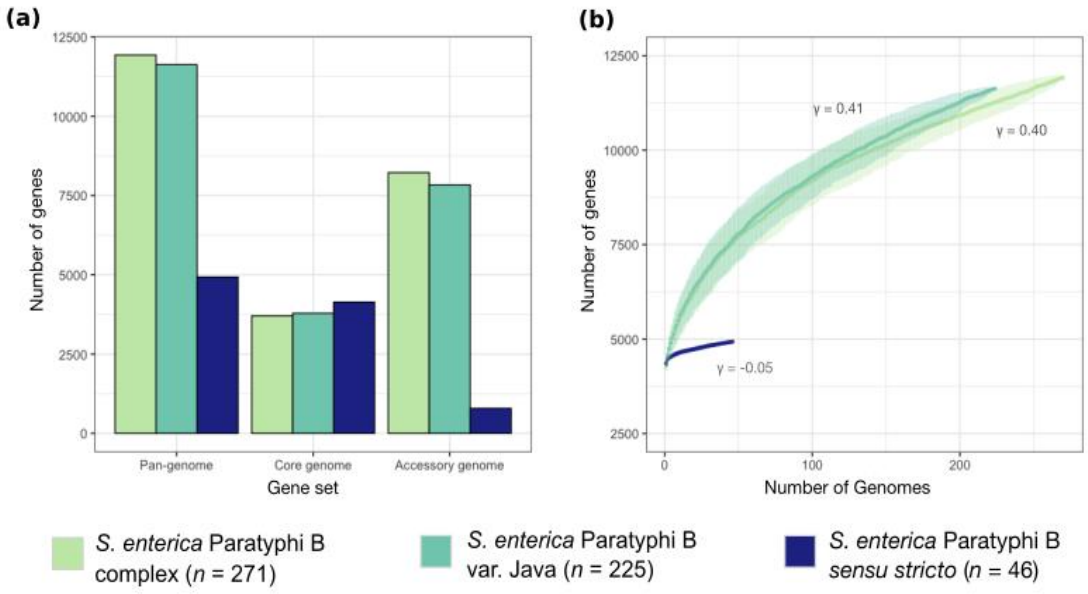

Figure 3.7 Pan-genome dynamics of the *S*. Paratyphi B complex.
(a) The pan, core and accessory genomes and (b) gene accumulation curves are depicted for the *S*. Paratyphi B complex and each biotype; see key for colour. Here, core genes are defined as genes present in ≥ 95% of total strains and accessory genes are present in < 95% of total strains. Error bars above and below the median are depicted by shading above and below the curve.

3.4 Study IV: Novel WGS approach based diagnostic assay

Of the total 62 blood culture positive bacterial strains isolated from blood samples of enteric fever patients, 22 were *S.* Paratyphi A and 40 were *S.* Typhi. Among these *S.* Typhi strains, 20 were belonged to H58 *S.* Typhi lineage based on WGS result of study I. A single-plex conventional PCR assay was initially optimized with these three sets of bacterial strains taken from study I, II (set A: 22 *S.* Paratyphi A; set B: 20 *S.* Typhi and set C: 20 H58 *S.* Typhi) to validate the accuracy of three target genes SSPA2308, STY2513, STY0307 as potential diagnostic markers for *S.* Paratyphi A, *S.* Typhi and H58 *S.* Typhi detection, respectively.

The optimized single-plex PCR assays correctly amplified SSPA2308, STY2513, STY0307 target genes for all *S.* Paratyphi A (22/22, 100%), *S.* Typhi (20/20,100%), H58 *S.* Typhi (20/20, 100%) positive DNA samples, respectively (**Figure 3.8).** In case of *S.* Paratyphi A positive DNA, it did not show false positive band for *S.* Typhi, H58

S. Typhi specific primer sets (lane 23, 24 in **Figure 3.8 A)**. Similarly, in case of *S.* Typhi positive DNA, it did not show false positive band for H58 *S.* Typhi, *S.* Paratyphi A specific primer sets (lane 21,22 in **Figure 3.8 B)**. As H58 Typhi is a defined lineage of *S.* Typhi, H58 Typhi positive DNA also amplified for STY2513 target gene of Typhi (lane 21) but not amplified for *S.* Paratyphi A target gene (lane 22) (**Figure 3.8 C)**

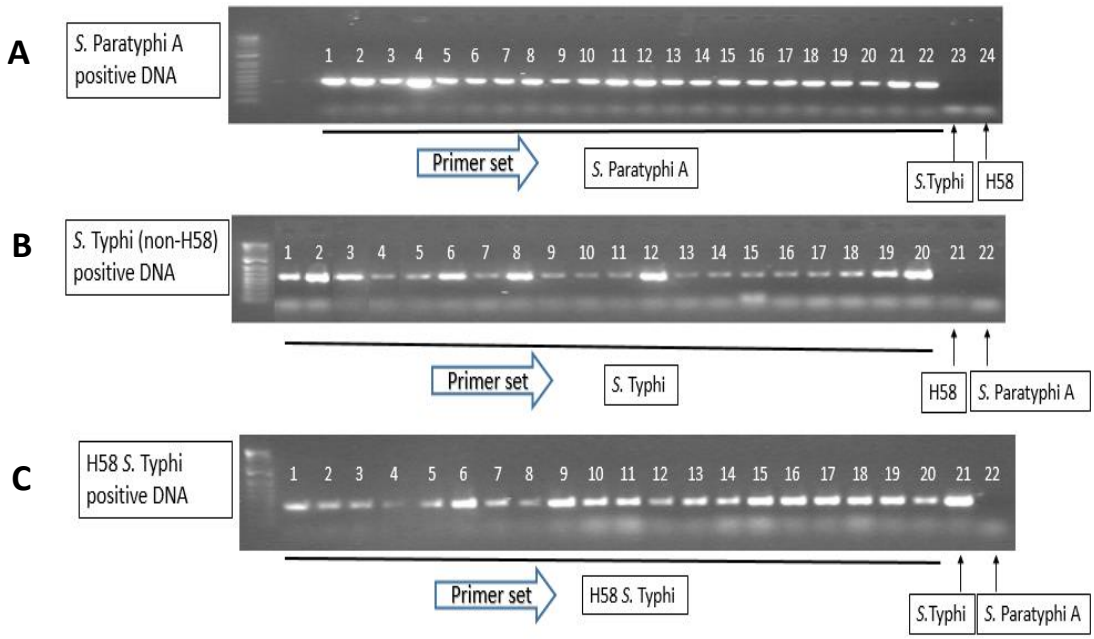

Figure 3.8 Single-plex PCR for the diagnosis of typhoidal pathogens.
Agarose gel electrophoresis result showed positive, negative bands for respective target genes separately (A) SSPA2308 gene: Lane 1-22: *S.* Paratyphi A positive, lane 23: *S.* Typhi negative, lane 24: H58 *S.* Typhi; (B) STY2513 gene: Lane 1-20: *S.* Typhi positive, lane 21: H58 *S.* Typhi negative, lane 22: *S.* Paratyphi A; (C) STY0307 gene: Lane 1-20: H58 *S.* Typhi positive, lane 21: H58 *S.* Typhi positive, lane 22: *S.* Paratyphi A negative.

Once single-plex PCR optimization has been done, in next step this research moved to set up multi-plex PCR with single sample from each set. But unfortunately, the PCR assay was not worked out with the same thermal condition of single-plex PCR and did not get positive bands for respective target genes (data not shown). Further optimization was made by slightly changing in the elongation step (mentioned detailed above method section 2.7.3) which showed single positive band for *S.* Typhi, *S.* Paratyphi A positive strain (band position 1-2 in **figure 3.9**) and double positive band for H58 lineage of *S.* Typhi (band position 3 in **figure 3.9**).

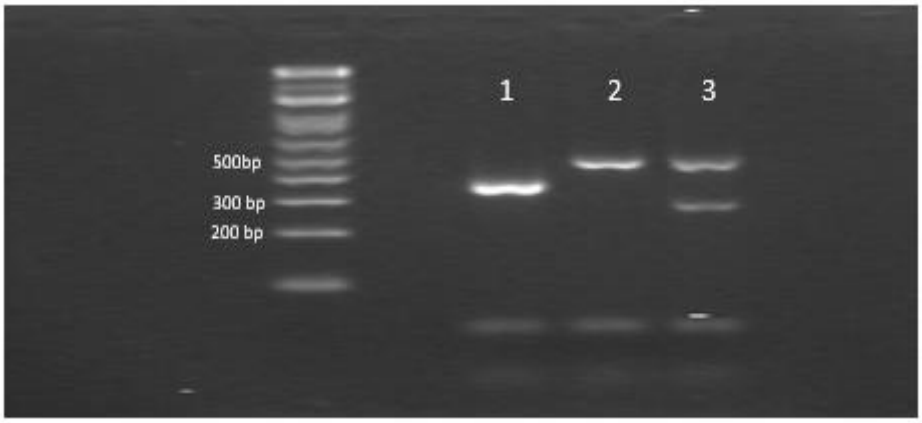

Figure 3.9 Multi-plex PCR for the diagnosis of typhoidal pathogens.
Agarose gel electrophoresis showed result for all three SSPA2308, STY2513, STY0307 target genes in single reaction. Lane 1: *S.* Paratyphi A positive band for *S.* Paratyphi A positive sample; Lane 2: *S.* Typhi positive band for *S.* Typhi positive sample and Lane 3: H58 *S.* Typhi positive bands (one band for *S.* Typhi, one for H58 *S.* Typhi) for H58 *S.* Typhi sample. 1000 bp molecular weight marker was positioned in the left side of the gel band (position for 200, 300, 500 bp were highlighted). The PCR amplicon sizes for SSPA2308, STY2513, STY0307 was 305 bp, 509 bp, 227 bp respectively.

CHAPTER 4

DISCUSSION

4.1 General discussion

Although enteric fever is one of the major cause of morbidity and mortality in LMICs, the characterization of *S.* Typhi and *S.* Paratyphi has not examined in greater detail at molecular level. Also, there is still lack of systematic surveillance data representing the actual disease burden and epidemiology of enteric fever in low resource settings like Bangladesh. While the introduction of MLST approach has led to rapid progress in population genetic studies for many bacterial pathogens [105], the analysis of monomorphic pathogen such as *S.* Typhi, *S.* Paratyphi A populations have lagged behind which necessitates the advantage of WGS data to provide an in-depth comprehensive overview of the individual sequence variation. In this dissertation, high resolution SNP based phylogenetic analysis was performed using a collection of Bangladeshi *S.* Typhi and *S.* Paratyphi A, B strains which proven to be the most reliable typing method and provide an improved discriminatory power than the previously used typing scheme of these *Salmonella* serovars. In order to better understand both global and regional strain circulation pattern, *S.* Typhi and *S.* Paratyphi A, B sequence data generated in this dissertation was combined with previously published global sequence data.

'Study I' demonstrated detail phylogenetic relationship of whole genome sequenced 202 Bangladeshi *S.* Typhi strains, representative of the past 13 years, harboring the diverse population structure of Bangladesh with nine distinct genotypes. This contrasts with the genomic data from '**Study II'** where less diverse population structure was observed among Bangladeshi *S.* Paratyphi A population. Previously, a global population structure study classified *S.* Paratyphi A into seven lineages with lineage A and C being the most dominant globally [112] which in lines with the phylogenetic analysis of '**Study II'** showing Bangladeshi *S.* Paratyphi A strains clustered within these two dominant global lineages A, C along with lineage F. Similarly, the population structure of Bangladeshi *S.* Java in '**Study III'** was found to

be less diverse, harbouring only three STs and clustering within either PG3 or PG4 in global phylogeny.

Furthermore, '**Study III'** demonstrated the high-resolution view of *S.* Paratyphi B complex, as this is the first WGS study in Bangladesh to distinguish between the *S.* Paratyphi B biotypes Java and *sensu stricto*. *S.* Paratyphi B complex has long been source of confusion for microbiologists, as these two biotypes have until now been indistinguishable by the traditional serotyping approach [8, 116]. The addition of the Bangladeshi *Salmonella* strains serotyped as Paratyphi B to the previously reported *S.* Paratyphi B complex global phylogeny resulted phylogenetic clustering with the non-typhoidal *S.* Java clades PG 3 and PG4 which were divergent from the typhoidal *S.* Paratyphi B biotype *sensu stricto* lineage PG1 in the context of disease phenotype. Furthermore, coupling of the WGS approach with the existing nationwide enteric disease surveillance study across Bangladesh reported low prevalence of gastroenteritis causing biotype *S.* Java collected from the stool sample of diarrhoeal patients compared to other enteric pathogens. In general, *S.* Paratyphi B *sensu stricto* infection is relatively uncommon in contrast to Chile where the substantial number of paratyphoid B fever caused by this biotype were reported in the 1980s [9, 21, 196]. This report correlates with this study suggesting the absence of invasive paratyphoid B fever with dominance of non-invasive gastroenteritis caused by *S.* Java in Bangladesh.

The phylogeographical data from *S.* Typhi and *S.* Paratyphi A, B genomic studies in this thesis revealed random distribution throughout all the study sites with little evidence of any regional geographic restriction across Bangladesh. Furthermore, there did not appear to be any stratification by patient age group or sex suggesting all patient groups were affected by the same range of lineages in these study settings, similar to previous report on *S.* Typhi from Nepal [53]. Furthermore, global contextualization of Bangladeshi *S.* Typhi and *S.* Paratyphi A strains revealed intermingling with strains from neighboring India, Pakistan and Nepal which suggests circulation of Bangladeshi strains throughout South Asia. Whereas the *S.* Paratyphi B genomic data showed that all except two of the Bangladeshi *S.* Java strains clustered with the strains from UK suggesting strong evidence of intercontinental long-range transmission that might be because of large diaspora resident in the UK or the frequent travel back to Bangladesh. Random global distribution, biasness of Bangladeshi *S.*

Java strain towards only PG4 clade and rareness of the presence of other STs, PGs in Bangladeshi population will provide information to track ongoing local or global transmission for future surveillance efforts.

However, the formation and clonal transmission of multiple localised monophyletic *S.* Typhi lineages within Bangladesh indicate ongoing local evolution, which perhaps explain the dominance of novel Bangladesh-associated genotypes 4.3.1.3 and 3.3.2 including 3.3.2.Bd1, 3.3.2.Bd2 detected in '**Study I'**. The updated global *S.* Typhi population structure enabled to redefine these genotypes based on unique SNPs that have been added to the current GenoTyphi scheme (https://github.com/katholt/genotyphi) [107]. The WGS based SNP genotyping framework GenoTyphi was first developed in 2016 to provide robust and phylogenetically informative nomenclature for *S.* Typhi genotypes [101]. Later, subsequent genomic surveillance studies reported epidemiologically important subpopulations with the availability of new sequence data which will facilitate the identification and long-term monitoring of regional, global high risk clones [52, 53, 55, 57, 64, 66, 110, 138, 193].

In 2015, a landmark study addressing phylogeography of H58 haplotype reported that South Asia was an early hub for H58 from which it rapidly disseminated to many countries across Asia and Africa over the last 30 years [52]. In '**Study I'**, H58 lineage I appeared to be dominant compared to H58 lineage II which was consistent with previous studies in which H58 lineage II circulation was commonly found in neighboring Nepal and India [53, 61, 197]. On the contrary, the newly defined H58 genotype 4.3.1.3 has been exclusively seen in Bangladesh except single introduction of South African strain suggesting scant evidence of expansion of this lineage outside Bangladesh and only sporadic evidence of intercontinental transmission events [109]. However, Bangladeshi *S.* Paratyphi A strains contributed to the clonal expansion of lineage A which has been imported from neighboring countries on multiple occasions whereas importation of lineages C, F occurred on single transmission event. This will provide information to track ongoing local or global transmission for future surveillance efforts.

'Study III' illustrates the population structure of Bangladeshi *S.* Java in finer detail by redefining the original definition of the existing *S.* Paratyphi B complex global phylogeny [198]. Bayesian Hierarchical clustering resolved the intermingling of PGs 3, 4 by updating the phylogeny upto two BAPS levels. The possible explanations for this discrepancy might be the differences in the aims of these two studies. Previously, the phylogenetic tree reported in Connor *et al*. was based on core SNPs alignment by placing *S.* Paratyphi B strains into the wider genomic context of other *Salmonella* serovars to define distinct lineages for *S.* Paratyphi B biotypes *sensu stricto* and Java [8]. Whereas the phylogeny in '**Study III'** was generated from SNPs alignment relative to the reference *S.* Paratyphi B SPB7 genome and included only *S.* Paratyphi B complex global strain together with Bangladeshi *S.* Paratyphi B sequenced genomes to further classify which *S.* Paratyphi B biotype they belonged to [198].

The global *S.* Typhi genomic framework previously showed strong association of MDR typhoid infection with the highly clonal H58 lineages mediated via IncHI1 plasmid [51-53, 99, 123, 136]. Surprisingly, acquisition of MDR genes were found among 4.3.1.1 Bangladeshi strains in absence of IncHI1 plasmid. These MDR island was inserted into *S.* Typhi chromosome, which has been observed previously in different countries in Asia [52, 53] and also in Africa [56]. Such chromosomal integration within Tn2670 like composite transposon is of great concern, as it facilitates more stable transmission of MDR phenotype compared to plasmids driven by the horizontal gene transfer. This *S.* Typhi genomic study reported for the first time two major patterns of plasmid mediated resistant mechanism among 4.3.1.3 strains which might be considered as a cause of initial clonal expansion of this recent H58 lineage.

Despite sporadic case report of IncHI1 like pAKU_1 MDR plasmid in *S.* Paratyphi A suggests a very low prevalence of MDR but an increasing trend of fluoroquinolone resistance due to chromosomal QRDR mutations of *gyr*, *par* genes [99]. However, genomic data from '**Study I, II'** showed a high frequency of QRDR mutation associated with reduced fluoroquinolone susceptibility among both typhoidal *S.* Typhi and *S.* Paratyphi A which might be the result of an increase in over-the-counter sale of this drug over the last decade for treatment [109, 199]. Despite increasing reports of antibiotic usage without prescription in Bangladesh, there has been a stark lack of

evidence for widespread extrinsic resistance gene and plasmid acquisition amongst Bangladeshi *S.* Java strains [198]. While the high frequency of AMR genes was reported in global poultry-associated *S.* Java strains [8, 25], three Bangladeshi *S.* Java strains carried fluoroquinolone resistance genes. This is matter of concern because fluoroquinolone-resistant *Salmonellae* are on the WHO priority list of bacteria. This result highlights the need of continued genomic surveillance to monitor the ongoing AMR evolution [200].

From 2012 onwards, the alarming increase of QRDR single mutant non-H58 *S.* Typhi strains coincided with the reduced prevalence of H58 MDR *S.* Typhi across all the three sites inside urban Dhaka [109]. Despite high prevalence of single or double chromosomal mediated QRDR mutations found among both H58 and non-H58 strains, a limited evidence of plasmid mediated quinolone resistance (PMQR) via IncFIB(K) plasmid carrying *qnrS1* resistance gene in absence of triple QRDR mutant was detected in only 4.3.1.3 strains. Multiple subsequent studies from Nepal and India have reported an emerging triple QRDR mutant H58 subclade associated with full ciprofloxacin resistance[53, 110, 138] . Whereas a new sublineage of H58 lineage Bd named 'Bdq' carrying *qnrS1* in addition to multiple QRDR mutations was detected in Bangladesh. This H58 subclade possessing high minimum inhibitory concentration (MIC>4.0 μg/ml) against fluoroquinolone is responsible for treatment failure which suggests ceftriaxone and azithromycin can be used as alternative treatment option for enteric fever [108].

Fortunately, ceftriaxone and azithromycin resistant *S.* Typhi strain did not detect in '**Study I',** while acquisition of azithromycin resistance *acrB-R717L* gene in one *S.* Paratyphi A isolated from Dhaka in 2018 is of great concern [109]. Furthermore, azithromycin resistant *S.* Typhi strains acquiring *acrB-R717Q* from genotype 4.3.1.1 was previously reported in Bangladesh [64, 108] which may possibly warrant for reconsideration of current empirical antimicrobial use for treatment. Fortunately, the major outbreak of ceftriaxone resistance that was first observed in Pakistan in 2016 among genotype 4.3.1.1 strain have not detected in Bangladesh [66]. Though single ceftriaxone resistant *S.* Typhi strain acquiring *blaCTX-M* gene from genotype 3.3.0 in 2000 was detected previously in Bangladesh, this outbreak has not spread to Bangladesh during the time of this study [108]. All these concluded consideration of

ceftriaxone as first line of emperical treatment at present. Antimicrobial stewardship has become more challenging with changing of antibiotic treatment practice and altering the selective pressure profile.

'Study II' conducted comparative analysis of cryptic pHCM2 plasmid among Bangladeshi *S.* Typhi, *S.* Paratyphi A and non-typhoidal *S.* Java populations [109, 198, 199]. This plasmid was reported for the first time in Bangladeshi *S.* Paratyphi A population in this study which was absent in the global *S.* Paratyphi A collection. The lack of pHCM2 in global *S.* Paratyphi A strains emphasizes the point of interest to better understand the source of origin and genetic difference from other *Salmonella* serovars previously detected in Bangladesh. The pHCM2 cryptic plasmid was originally isolated from a typhoid patient in Vietnam in 1992 [201]. The pHCM2 phylogenetic analysis showed high degree of genetic similarity between typhoidal *S.* Typhi and *S.* Paratyphi A which was different from the nontyphoidal *S.* Paratyphi B biotype Java at the level of single nucleotide changes and plasmid length suggesting the evolution of this plasmid in *S.* Paratyphi A has occurred from *S.* Typhi. Further studies are needed to monitor the trend over time period and to understand the effect of this cryptic plasmid on enteric infections.

In this dissertation, genomic variation and evolution between typhoidal and non-typhoidal *Salmonella* strains were further investigated by the pan-genome analysis [198, 199]. Pan-genomic composition of core and accessory genes among Bangladeshi *S.* Paratyphi A and *S.* Java was studied to better understand the population genetic structure which provide cues about the mechanisms underlying host adaptation and evolution. Previous studies based on whole genome comparative analyses carried out at the population level showed highly conservative genomic structure among typhoidal *Salmonella* serovars because of consequences of active pseudogenisation and human host restricted nature [97, 98, 124, 132]. This is in line with previous reports in which the lack of imported new genes with declining trend of core gene indicated close *S.* Paratyphi A pan-genome. The similar pattern was also observed among global typhoidal *S.* Paratyphi B biotype *sensu stricto* strains whereas pan-genome analysis of non-typhoidal *S.* Java strains revealed open pan-genome due to presence of higher numbers of accessory genes [198]. This could be a possible reason behind the broad range of host adaptation from human to animal of *S.* Java [97]. In

addition, comparative pan-genome analysis revealed distinguishing gene sets for both biotypes of *S.* Paratyphi B complex and also for PGs 3,4 clades belonging Bangladeshi *S.* Java strains which might be used as suitable markers for future detection of these clades.

'Study IV' was designed to evaluate the performance of rapid, reliable and, cost effective WGS based PCR assay as diagnostic approach in poor resource settings like Bangladesh that can be replaced conventional serotyping of blood or stool culture positive bacteral isolates for pathogen detection. Firstly, a single-plex PCR assay was optimized with blood culture positive *S.* Typhi, *S.* Paratyphi A and H58 *S.* Typhi strain collection and showed 100% sensitivity of this assay by comparing with the previously recorded blood culture result. Secondly, multi-plex PCR optimization has been done with single set of samples which will evalute with bulk number of sample set. Further studies are needed to continue this novel multi-plex PCR assay for regular diagnostic or surveillance purpose in future.

Many earlier studies on nucleic acid based amplification method were selected *S.* Typhi PCR targets based on antigenic properties such as O antigen somatic genes (*tyv, prt*) [147], H antigen flagellar gene (*fliC-d)* [202], Vi capsular antigen gene (*viaB)* [203] which were also found in other *Salmonella* serovars. Recent advances in WGS of *S.* Typhi CT18, *S.* Paratyphi A AKU_12607 reference genome have contributed to identify unique conserved regions specific to these serovars to overcome limitations of poor sensitivity and specificity of previously used target genes in earlier studies [45, 204]. This study tested only typhoidal pathogens with the SSPA2308, STY2513, STY0307 target genes specific to *S.* Paratyphi A, *S.* Typhi, H58 *S.* Typhi for their detection. Though the sensitivity of these target genes was calculated with a panel of sample set but specificity was checked with only single sample set which is not enough to calculate the specificity. In future, other febrile illness patients and healthy individuals will also be included to evaluate the better performance of this technique. The target genes used in this study not only have the potential to address cost effective point of care (POC) diagnosis to distinguish *S.* Typhi from *S.* Paratyphi A detection at the primary health care level in low resource settings but also important to detect globally widespread disseminating *S.* Typhi H58 lineage. The genome based diagnostic approach has added a new dimension to design

unique diagnostic marker for MDR associated H58 lineage detection which has the potential to inform local treatment algorithms based on regional antibiotic susceptibilities. In conclusion, this genome based PCR assay can be considered as rapid, sensitive, reliable, accurate, cost effective for the diagnosis of typhoidal pathogens directly from blood samples.

Each study described in this dissertation has some limitations. The *S.* Typhi, *S.* Paratyphi A and *S.* Paratyphi B biotype Java genomes sequenced in study I, II, III was relatively low in number. However, *S.* Typhi, *S.* Paratyphi A strains analysed in study I, II were from three study sites inside Dhaka and only four divisions across Bangladesh, respectively that may not be representative of the overall the actual incidence, population structure of enteric fever in Bangladesh. Although the overall dataset covering all divisions of Bangladesh in study III, still this study could not observe any significant MLST distribution across ten study sites due to low prevalence of *S.* Java positive cases in each surveillance sites. However, these genomic data improve the knowledge regarding genomic variation, evolutionary changes for planning future treatment or control strategies.

4.2 Future work and conclusion

The results of the all of the four studies carried out in this dissertation highlight the importance of WGS based surveillance studies in Bangladesh and other resource limited settings where the disease burden is high and many different AMR phenotypes have been observed. Continued genomic surveillance might help us to reveal novel resistance mechanism and also provide detailed information to design better, rapid, cost effective diagnostic approaches.

Future studies on evaluating the performance of the SNP based diagnostic assay and implementing in regional hospital settings for routine diagnosis will be crucial to track and control the spread of pathogen at early stage. In future, multi-plex PCR will be carried out with DNA directly extracted from blood specimen of suspected enteric fever patients without doing blood culture to reduce the time of identification. In addition, a rapid dipstick lateral flow assay developed by Professor Ankur Mutreja's team at University of Cambridge will be introduced at icddr,b in replace of traditional agarose gel electrophoresis for visualization of amplified PCR product.

However, WGS approaches open up new opportunities for disease prevention programme through use of improved surveillance framework to differentiate between the typhoidal and non-typhoidal *Salmonella* serovars so that public health interventions can be implemented and further serovar replacement can be tracked early. The potential application of WGS offers to implement innovative molecular surveillance for routine enteric fever diagnosis and paves the pathway for regional and global high-risk AMR clone identification, resistance mechanisms and their evolution to inform disease control measures, preventive strategies and treatment guidelines.

REFERENCES

1. Deen, J., et al., *Community-acquired bacterial bloodstream infections in developing countries in south and southeast Asia: a systematic review.* The Lancet Infectious Diseases, 2012. **12**(6): p. 480-487.
2. Crump, J.A., et al., *Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive Salmonella Infections.* Clinical microbiology reviews, 2015. **28**(4): p. 901-937.
3. Browne, A.J., et al., *Drug-resistant enteric fever worldwide, 1990 to 2018: a systematic review and meta-analysis.* BMC Med, 2020. **18**(1): p. 1.
4. Brenner, F.W., et al., *Salmonella Nomenclature.* Journal of Clinical Microbiology, 2000. **38**(7): p. 2465.
5. Sanderson, K.E., et al., *Chapter 71 - Salmonella Typhi and Salmonella Paratyphi A*, in *Molecular Medical Microbiology (Second Edition)*, Y.-W. Tang, et al., Editors. 2015, Academic Press: Boston. p. 1275-1306.
6. Akyala, A.I. and S. Alsam, *Extended Spectrum Beta Lactamase Producing Strains of Salmonella sp ecies - A Systematic Review.* Journal of Microbiology Research, 2015. **5**: p. 57-70.
7. Selander, R., et al., *Evolutionary genetic relationships of clones of Salmonella serovars that cause human typhoid and other enteric fevers.* Infection and immunity, 1990. **58**: p. 2262-75.
8. Connor, T., et al., *What's in a name? Species wide whole genome sequencing resolves invasive and non-invasive Salmonella Paratyphi B.* mBio, 2016. **7**(4).
9. Higginson, E.E., et al., *Improving Our Understanding of Salmonella enterica Serovar Paratyphi B through the Engineering and Testing of a Live Attenuated Vaccine Strain.* mSphere, 2018. **3**(6): p. e00474-18.
10. Radhakrishnan, A., et al., *Introductory article on global burden and epidemiology of typhoid fever.* 2018. **99**(3_Suppl): p. 4-9.
11. Darton, T.C., et al., *The STRATAA study protocol: a programme to assess the burden of enteric fever in Bangladesh, Malawi and Nepal using prospective population census, passive surveillance, serological studies and healthcare utilisation surveys.* BMJ Open, 2017. **7**(6): p. e016283.
12. Stanaway, J.D., et al., *The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017.* 2019. **19**(4): p. 369-381.
13. Saha, S., et al., *Epidemiology of Typhoid and Paratyphoid: Implications for Vaccine Policy.* Clin Infect Dis, 2019. **68**(Suppl 2): p. S117-s123.
14. Reddy, E.A., A.V. Shaw, and J.A. Crump, *Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis.* Lancet Infect Dis, 2010. **10**(6): p. 417-32.
15. Mogasale, V., et al., *Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment.* Lancet Glob Health, 2014. **2**(10): p. e570-80.
16. Stoll, B.J., et al., *Enteric fever in patients admitted to a diarrhoeal disease hospital in Bangladesh.* Trans R Soc Trop Med Hyg, 1983. **77**(4): p. 548-51.
17. Dewan, A.M., et al., *Typhoid Fever and its association with environmental factors in the Dhaka Metropolitan Area of Bangladesh: a spatial and time-series approach.* PLoS Negl Trop Dis, 2013. **7**(1): p. e1998.

18. Meiring, J.E., et al., *Burden of enteric fever at three urban sites in Africa and Asia: a multicentre population-based study.* Lancet Glob Health, 2021. **9**(12): p. e1688-e1696.
19. Basnyat, B., et al., *Enteric (typhoid) fever in travelers.* Clin Infect Dis, 2005. **41**(10): p. 1467-72.
20. Ochiai, R.L., et al., *Salmonella paratyphi A rates, Asia.* Emerg Infect Dis, 2005. **11**(11): p. 1764-6.
21. Levine, M.M., et al., *Ty21a Live Oral Typhoid Vaccine and Prevention of Paratyphoid Fever Caused by Salmonella enterica Serovar Paratyphi B.* Clinical Infectious Diseases, 2007. **45**(Supplement_1): p. S24-S28.
22. Bzdil, J., O. Holý, and J. Toporcak, *Salmonella enterica subspecies enterica serovar Paratyphi B as a disease-causing agent in reptiles in the Czech Republic.* Veterinární Medicína, 2017. **62**: p. 410-415.
23. Gobin, M., et al., *National outbreak of Salmonella Java phage type 3b variant 9 infection using parallel case-control and case-case study designs, United Kingdom, July to October 2010.* Euro Surveill, 2011. **16**(47): p. 20023.
24. al-Nakhli, H.M., Z.H. al-Ogaily, and T.J. Nassar, *Representative Salmonella serovars isolated from poultry and poultry environments in Saudi Arabia.* Rev Sci Tech, 1999. **18**(3): p. 700-9.
25. Barua, H., et al., *Poultry as a possible source of non-typhoidal Salmonella enterica serovars in humans in Bangladesh.* Vet Microbiol, 2014. **168**(2-4): p. 372-80.
26. Arya, G., et al., *Epidemiology, Pathogenesis, Genoserotyping, Antimicrobial Resistance, and Prevention and Control of Non-Typhoidal Salmonella Serovars.* Current Clinical Microbiology Reports, 2017. **4**(1): p. 43-53.
27. Dougan, G. and S. Baker, *Salmonella enterica serovar Typhi and the pathogenesis of typhoid fever.* Annu Rev Microbiol, 2014. **68**: p. 317-36.
28. Parry, C.M., et al., *Typhoid fever.* N Engl J Med, 2002. **347**(22): p. 1770-82.
29. Rahman, T., I. Hosen, and S. Chakraborty, *A Rapid Glimpse on Typhoid Fever: An Updated Mini Review.* Journal of Life Medicine, 2013. **01**: p. 83-92.
30. Crawford, R.W., et al., *Gallstones play a significant role in Salmonella spp. gallbladder colonization and carriage.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4353-8.
31. Khanam, F., et al., *Salmonella Typhi stool shedding by enteric fever patients and asymptomatic chronic carriers in an endemic urban setting.* J Infect Dis, 2021.
32. Parry, C.M., et al., *The utility of diagnostic tests for enteric fever in endemic locations.* Expert Rev Anti Infect Ther, 2011. **9**(6): p. 711-25.
33. Eng, S.-K., et al., *Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance.* Frontiers in Life Science, 2015. **8**(3): p. 284-293.
34. Gasem, M.H., et al., *Culture of Salmonella typhi and Salmonella paratyphi from blood and bone marrow in suspected typhoid fever.* Trop Geogr Med, 1995. **47**(4): p. 164-7.
35. Sheikh, A., et al., *Salmonella enterica serovar Typhi-specific immunoglobulin A antibody responses in plasma and antibody in lymphocyte supernatant specimens in Bangladeshi patients with suspected typhoid fever.* Clin Vaccine Immunol, 2009. **16**(11): p. 1587-94.
36. Khanam, F., et al., *Evaluation of a Typhoid/Paratyphoid Diagnostic Assay (TPTest) Detecting Anti-Salmonella IgA in Secretions of Peripheral Blood*

Lymphocytes in Patients in Dhaka, Bangladesh. PLOS Neglected Tropical Diseases, 2013. **7**(7): p. e2316.
37. Gibani, M.M., C. Britto, and A.J. Pollard, *Typhoid and paratyphoid fever: a call to action.* Curr Opin Infect Dis, 2018. **31**(5): p. 440-448.
38. Mawazo, A., G.M. Bwire, and M.I.N. Matee, *Performance of Widal test and stool culture in the diagnosis of typhoid fever among suspected patients in Dar es Salaam, Tanzania.* BMC research notes, 2019. **12**(1): p. 316-316.
39. Bhutta, Z.A., *Current concepts in the diagnosis and treatment of typhoid fever.* Bmj, 2006. **333**(7558): p. 78-82.
40. Coovadia, Y.M., et al., *Comparison of passive haemagglutination test with Widal agglutination test for serological diagnosis of typhoid fever in an endemic area.* J Clin Pathol, 1986. **39**(6): p. 680-3.
41. Oracz, G., et al., *Rapid diagnosis of acute Salmonella gastrointestinal infection.* Clin Infect Dis, 2003. **36**(1): p. 112-5.
42. Jesudason, M., E. Esther, and E. Mathai, *Typhidot test to detect IgG & IgM antibodies in typhoid fever.* Indian J Med Res, 2002. **116**: p. 70-2.
43. Keddy, K.H., et al., *Sensitivity and specificity of typhoid fever rapid antibody tests for laboratory diagnosis at two sub-Saharan African sites.* Bull World Health Organ, 2011. **89**(9): p. 640-7.
44. Baker, S., M. Favorov, and G. Dougan, *Searching for the elusive typhoid diagnostic.* BMC Infect Dis, 2010. **10**: p. 45.
45. Goay, Y.X., et al., *Identification of Five Novel Salmonella Typhi-Specific Genes as Markers for Diagnosis of Typhoid Fever Using Single-Gene Target PCR Assays.* Biomed Res Int, 2016. **2016**: p. 8905675.
46. Ali, A., et al., *Multiplex PCR for differential diagnosis of emerging typhoidal pathogens directly from blood samples.* Epidemiol Infect, 2009. **137**(1): p. 102-7.
47. Hirose, K., et al., *Selective amplification of tyv (rfbE), prt (rfbS), viaB, and fliC genes by multiplex PCR for identification of Salmonella enterica serovars Typhi and Paratyphi A.* J Clin Microbiol, 2002. **40**(2): p. 633-6.
48. Khokhar, F.A., et al., *Multiplex PCR Assay to Detect High Risk Lineages of Salmonella Typhi and Paratyphi A.* medRxiv, 2021: p. 2021.09.14.21263553.
49. Dyson, Z.A., et al., *Antibiotic Resistance and Typhoid.* Clinical Infectious Diseases, 2019. **68**(Supplement_2): p. S165-S170.
50. Kasper, M.R., et al., *Emergence of multidrug-resistant Salmonella enterica serovar Typhi with reduced susceptibility to fluoroquinolones in Cambodia.* Diagnostic microbiology and infectious disease, 2010. **66**(2): p. 207-209.
51. Holt, K.E., et al., *Temporal Fluctuation of Multidrug Resistant Salmonella Typhi Haplotypes in the Mekong River Delta Region of Vietnam.* PLOS Neglected Tropical Diseases, 2011. **5**(1): p. e929.
52. Wong, V.K., et al., *Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events.* Nat Genet, 2015. **47**(6): p. 632-9.
53. Britto, C.D., et al., *Laboratory and molecular surveillance of paediatric typhoidal Salmonella in Nepal: Antimicrobial resistance and implications for vaccine policy.* PLoS neglected tropical diseases, 2018. **12**(4): p. e0006408.
54. Mengo, D.M., et al., *Trends in Salmonella enteric serovar Typhi in Nairobi, Kenya from 2004 to 2006.* J Infect Dev Ctries, 2010. **4**(6): p. 393-6.

55. Feasey, N.A., et al., *Rapid emergence of multidrug resistant, H58-lineage Salmonella typhi in Blantyre, Malawi.* PLoS neglected tropical diseases, 2015. **9**(4): p. e0003748.
56. Hendriksen, R.S., et al., *Genomic signature of multidrug-resistant Salmonella enterica serovar Typhi isolates related to a massive outbreak in Zambia between 2010 and 2012.* Journal of clinical microbiology, 2015. **53**(1): p. 262-272.
57. Park, S.E., et al., *The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa.* Nature communications, 2018. **9**(1): p. 5094.
58. Wong, V.K., et al., *Molecular surveillance identifies multiple transmissions of typhoid in West Africa.* PLoS neglected tropical diseases, 2016. **10**(9): p. e0004781.
59. Patel, T.A., et al., *Imported enteric fever: case series from the hospital for tropical diseases, London, United Kingdom.* Am J Trop Med Hyg, 2010. **82**(6): p. 1121-6.
60. Rahman, M., et al., *Emergence of multidrug-resistant Salmonella enterica serotype Typhi with decreased ciprofloxacin susceptibility in Bangladesh.* Epidemiology & Infection, 2006. **134**(2): p. 433-438.
61. Thanh, D.P., et al., *A novel ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure.* Elife, 2016. **5**: p. e14003.
62. Parry, C., et al., *Quinolone-resistant Salmonella typhi in Vietnam.* Lancet, 1998. **351**(9111): p. 1289.
63. Balasegaram, S., et al., *Guidelines for the public health management of typhoid and paratyphoid in England: practice guidelines from the National Typhoid and Paratyphoid Reference Group.* J Infect, 2012. **65**(3): p. 197-213.
64. Hooda, Y., et al., *Molecular mechanism of azithromycin resistance among typhoidal Salmonella strains in Bangladesh identified through passive pediatric surveillance.* PLoS Negl Trop Dis, 2019. **13**(11): p. e0007868.
65. Djeghout, B., et al., *Ceftriaxone-resistant Salmonella Typhi carries an IncI1-ST31 plasmid encoding CTX-M-15.* J Med Microbiol, 2018. **67**(5): p. 620-627.
66. Klemm, E.J., 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**(1): p. e00105-18.
67. Akram, J., et al., *Extensively Drug-Resistant (XDR) Typhoid: Evolution, Prevention, and Its Management.* BioMed Research International, 2020. **2020**.
68. Andrews, J.R., et al., *Extensively Drug-Resistant Typhoid — Are Conjugate Vaccines Arriving Just in Time?* New England Journal of Medicine, 2018. **379**(16): p. 1493-1495.
69. Anwar, E., et al., *Vaccines for preventing typhoid fever.* Cochrane Database Syst Rev, 2014(1): p. Cd001261.
70. Theiss-Nyland, K., et al., *Assessing the Impact of a Vi-polysaccharide Conjugate Vaccine in Preventing Typhoid Infection Among Bangladeshi Children: A Protocol for a Phase IIIb Trial.* Clin Infect Dis, 2019. **68**(Suppl 2): p. S74-s82.
71. Qadri, F., et al., *Protection by vaccination of children against typhoid fever with a Vi-tetanus toxoid conjugate vaccine in urban Bangladesh: a cluster-randomised trial.* The Lancet, 2021. **398**(10301): p. 675-684.

72. Sahastrabuddhe, S., et al., *Increasing rates of Salmonella Paratyphi A and the current status of its vaccine development.* Expert Rev Vaccines, 2013. **12**(9): p. 1021-31.
73. Dobinson, H.C., et al., *Evaluation of the Clinical and Microbiological Response to Salmonella Paratyphi A Infection in the First Paratyphoid Human Challenge Model.* Clin Infect Dis, 2017. **64**(8): p. 1066-1073.
74. Mardis, E.R., *Next-generation DNA sequencing methods.* Annu Rev Genomics Hum Genet, 2008. **9**: p. 387-402.
75. Liu, L., et al., *Comparison of next-generation sequencing systems.* J Biomed Biotechnol, 2012. **2012**: p. 251364.
76. Pop, M. and S.L. Salzberg, *Bioinformatics challenges of new sequencing technology.* Trends in genetics : TIG, 2008. **24**(3): p. 142-149.
77. Lang, D., et al., *Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore.* GigaScience, 2020. **9**(12).
78. Riaz, N., et al., *Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants.* BMC Genomics, 2021. **22**(1): p. 148.
79. Hess, J.F., et al., *Library preparation for next generation sequencing: A review of automation strategies.* Biotechnology Advances, 2020. **41**: p. 107537.
80. Loman, N.J. and M.J. Pallen, *Twenty years of bacterial genome sequencing.* Nature Reviews Microbiology, 2015. **13**(12): p. 787-794.
81. Zhou, Q., et al., *QC-Chain: Fast and Holistic Quality Control Method for Next-Generation Sequencing Data.* PLOS ONE, 2013. **8**(4): p. e60234.
82. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
83. He, Y., et al., *De novo assembly methods for next generation sequencing data.* Tsinghua Science and Technology, 2013. **18**(5): p. 500-514.
84. Ekblom, R. and J.B. Wolf, *A field guide to whole-genome sequencing, assembly and annotation.* Evol Appl, 2014. **7**(9): p. 1026-42.
85. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.* Genome Res, 2010. **20**(9): p. 1297-303.
86. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
87. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.* arXiv preprint arXiv:1303.3997, 2013.
88. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
89. Danecek, P. and S.A. McCarthy, *BCFtools/csq: haplotype-aware variant consequences.* Bioinformatics, 2017. **33**(13): p. 2037-2039.
90. Carver, T., et al., *Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data.* Bioinformatics, 2012. **28**(4): p. 464-9.
91. Shumway, M., G. Cochrane, and H. Sugawara, *Archiving next generation sequencing data.* Nucleic acids research, 2010. **38**(Database issue): p. D870-D871.
92. Ong, S.Y., et al., *The Genomic Blueprint of Salmonella enterica subspecies enterica serovar Typhi P-stx-12.* Stand Genomic Sci, 2013. **7**(3): p. 483-96.

93. Yap, K.P. and K.L. Thong, *Salmonella Typhi genomics: envisaging the future of typhoid eradication.* Trop Med Int Health, 2017. **22**(8): p. 918-925.
94. Parkhill, J., et al., *Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.* Nature, 2001. **413**(6858): p. 848-852.
95. Xu, D., et al., *Genome Sequence of Salmonella enterica Serovar Typhi Oral Vaccine Strain Ty21a.* Genome announcements, 2013. **1**(4): p. e00650-13.
96. Deng, W., et al., *Comparative Genomics of Salmonellaenterica Serovar Typhi Strains Ty2 and CT18*. Journal of Bacteriology, 2003. **185**(7): p. 2330-2337.
97. McClelland, M., et al., *Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid.* Nature Genetics, 2004. **36**(12): p. 1268-1274.
98. Holt, K.E., et al., *Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi.* BMC genomics, 2009. **10**: p. 36-36.
99. Holt Kathryn, E., et al., *Multidrug-Resistant Salmonella enterica Serovar Paratyphi A Harbors IncHI1 Plasmids Similar to Those Found in Serovar Typhi.* Journal of Bacteriology, 2007. **189**(11): p. 4257-4264.
100. Borowiak, M., et al., *Complete Genome Sequence of Salmonella enterica subsp. enterica Serovar Paratyphi B Sequence Type 28 Harboring mcr-1*. Genome announcements, 2017. **5**(37): p. e00991-17.
101. Wong, V.K., et al., *An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid.* Nature Communications, 2016. **7**(1): p. 12827.
102. Achtman, M., *Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens.* Annu Rev Microbiol, 2008. **62**: p. 53-70.
103. Sen, B., et al., *Phage typing, biotyping & antimicrobial resistance profile of Salmonella enterica serotype Typhi from Kolkata.* Indian J Med Res, 2007. **125**(5): p. 685-8.
104. Kim, S., et al., *Clustering analysis of Salmonella enterica serovar Typhi isolates in Korea by PFGE, ribotying, and phage typing.* Foodborne Pathog Dis, 2009. **6**(6): p. 733-8.
105. Spratt, B.G., *Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet.* Current Opinion in Microbiology, 1999. **2**(3): p. 312-316.
106. Roumagnac, P., et al., *Evolutionary history of Salmonella typhi.* Science, 2006. **314**(5803): p. 1301-4.
107. Dyson, Z.A. and K.E. Holt, *Five years of GenoTyphi: updates to the global Salmonella Typhi genotyping framework.* The Journal of Infectious Diseases, 2021.
108. Tanmoy, A.M., et al., *Salmonella enterica Serovar Typhi in Bangladesh: Exploration of Genomic Diversity and Antimicrobial Resistance.* mBio, 2018. **9**(6).
109. Rahman, S.I.A., 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**(2): p. e0008036.

110. Pham Thanh, D., et al., *A novel ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure.* Elife, 2016. **5**: p. e14003.
111. Dyson, Z.A., et al., *Whole Genome Sequence Analysis of Salmonella Typhi Isolated in Thailand before and after the Introduction of a National Immunization Program.* PLoS Negl Trop Dis, 2017. **11**(1): p. e0005274.
112. Zhou, Z., et al., *Transient Darwinian selection in Salmonella enterica serovar Paratyphi A during 450 years of global spread of enteric fever.* Proceedings of the National Academy of Sciences, 2014. **111**(33): p. 12199-12204.
113. Kuijpers, L.M.F., et al., *Genomic analysis of Salmonella enterica serotype Paratyphi A during an outbreak in Cambodia, 2013-2015.* Microbial genomics, 2016. **2**(11): p. e000092-e000092.
114. Yan, M., et al., *A Large-Scale Community-Based Outbreak of Paratyphoid Fever Caused by Hospital-Derived Transmission in Southern China.* PLoS Negl Trop Dis, 2015. **9**(7): p. e0003859.
115. Tanmoy, A.M., et al., *Paratype: A genotyping framework and an open-source tool for Salmonella Paratyphi A.* medRxiv, 2021: p. 2021.11.13.21266165.
116. Barker, R.M., et al., *Types of Salmonella paratyphi B and their phylogenetic significance.* J Med Microbiol, 1988. **26**(4): p. 285-93.
117. Ezquerra, E., et al., *Genotypic typing and phylogenetic analysis of Salmonella paratyphi B and S. java with IS200.* Journal of general microbiology, 1993. **139 10**: p. 2409-14.
118. Achtman, M., et al., *Multilocus Sequence Typing as a Replacement for Serotyping in Salmonella enterica.* PLOS Pathogens, 2012. **8**(6): p. e1002776.
119. Fabre, L., et al., *CRISPR typing and subtyping for improved laboratory surveillance of Salmonella infections.* PLoS One, 2012. **7**(5): p. e36995.
120. Baker, S. and G. Dougan, *The genome of Salmonella enterica serovar Typhi.* Clin Infect Dis, 2007. **45 Suppl 1**: p. S29-33.
121. Kidgell, C., et al., *Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old.* Infect Genet Evol, 2002. **2**(1): p. 39-45.
122. Ochman, H., S. Elwyn, and N.A. Moran, *Calibrating bacterial evolution.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(22): p. 12638-12643.
123. Holt, K.E., et al., *High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi.* Nat Genet, 2008. **40**(8): p. 987-93.
124. Baddam, R., et al., *Genome dynamics and evolution of Salmonella Typhi strains from the typhoid-endemic zones.* Scientific Reports, 2014. **4**(1): p. 7457.
125. Didelot, X., et al., *A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination?* Genome research, 2007. **17**(1): p. 61-68.
126. Li, H., et al., *Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects.* Signal Transduction and Targeted Therapy, 2020. **5**(1): p. 1.
127. McClelland, M., et al., *Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid.* Nat Genet, 2004. **36**(12): p. 1268-74.

128. Vila Nova, M., et al., *Genetic and metabolic signatures of Salmonella enterica subsp. enterica associated with animal sources at the pangenomic scale.* BMC Genomics, 2019. **20**(1): p. 814.
129. Tanner, J.R. and R.A. Kingsley, *Evolution of Salmonella within Hosts.* Trends Microbiol, 2018. **26**(12): p. 986-998.
130. Bäumler, A.J., et al., *Evolution of host adaptation in Salmonella enterica.* Infect Immun, 1998. **66**(10): p. 4579-87.
131. Jacobsen, A., et al., *The Salmonella enterica pan-genome.* Microb Ecol, 2011. **62**(3): p. 487-504.
132. Liang, W., et al., *Pan-Genomic Analysis Provides Insights into the Genomic Variation and Evolution of Salmonella Paratyphi A.* PLOS ONE, 2012. **7**(9): p. e45346.
133. Laing, C.R., M.D. Whiteside, and V.P.J. Gannon, *Pan-genome Analyses of the Species Salmonella enterica, and Identification of Genomic Markers Predictive for Species, Subspecies, and Serovar.* Frontiers in Microbiology, 2017. **8**(1345).
134. Munita, J.M. and C.A. Arias, *Mechanisms of Antibiotic Resistance.* Microbiology spectrum, 2016. **4**(2): p. 10.1128/microbiolspec.VMBF-0016-2015.
135. Phan, M.D., et al., *Variation in Salmonella enterica serovar typhi IncHI1 plasmids during the global spread of resistant typhoid fever.* Antimicrob Agents Chemother, 2009. **53**(2): p. 716-27.
136. Holt, K.E., et al., *Emergence of a Globally Dominant IncHI1 Plasmid Type Associated with Multiple Drug Resistant Typhoid.* PLOS Neglected Tropical Diseases, 2011. **5**(7): p. e1245.
137. Rowe, B., L.R. Ward, and E.J. Threlfall, *Multidrug-resistant Salmonella typhi: a worldwide epidemic.* Clin Infect Dis, 1997. **24 Suppl 1**: p. S106-9.
138. Britto, C.D., et al., *Persistent circulation of a fluoroquinolone-resistant Salmonella enterica Typhi clone in the Indian subcontinent.* Journal of Antimicrobial Chemotherapy, 2019. **75**(2): p. 337-341.
139. Carey, M.E., et al., *Spontaneous Emergence of Azithromycin Resistance in Independent Lineages of Salmonella Typhi in Northern India.* Clin Infect Dis, 2021. **72**(5): p. e120-e127.
140. Iqbal, J., et al., *A Race against Time: Reduced Azithromycin Susceptibility in Salmonella enterica Serovar Typhi in Pakistan.* mSphere, 2020. **5**(4): p. e00215-20.
141. Duy, P.T., et al., *The emergence of azithromycin-resistant Salmonella Typhi in Nepal.* JAC-antimicrobial resistance, 2020. **2**(4).
142. Qian, H., et al., *Discovery of seven novel mutations of gyrB, parC and parE in Salmonella Typhi and Paratyphi strains from Jiangsu Province of China.* Scientific reports, 2020. **10**(1): p. 7359-7359.
143. Menezes, G.A., et al., *Antimicrobial resistance trends in blood culture positive Salmonella Paratyphi A isolates from Pondicherry, India.* Indian J Med Microbiol, 2016. **34**(2): p. 222-7.
144. Nair, S., et al., *Genomic surveillance detects Salmonella enterica serovar Paratyphi A harbouring blaCTX-M-15 from a traveller returning from Bangladesh.* PLOS ONE, 2020. **15**(1): p. e0228250.
145. Pokharel, B.M., et al., *Multidrug-resistant and extended-spectrum beta-lactamase (ESBL)-producing Salmonella enterica (serotypes Typhi and*

Paratyphi A) from blood isolates in Nepal: surveillance of resistance and a search for newer alternatives. Int J Infect Dis, 2006. **10**(6): p. 434-8.
146. Threlfall, J., et al., *Multidrug-resistant Salmonella Java.* Emerging infectious diseases, 2005. **11**(1): p. 170-171.
147. Hirose, K., et al., *Selective amplification of tyv (rfbE), prt (rfbS), viaB, and fliC genes by multiplex PCR for identification of Salmonella enterica serovars Typhi and Paratyphi A.* Journal of clinical microbiology, 2002. **40**(2): p. 633-636.
148. Khanam, F., et al., *Typhoid Fever in Young Children in Bangladesh: Clinical Findings, Antibiotic Susceptibility Pattern and Immune Responses.* PLOS Neglected Tropical Diseases, 2015. **9**(4): p. e0003619.
149. Khan, A.I., et al., *Epidemiology of Cholera in Bangladesh: Findings From Nationwide Hospital-based Surveillance, 2014-2018.* Clin Infect Dis, 2020. **71**(7): p. 1635-1642.
150. Sack, R.B., et al., *A 4-year study of the epidemiology of Vibrio cholerae in four rural areas of Bangladesh.* J Infect Dis, 2003. **187**(1): p. 96-101.
151. Wang, S.K., et al., *Yields from blood cultures of patients with suspected paratyphoid fever A.* Braz J Microbiol, 2010. **41**(2): p. 283-5.
152. Wahud, A., et al., *Bacteriological profiles of Blood culture isolates by BacT/ALERT 3D automated system.* Journal of Shaheed Suhrawardy Medical College, 2012. **1**.
153. Hanum, L., et al., *Comparison of CTAB Method and Wizard Genomic DNA Purification System Kit from Promega on DNA Isolation of Local Varities of Rice of South Sumatera.* Science and Technology Indonesia, 2018. **3**: p. 26-29.
154. García-Alegría, A.M., et al., *Quantification of DNA through the NanoDrop Spectrophotometer: Methodological Validation Using Standard Reference Material and Sprague Dawley Rat and Human DNA.* International Journal of Analytical Chemistry, 2020. **2020**: p. 8896738.
155. Holm, J.B., et al., *Ultrahigh-Throughput Multiplexing and Sequencing of >500-Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform.* mSystems, 2019. **4**(1): p. e00029-19.
156. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature Methods, 2012. **9**(4): p. 357-359.
157. Ponstingl, H. and Z. Ning, *SMALT - A new mapper for DNA sequencing reads*. Vol. 1. 2010.
158. Croucher, N.J., et al., *Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.* Nucleic Acids Research, 2014. **43**(3): p. e15-e15.
159. Page, A.J., et al., *SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments.* Microbial genomics, 2016. **2**(4): p. e000056-e000056.
160. Paradis, E., J. Claude, and K. Strimmer, *APE: Analyses of Phylogenetics and Evolution in R language.* Bioinformatics, 2004. **20**(2): p. 289-90.
161. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-1313.
162. Tonkin-Hill, G., et al., *RhierBAPS: An R implementation of the population clustering algorithm hierBAPS.* Wellcome Open Research, 2018. **3**: p. 93.
163. Argimón, S., et al., *Microreact: visualizing and sharing data for genomic epidemiology and phylogeography.* Microbial genomics, 2016. **2**(11): p. e000093-e000093.

164. Yu, G., et al., *ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data.* Methods in Ecology and Evolution, 2017. **8**(1): p. 28-36.
165. Hadfield, J., et al., *Phandango: an interactive viewer for bacterial population genomics.* Bioinformatics (Oxford, England), 2018. **34**(2): p. 292-293.
166. Letunic, I. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.* Nucleic acids research, 2016. **44**(W1): p. W242-W245.
167. Inouye, M., et al., *SRST2: Rapid genomic surveillance for public health and hospital microbiology labs.* Genome Medicine, 2014. **6**(11): p. 90.
168. Gupta, S.K., et al., *ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes.* Antimicrob Agents Chemother, 2014. **58**(1): p. 212-20.
169. Carattoli, A. and H. Hasman, *PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS).* Methods Mol Biol, 2020. **2075**: p. 285-294.
170. Hawkey, J., et al., *ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data.* BMC Genomics, 2015. **16**(1): p. 667.
171. Wick, R.R., et al., *Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.* PLOS Computational Biology, 2017. **13**(6): p. e1005595.
172. Wick, R.R., et al., *Bandage: interactive visualization of de novo genome assemblies.* Bioinformatics, 2015. **31**(20): p. 3350-3352.
173. Hunt, M., et al., *ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads.* Microbial genomics, 2017. **3**(10): p. e000131-e000131.
174. McArthur, A.G., et al., *The comprehensive antibiotic resistance database.* Antimicrob Agents Chemother, 2013. **57**(7): p. 3348-57.
175. Chen, L., et al., *VFDB: a reference database for bacterial virulence factors.* Nucleic Acids Res, 2005. **33**(Database issue): p. D325-8.
176. Yoshida, C.E., et al., *The Salmonella In Silico Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft Salmonella Genome Assemblies.* PLoS One, 2016. **11**(1): p. e0147101.
177. Argimón, S., et al., *A global resource for genomic predictions of antimicrobial resistance and surveillance of Salmonella Typhi at pathogenwatch.* Nature Communications, 2021. **12**(1): p. 2879.
178. Donkor, E., N. Dayie, and T. Adiku, *Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA).* Journal of Bioinformatics and Sequence Analysis, 2014. **6**: p. 1-6.
179. Carver, T.J., et al., *ACT: the Artemis Comparison Tool.* Bioinformatics, 2005. **21**(16): p. 3422-3.
180. Sullivan, M.J., N.K. Petty, and S.A. Beatson, *Easyfig: a genome comparison visualizer.* Bioinformatics, 2011. **27**(7): p. 1009-10.
181. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
182. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072-5.
183. Page, A.J., et al., *Roary: rapid large-scale prokaryote pan genome analysis.* Bioinformatics, 2015. **31**(22): p. 3691-3.

184. Snipen, L. and K.H. Liland, *micropan: an R-package for microbial pan-genomics.* BMC Bioinformatics, 2015. **16**(1): p. 79.
185. Tettelin, H., et al., *Comparative genomics: the bacterial pan-genome.* Curr Opin Microbiol, 2008. **11**(5): p. 472-7.
186. Tonkin-Hill, G., et al., *Producing polished prokaryotic pangenomes with the Panaroo pipeline.* Genome Biol, 2020. **21**(1): p. 180.
187. Hadfield, J., et al., *Phandango: an interactive viewer for bacterial population genomics.* Bioinformatics, 2018. **34**(2): p. 292-293.
188. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome research, 2003. **13**(11): p. 2498-2504.
189. Quevillon, E., et al., *InterProScan: protein domains identifier.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W116-20.
190. Eichinger, V., et al., *EffectiveDB--updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems.* Nucleic Acids Res, 2016. **44**(D1): p. D669-74.
191. Team, R.C., *R: A language and environment for statistical computing.* 2013.
192. Dixon, P., *VEGAN, a package of R functions for community ecology.* Journal of Vegetation Science, 2003. **14**(6): p. 927-930.
193. StataCorp, L., *Stata/SE version 12.1 [computer software]*. College Station, TX: Stata Corp, LLP, 2011.
194. Maude, R.R., et al., *A prospective study of the importance of enteric fever as a cause of non-malarial febrile illness in patients admitted to Chittagong Medical College Hospital, Bangladesh.* BMC Infect Dis, 2016. **16**(1): p. 567.
195. Ye, J., et al., *Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction.* BMC Bioinformatics, 2012. **13**(1): p. 134.
196. Black, R.E., et al., *Efficacy of one or two doses of Ty21a Salmonella typhi vaccine in enteric-coated capsules in a controlled field trial. Chilean Typhoid Committee.* Vaccine, 1990. **8**(1): p. 81-4.
197. Holt, K.E., et al., *High-resolution genotyping of the endemic Salmonella Typhi population during a Vi (typhoid) vaccination trial in Kolkata.* PLoS Negl Trop Dis, 2012. **6**(1): p. e1490.
198. Rahman, S.I.A., et al., *Genome-wide analysis provides a deeper understanding of the population structure of the Salmonella enterica serotype Paratyphi B complex in Bangladesh.* Microb Genom, 2021. **7**(9).
199. Rahman, S.I.A., et al., *Genetic diversity of Salmonella Paratyphi A isolated from enteric fever patients in Bangladesh from 2008 to 2018.* PLOS Neglected Tropical Diseases, 2021. **15**(10): p. e0009748.
200. Cuypers, W.L., et al., *Fluoroquinolone resistance in Salmonella: insights by whole-genome sequencing.* Microb Genom, 2018. **4**(7).
201. Kidgell, C., et al., *Characterisation and distribution of a cryptic Salmonella typhi plasmid pHCM2.* Plasmid, 2002. **47**(3): p. 159-71.
202. Song, J.H., et al., *Detection of Salmonella typhi in the blood of patients with typhoid fever by polymerase chain reaction.* J Clin Microbiol, 1993. **31**(6): p. 1439-43.
203. Virlogeux, I., et al., *Role of the viaB locus in synthesis, transport and expression of Salmonella typhi Vi antigen.* Microbiology, 1995. **141**(12): p. 3039-3047.

204. Maude, R.R., et al., *A prospective study of the importance of enteric fever as a cause of non-malarial febrile illness in patients admitted to Chittagong Medical College Hospital, Bangladesh.* BMC infectious diseases, 2016. **16**(1): p. 567-567.

APPENDIX A

Reagents and Preparation of Different Media

Reagents

1. *Salmonella*-specific O and flagellar H antiserum (Denka Sieken Tokyo, Japan)
2. 20% glycerol (Sigma Aldrich)
3. Wizard Genomic DNA Kit (Promega, A1120)
4. NexteraXT DNA sample Preparation Kit (Illumina, FC-131-1024)
5. Nuclease free water (Thermo Scientific)
6. 2X PCR master mix (Thermo scientific)
7. Agarose (Bio-Rad Laboratory, Richmond, CA, USA)
8. Gel Red, Biotium, Hayward, CA, USA

Preparation of different media

1. MacConkey agar plate (1000 mL, pH=7.2)

Peptone	20.0 gm
Lactose	10.0 gm
NaCl	5.0 gm
Bile salts	1.5 gm
Neutral red	0.05 gm
Crystal violet	1.0 gm
Bacto agar	15.0 gm

2. *Salmonella-Shigella* (SS) agar plates (1000 mL, pH=7.2)

Beef extract	5.0 gm
Enzymatic digest of Casein	2.50 gm
Enzymatic digest of Animal tissue	2.50 gm
Lactose	10.0 gm
Bile salts	8.50 gm
Sodium citrate	8.50 gm
Sodium thiosulfate	8.50 gm
Ferric citrate	1.0 gm
Brilliant green	0.00033 gm
Neutral red	0.025
Agar	13.50

3. Kligler's Iron Agar (1000 mL, pH=7.4)

Beef extract	3.0 gm
Yeast extract	3.0 gm
Peptone	15.0 gm
Proteose peptone	5.0 gm
Lactose	10.0 gm
Glucose	1.0 gm
Ferrous sulfate	0.2 gm
Sodium chloride	5.0 gm
Sodium thiosulfate	0.3 gm
Agar	12.0 gm
Phenol red	0.024 gm

4. Motility Indole Urea (MIU) (1000 mL, pH=7.4)

Peptone	30.0 gm
KH_2PO_4	2.0 gm
Sodium chloride	5.0 gm
Phenol red	0.005 gm
Urea	20.0 gm
Bacto agar	4.0 gm

5. Citrate agar (1000 mL, pH=7.4)

NaCl	5.0 gm
$MgSO_4$	0.2 gm
NH_4PO_4	1.0 gm
Sodium citrate	2.0 gm
Bacto agar	20.0 gm

6. Brain Heart Infusion (BHI) broth (1000 mL,pH=7.4)

Calf brain, infusion form	200 gm
Beef heart, infusion form	250 gm
Proteose peptone	10 gm
Dextrose	2 gm
Sodium chloride	5 gm
Disodium phosphate	2.5 gm

PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Population structure and antimicrobial resistance patterns of *Salmonella* Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016

Sadia Isfat Ara Rahman[1☯], Zoe A. Dyson[2,3,4☯], Elizabeth J. Klemm[4], Farhana Khanam[1], Kathryn E. Holt[3,5], Emran Kabir Chowdhury[6], Gordon Dougan[2,4‡*], Firdausi Qadri[1‡]

1 Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh, **2** Department of Medicine, University of Cambridge, Cambridge, United Kingdom, **3** Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria, Australia, **4** Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom, **5** Department of Infection Biology, Faculty of Infections and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **6** Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh

☯ These authors contributed equally to this work.
‡ These authors are joint senior authors on this work.
* gd312@medschl.cam.ac.uk

OPEN ACCESS

Citation: Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE, Chowdhury EK, et al. (2020) Population structure and antimicrobial resistance patterns of *Salmonella* Typhi isolates in urban Dhaka, Bangladesh from 2004 to 2016. PLoS Negl Trop Dis 14(2): e0008036. https://doi.org/10.1371/journal.pntd.0008036

Editor: Abiola Senok, Mohammed Bin Rashid University of Medicine and Health Sciences, UNITED ARAB EMIRATES

Received: July 3, 2019

Accepted: January 8, 2020

Published: February 27, 2020

Data Availability Statement: Genomic data are submitted to the European Nucleotide Archive (ENA) and accession numbers are available in the supplementary tables.

Funding: This work was supported by the Wellcome Trust Sanger Institute, Cambridge, UK and International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) as well as the Cambridge BRC AMR Theme (NIHR). This study

Abstract

Background

Multi-drug resistant typhoid fever remains an enormous public health threat in low and middle-income countries. However, we still lack a detailed understanding of the epidemiology and genomics of *S*. Typhi in many regions. Here we have undertaken a detailed genomic analysis of typhoid in urban Dhaka, Bangladesh to unravel the population structure and antimicrobial resistance patterns in *S*. Typhi isolated between 2004–2016.

Principal findings

Whole genome sequencing of 202 *S*. Typhi isolates obtained from three study locations in urban Dhaka revealed a diverse range of *S*. Typhi genotypes and AMR profiles. The bacterial population within Dhaka were relatively homogenous with little stratification between different healthcare facilities or age groups. We also observed evidence of exchange of Bangladeshi genotypes with neighboring South Asian countries (India, Pakistan and Nepal) suggesting these are circulating throughout the region. This analysis revealed a decline in H58 (genotype 4.3.1) isolates from 2011 onwards, coinciding with a rise in a diverse range of non-H58 genotypes and a simultaneous rise in isolates with reduced susceptibility to fluoroquinolones, potentially reflecting a change in treatment practices. We identified a novel *S*. Typhi genotype, subclade 3.3.2 (previously defined only to clade level, 3.3), which formed two localized clusters (3.3.2.Bd1 and 3.3.2.Bd2) associated with different mutations in the Quinolone Resistance Determining Region (QRDR) of gene *gyrA*.

was supported by the grants from the Wellcome Trust, the National Institutes of Health, including the National Institute of Allergy and Infectious Diseases (niaid.nih.gov, AI100023 [FQ]); AI058935 [FQ]) as well as a Fogarty International Center Training Grant in Vaccine Development and Public Health (fic.nih.gov, TW005572 [FK and FQ]), Bill and Melinda Gates Foundation (gatesfoundation.org, Grant no. OPP50419) and also SIDA fund (sida.se, 54100020, 51060029). icddr,b is grateful to the Governments of Bangladesh, Canada, Sweden and UK for providing core support. ZAD was supported by a grant funded by the Wellcome Trust (wellcome.ac.uk, 106158/Z/14/Z). KEH is supported by a Senior Medical Research Fellowship from the Viertel Foundation of Australia (viertel.org.au). This work was also supported by The Wellcome Trust (STRATAA) [106158 and 098051], the Gates Foundation (TyVAC) [OPP1151153]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Significance

Our analysis of *S*. Typhi isolates from urban Dhaka, Bangladesh isolated over a twelve year period identified a diverse range of AMR profiles and genotypes. The observed increase in non-H58 genotypes associated with reduced fluoroquinolone susceptibility may reflect a change in treatment practice in this region and highlights the importance of continued molecular surveillance to monitor the ongoing evolution of AMR in Dhaka. We have defined new genotypes and lineages of Bangladeshi *S*. Typhi which will facilitate the identification of these emerging AMR clones in future surveillance efforts.

Author summary

Typhoid fever, caused by *Salmonella enterica* serovar Typhi, is an acute and often life-threatening febrile illness in developing countries. Until recently, there have been limited studies focusing on the epidemiology and disease burden of typhoid in poor resource settings including Bangladesh. This study highlights the urgent need for sustained genomics based surveillance studies to monitor the population structure and ongoing evolution of AMR. Our data revealed a diverse range of *S*. Typhi genotypes and AMR patterns among 202 isolates collected from three urban areas in Dhaka, Bangladesh. Moreover, we defined a novel genotype, subclade 3.3.2 (previously typed only to clade level, 3.3) with two Bangladesh-localized clades 3.3.2.Bd1 and 3.3.2.Bd2 showing reduced susceptibility to fluoroquinolones. Our data shows a significant increase in non-H58 genotypes carrying QRDR mutations from 2012 onwards, replacing MDR H58 genotypes. Our data suggest that a shift in treatment practice towards third generation cephalosporins to control typhoid may be beneficial, in addition to the introduction of vaccination programs and improvements in water sanitation and hygiene (WASH) in urban Dhaka, Bangladesh.

Introduction

Salmonella enterica serovar Typhi (*S*. Typhi), the causative agent of typhoid fever, is a facultative intracellular and human restricted pathogen predominantly transmitted by the fecal-oral route. Typhoid remains an enormous public health threat in many developing countries due to inadequate access to safe water, poor sanitation systems and inappropriate use of antimicrobial drugs. It is estimated that typhoid fever affects 12–27 million people globally each year whereas in Bangladesh the overall incidence is estimated at between 292–395 cases per 100,000 people per year [1–5]. Multi-drug resistant (MDR) *S*. Typhi, defined as resistance to the first-line antibiotics ampicillin, chloramphenicol and trimethoprim-sulfamethoxazole, was first observed in the 1970s [6–9]. The more recent emergence of MDR *S*. Typhi with nalidixic acid resistance and reduced susceptibility to fluoroquinolones complicates treatment options. However, third generation cephalosporins such as ceftriaxone and cefixime have proven to be effective choices for treatment as resistance to cephalosporin in *S*. Typhi is still relatively rare. The first major outbreak with ceftriaxone resistance (extensively drug resistant, defined as resistant to three first-line drugs, fluoroquinolones and third-generation cephalosporin) was observed in Pakistan from 2016 onwards [7, 10, 11].

The acquisition of antimicrobial resistance (AMR) genes by *S*. Typhi was historically associated with self-transmissible IncHI1 plasmids that harbor composite transposons [9, 12]. The global burden of MDR typhoid is driven to a significant degree by the dissemination of the highly clonal, expanding haplotype H58 (genotype 4.3.1), which is now dominant in many

endemic settings throughout Africa and Asia [7, 9, 13, 14]. H58 *S*. Typhi encoding nonsynonymous mutations in the quinolone resistance determining region (QRDR) of DNA gyrase genes *gyrA* and *gyrB* and DNA topoisomerase IV genes *parC* and *parE* have shown reduced susceptibility to fluoroquinolones [7, 15]. Studies on typhoid in Nepal have reported the evolution of fluoroquinolone resistant H58 lineage II isolates carrying three QRDR mutations (*gyrA-S83F*, *gyrA-D87N*, and *parC-S80I*) responsible for the failure of a gatifloxacin treatment trial [6, 15]. More recently, in Bangladesh, a new H58 lineage I triple QRDR mutant carrying three different mutations (*gyrA-S83F*, *gyrA-D87G*, and *parC-E84K*) has been observed [14]. This study also defined a H58 "lineage Bd" containing the sublineage "Bdq" that is characterized by a high median minimal inhibitory concentration (MIC) to ciprofloxacin (4 μg/mL) potentially involving a *qnrS* gene in addition to a *gyrA*-S83Y mutation [14].

The lack of credible surveillance data representing the actual disease burden of typhoid fever in Bangladesh presents a barrier to the implementation of control strategies. Thus, there is an urgent need for sustained genomics-based surveillance studies in poor resource settings like Bangladesh to monitor the pathogen population structure, transmission dynamics, AMR patterns and the impacts of control strategies such as vaccination programs. Here, we have used whole genome sequencing (WGS) to better understand the population structure of *S*. Typhi isolated from three different urban areas of Dhaka, Bangladesh, between 2004 and 2016.

Methods

Ethics statement

Ethical approval was obtained from Research Review Committee (RRC) and the Ethical Review Committee (ERC) of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). Informed written consent and clinical information were taken from legal guardians of child participants and adult participants.

Study settings and blood sample collection

Dhaka is the capital city of Bangladesh and the most densely populated city with a population of over 18 million [16]. icddr,b is an international health research organization located at the Mohakhali area in Dhaka which runs two urban field sites at Kamalapur and Mirpur. Kamalapur is situated in the southeast part of Dhaka, whereas Mirpur is located in the northeast part of Dhaka metropolitan area. Both of these field sites are frequented by typhoid fever patients where sanitation systems are poor and access of safe drinking water is limited. This study was designed with these three urban areas inside Dhaka city: icddr,b Kamalapur field site, icddr,b Mirpur field site and icddr,b Dhaka hospital, Mohakhali. Suspected typhoid fever patients were enrolled from the three sites based on the criteria of fever at least 38˚C with a minimum duration of 3 days. Blood samples (3 mL for children <5 years of age and 5 mL for others) from typhoid suspected patients were collected between 2004 to 2016 and were cultured using the automated BacT/Alert method [17, 18] for confirmation of typhoid fever.

Bacterial isolation from blood culture

Specimens from positive blood culture bottles were sub-cultured on MacConkey agar plate and incubated at 37˚C for 18–24 hours. *S*. Typhi colonies were identified using standard biochemical test and slide agglutination test with *Salmonella*-specific antisera (Denka Sieken Tokyo, Japan) [17–19]. On the basis of blood culture confirmation result, we included all the available stored 202 *S*. Typhi isolated from 2004 to 2016 in this study and subjected these to whole genome sequencing analysis.

DNA extraction and whole genome sequencing

Genomic DNA was extracted using the Wizard Genomic DNA Kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. Index-tagged paired-end Illumina sequencing libraries with an insert size of 500 bp were prepared as previously described [20] and combined into pools each containing 96 uniquely tagged libraries. WGS was performed at the Wellcome Trust Sanger Institute using the Illumina Hiseq2500 platform (Illumina, San Diego, CA, USA) to generate 125 bp paired-end reads. Sequence data quality was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to remove adapter sequences and low quality reads. Illumina sequence data was submitted to the European Nucleotide Archive and a full list of accession numbers for each isolate is summarized in **S1 Table**. Sequence data from 88 Bangladeshi *S*. Typhi from Wong *et al*. 2016 [21], and a further 528 from Tanmoy *et al*. 2018 [14], were also included for context in this study (raw sequence data are available in European Nucleotide Archive under study accessions ERP001718 and PRJEB27394, respectively).

Read alignment and SNP analysis

S. Typhi Illumina reads were mapped to the CT18 (accession no. AL513382) reference chromosome sequence [22] using the RedDog mapping pipeline (v1beta.10.3; https://github.com/katholt/reddog), specifying a 'phylogeny' run and 0% conservation, and extended run time & memory for a 'large' run, with all other parameters set to default values as specified in the python configuration script 'RedDog_config_massive.py'. RedDog uses Bowtie (v2.2.9) [23] to map reads to the reference genome, and SAMtools (v1.3.1) [24, 25] to identify SNPs that have a phred quality score above 30, and to filter out those SNPs supported by less than five reads, or with 2.5x the average read depth that represent putative repeated sequences, or those that have ambiguous base calls. For each SNP that passes these criteria in any one isolate, consensus base calls for the SNP locus were extracted from all genomes mapped, with those having phred quality scores under 20 being treated as unknown alleles and represented with a gap character. The isolates were assigned to previously defined genotypes according to an extended *S*. Typhi genotyping framework [21] using the GenoTyphi python script (genotyphi.py) on bam files output from the RedDog mapping pipeline with the—mode bam and—bam options specified (all other parameters were set to the default settings). GenoTyphi is available at https://github.com/katholt/genotyphi [6, 21]. Unique SNPs defining novel genotypes and lineages (but absent from other genotypes of *S*. Typhi) detected in the present study were manually extracted from allele tables output by RedDog (file alleles_var.csv) using R, with SNPs responsible for non-synonymous mutations (defined by RedDog output file alleles_var_cons0.0_consequences.txt) in highly conserved genes without deletions prioritized for lineage definitions.

Chromosomal SNPs with confident homozygous calls (phred score above 20) in >95% of the genomes mapped (representing a 'soft' core genome) were concatenated to form an alignment of alleles at 6,089 variant sites using the RedDog python script parseSNPtable.py with parameters -m cons,aln and -c 0.95 and SNPs called in prophage regions and repetitive sequences (354 kb; ~7.4% of bases in the CT18 reference chromosome, as defined previously, and provide in **S2 Table**) [21] were excluded using the -x and -m filter parameters. Further SNPs occurring in recombinant regions as detected by Gubbins (v2.3.2) [26] specifying a maximum of 10 iterations (—iterations 10) and the recombination convergence mode (—converge_method recombination) were excluded resulting in a final alignment of 4,395 chromosomal SNPs out of a total alignment length of 4,462,203 bp for 818 Bangladeshi isolates. SNP alleles from *S*. Paratyphi A AKU1_12601 (accession no: FM2000053) were included

at the RedDog mapping stage for the purpose of outgroup rooting the phylogenetic tree using FigTree (v1.4.3) (available at: http://tree.bio.ed.ac.uk/software/figtree/).

To provide global context, 1,560 additional *S*. Typhi genomes belonging to the genotypes found in Bangladesh [6, 7, 9, 14, 15, 21, 27] were subjected to both SNP calling, recombination filtering, and genotyping as described above, resulting in an alignment of 14,852 chromosomal SNPs.

SNP distances were calculated using the dna.dist() function in the Analysis of Phylogenetics and Evolution (ape) R package [28]. Shannon diversity was calculated using the diversity() function in the R package vegan [29]. Base R function chisq.test() was used to conduct a Chi-Squared test of age groups and *S*. Typhi genotypes [30].

Phylogenetic analysis

Maximum likelihood (ML) phylogenetic trees were inferred from the aforementioned chromosomal SNP alignments using RAxML (v8.2.8) [31] (command raxmlHPC-PTHREADS-SSE3). A generalized time-reversible model and a Gamma distribution was used to model site-specific rate variation (GTR+ Γ substitution model; GTRGAMMA in RAxML) with 100 bootstrap pseudoreplicates used to assess branch support for the ML phylogeny. RAxML runs were repeated 10 times and likelihood scores were compared across runs to confirm that all runs had converged on a similar likelihood; we selected the single tree with the highest likelihood score as the best tree. The resulting phylogenies were visualized and annotated using Microreact [32] and the R package ggtree [33] using functions ggtree() and gheatmap().

Antimicrobial resistance gene identification and plasmid replicon analysis

The mapping based allele typer SRST2 [34] was used in conjunction with the ARGannot [35] and PlasmidFinder [36] databases supplied with SRST2 to detect acquired AMR genes and plasmid replicons, respectively using the slurm_srst2.py python script with the—gene_db parameter specified. Where detected, isolates possessing IncHI1 plasmids were subjected to plasmid multilocus sequence typing (PMLST) [37] using SRST2 [34] (slurm_srst2.py python script specifying the—mlst_db and—mlst_definitions parameters specified). Where AMR genes were present without evidence of a known resistance plasmid replicon based on SRST2 output, raw sequence reads were assembled *de novo* using Unicycler (v0.3.0b) [38] with default parameters and examined visually using the assembly graph viewer Bandage (0.8.1) [39] to inspect the composition and insertion sites of resistance-associated composite transposons using the create/view blast search option against the integrated MDR loci from previously reported *S*. Typhi reference genomes ERL12960 (accession number ERR343327) and 12148 (accession number ERR343322) [9]. ISMapper (v2.0) [40] python script ismap.py was run with default parameters to screen for insertion sites the transposases IS*1* (accession number J01730) and IS*26* (accession number X00011), relative to the CT18 reference sequence in order to identify the location of these in the chromosome of each Bangladeshi *S*. Typhi genome. Point mutations in the QRDR of genes *gyrA* and *parC* associated with reduced susceptibility to fluoroquinolones [15] were determined using GenoTyphi [6, 21] from RedDog output bam files as specified above (**Read alignment and SNP analysis** section).

Results

Population structure of *S.* Typhi in Bangladesh

A total collection of 202 Bangladeshi *S.* Typhi isolated from three different urban areas of Dhaka between 2004 and 2016 (**S1 Table**) were subjected to Illumina whole genome sequencing, and were analyzed together with 616 additional Bangladeshi *S.* Typhi whole genome sequences from previous studies (isolated between 1998 and 2016, see **S3 Table**) to provide a robust genomic context. The *S.* Typhi genomes were subjected to phylogenetic and GenoTyphi analysis, revealing that the pathogen population structure in Bangladesh is diverse with 17 distinct genotypes as summarized in **S4 Table** and **Fig 1**. Eight *S.* Typhi subclades (genotypes 2.0.1, 2.1.7, 2.3.3, 3.0.1, 3.1.2, 3.2.2, 3.3.0, 4.3.1) and two undifferentiated isolates of major lineage 2 (genotype 2.0.0) were observed among our 202 samples collected in urban Dhaka between 2004 and 2016 (**Table 1**). The majority of our isolates (n = 83, 41.1%) belonged to genotype 4.3.1 (haplotype H58), which has rapidly expanded in South Asia from the early 1990s. Previously defined major sublineages of H58 (lineage I and lineage II) [9] were present among our 202 isolates, with H58 lineage I (genotype 4.3.1.1) isolates appearing dominant (n = 63, 31.2%). Only a single H58 lineage II (genotype 4.3.1.2) isolate (n = 1, 0.50%) was present; this is somewhat surprising as this lineage is common in neighboring countries Nepal and India [6, 15]. In addition to H58 lineage I and II isolates, 19 genetically undifferentiated H58 isolates (genotype 4.3.1 according to the framework of Wong *et al* 2016) were also observed. Further analysis of these revealed close clustering with 119 "H58 lineage Bd" isolates reported by Tanmoy *et al.* 2018 [14], forming a monophyletic sister clade to H58 lineages I and II that we herein define as 4.3.1.3 (see **Fig 1**). The 19 4.3.1.3 isolates from our collection possess the characteristic SNPs STY0054-C711A (position 561056 in CT18) and STY2973-A1421C (position 2849843 in CT18) that Tanmoy *et al* used to define this lineage [14]. We have therefore added the SNP STY0054-C711A to the GenoTyphi script (available at: http://github.com/katholt/genotyphi), to facilitate the detection of this lineage (genotype 4.3.1.3) in future WGS surveillance studies.

Genotypes 2.3.3 (n = 30, 14.85%), 3.2.2 (n = 44, 21.80%), 3.3.0 (n = 28, 13.86%) were also relatively common in urban Dhaka. Of the 818 Bangladeshi *S.* Typhi analysed, 119 (14.5%) formed a monophyletic sublineage within clade 3.3 that did not belong to any of the previously defined subclades (i.e. assigned genotype 3.3.0 in the existing scheme) and were closely related to other clade 3.3 isolates found in Nepal (median distance of ~70 SNPs) [6, 15] (see **S1 Fig** and interactive phylogeny available at http://microreact.org/project/5GzJ7Umoz). We herein assign these Bangladesh and Nepal *S.* Typhi to a novel genotype, subclade 3.3.2 (labeled in **Fig 1**), which can be identified by the presence of a marker SNP STY3641-A224G (position 3498544 in CT18) that confers an amino acid mutation (Q75G) within the ST3641 encoded protein. This genotype has also been added to the GenoTyphi script to facilitate detection of subclade 3.2.2 from WGS data in future studies.

Intra-country transmission dynamics within Bangladesh

Geographical location data were available for the 202 novel *S.* Typhi, which were collected from three different sites inside urban Dhaka (icddr,b Kamalapur field site, icddr,b Mirpur field site and icddr,b Dhaka hospital Mohakhali). Detailed genotypic distribution of 202 isolates from these three study sites are shown in **Fig 2**, and an interactive version of the phylogeny and map are available online at https://microreact.org/project/sP2Uwk_DI). Our data showed that genotypes 2.3.3, 3.3.2, 3.2.2, 4.3.1.1 and 4.3.1.3 were present in all three study sites (**Fig 2A**) and were intermingled in the phylogeny (**Fig 2B**) suggesting circulation across the

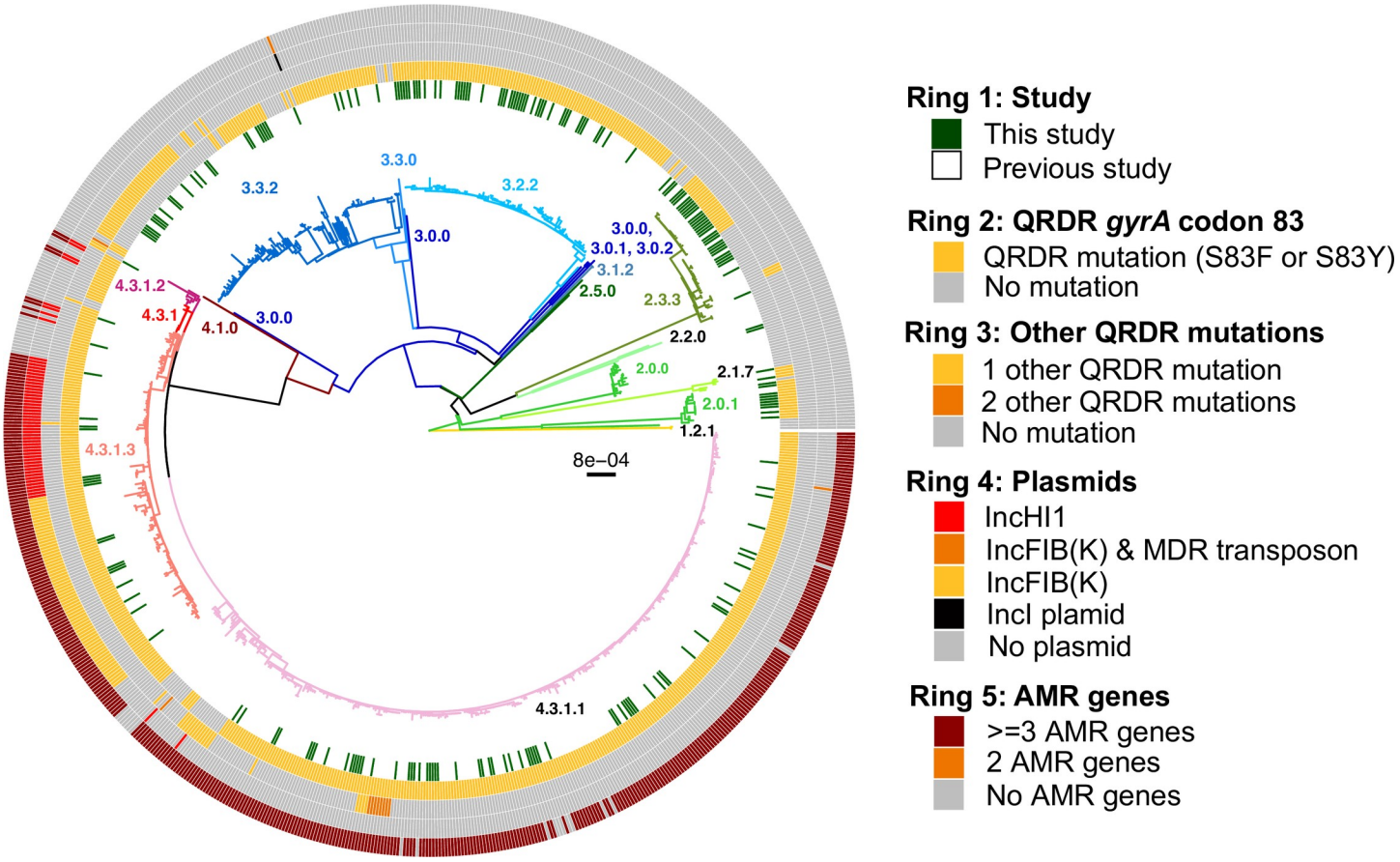

Fig 1. Urban Dhaka *S.* Typhi population structure with antimicrobial resistance. Maximum likelihood outgroup-rooted tree of 816 Bangladeshi *S.* Typhi isolates. Branch colors indicate the genotype (as labelled). Inner ring indicates the position of the 202 *S.* Typhi from this study, the second and third rings indicate the presence of QRDR mutations, fourth ring indicates plasmids and the outer fifth ring indicates AMR genes colored as per the inset legend.

https://doi.org/10.1371/journal.pntd.0008036.g001

Table 1. Genotypes present among 202 novel *S.* Typhi isolates from Dhaka, Bangladesh.

Genotype	No. of isolates (%)
2.0.0	2 (0.99%)
2.0.1	10 (4.95%)
2.1.7	2 (0.99%)
2.3.3	30 (14.85%)
3.0.1	1 (0.50%)
3.1.2	2 (0.99%)
3.2.2	44 (21.80%)
3.3.2*	28 (13.86%)
4.3.1.3*	19 (9.41%)
4.3.1.1	63 (31.19%)
4.3.1.2	1 (0.50%)

*Novel genotypes described in this manuscript.

https://doi.org/10.1371/journal.pntd.0008036.t001

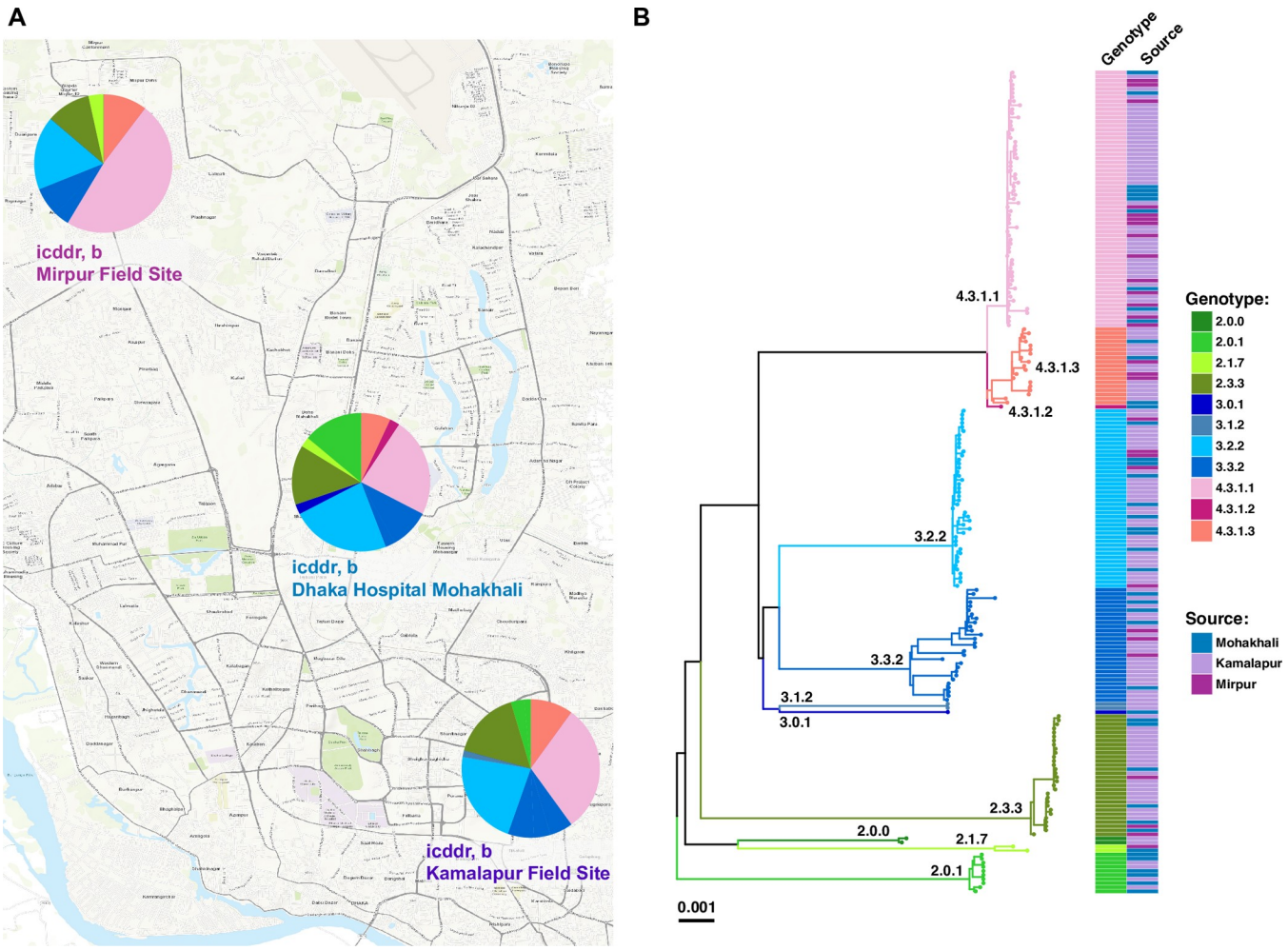

Fig 2. Spatial analysis of 202 *S.* Typhi from this study. (A) Map of Dhaka indicating the prevalence of *S.* Typhi genotypes present at each of the three study sites of icddr,b. Map sourced from LandsatLook Viewer: https://landsatlook.usgs.gov/ [41]. Online interactive phylogeny available via Microreact [32] at: https://microreact.org/project/sP2Uwk_DI. **(B)** Maximum likelihood tree of 202 urban Dhaka *S.* Typhi isolates from the three study sites. Branch colors indicate the genotype (as labelled) and the colored heatmap (on the right) shows, for each isolate, its genotype, and source of the isolates colored as per the inset legend.

https://doi.org/10.1371/journal.pntd.0008036.g002

city. Two genotypes, 2.0.1 and 2.1.7, were restricted to two of the three study locations (Kamalapur and icddr,b hospital, and Mirpur and icddr,b hospital, respectively); genotype 2.0.0, 3.1.2 and genotype 3.0.1, 4.3.1.2 were found only in Kamalapur and icddr,b hospital, respectively.

Information on patient ages were available for 185 (91.6%) of the 202 *S.* Typhi, facilitating stratification by age groups (young children under 5 years of age, older children from 5 to 15 years of age, and adults above 15 years of age). This stratification did not reveal any significant differences ($p = 0.344$ using Chi-squared test) as all age groups appeared to be infected with a diverse range of *S.* Typhi genotypes (**Fig 3**).

Inter-country transmission patterns and population structure of urban Dhaka *S.* Typhi isolates

To provide a global context for the 818 Bangladeshi *S.* Typhi and to better understand inter-country transmission patterns, we constructed a global phylogeny including an additional 1,560 *S.* Typhi from over 30 countries [6, 7, 14, 15, 21, 27] from the global WGS collection,

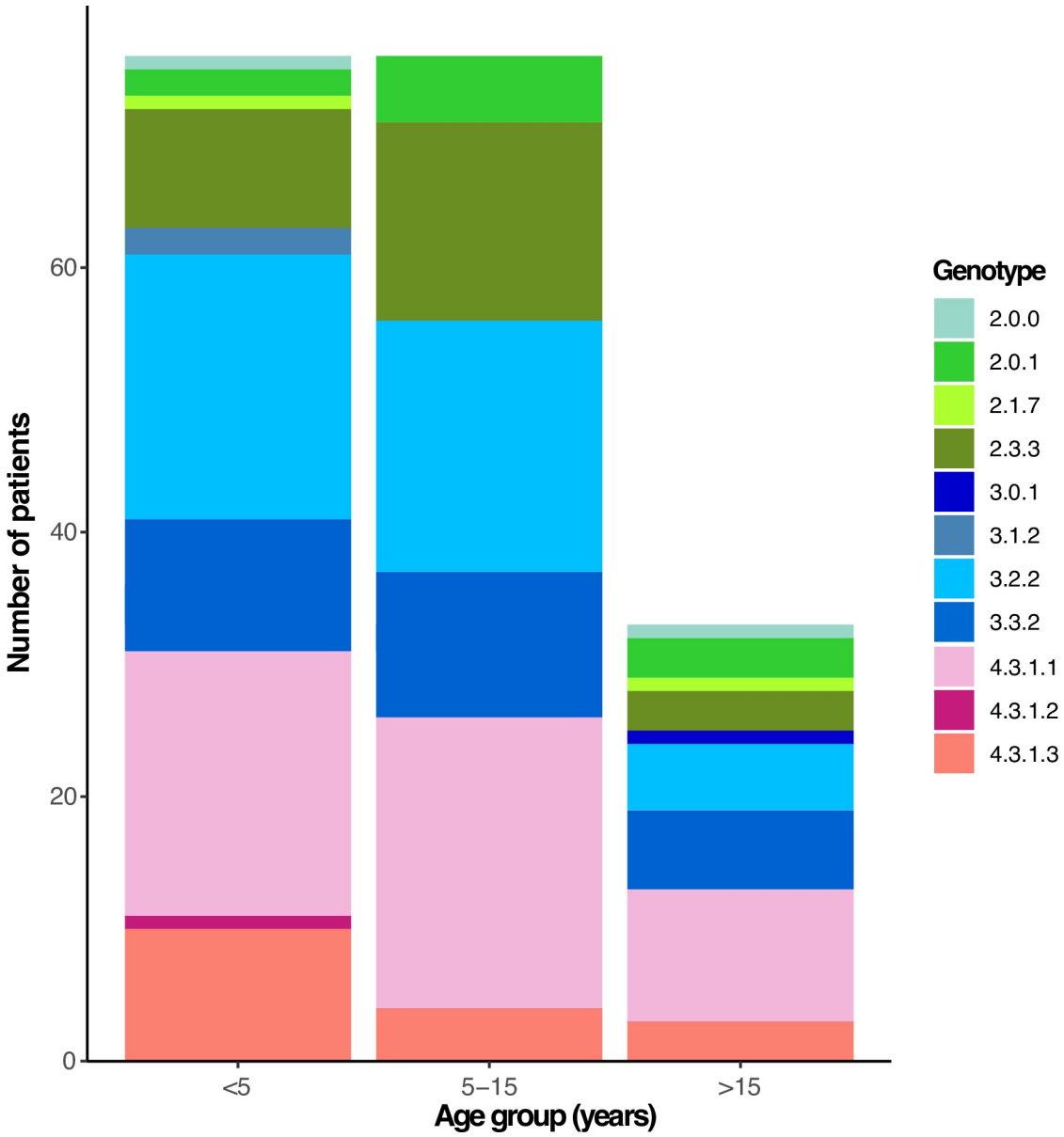

Fig 3. Urban Dhaka genotypes observed in pediatric and adult patients. Individual *S*. Typhi genotypes are colored as described in the inset legend.

https://doi.org/10.1371/journal.pntd.0008036.g003

belonging to the 17 genotypes that were detected in Bangladesh (Fig 4, interactive phylogeny available at https://microreact.org/project/5GzJ7Umoz). A single South African *S*. Typhi isolated in 2010 was intermingled among the genotype 4.3.1.3 isolates, ~2 SNPs away from its closest Bangladesh relative, suggesting that the H58 lineage Bd has been transferred from South Asia to Africa on at least one occasion. Predominantly, the Bangladeshi H58 and non-H58 *S*. Typhi were related to isolates from India, Pakistan, and Nepal (see Fig 4 and interactive phylogeny at https://microreact.org/project/5GzJ7Umoz), suggesting regional circulation of these genotypes throughout South Asia. Notably, we found that Bangladesh isolates formed several unique monophyletic lineages within this tree, consistent with local establishment and

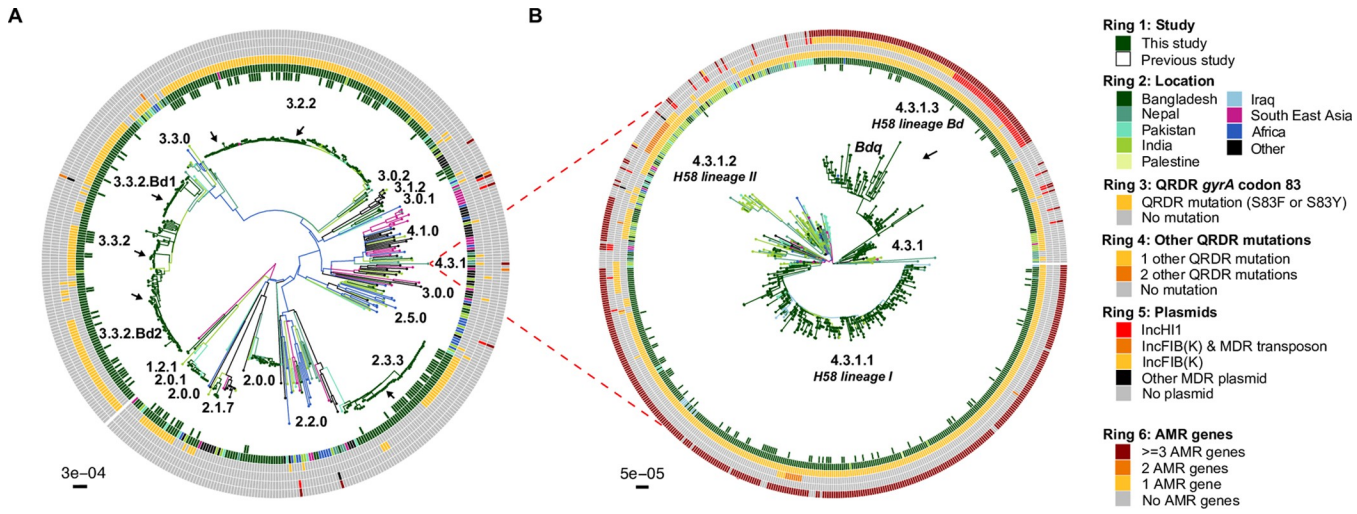

Fig 4. Global population structure of Bangladesh *S.* Typhi. (A) Global population structure of Non-H58 (4.3.1) Bangladesh genotypes. (B) Global population structure of H58. Branch colors indicate country/region of origin, as do the inner two rings. The third ring indicates mutations in codon 83 of gene *gyrA*, fourth ring indicate the number of additional QRDR mutations. The fourth ring indicates the presence of any plasmids and the sixth ring indicates the presence of AMR genes. All branches and rings are colored as per the inset legend. Arrows indicate localized lineages of Bangladeshi isolates.

https://doi.org/10.1371/journal.pntd.0008036.g004

ongoing clonal transmission within Bangladesh; e.g. 2.3.3, 3.2.2, 3.3.2 (Fig 4A) and 4.3.1.1, 4.3.1.3 (Fig 4B).

Antimicrobial resistance and plasmid replicons in *S.* Typhi in urban Dhaka

To better understand the AMR burden among *S.* Typhi in urban Dhaka, we subjected our 202 isolates (Table 2) and the additional 616 *S.* Typhi from previous studies [9, 14, 21] (S5 Table) to screening for both genes and mutations associated with AMR. Only 17 of our 202 *S.* Typhi (9.42%) lacked any known molecular determinants of AMR and were thus predicted to be fully susceptible to antibiotics (Fig 1 and Table 2).

Table 2. Genetic determinants of antimicrobial resistance in 202 *S.* Typhi isolates from urban Dhaka, Bangladesh.

Resistance patterns	H58 isolates (n = 83)	Non H58 isolates (n = 119)	Total isolates (n = 202)
Acquired AMR genes / QRDR mutations	**74 (89.16%)**	**0 (0.0%)**	**74 (36.63%)**
bla_{TEM-1}, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB* / *gyrA*-S83F	47 (56.63%)	0 (0.00%)	47 (23.27%)
bla_{TEM-1}, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB*, *tetB* / *gyrA*-S83Y	8 (9.64%)	0 (0.00%)	8 (3.96%)
bla_{TEM-1}, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB* / *gyraccessA*-S83F, *parC*-E84K	1 (1.20%)	0 (0.00%)	1 (0.50%)
bla_{TEM-1}, *catA1*, *dfrA7*, *sul1* / *gyrA*-S83F	1 (1.20%)	0 (0.00%)	1 (0.50%)
catA1, *dfrA7*, *sul1* / *gyrA*-S83F	8 (9.64%)	0 (0.00%)	8 (3.96%)
bla_{TEM-1}, *sul2*, *qnrS*, *tetA* / *gyrA*-S83Y	9 (10.84%)	0 (0.00%)	9 (4.46%)
QRDR mutations only	**83 (100%)**	**102 (85.71%)**	**185 (91.6%)**
gyrA-S83F	62 (74.70%)	85 (83.33%)	147 (72.77%)
gyrA-S83Y	19 (22.90%)	3 (2.94%)	22 (10.89%)
gyrA-D87G	1 (1.20%)	0 (0.00%)	1 (0.50%)
gyrA-D87N	0 (0.00%)	12 (11.76%)	12 (5.94%)
gyrA-D87Y	0 (0.00%)	2 (1.96%)	2 (0.99%)
gyrA-S83F, *parC*-E84K	1 (1.20%)	0 (0.00%)	1 (0.50%)
Susceptible to all antibiotics	**0 (0.00%)**	**17 (16.67%)**	**17 (8.42%)**

https://doi.org/10.1371/journal.pntd.0008036.t002

Many of our H58 isolates (n = 57, 68.67%) were predicted to be MDR, carrying genes associated with resistance to the first-line drugs chloramphenicol, trimethoprim-sulfomethoxazole and ampicillin. The majority of these were H58 lineage I isolates (genotype 4.3.1.1, n = 48, 64.86%) carrying genes *catA1*, *dfrA7*, *sul1*, *sul2*, $bla_{\text{TEM-1}}$ and *strAB* (S2 Fig) in a composite transposon (Fig 5) conferring resistance to all three first-line drugs and also streptomycin. These isolates lacked the IncHI1 plasmid, suggesting that the AMR genes have been integrated into the chromosome. A small proportion of genotype 4.3.1.1 *S*. Typhi (n = 8, 10.8%) carried an alternative form of the typical transposon encoding just three AMR genes (*catA1*, *dfrA7*, *sul1*; see Fig 5). Examination of the assembly graphs and nucleotide sequence comparisons [42] of the genomes carrying the 3-gene vs 7-gene locus revealed that integration of both transposons were mediated by IS*1* transposition associated with an 8 bp target site duplication (TACGCTTT; see Fig 5). However as IS*1* was present at multiple locations in the chromosome sequences of these isolates, we were unable to resolve the precise location of the MDR integration site, and insertions at either of two previously reported chromosomal integration sites (near *cyaA* or within *yidA* [9]) were equally possible in our isolates.

Of the 19 H58 lineage Bd isolates detected in our collection of 202 isolates, two different plasmid mediated AMR patterns emerged (S2 Fig). The first pattern (n = 8) was characterized by the presence of the IncHI1 plasmid (plasmid sequence type, PST6) [9] carrying eight AMR genes (bla_{TEM-1}, *catA1*, *dfrA7*, *sul1*, *sul2*, *strAB*, and *tetB*) conferring resistance to the first line drugs plus streptomycin and tetracyclines. The second pattern (n = 9) was characterized by the presence of an IncFIB(K) plasmid carrying the AMR genes (bla_{TEM-1}, *sul2*, *tetA*) conferring resistance to ampicillin, sulfonamides, tetracyclines, and also *qnrS* together with the *gyrA*-S83Y mutation confers resistance to fluoroquinolones. These isolates were intermingled with IncFIB (K)-carrying isolates described by Tanmoy *et al* 2018 (which they termed "sublineage Bdq"). This IncFIB(K)-carrying cluster appears to have emerged from the main 4.3.1.3 group that typically carries the IncHI1 plasmid (separated by a median of ~11 SNPs), but the IncHI1 MDR plasmid has been replaced in this group by the IncFIB(K) fluoroquinolone resistance plasmid (see S2 Fig). The IncFIB(K)-containing Bdq cluster appears to have undergone a clonal expansion in Bangladesh, but has not replaced the IncHI1 form of 4.3.1.3 in urban Dhaka (see below).

Overall, non-synonymous mutations in the QRDR of genes *gyrA* and *parC* associated with reduced susceptibility to fluoroquinolones (FQ) were common among our 202 *S*. Typhi isolates (n = 185, 91.6%) harboring at least one QRDR mutation (**Table 2**). Unlike the acquisition of MDR genes, the QRDR mutations (mainly in gene *gyrA* at codon 83) were common in both non-H58 (n = 102, 85.71%) as well as H58 isolates (n = 83, 100%). Examination of genotype 3.3.2 revealed two distinct monophyletic lineages of Bangladeshi *S*. Typhi of this genotype each with a different QRDR mutation. Here we defined these two Bangladeshi lineages as 3.3.2.Bd1 and 3.3.2.Bd2, carrying the *gyrA*-S83F and *gyrA*-D87N mutations, respectively (**Fig 4A**). Markers for these two lineages (SNPs STY2588-G378A, position 2424394 in CT18; and STY2441-G439A, position 2272144 in CT18; respectively) have been added to the GenoTyphi script to facilitate their detection. No QRDR triple mutants were observed among our collection; however, a single double mutant *S*. Typhi of genotype 4.3.1.1 was identified carrying both *gyrA*-S83F and *parC*-E84K mutations. Hence, the only isolates we predict to be ciprofloxacin resistant are the IncFIB(K) group carrying *qnrS* and *gyrA*-S83Y.

Temporal trends in genotypic distribution and AMR pattern in urban Dhaka

We examined the genotypic distribution and AMR patterns over time during our study period of 2004 to 2016; note our sampling includes all available blood culture positive *S*. Typhi from

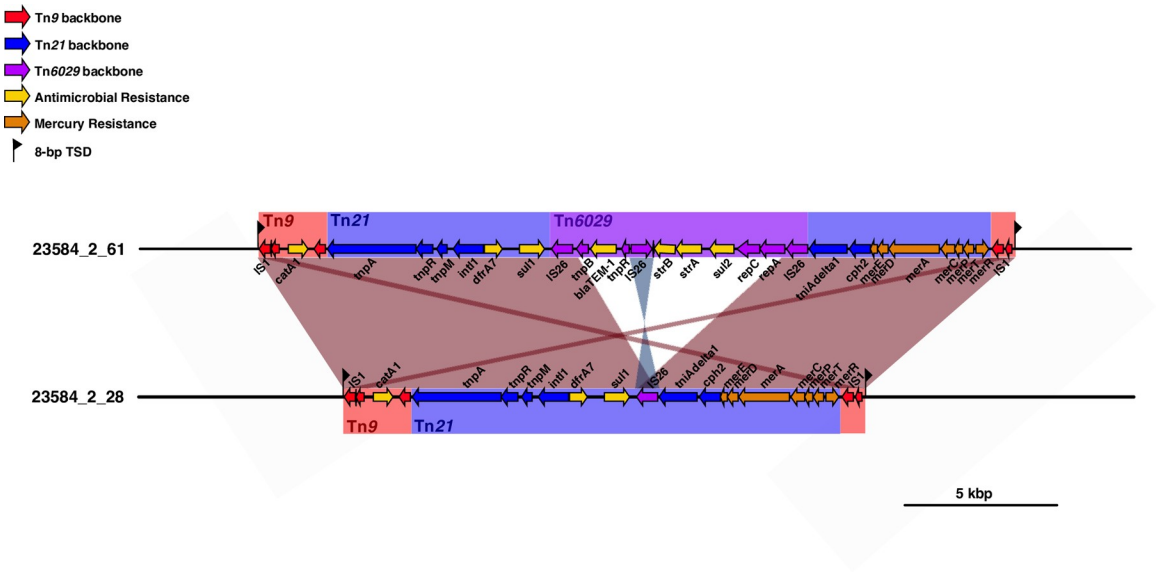

Fig 5. Insertion sites of transposons observed in *S*. Typhi from urban Dhaka, Bangladesh. Genes and transposons are indicated according to the inset legend. TSD indicates target site duplication, and *Tn* indicates transposons.

https://doi.org/10.1371/journal.pntd.0008036.g005

the three sites in urban Dhaka. Prior to 2011, H58 lineages (4.3.1.1 and 4.3.1.3) dominated the population across the three sites sampled (median 66.67% per year; see Fig 6A). During this time most isolates were MDR (median 55.56% per year, see Fig 6B). After this period, there appears to be a relative decrease in the frequency of H58 genotypes (4.3.1.1, 4.3.1.2, 4.3.1.3; median 28.03% per year in 2011–2016) and an increase in non-H58 genotypes 2.3.3, 3.2.2,

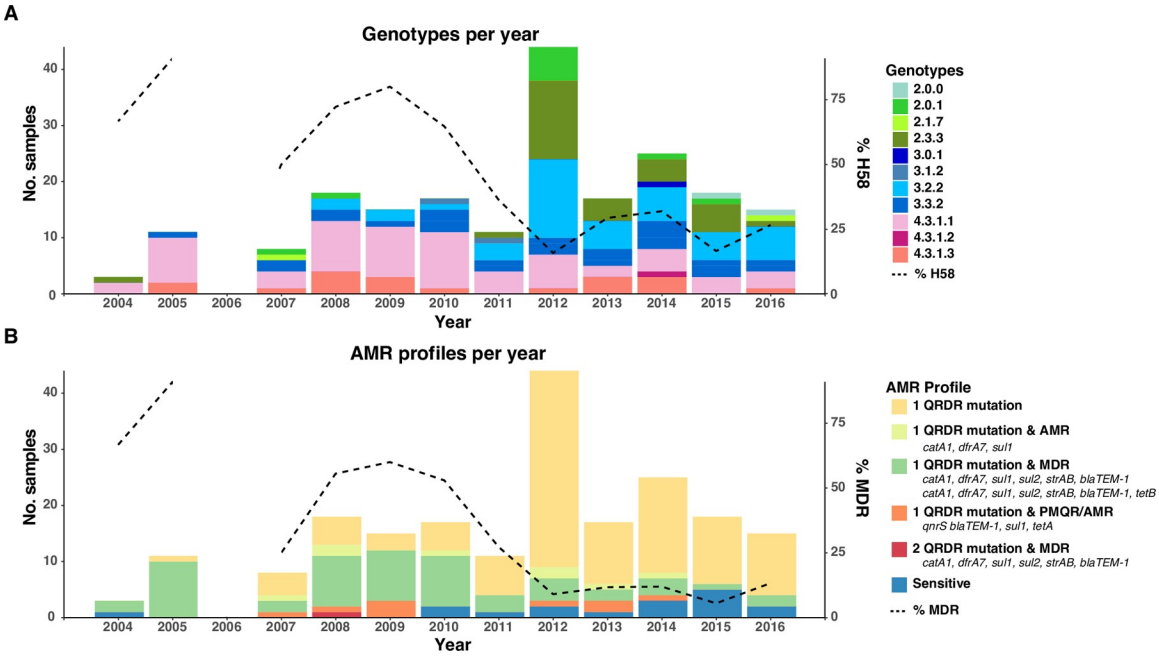

Fig 6. Timeline of urban Dhaka *S*. Typhi genotypes and AMR profiles from 2004–2016. (A) Genotypes observed per annum. Overlaid line indicates the percentage of H58 isolates per year. (B) AMR profiles observed per annum. Overlaid line indicates the percentage of MDR isolates per year. Genotypes and AMR profiles colored as per inset legend.

https://doi.org/10.1371/journal.pntd.0008036.g006

3.3.2 (from median 22.2% per year in 2004–2010 to 65.2% per year in 2011–2016; overall diversity increasing from 1.34 Shannon diversity in 2004–2010 to 1.83 in 2011–2016). The decline in H58 *S*. Typhi coincided with a decline in MDR (median 11.88% per year in 2011–2016), with the most common profile being non-MDR with a single QRDR mutation (*gyrA*-S83F being the most prevalent). Notably, the IncFIB(K)/*qnrS* lineage Bdq remained at low frequency throughout the study period (detected from 2007 to 2014, median 4.78% per year).

Discussion

We show here that the population structure of *S*. Typhi in Bangladesh is diverse, harboring 17 distinct genotypes, with 9 genotypes circulating within urban Dhaka between 2004–2016 (**Fig 1**). There was little evidence of any local geographic restriction of *S*. Typhi genotypes within the city of Dhaka (**Fig 2**) and no obvious stratification of *S*. Typhi genotypes by patient age (**Fig 3**), consistent with what has been observed in other South Asian settings [6, 15, 43].

S. Typhi circulating in Bangladesh are closely related to isolates from neighboring India, Pakistan and Nepal suggesting circulation throughout South Asia. However, the formation of multiple localized lineages indicates establishment and ongoing local transmission of multiple genotypes in parallel within Bangladesh; and warranted definition of novel Bangladesh-specific subclades for future tracking via WGS surveillance. Firstly, the most common genotype in our collection was H58 lineage I (31.2%) and H58 lineage II was rare (0.5%), in contrast to neighboring India and Nepal where lineage II are highly prevalent and lineage I is rarely reported [6, 9, 15]. Secondly, this distinction between Bangladeshi and Indian/Nepali pathogen populations was also evident for the newly defined H58 lineage Bd (4.3.1.3) [14], which so far has been almost exclusively found in Bangladesh (the exceptions being singleton isolates reported in Nepal and South Africa). Thirdly, we identified a novel subclade of *S*. Typhi, genotype 3.3.2, which included a Bangladesh-specific monophyletic group with relatives in Nepal, that we further divided into 3.3.2.Bd1 and 3.3.2.Bd2 based on distinct QRDR mutations conferring reduced susceptibility to fluoroquinolones **(Fig 4A**). These novel Bangladesh-associated lineages (4.3.1.3, 3.3.2.Bd1, 3.3.2.Bd2) have been added to the GenoTyphi genotyping scheme, which will facilitate their detection and tracking in future surveillance efforts, and over time will reveal whether they remain localized to Bangladesh or being to disseminate through Asia and Africa as has been observed for H58 lineages I and II [6, 13, 27, 44].

The sustained, very high frequency of *S*. Typhi carrying mutations associated with reduced susceptibility to fluoroquinolones that we detected in this study (>66% per year; **see Fig 6B)** is likely the result of an increase in over-the-counter sale (without prescription) of this antibiotic class over the last decade [45]. This, along with a decrease in MDR is similar to reports from both India and Nepal [6, 9]. However, while the prevalence of reduced susceptibility was very high, we found limited evidence of evolution towards full resistance, with the *qnrS*-positive clade (associated with ciprofloxacin MIC of 4 μg/mL, [14]) remaining at low frequency throughout the study period. A H58 lineage I (4.3.1.1) QRDR triple mutant has been previously reported in Bangladesh [14], however, this was not observed among our collection from urban Dhaka; we only detected a single QRDR double mutant in H58 lineage I in 2008 (see **Table 2** and **Fig 6**).

Notably, the reduced prevalence of MDR coincided with a reduction in H58 (4.3.1) isolates across our 3 study sites in urban Dhaka and a significant diversification in the pathogen population, particularly driven by increased prevalence of QRDR single-mutant *S*. Typhi genotypes 2.3.3, 3.2.2 and 3.3.2. This unexpected change in population structure that cannot be explained by selection for AMR suggests unknown selective pressures may be influencing the pathogen population in urban Dhaka, and highlights the need for ongoing genomic surveillance.

Further, while MDR was less frequent after 2010, the presence of multiple forms of the MDR chromosomal insertion is highly concerning, as such insertions may facilitate more stable transmission of the MDR phenotype [6, 9]. Similarly, the persistent presence of plasmid-mediated quinolone resistance (PMQR) via an IncFIB(K) plasmid carrying a *qnrS1* gene is concerning, despite the relatively low frequency (4.5%) at which it is observed currently. Fortunately, we did not detect any ceftriaxone resistant isolates in this study which is now one of the drugs of choice for typhoid treatment in this region. The first case of carbapenem resistance was reported in a *Salmonella enterica* serovar Typhimurium isolate which was also a co-producer of extended-spectrum beta-lactamase (SHV-12), suggesting the occurrence of extended-spectrum beta-lactamase (ESBL) resistance is very rare among typhoidal *Salmonellae* [46]. Moreover, only a single Bangladesh *S*. Typhi isolate of genotype 3.3.2 from 2000 has been reported to have acquired the extended-spectrum beta-lactamases ESBL resistance gene $bla_{CTX\text{-}M\text{-}15}$ **(S2 Fig)** [14]. We detected no evidence of the ceftriaxone resistant H58 Pakistan Extensively Drug Resistant (XDR; genotype 4.3.1.1.P1) isolates [7], perhaps suggesting that ESBL resistance is not widespread in *S*. Typhi isolated from Bangladesh at the time of this study.

Our study has some limitations as all of our 202 isolates analysed here were from study sites in urban Dhaka, and subsequently may not be representative of the overall the population structure and antimicrobial resistance patterns of typhoid in Bangladesh. However, our 202 strains are intermingled with those collected from previous studies of typhoid in both urban Dhaka, and Bangladesh (**Fig 1**) [14, 21, 47] suggesting that our opportunistic sampling is reasonably representative of the population structure in this region.

Conclusion

This study demonstrates the importance of molecular based surveillance studies in endemic regions, especially in urban Dhaka in Bangladesh, where the disease burden is high and many different AMR phenotypes were observed. The change in both population structure and AMR patterns over twelve years (2004 to 2012) in urban Dhaka shows increased prevalence of populations with reduced susceptibility fluoroquinolones, emphasizing the ongoing evolution of AMR in this setting as well as the urgent need for WGS based surveillance in Bangladesh to inform both treatment guidelines and control strategies.

Supporting information

S1 Fig. Population structure and country of origin of genotype 3.3.2 isolates. Maximum likelihood phylogeny of genotype 3.3.2 isolates. Colored bar indicates country of origin as per the inset legend.
(TIF)

S2 Fig. Bangladesh *S*. Typhi population structure showing acquired AMR genes, and plasmids. Branches are colored by genotype as labelled, the heatmap the molecular determinants of antimicrobial resistance and the presence of plasmids colored as per the inset legend.
(TIF)

S1 Table. Data of 202 S. Typhi isolates from 3 different study sites in Dhaka, Bangladesh between 2004–2016.
(XLSX)

S2 Table. Excluded repetitive and phage regions in CT18 (AL513382) reference genome.
(XLSX)

S3 Table. Data of 818 S. Typhi isolates from Bangladesh including previous studies between 1998–2016.
(XLSX)

S4 Table. Genotyping results for the 818 S. Typhi isolates from Bangladesh.
(DOCX)

S5 Table. Genetic determinants of antimicrobial resistance in S. Typhi isolates from Bangladesh.
(DOCX)

Acknowledgments

We acknowledge the support of dedicated field and laboratory workers at the icddr,b involved in this study. We would like to thank Doli Goswami, Md. Lokman Hossain and Abdullah Brooks for their contribution in Kamalapur field study. We would also like to thank the members of Pathogen Informatics Team and core sequencing teams at the Wellcome Trust Sanger Institute.

Author Contributions

Conceptualization: Sadia Isfat Ara Rahman, Zoe A. Dyson, Gordon Dougan, Firdausi Qadri.

Data curation: Farhana Khanam.

Formal analysis: Sadia Isfat Ara Rahman, Zoe A. Dyson, Elizabeth J. Klemm, Kathryn E. Holt.

Funding acquisition: Farhana Khanam, Gordon Dougan, Firdausi Qadri.

Methodology: Sadia Isfat Ara Rahman.

Resources: Farhana Khanam.

Supervision: Emran Kabir Chowdhury, Gordon Dougan, Firdausi Qadri.

Visualization: Sadia Isfat Ara Rahman, Zoe A. Dyson, Elizabeth J. Klemm.

Writing – original draft: Sadia Isfat Ara Rahman, Zoe A. Dyson, Gordon Dougan, Firdausi Qadri.

Writing – review & editing: Sadia Isfat Ara Rahman, Zoe A. Dyson, Elizabeth J. Klemm, Farhana Khanam, Kathryn E. Holt, Emran Kabir Chowdhury, Gordon Dougan, Firdausi Qadri.

References

1. Mogasale V, Maskery B, Ochiai RL, Lee JS, Mogasale VV, Ramani E, et al. Burden of typhoid fever in low-income and middle-income countries: a systematic, literature-based update with risk-factor adjustment. Lancet Glob Health. 2014; 2(10):e570–e80. https://doi.org/10.1016/S2214-109X(14)70301-8 PMID: 25304633
2. Crump JA, Mintz ED. Global trends in typhoid and paratyphoid fever. Clin Infect Dis. 2010; 50(2):241–6. https://doi.org/10.1086/649541 PMID: 20014951
3. Naheed A, Ram PK, Brooks WA, Hossain MA, Parsons MB, Talukder KA, et al. Burden of typhoid and paratyphoid fever in a densely populated urban community, Dhaka, Bangladesh. Int J Infect Dis. 2010; 14:e93–e9. https://doi.org/10.1016/j.ijid.2009.11.023 PMID: 20236850
4. Theiss-Nyland K, Qadri F, Colin-Jones R, Zaman K, Khanam F, Liu X, et al. Assessing the Impact of a Vi-polysaccharide Conjugate Vaccine in Preventing Typhoid Infection Among Bangladeshi Children: A Protocol for a Phase IIIb Trial. Clin Infect Dis. 2019; 68(Supplement_2):S74–S82.

5. Stanaway JD, Reiner RC, Blacker BF, Goldberg EM, Khalil IA, Troeger CE, 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(4):369–81. https://doi.org/10.1016/S1473-3099(18)30685-6 PMID: 30792131
6. 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(4):e0006408. https://doi.org/10.1371/journal.pntd.0006408
7. 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(1):e00105–18. https://doi.org/10.1128/mBio.00105-18
8. Parry CM, Hien TT, Dougan G, White NJ, Farrar JJ. Typhoid Fever. N Engl J Med. 2002; 347 (22):1770–82. https://doi.org/10.1056/NEJMra020201 PMID: 12456854.
9. 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(6):632. https://doi.org/10.1038/ng.3281 PMID: 25961941
10. Levine MM, Simon R. The Gathering Storm: Is Untreatable Typhoid Fever on the Way? mBio. 2018; 9 (2). Epub 2018/03/22. https://doi.org/10.1128/mBio.00482-18 PMID: 29559573; PubMed Central PMCID: PMC5874914.
11. Godbole GS, Day MR, Murthy S, Chattaway MA, Nair S. First Report of CTX-M-15 Salmonella Typhi From England. Clin Infect Dis. 2018; 66(12):1976–7. Epub 2018/02/23. https://doi.org/10.1093/cid/ciy032 PMID: 29471386.
12. 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(7):e1245. https://doi.org/10.1371/journal.pntd.0001245 PMID: 21811646
13. Park SE, Pham DT, Boinett C, Wong VK, Pak GD, Panzner U, et al. The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. Nat Commun. 2018; 9(1):5094. https://doi.org/10.1038/s41467-018-07370-z PMID: 30504848
14. Tanmoy AM, Westeel E, De Bruyne K, Goris J, Rajoharison A, Sajib MS, et al. Salmonella enterica Serovar Typhi in Bangladesh: Exploration of Genomic Diversity and Antimicrobial Resistance. mBio. 2018; 9(6):e02112–18. https://doi.org/10.1128/mBio.02112-18 PMID: 30425150
15. Thanh DP, Karkey A, Dongol S, Thi NH, Thompson CN, Rabaa MA, et al. A novel ciprofloxacin-resistant subclade of H58 Salmonella Typhi is associated with fluoroquinolone treatment failure. eLife. 2016; 5: e14003. https://doi.org/10.7554/eLife.14003 PMID: 26974227
16. World Population Review 2019 [cited 2019 15th April]. Available from: http://worldpopulationreview.com/world-cities/dhaka-population/
17. Khanam F, Sheikh A, Sayeed MA, Bhuiyan MS, Choudhury FK, Salma U, et al. Evaluation of a typhoid/paratyphoid diagnostic assay (TPTest) detecting anti-Salmonella IgA in secretions of peripheral blood lymphocytes in patients in Dhaka, Bangladesh. PLoS Negl Trop Dis. 2013; 7(7):e2316. https://doi.org/10.1371/journal.pntd.0002316 PMID: 23951368
18. Sheikh A, Bhuiyan MS, Khanam F, Chowdhury F, Saha A, Ahmed D, et al. Salmonella enterica serovar Typhi-specific immunoglobulin A antibody responses in plasma and antibody in lymphocyte supernatant specimens in Bangladeshi patients with suspected typhoid fever. Clin Vaccine Immunol. 2009; 16 (11):1587–94. https://doi.org/10.1128/CVI.00311-09 PMID: 19741090
19. Talawadekar N, Vadher P, Antani D, Kale V, Kamat S. Chloramphenicol resistant Salmonella species isolated between 1978 and 1987. J Postgrad Med. 1989; 35(2):79. PMID: 2621666
20. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. Science. 2011; 331(6016):430–4. https://doi.org/10.1126/science.1198545 PMID: 21273480
21. 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: ncomms12827.
22. Parkhill J, Dougan G, James K, Thomson N, Pickard D, Wain J, et al. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature. 2001; 413(6858):848. https://doi.org/10.1038/35101607 PMID: 11677608
23. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9(4):357. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

24. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27(21):2987–93. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

26. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2014; 43(3):e15-e.

27. Wong VK, Holt KE, Okoro C, Baker S, Pickard DJ, Marks F, et al. Molecular surveillance identifies multiple transmissions of typhoid in West Africa. PLoS Negl Trop Dis. 2016; 10(9):e0004781. https://doi.org/10.1371/journal.pntd.0004781 PMID: 27657909

28. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20(2):289–90. Epub 2004/01/22. https://doi.org/10.1093/bioinformatics/btg412 PMID: 14734327.

29. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003; 14(6):927–30.

30. Core Team R. A language and environment for statistical computing. R Found Stat Comput Vienna, Austria.

31. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90. https://doi.org/10.1093/bioinformatics/btl446 PMID: 16928733

32. Argimón S, Abudahab K, Goater RJ, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016; 2(11).

33. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017; 8(1):28–36.

34. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014; 6(11):90. https://doi.org/10.1186/s13073-014-0090-6 PMID: 25422674

35. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014; 58(1):212–20. https://doi.org/10.1128/AAC.01310-13 PMID: 24145532

36. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014; 58(7):3895–903. Epub 2014/04/30. https://doi.org/10.1128/AAC.02412-14 PMID: 24777092; PubMed Central PMCID: PMC4068535.

37. Phan M-D, Kidgell C, Nair S, Holt KE, Turner AK, Hinds J, et al. Variation in Salmonella enterica serovar typhi IncHI1 plasmids during the global spread of resistant typhoid fever. Antimicrob Agents Chemother. 2009; 53(2):716–27. https://doi.org/10.1128/AAC.00645-08 PMID: 19015365

38. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017; 13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595 PMID: 28594827

39. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015; 31(20):3350–2. https://doi.org/10.1093/bioinformatics/btv383 PMID: 26099265

40. Hawkey J, Hamidian M, Wick RR, Edwards DJ, Billman-Jacobe H, Hall RM, et al. ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. BMC Genomics. 2015; 16(1):667.

41. U.S. Geological Survey. LandsatLook Viewer [cited 2019 Jul 3]. Available from: https://landsatlook.usgs.gov/

42. Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010; 26(18):2334–5. Epub 2010/07/14. https://doi.org/10.1093/bioinformatics/btq413 PMID: 20624783; PubMed Central PMCID: PMC2935412.

43. Baker S, Holt KE, Clements AC, Karkey A, Arjyal A, Boni MF, et al. Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission. Open Biol. 2011; 1(2):110008. Epub 2012/05/31. https://doi.org/10.1098/rsob.110008 PMID: 22645647; PubMed Central PMCID: PMC3352080.

44. Feasey NA, Gaskell K, Wong V, Msefula C, Selemani G, Kumwenda S, et al. Rapid emergence of multidrug resistant, H58-lineage Salmonella typhi in Blantyre, Malawi. PLoS Negl Trop Dis. 2015; 9(4):

e0003748. Epub 2015/04/25. https://doi.org/10.1371/journal.pntd.0003748 PMID: 25909750; PubMed Central PMCID: PMC4409211.

45. Rahman M, Siddique AK, Shoma S, Rashid H, Salam MA, Ahmed QS, et al. Emergence of multidrug-resistant Salmonella enterica serotype Typhi with decreased ciprofloxacin susceptibility in Bangladesh. Epidemiol Infect. 2006; 134(2):433–8. Epub 2006/02/24. https://doi.org/10.1017/S0950268805004759 PMID: 16490150; PubMed Central PMCID: PMC2870378.

46. Banerjee K, Sekar P, Krishnan P, Wattam AR, Roy S, Hays JP, et al. Whole genome sequence analysis of NDM-1, CMY-4, and SHV-12 coproducing Salmonella enterica serovar Typhimurium isolated from a case of fatal burn wound infection. Infect Drug Resist. 2018; 11:2491–5. Epub 2018/12/18. https://doi.org/10.2147/IDR.S167027 PMID: 30555248; PubMed Central PMCID: PMC6278884.

47. Ingle DJ, Nair S, Hartman H, Ashton PM, Dyson ZA, Day M, et al. Informal genomic surveillance of regional distribution of Salmonella Typhi genotypes and antimicrobial resistance via returning travellers. PLoS Negl Trop Dis. 2019; 13(9):e0007620. Epub 2019/09/13. https://doi.org/10.1371/journal.pntd.0007620 PMID: 31513580; PubMed Central PMCID: PMC6741848.

PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Genetic diversity of *Salmonella* Paratyphi A isolated from enteric fever patients in Bangladesh from 2008 to 2018

Sadia Isfat Ara Rahman[1], To Nguyen Thi Nguyen[2], Farhana Khanam[1], Nicholas R. Thomson[3,4‡], Zoe A. Dyson[3,4,5,6], Alyce Taylor-Brown[3], Emran Kabir Chowdhury[7], Gordon Dougan[3,5‡], Stephen Baker[5,8‡], Firdausi Qadri[1‡]*

1 Infectious Diseases Division, International Centre for Diarrheal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh, **2** Oxford University Clinical Research Unit, The Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam, **3** Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom, **4** Department of Infection Biology, Faculty of Infections and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **5** Department of Medicine, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, United Kingdom, **6** Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria, Australia, **7** Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh, **8** University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, United Kingdom

‡ These authors are joint senior authors on this work.
* fqadri@icddrb.org

OPEN ACCESS

Citation: Rahman SIA, Nguyen TNT, Khanam F, Thomson NR, Dyson ZA, Taylor-Brown A, et al. (2021) Genetic diversity of *Salmonella* Paratyphi A isolated from enteric fever patients in Bangladesh from 2008 to 2018. PLoS Negl Trop Dis 15(10): e0009748. https://doi.org/10.1371/journal.pntd.0009748

Editor: Abiola Senok, Mohammed Bin Rashid University of Medicine and Health Sciences, UNITED ARAB EMIRATES

Received: March 4, 2021

Accepted: August 22, 2021

Published: October 14, 2021

Data Availability Statement: Genomic data are submitted to the European Nucleotide Archive (ENA) and accession numbers are available in S1 Table (ERS3422841 to ERS3422879, ERS3422933 to ERS3422934, ERS4218304 to ERS4218308, ERS4218310 to ERS4218330).

Funding: This work was supported by the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) and also by the

Abstract

Background

The proportion of enteric fever cases caused by *Salmonella* Paratyphi A is increasing and may increase further as we begin to introduce typhoid conjugate vaccines (TCVs). While numerous epidemiological and genomic studies have been conducted for *S*. Typhi, there are limited data describing the genomic epidemiology of *S*. Paratyphi A in especially in endemic settings, such as Bangladesh.

Principal findings

We conducted whole genome sequencing (WGS) of 67 *S*. Paratyphi A isolated between 2008 and 2018 from eight enteric disease surveillance sites across Bangladesh. We performed a detailed phylogenetic analysis of these sequence data incorporating sequences from 242 previously sequenced *S*. Paratyphi A isolates from a global collection and provided evidence of lineage migration from neighboring countries in South Asia. The data revealed that the majority of the Bangladeshi *S*. Paratyphi A isolates belonged to the dominant global lineage A (67.2%), while the remainder were either lineage C (19.4%) or F (13.4%). The population structure was relatively homogenous across the country as we did not find any significant lineage distributions between study sites inside or outside Dhaka. Our genomic data showed presence of single point mutations in *gyrA* gene either at codon 83 or 87 associated with decreased fluoroquinolone susceptibility in all Bangladeshi *S*. Paratyphi A isolates. Notably, we identified the pHCM2- like cryptic plasmid which was highly similar to *S*.

grants from the Bill and Melinda Gates Foundation (gatesfoundation.org, grant no. OPP1135223 [FQ]) (TyVac), Wellcome Trust (welcome.ac.uk, grant 206194 [FQ]), the National Institutes of Health, including the National Institute of Allergy and Infectious Diseases (niaid.nih.gov, AI106878 [FQ]); AI058935 [FQ] and AI103055 [FQ])). The icddr,b is grateful to the Governments of Bangladesh, Canada, Sweden and UK for providing core support. SB was supported by a Wellcome senior research fellowship (215515/Z/19/Z). GD was supported by NIHR BRC funding for AMR research. ZAD was supported by a grant funded by the Wellcome Trust (STRATAA; 106158/Z/14/Z), and received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement TyphiNET No 845681. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests: The authors have declared that no competing interests exist.

Typhi plasmids circulating in Bangladesh and has not been previously identified in *S*. Paratyphi A organisms.

Significance

This study demonstrates the utility of WGS to monitor the ongoing evolution of this emerging enteric pathogen. Novel insights into the genetic structure of *S*. Paratyphi A will aid the understanding of both regional and global circulation patterns of this emerging pathogen and provide a framework for future genomic surveillance studies.

Author summary

Salmonella enterica serovar Paratyphi A is an understudied cause of enteric fever which has started replacing *Salmonella* Typhi in some endemic regions. Currently, there are limited genomic epidemiology data to understand the true disease burden of disease caused by *S*. Paratyphi A in poor resource settings like Bangladesh. Our genomic data revealed that the population structure of *S*. Paratyphi A in Bangladesh comprised of only previously defined lineages A, C and F and carried single point mutation in *gyrA* gene associated with decreased fluoroquinolone susceptibility with no evidence of multi-drug resistance. Additionally, the global context revealed the clustering of organisms associated with neighboring countries in South Asia. We identified the pHCM2-like cryptic plasmid among our collection, which was not found in any *S*. Paratyphi A isolates previously. A comparative plasmid analysis showed that *S*. Paratyphi A pHCM2-like plasmid was highly similar to the plasmid found in *S*. Typhi but significantly different to a similar plasmid from non-invasive *Salmonella* Paratyphi B variant Java isolate from Bangladesh. This present study highlights the importance of sustained genomic surveillance to monitor the ongoing evolution of antimicrobial resistant determinants and plasmids in Bangladesh and transmission between different countries.

Introduction

Typhoid and paratyphoid fever are severe systemic infections caused respectively by the human-restricted bacteria *Salmonella enterica* serovar Typhi (*S*. Typhi), and the various *Salmonella* Paratyphi pathovars (*S*. Paratyphi A, B, and C). Collectively, these infections are called enteric fever, with *S*. Typhi accounting for approximately 80% of all enteric fever cases globally. Recent data suggests an increased prevalence of *S*. Paratyphi A (the dominant Paratyphi pathovar) infection in parts of Asia, which has been estimated as ~35% of cases in India and Nepal and >60% of enteric fever in China [1–4]. During a community-based study in densely populated area of Dhaka, Bangladesh, the incidence of *S*. Paratyphi A infections doubled from 0.2 per 1,000 person-years in 2001 to 0.4 episodes per 1,000 person-years in 2004 [5].

S. Paratyphi A is recognized as an emerging pathogen and is also commonly reported among travelers returning from endemic regions, as enteric fever vaccines only protect against disease caused by *S*. Typhi [3]. The key antigen of current typhoid conjugate vaccines (TCVs) is the Vi polysaccharide [6]; however, *S*. Paratyphi A lacks the Vi polysaccharide, with the focus of *S*. Paratyphi A vaccine research being the specific O-antigen. A number of promising live attenuated and conjugate *S*. Paratyphi A vaccines are in the early phase of clinical development [2,7]. However, until there is a licensed *S*. Paratyphi A vaccine, or a global improvement

in water, sanitation, and hygiene (WASH) conditions, antimicrobial therapy is the only strategy to control paratyphoid fever. Notably, in the case of *S*. Typhi, the widespread use of antimicrobials has resulted in the emergence of antimicrobial resistance (AMR), with a high prevalence of multi-drug resistance (MDR, defined as resistance to ampicillin, chloramphenicol and trimethoprim-sulfamethoxazole) associated with a self-transmissible IncHI1 plasmid, and chromosome-mediated resistance to fluoroquinolones. Correspondingly, there are less data regarding plasmid-mediated AMR in *S*. Paratyphi A, with some reports suggesting a very low prevalence of MDR but an increasing trend of fluoroquinolone resistance due to chromosomal mutations in the DNA gyrase (*gyrA*, *gyrB*) genes and the topoisomerase IV (*parC*, *parE*) genes [4,7–11]. Additionally, apart from the rising trend of fluoroquinolone resistance in *S*. Paratyphi A, there have been sporadic reports of azithromycin resistant isolates [8], and a $bla_{CTX\text{-}M\text{-}15}$ extended-spectrum beta-lactamases (ESBL) producing *S*. Paratyphi A isolate was isolated from a traveler returning to UK from Bangladesh in 2017 [12].

Data regarding the molecular epidemiology, including the population structure and transmission dynamics, of this neglected emergent pathogen is limited from endemic countries like Bangladesh. Whole genome sequencing (WGS) is an excellent tool for generating new insights into bacterial pathogens, and sequencing data will aid in the identification of specific targets for better diagnostics and designing potential vaccine candidates. Here, we exploited WGS to better understand the population of *S*. Paratyphi A isolated from patients in Bangladesh in the last 10 years until 2018. We employed this method to comprehensively assess the genomic variation, AMR determinants, and auxiliary plasmid profiles of 67 *S*. Paratyphi A isolated between 2008 and 2018 from eight enteric disease surveillance study sites in Bangladesh.

Methods

Ethics statement

Ethical approval was obtained from the Research Review Committee (RRC) and Ethical Review Committee (ERC) of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). Informed written consent was taken from adult participants and the legal guardians of child participants under 18 years old for the studies from which isolates were collected.

Study settings

icddr,b is an international health research organization located at Mohakhali area in Dhaka which contributed to various enteric disease surveillance studies across Bangladesh. We designed this genomic study with eight different enteric disease surveillance study sites in Bangladesh that conducted by icddr,b. A nationwide enteric disease surveillance study [13] was conducted in ten hospitals located in eight different districts of Bangladesh in collaboration with the Institute of Epidemiology, Disease Control and Research (IEDCR), from which we included all the available *S*. Paratyphi A isolated between 2014 and 2018 from five hospitals (Site 1: Naogaon Sadar Hospital, Site 2: Potuakhali General Hospital, Site 3: BITID, Chittagong, Site 4: Uttara Adhunik Hospital, Site 7: Dhaka Medical College) for this genomic study (**Fig 1**). In addition, the Typhoid Immunization Surveillance study (TIS study) [14–16] was conducted within three urban areas in Dhaka city and, we included *S*. Paratyphi A isolated between 2008 and 2016 from these three sites (Site 5: icddr,b Mirpur field site, Site 6: icddr,b Mohakhali Hospital and Site 8: icddr,b Kamalapur field site) (**Fig 1**). For both surveillance studies, suspected enteric fever patients were enrolled from study sites based on the criteria of fever of at least 38˚C for a minimum duration of three days.

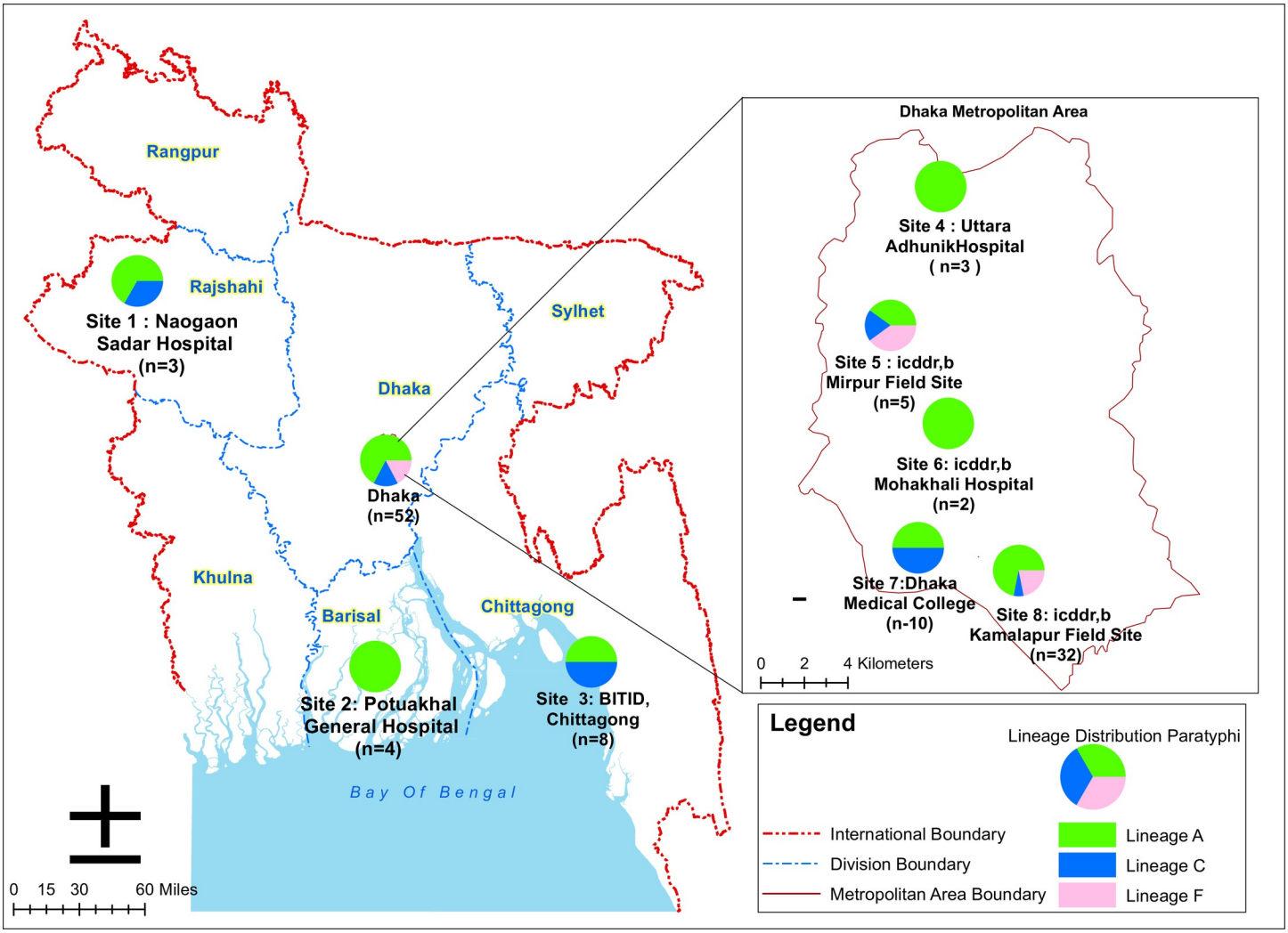

Fig 1. Geographical distrbution of *S*. Paratyphi A lineages in each study site of Bangladesh. Map of Bangladesh showing the lineage distribution at each of the eight study sites across Bangladesh (Direct base layer link for Bangladesh and Dhaka map available athttps://geodash.gov.bd/layers/geonode:level_1_administrative_areas and https://geodash.gov.bd/layers/geonode:dhaka_administrative_boundaries respectively, maps sourced from https://geodash.gov.bd and modified by using ArcGIS10.8.1 software). Pie charts at each site depict the lineage distribution. The number of *S*. Paratyphi A positive cases (*n*) at each study site are also shown in the key.

https://doi.org/10.1371/journal.pntd.0009748.g001

Sample collection and bacterial isolation

Blood samples (3mL for children <5 years of age and 5mL for others) from enteric fever suspected patients were collected upon enrollment, inoculated into blood culture bottle and incubated in automated BacT/Alert machines [14]. All positive blood culture bottles were subcultured onto MacConkey agar plates and incubated at 37˚C for 18–24 hours. Non-lactose fermenting colony(s) from the MacConkey agar plates were inoculated to identify the *Salmonella* spp. by using standard biochemical tests and candidate *S*. Paratyphi A isolates were further confirmed by serotyping with *Salmonella*-specific O and flagellar H antiserum (Denka Sieken Tokyo, Japan) [14,15]. All *S*. Paratyphi A isolates were stored with 20% glycerol at -70˚C for further use. In this study, we included all available stored *S*. Paratyphi A isolates (n = 67) for WGS.

Whole genome sequencing and SNP calling

Genomic DNA was extracted from the *S*. Paratyphi A isolates using the Wizard Genomic DNA Kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. Index-tagged paired-end Illumina sequencing libraries, with an insert size of 500 bp, were prepared as previously described [17]. WGS was performed at the Wellcome Sanger Institute using the Illumina Hiseq2500 platform (Illumina, San Diego, CA, USA) to generate 150 bp paired end reads. The read quality of each isolate was screened using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Sequence reads were submitted to the European Nucleotide Archive and a full list of accession numbers with metadata for each isolate is summarized in **S1 Table**. Additionally, 242 global *S*. Paratyphi A sequences were also included in the analysis to provide global context (**S2 Table**) [18].

Paired-end reads were mapped against the reference genome of *S*. Paratyphi A AKU_12607 (accession number: FM200053) using SMALT (v0.7.6) [19] to produce a BAM file. SAMtools (v1.9) [20] was used to identify SNPs with Phred scores above 30 and remove low confidence alleles with consensus base quality $\leq$20, read depth $\leq$5 or heterozygous base calls. SNPs located in prophage regions, repetitive sequences or recombinant regions as detected by Gubbins (v2.4.1) [21] were also excluded. Then SNP-sites (v2.5.1) [22] extracted the SNPs from the multi-FASTA alignment and SNP distances were calculated using the dna.dist function in the Analysis of Phylogenetics and Evolution (ape) R package (v5.4–1) [23].

Phylogenetic and statistical analysis

RAxML (v8.2.8) [24] was used to construct a maximum likelihood (ML) phylogenetic tree from the SNP alignment. A generalized time-reversible model and a gamma distribution was used to model site-specific rate variation (GTR+ Γ substitution model; GTRGAMMA in RAxML) with 100 bootstrap pseudo-replicates used to assess branch support for the ML phylogeny. The resulting phylogenies were visualized and annotated using FigTree (v1.4.4) (http://tree.bio.ed.ac.uk/software/figtree), and the R package ggtree (v2.4.1) [25]. We performed Fisher's exact tests implemented in STATA [26] to determine statistical differences of lineage distribution between study sites inside and outside Dhaka.

Comparative pan-genome analysis

Raw sequence reads were assembled *de novo* using Unicycler (v0.4.7) [27] and annotated with Prokka (v1.12) [28]. The quality of genome assemblies were assessed by QUAST (v5.0.2) [29] and the assembly metrics i.e. number of contigs ($\geq$1,000bp) and total assembly length (contigs $\geq$1,000bp) of both Bangladeshi and global Paratyphi A isolates are summarised in **S3 Table.** We included genome assemblies that meet the threshold less than 100 contigs by using the formula (mean + 0.5 standard deviation) for comparative pan-genome analysis. Roary (v3.12.0) [30] was performed on annotated assembled genomes to identify the pan-genome, using a blastp percentage sequence identity of 95% and a core definition of present in $\geq$ 95% of the included isolates. The Heaps function within the Micropan R package [31] was used to plot the pan-genome curve which calculates the curve fit constant according to Heaps law [32]: $n = k^{*}N^{-\alpha}$, where n is pan-genome size, N is the number of genomes and k is the curve-specific constant. The exponential term, α determines whether the pan-genome of a bacterial variant is closed ($\alpha>1$) or open ($\alpha<1$).

Antimicrobial resistance gene, plasmid identification and comparative pHCM2 plasmid analysis

ARIBA (v2.14.4) [33], in conjunction with the comprehensive antibiotic resistance database (CARD) [34], and the PlasmidFinder database [35] were used to detect AMR genes and

plasmid replicons, respectively. AMR mutations in *S*. Paratyphi A genomes, including point mutations in the quinolone resistance-determining region (QRDR) of genes *gyrA*, *parC* and mutation in *acrB* associated with azithromycin resistance were identified using the genoparatyphi python script (available at: https://github.com/zadyson/genoparatyphi/).

To investigate the genetic diversity, evolution, and circulation of the pHCM2 plasmids present in Bangladeshi *S*. Typhi and *S*. Paratyphi populations, we constructed pHCM2 plasmid phylogeny including 17 *S*. Paratyphi A genomes from this present study and 334 *S*. Typhi and a single *S*. Paratyphi B variant Java (*S*. Java) genome from previous studies in Bangladesh [36–38]. A full list of accession numbers for each genome used for pHCM2 phylogeny is summarized in **S4 Table**. Briefly, all the raw sequencing reads from these organisms were mapped to the reference *S*. Typhi CT18 pHCM2 plasmid (accession number: AL513384.1) using local sensitive mapping Bowtie2 (v2.4.2) [39] and SNPs were identified using SAMTools (v1.3.1) [20]. SNPs that did not meet the quality criteria (Phred score $\geq$ 30, depth coverage $\geq$5) were excluded from plasmid analysis. RAxML v8.2 [24] was used for constructing a maximum likelihood (ML) plasmid phylogeny based on the total of 1788 core SNPs calling from mapping. RAxML (v8.2.8) used the generalized time-reversible evolutionary model with gamma-distributed rate variation (GTR+ Γ). One hundred bootstrap pseudo-replicate analyses were performed to assess the robustness of the ML tree topology. Moreover, we also compared the annotated pHCM2 plasmid sequence of *S*. Paratyphi A and *S*. Paratyphi B variant Java and to the reference pHCM2 plasmid by BLAST [40] to identify sequence similarity with the database and visualized these comparisons using Artemis Comparison Tool (ACT) (v18.0.2) [41] and Easyfig (available at https://mjsull.github.io/Easyfig/). pHCM2 plasmid integrity was confirmed by visualizing the assembly graph of pHCM2 plasmid carrying *S*. Paratyphi A genome in assembly graph viewer Bandage tool (v0.8.1) (available at https://rrwick.github.io/Bandage/).

Results

The population structure of *S*. Paratyphi A isolates in Bangladesh

We subjected the genome sequences of the 67 Bangladeshi *S*. Paratyphi A isolates and an additional 242 global *S*. Paratyphi A genome sequences to phylogenetic analysis. We obtained a final set of 5,419 chromosomal SNPs from a total alignment length of 4,794,508 bp for the total 309 *S*. Paratyphi A genomes. The resulting global phylogeny could be subdivided into the seven previously defined distinct lineages (A to G); the *S*. Paratyphi A isolated from Bangladesh were restricted to only lineages A (67.2%), C (19.4%) and F (13.4%) (**Fig 2**). *S*. Paratyphi A isolates in this study were most closely related to isolates originating in Nepal, India, Pakistan, and Myanmar, suggesting regional circulation of these lineages across South Asia. Notably, the majority (42/67, 62.6%) of the contemporary Bangladeshi sequences formed a monophyletic sub-lineage within lineage A that we herein defined as sub-lineage A3. This dominant Bangladeshi sub-lineage was closely related to sub-lineage A1 isolates from Nepal (median distance ~70 SNPs). Moreover, three Bangladeshi isolates were intermingled with lineage A Indian isolates on two independent branches separated by median of 25 and 9 SNPs respectively. Aside from lineage A, a single Nepalese *S*. Paratyphi A isolate from lineage F was grouped within nine Bangladeshi isolates and separated by median of 44 SNPs. The remaining 13 isolates clustered within sub-lineage C4 as did organisms from Pakistan and Myanmar (median distance 57 SNPs).

We additionally assessed the distribution of lineage A, C, and F among 67 isolates collected from eight study sites covering four divisions within Bangladesh (**Fig 1**). This geographical location data outlined a heterogeneous distribution of lineages between study sites, with the exception of some study sites with a limited number of isolates. Lineage A was distributed

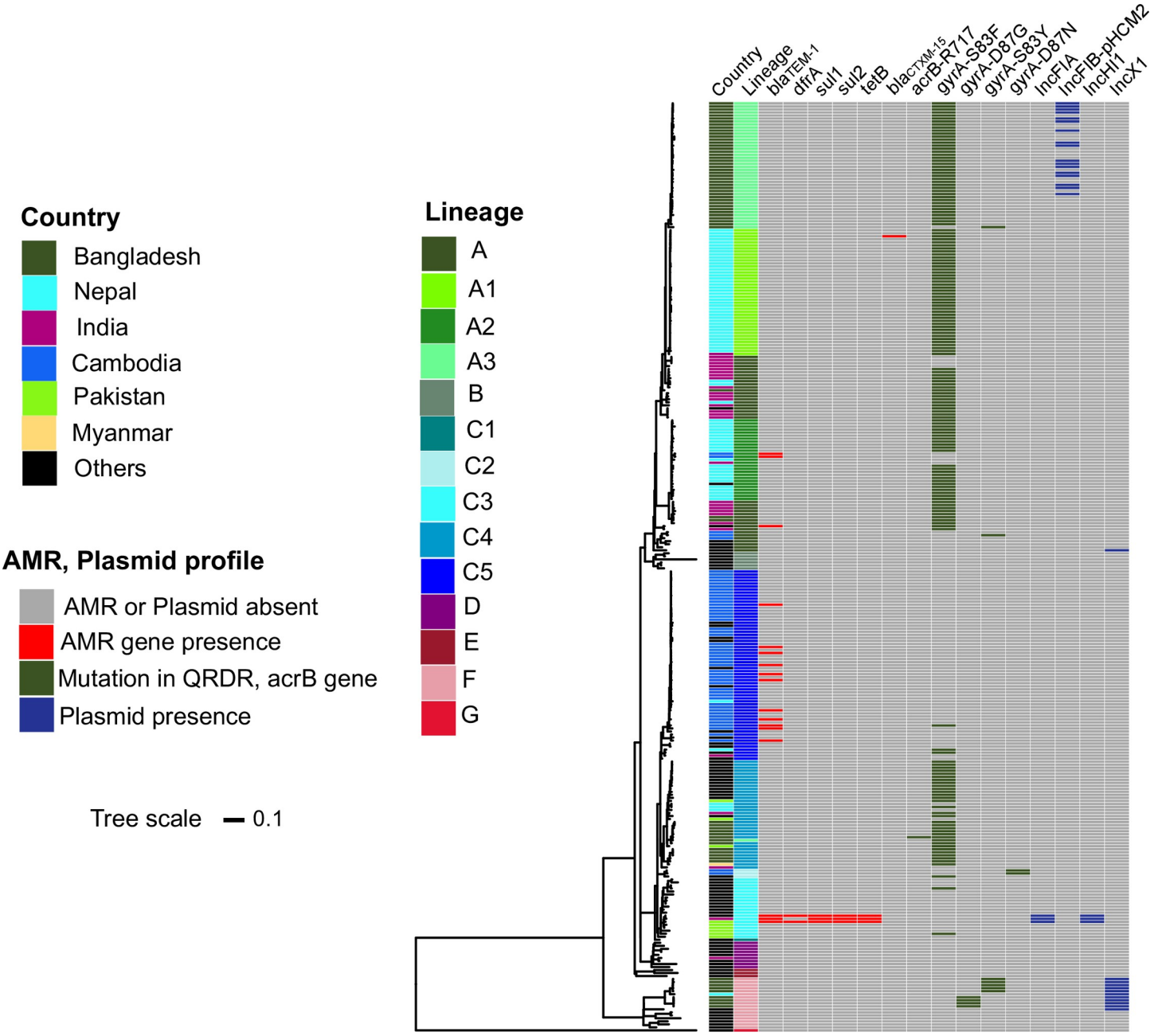

Fig 2. Global population structure of Bangladesh *S*. Paratyphi A with AMR genes and plasmid replicons presence heatmap. Maximum-likelihood phylogenetic tree of 309 *S*. Paratyphi A genome sequences from the global collection including Bangladeshi isolates from this study. The coloured heatmap shows country, lineages, AMR genes, mutations in the QRDR and *acrB* genes, plasmid replicons for each isolate; see legend for colours.

https://doi.org/10.1371/journal.pntd.0009748.g002

throughout all eight study sites, with three sites (2, 4, and 6) harboring isolates from lineage A only. All these three lineages (A, C, F) were detected in the icddr,b Kamalapur field site (site 8; n = 32) and the Mirpur field sites (site 5; n = 5), located in Dhaka. Two lineages (A, C) were restricted to the remaining three hospital sites (1, 3, and 7). We did not observe any significant difference ($p = 0.15$; Fisher's exact test) in the distribution of lineages between the five study sites within Dhaka (sites 4, 5, 6, and 7; n = 52) and the three study sites outside Dhaka (sites 1, 2, and 3; n = 15).

Comparative pan-genome analysis

We first conducted a pan-genome analysis on all 309 *S*. Paratyphi A genomes to investigate the gene distribution among the serovar. The pan-genome comprised 18,802 genes including 4,054 core genes (detected in $\geq$ 95% of all genomes) and 14,748 accessory genes (detected in $<$ 95% of all genomes). The seemingly overestimated accessory gene set, driven by the presence of 14,748 cloud genes (detected in only $<$ 15% of total genome), was due to fragmentation of genome assemblies. To minimize the effect of low-quality sequence data, we further assessed the quality of genome assemblies (i.e explained details in method; **S3 Table**), and reconsidered the comparative pan-genome analysis using 295 quality-filtered assemblies. The revised pan-genome contained 9,093 genes, divided into 4,125 core genes and 4,968 accessory genes. In a typical *S*. Paratyphi A genome (average 4,300 coding sequences, n = 295), the core genes account for 96.1% of the coding sequences. We also observed the gene accumulation curve of *S*. Paratyphi A pan-genome (n = 295) slowly flattened with the addition of new genomes, with a curve fitting parameter α value of 1.000015, indicating a closed pan-genome (**S1 Fig**). Based on the pan-genome size and the proportion of a typical genome comprised of core genes, we concluded that the *S*. Paratyphi A possesses a conservative genomic structure with little evidence of importation of new genes.

AMR genes and plasmid profiles associated with Bangladeshi *S*. Paratyphi A

We explored the Bangladeshi *S*. Paratyphi A genome sequences together with a global collection of *S*. Paratyphi A from previous studies for genes and mutations associated with AMR and also to identify plasmid replicons (**Fig 2**). Notably, and contrary to previously published data from *S*. Typhi from Bangladesh, the AMR genes, *catA*, *dfrA7*, *sul1*, *sul2*, *strA*, *strB*, and $bla_{TEM\text{-}1}$, which confer an MDR phenotype were not detected in our Paratyphi A collection. Despite a lack of genes associated with MDR, non-synonymous mutations in the QRDR (either codon 83 or 87) of *gyrA*, associated with reduced susceptibility to fluoroquinolones, were found in all 67 of the Bangladeshi *S*. Paratyphi A isolates. Among the *gyrA* QRDR mutations detected, S83F (85.1%; 57/67) was the most common, followed by S83Y (9%; 6/67), and D87G (6%; 4/67). No double or triple QRDR mutations in *gyrA*, *gyrB*, *parC*, *parE* genes were observed in any *S*. Paratyphi A isolates of this study. In addition, a single *S*. Paratyphi A organism isolated from Dhaka Medical College (site 7) in 2018 carried an R717L mutation in *acrB*, which conferred resistance to azithromycin. Plasmid associated sequences were limited in these genome sequences, an IncX1 plasmid and IncFIB-pHCM2-like plasmid were identified in 12 and 17 isolates, respectively (**Fig 2**).

The origins of pHCMC2-like plasmid in Bangladeshi *S*. Paratyphi A

Aiming to better understand the emergence of pHCM2-like plasmid in 17 Bangladeshi *S*. Paratyphi A isolated from this present study, we conducted further fine-detailed comparisons of these genomes with the reference *S*. Typhi CT18 pHCM2 plasmid and a non-typhoidal *Salmonella* serovar (*S*. Java). We constructed a pHCM2 phylogenetic tree incorporating plasmid sequences from the 17 *S*. Paratyphi A, 334 *S*. Typhi, and one *S*. Java isolated from Bangladesh (**Fig 3 and S4 Table**) to understand the recent evolutionary history of this cryptic pHCM2 plasmid. The pHCM2 plasmid phylogeny and BLAST analysis revealed that the pHCM2 sequences from *S*. Typhi (106,516 bp) and *S*. Paratyphi A (106,706 bp) were highly related, with 99.98% nucleotide identity (average 19 SNPs variation). Importantly, the *S*. Paratyphi A pHCM2- like plasmid assembled into a single contig that was predicted to be circular by analysis using Bandage (S2 Fig). Conversely, a more distantly related pHCM2-like plasmid has been

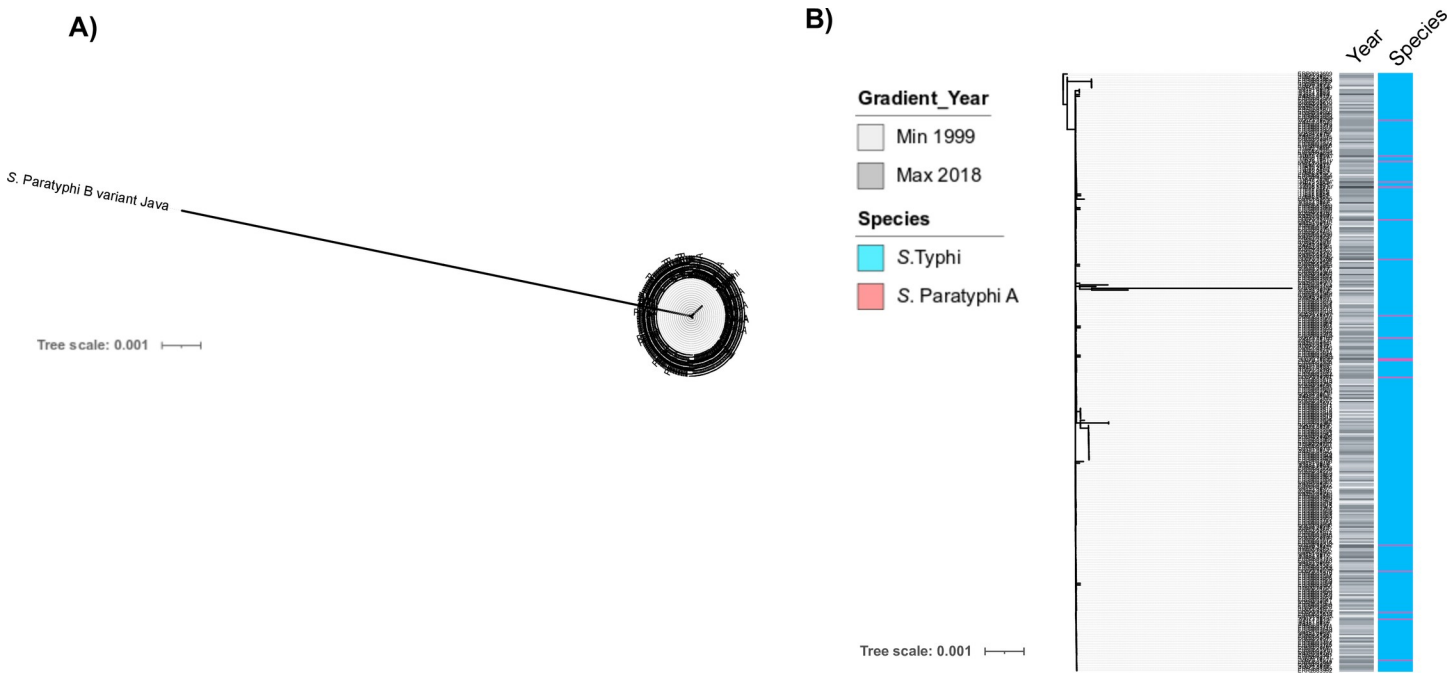

Fig 3. The local evolutionary history of the pHCM2 plasmid in Bangladesh. (A) Unrooted of maximum likelihood whole pHCM2 phylogeny including *S*. Paratyphi A isolates from this study and *S*. Typhi, *S*. Paratyphi B variant Java from previous studies in Bangladesh (B) Maximum likelihood of pHCM2 phylogeny including only *S*. Paratyphi A and Typhi isolates. The coloured heatmap shows species and year of collection for each isolate.

https://doi.org/10.1371/journal.pntd.0009748.g003

previously detected in non-typhoidal *S*. Java (107,362 bp), and shares 98.8% identity with *S*. Typhi CT18 pHCM2 plasmid across only 90% of the sequence (average 1,697 SNPs variation).

Next, we compared the annotated pHCM2 reference plasmid sequence of *S*. Typhi CT18 with *S*. Paratyphi A and *S*. Java to identify the genetic difference of pHCM2 between typhoidal and non-typhoidal *Salmonella* and identified 12 non-identical regions against pHCM2 plasmid reference, including three insertions and one deletion (**S2 Fig**, **S5 Table**). Most of these non-identical regions were hypothetical proteins and only a few were related to DNA metabolism and phage. Furthermore, we conducted pan-genome analysis on all 309 *S*. Paratyphi A and identified 127 proteins (113 hypothetical proteins and 14 annotated proteins) which were present in only the 17 pHCM2-like plasmid-harboring Bangladeshi *S*. Paratyphi A isolates but not present in other global *S*. Paratyphi A isolates (**S6 Table**).

Discussion

In this study, we performed whole genome sequencing of *S*. Paratyphi A collection isolated during nationwide enteric disease surveillance in Bangladesh and combined these sequence data with an additional 242 sequences from a global collection of *S*. Paratyphi A genomes [18]. Previously, a global population structure study by Zhou *et al*. 2018 classified *S*. Paratyphi A into seven lineages and identified lineages A and C as the most dominant globally [42]. Our phylogenetic analysis confirmed these data, finding lineage A to be the most common in Bangladesh, with the overall population comprised of only three lineages (A, C, and F). The addition of 67 Bangladeshi *S*. Paratyphi A isolates into the global phylogeny suggests that there may have been an ongoing clonal expansion of lineage A. This lineage appears to have been imported from neighboring countries on multiple occasions, which contrasts with onservations from lineages C and F which appeare to have been introduced on one occasion in this

dataset, all of which provides information to track ongoing local or international transmission for surveillance efforts. Additionally, the apparently random distribution of these three lineages across eight nationwide surveillance sites provides limited evidence for regional geographic restriction of *S*. Paratyphi A lineages in Bangladesh.

Over the past two decades, the emergence of resistance to the first line antimicrobials for *S*. Typhi infections has led to change in treatment. Firstly, fluoroquinolones were adopted, followed by third generation cephalosporins and the antibiotic azithromycin. High incidence of MDR typhoid, associated with self-transmissible IncHI1 plasmids, has been previously reported in many endemic settings throughout Asia and Africa [18,37,43]. Notably, the absence of MDR genes in the genome sequences of these 67 Bangladeshi *S*. Paratyphi A isolates is in stark contrast with *S*. Typhi. Indeed, studies from Nepal and Pakistan have identified cases of *S*. Paratyphi A associated with MDR phenotypes and resistance to third-generation cephalosporins mediated by extended spectrum beta lactamases [44,45]. The potential exception to the surprising lack of AMR, was the resistance (or reduced susceptibility) to fluoroquinolones. Decreased fluoroquinolone susceptibility is associated with chromosomal mutations in the QRDR. Our data shows that single chromosomal *gyrA* point mutations in QRDR was observed among all Bangladeshi *S*. Paratyphi A isolates. In addition to QRDR mutation, one recent acquisition of R717L mutation in *acrB* gene associated with azithromycin isolated from Dhaka in 2018 is of major concern. Similar mutations have been recently identified in *S*. Typhi and *S*. Paratyphi A in Nepal, India, Pakistan and also Bangladesh [8,46,47]. This observation reminds us of the need for constant review of empirical antimicrobials used for the treatment of enteric fever and highlights the need of continued genomic surveillance to monitor ongoing AMR trends in invasive *Salmonella*.

We additionally observed non-AMR associated IncFIB-pHCM2-like cryptic plasmids and IncX1 plasmids among the contemporary Bangladeshi *S*. Paratyphi A isolates, which have also been previously reported in the *S*. Typhi population [36,37], although the pHCM2-like plamid was not detected in any global *S*. Paratyphi A isolates [18]. The abundance of this cryptic pHCM2 plasmid among Bangladeshi *S*. Typhi and *S*. Paratyphi A population has still not been well understood. To our knowledge this is the first time that a pHCM2-like plasmid has been described in *S*. Paratyphi A and our phylogenetic analysis might suggest that the ancestral origins of pHCM2-like plasmid in the *S*. Paratyphi A isolates were likely in *S*. Typhi, and potentially entered into the *S*. Paratyphi A population as early as 2011; the earliest origination of pHCM2 in *S*. Typhi in Bangladesh has been reported in 1999 [36,37]. However, these 17 Bangladeshi *S*. Paratyphi A isolates were only from Dhaka city and restricted within only sub-lineage A3 which indicate that they may have been transferred on a single occasion. A previous study reported that pHCM2 in *S*. Typhi CT18 shared 56% sequence identity with the virulence-associated pMT1 plasmid in *Yersinia* pestis [48]. Here we determined that the pHCM2 sequences from the Bangladeshi *S*. Paratyphi A, *S*. Typhi, and *S*. Java shared approximately 51–59% sequence identity and 96% coverage with pMT1. pHCM2 carries bacteriophage genes and genes related to DNA metabolism and replication. It has been suggested that *Salmonella* enterica serovar Typhimurium rough strain-specific phage SSU5 (SSU5), may be an ancestral form of pHCM2 [49]. Here we found the pHCM2 plasmid shared 81–82% sequence similarity with SSU5 phage genes. We hypothesize that these phage gene regions may have a role in horizontal gene transfer and facilitate pathogen adaptation [50]. Though our pan-genome analysis concluded the *S*. Paratyphi A genome was highly conserved in nature, the presence of the pHCM2-like plasmid in Bangladeshi Paratyphi A expanded the known accessory genome, which might play a role in evolution of this pathogen. The pHCM2 cryptic plasmid was first identified in *S*. Typhi isolates associated with a typhoid case in Vietnam in 1992, before it was subsequently detected in Hong Kong, Cambodia, and Pakistan [51]. The lack of pHCM2-like

plasmid in the global collection of *S*. Paratyphi A isolates emphasize our point of interest to understand the source of origin and genetic difference of this plasmid from other *Salmonella* serovars previously detected in Bangladesh. Comparative analysis revealed that pHCM2 from *S*. Paratyphi A and *S*. Typhi shared a higher degree of genetic similarity than with that from *S*. Java. Further studies are required to monitor the trend over time and to understand the physiological effect of this cryptic plasmid on enteric infection caused by *S*. Typhi and *S*. Paratyphi A.

Our study has some limitations. The *S*. Paratyphi A genomes sequenced in this present study were collected from only four out of eight divisions of Bangladesh which may not be representative of the overall population structure of Bangladesh. In addition, we were unable to reach into any statistical significance of the lineage distribution over study sites due to the low number of sequence data generated from this study. However, these data improve our knowledge regarding genomic variation in *S*. Paratyphi A and potential genetic interplay with *S*. Typhi.

S. Paratyphi A is an emerging pathogen and may become the leading cause of enteric fever in Asia if typhoid vaccination is introduced in low and middle income countries (LMICs). Therefore, we need a better understanding of the circulation of this pathogen and to investigate if new vaccine approaches are warranted. In the absence of vaccines, monitoring AMR trends, and their associated mechanisms, is important in planning future approaches to therapy for enteric fever. Our data highlight the importance of sustained genome-based surveillance for emerging enteric pathogens like *S*. Paratyphi A in endemic regions.

Supporting information

S1 Fig. Pan-genome accumulation curve of *S*. Paratyphi. A: The gene accumulation curve with curve fitting parameter α value of 1.000015 are depicted for global *S*. Paratyphi A genomes (n = 295). Error bars above and below the median are depicted by a vertical line above and below the curve.
(TIF)

S2 Fig. Comparison of pHCM2 plasmid reference genome (AL513383) with pHCM2-like plasmid of *S*. Paratyphi A and *S*. Java. (A) The assembly graph of one representative pHCM2-like plasmid harbouring *S*. Paratyphi A genome was visualised in Bandage tool. Each grey line in assembly graph represents a node or assembled contig and a closed ring of pHCM2 plasmid region presented in single node is highlighted as blue colour which was analysed from Bandage's integrated BLAST search with reference *S*. Typhi CT18 pHCM2 plasmid. (B) Full pHCM2 plasmid sequence comparison of reference AL513384.1 with *S*. Paratyphi A and *S*. Java in Artemis Comparison Tool (ACT) and Easyfig including (C) three insertions and one deletion event in *S*. Paratyphi A relative to pHCM2 plasmid reference. Orange arrows indicted CDSs and grey shading between the sequences represents BLAST nucleotide identity (see key).
(TIF)

S1 Table. Metadata including accession numbers, AMR and plasmid profile of 67 *S*. Paratyphi A isolates of this present study.
(XLSX)

S2 Table. Metadata for the 242 global *S*. Paratyphi A genome collection from previous studies.
(XLSX)

S3 Table. Genome assembly metrics of total 309 Paratyphi A including *S*. Paratyphi A isolates of this present study.
(XLSX)

S4 Table. List of lane accession numbers of *S*. Typhi and *S*. Java isolates carrying pHCM2 from previous studies in Bangladesh.
(XLSX)

S5 Table. Gene list of non-identical regions of annotated pHCM2 plasmid sequence of *S*. Paratyphi A and *S*. Java isolated from Bangladesh and to the reference pHCM2 plasmid.
(XLSX)

S6 Table. List of unique genes including annotated proteins present in 17 Bangladeshi *S*. Paratyphi A isolates carrying pHCM2-like plasmid but not present in other global Paratyphi A isolates.
(XLSX)

Acknowledgments

We acknowledge the support of dedicated field and laboratory workers at the icddr,b and the different field sites involved in this study. The icddr,b is grateful to the Governments of Bangladesh, Canada, Sweden and UK for providing core support. We are grateful to Matthew Dorman and Sally Kay for their role in receipt of, testing, and submission of samples for sequencing. We would also like to thank the members of Pathogen Informatics Team and core sequencing teams at the Wellcome Sanger Institute.

Author Contributions

Conceptualization: Sadia Isfat Ara Rahman, Gordon Dougan, Stephen Baker, Firdausi Qadri.

Data curation: Sadia Isfat Ara Rahman, To Nguyen Thi Nguyen, Zoe A. Dyson, Alyce Taylor-Brown.

Formal analysis: Sadia Isfat Ara Rahman, To Nguyen Thi Nguyen, Zoe A. Dyson, Alyce Taylor-Brown.

Funding acquisition: Farhana Khanam, Stephen Baker, Firdausi Qadri.

Investigation: Farhana Khanam, Nicholas R. Thomson, Gordon Dougan, Stephen Baker, Firdausi Qadri.

Methodology: Sadia Isfat Ara Rahman.

Resources: Farhana Khanam, Firdausi Qadri.

Software: Zoe A. Dyson.

Supervision: Emran Kabir Chowdhury, Gordon Dougan, Stephen Baker, Firdausi Qadri.

Visualization: Sadia Isfat Ara Rahman, To Nguyen Thi Nguyen.

Writing – original draft: Sadia Isfat Ara Rahman, To Nguyen Thi Nguyen, Stephen Baker, Firdausi Qadri.

Writing – review & editing: Sadia Isfat Ara Rahman, To Nguyen Thi Nguyen, Farhana Khanam, Nicholas R. Thomson, Zoe A. Dyson, Alyce Taylor-Brown, Emran Kabir Chowdhury, Gordon Dougan, Stephen Baker, Firdausi Qadri.

References

1. Arndt MB, Mosites EM, Tian M, Forouzanfar MH, Mokhdad AH, Meller M, et al. Estimating the burden of paratyphoid a in Asia and Africa. PLoS neglected tropical diseases. 2014; 8(6):e2925. Epub 2014/06/06. https://doi.org/10.1371/journal.pntd.0002925 PMID: 24901439; PubMed Central PMCID: PMC4046978 by Sanofi Pasteur, a commercial company, as the lead for Dengue Epidemiological studies. This does not alter our adherence to all PLOS NTDs policies on sharing data and materials.
2. Sahastrabuddhe S, Carbis R, Wierzba TF, Ochiai RL. Increasing rates of Salmonella Paratyphi A and the current status of its vaccine development. Expert Rev Vaccines. 2013; 12(9):1021–31. Epub 2013/09/24. https://doi.org/10.1586/14760584.2013.825450 PMID: 24053396.
3. Connor BA, Schwartz E. Typhoid and paratyphoid fever in travellers. The Lancet Infectious diseases. 2005; 5(10):623–8. Epub 2005/09/27. https://doi.org/10.1016/S1473-3099(05)70239-5 PMID: 16183516.
4. Browne AJ, Kashef Hamadani BH, Kumaran EAP, Rao P, Longbottom J, Harriss E, et al. Drug-resistant enteric fever worldwide, 1990 to 2018: a systematic review and meta-analysis. BMC Medicine. 2020; 18 (1):1. https://doi.org/10.1186/s12916-019-1443-1 PMID: 31898501
5. Naheed A, Ram PK, Brooks WA, Hossain MA, Parsons MB, Talukder KA, et al. Burden of typhoid and paratyphoid fever in a densely populated urban community, Dhaka, Bangladesh. Int J Infect Dis. 2010; 14 Suppl 3:e93–9. Epub 2010/03/20. https://doi.org/10.1016/j.ijid.2009.11.023 PMID: 20236850.
6. Steele AD, Carey ME, Kumar S, MacLennan CA, Ma L-F, Diaz Z, et al. Typhoid Conjugate Vaccines and Enteric Fever Control: Where to Next? Clinical Infectious Diseases. 2020; 71(Supplement_2): S185–S90. https://doi.org/10.1093/cid/ciaa343%J Clinical Infectious Diseases
7. Dobinson HC, Gibani MM, Jones C, Thomaides-Brears HB, Voysey M, Darton TC, et al. Evaluation of the Clinical and Microbiological Response to Salmonella Paratyphi A Infection in the First Paratyphoid Human Challenge Model. Clin Infect Dis. 2017; 64(8):1066–73. Epub 2017/02/06. https://doi.org/10.1093/cid/cix042 PMID: 28158395; PubMed Central PMCID: PMC5439345.
8. 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(11):e0007868. Epub 2019/11/16. https://doi.org/10.1371/journal.pntd.0007868 PMID: 31730615; PubMed Central PMCID: PMC6881056.
9. Fangtham M, Wilde H. Emergence of Salmonella paratyphi A as a major cause of enteric fever: need for early detection, preventive measures, and effective vaccines. Journal of travel medicine. 2008; 15 (5):344–50. Epub 2008/11/14. https://doi.org/10.1111/j.1708-8305.2008.00237.x PMID: 19006508.
10. Qian H, Cheng S, Liu G, Tan Z, Dong C, Bao J, et al. Discovery of seven novel mutations of gyrB, parC and parE in Salmonella Typhi and Paratyphi strains from Jiangsu Province of China. Sci Rep. 2020; 10 (1):7359–. https://doi.org/10.1038/s41598-020-64346-0 PMID: 32355184.
11. Menezes GA, Harish BN, Khan MA, Goessens W, Hays JP. Antimicrobial resistance trends in blood culture positive Salmonella Paratyphi A isolates from Pondicherry, India. Indian J Med Microbiol. 2016; 34 (2):222–7. Epub 2016/04/16. https://doi.org/10.4103/0255-0857.180352 PMID: 27080779.
12. Nair S, Day M, Godbole G, Saluja T, Langridge GC, Dallman TJ, et al. Genomic surveillance detects Salmonella enterica serovar Paratyphi A harbouring blaCTX-M-15 from a traveller returning from Bangladesh. PLOS ONE. 2020; 15(1):e0228250. https://doi.org/10.1371/journal.pone.0228250 PMID: 32000262
13. Khan AI, Rashid MM, Islam MT, Afrad MH, Salimuzzaman M, Hegde ST, et al. Epidemiology of Cholera in Bangladesh: Findings From Nationwide Hospital-based Surveillance, 2014–2018. Clinical Infectious Diseases. 2019. https://doi.org/10.1093/cid/ciz1075 PMID: 31891368
14. Khanam F, Sheikh A, Sayeed MA, Bhuiyan MS, Choudhury FK, Salma U, et al. Evaluation of a typhoid/paratyphoid diagnostic assay (TPTest) detecting anti-Salmonella IgA in secretions of peripheral blood lymphocytes in patients in Dhaka, Bangladesh. PLoS neglected tropical diseases. 2013; 7(7):e2316. https://doi.org/10.1371/journal.pntd.0002316 PMID: 23951368
15. Sheikh A, Bhuiyan MS, Khanam F, Chowdhury F, Saha A, Ahmed D, et al. Salmonella enterica serovar Typhi-specific immunoglobulin A antibody responses in plasma and antibody in lymphocyte supernatant specimens in Bangladeshi patients with suspected typhoid fever. Clin Vaccine Immunol. 2009; 16 (11):1587–94. Epub 2009/09/11. https://doi.org/10.1128/CVI.00311-09 PMID: 19741090; PubMed Central PMCID: PMC2772369.
16. Khanam F, Sayeed MA, Choudhury FK, Sheikh A, Ahmed D, Goswami D, et al. Typhoid fever in young children in Bangladesh: clinical findings, antibiotic susceptibility pattern and immune responses. PLoS neglected tropical diseases. 2015; 9(4):e0003619. Epub 2015/04/08. https://doi.org/10.1371/journal.pntd.0003619 PMID: 25849611; PubMed Central PMCID: PMC4388457.

17. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid pneumococcal evolution in response to clinical interventions. science. 2011; 331(6016):430–4. https://doi.org/10.1126/science.1198545 PMID: 21273480

18. 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 neglected tropical diseases. 2018; 12(4):e0006408. https://doi.org/10.1371/journal.pntd.0006408 PMID: 29684021

19. Ponstingl H, Ning Z. SMALT-a new mapper for DNA sequencing reads. F1000 Posters. 2010; 1:313.

20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

21. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research. 2014; 43(3):e15–e. https://doi.org/10.1093/nar/gku1196 PMID: 25414349

22. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial genomics. 2016; 2(4):e000056. Epub 2017/03/30. https://doi.org/10.1099/mgen.0.000056 PMID: 28348851; PubMed Central PMCID: PMC5320690.

23. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004; 20(2):289–90. Epub 2004/01/22. https://doi.org/10.1093/bioinformatics/btg412 PMID: 14734327.

24. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

25. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017; 8(1):28–36.

26. StataCorp L. Stata statistical software: release 12.0. College Station, TX: StataCorp LP; 2011.

27. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS computational biology. 2017; 13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595 PMID: 28594827

28. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30(14):2068–9. https://doi.org/10.1093/bioinformatics/btu153 PMID: 24642063

29. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013; 29(8):1072–5. Epub 2013/02/21. https://doi.org/10.1093/bioinformatics/btt086 PMID: 23422339; PubMed Central PMCID: PMC3624806.

30. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015; 31(22):3691–3. https://doi.org/10.1093/bioinformatics/btv421 PMID: 26198102

31. Snipen L, Liland KH. micropan: an R-package for microbial pan-genomics. BMC Bioinformatics. 2015; 16(1):79. https://doi.org/10.1186/s12859-015-0517-0 PMID: 25888166

32. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Current opinion in microbiology. 2008; 11(5):472–7. Epub 2008/12/18. https://doi.org/10.1016/j.mib.2008.09.006 PMID: 19086349.

33. Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microbial genomics. 2017; 3(10):e000131. Epub 2017/11/28. https://doi.org/10.1099/mgen.0.000131 PMID: 29177089; PubMed Central PMCID: PMC5695208.

34. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013; 57(7):3348–57. Epub 2013/05/08. https://doi.org/10.1128/AAC.00419-13 PMID: 23650175; PubMed Central PMCID: PMC3697360.

35. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrobial agents and chemotherapy. 2014; 58(7):3895–903. https://doi.org/10.1128/AAC.02412-14 PMID: 24777092

36. Tanmoy AM, Westeel E, De Bruyne K, Goris J, Rajoharison A, Sajib MS, et al. Salmonella enterica Serovar Typhi in Bangladesh: Exploration of Genomic Diversity and Antimicrobial Resistance. mBio. 2018; 9(6):e02112–18. https://doi.org/10.1128/mBio.02112-18 PMID: 30425150

37. 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(2):e0008036. Epub 2020/02/28. https://doi.org/10.1371/journal.pntd.0008036 PMID: 32106221; PubMed Central PMCID: PMC7064254.

38. Rahman SIA, Taylor-Brown A, Khanam F, Khan AI, Horesh G, Dyson ZA, et al. Genome wide analysis provides a deeper understanding of the population structure of the Salmonella enterica serotype Paratyphi B complex in Bangladesh. Microbial genomics. 2021. https://doi.org/10.1099/mgen.0.000617 PMID: 34550065

39. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9 (4):357–9. Epub 2012/03/06. https://doi.org/10.1038/nmeth.1923 PMID: 22388286; PubMed Central PMCID: PMC3322381.

40. Donkor ES, Dayie NT, Adiku TK. Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA). Journal of Bioinformatics and Sequence Analysis. 2014; 6(1):1–6.

41. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. Bioinformatics. 2005; 21(16):3422–3. Epub 2005/06/25. https://doi.org/10.1093/bioinformatics/bti553 PMID: 15976072.

42. Zhou Z, McCann A, Weill FX, Blin C, Nair S, Wain J, et al. Transient Darwinian selection in Salmonella enterica serovar Paratyphi A during 450 years of global spread of enteric fever. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(33):12199–204. Epub 2014/08/06. https://doi.org/10.1073/pnas.1411012111 PMID: 25092320; PubMed Central PMCID: PMC4143038.

43. Park SE, Pham DT, Boinett C, Wong VK, Pak GD, Panzner U, et al. The phylogeography and incidence of multi-drug resistant typhoid fever in sub-Saharan Africa. Nature communications. 2018; 9(1):5094. Epub 2018/12/07. https://doi.org/10.1038/s41467-018-07370-z PMID: 30504848; PubMed Central PMCID: PMC6269545.

44. Pokharel BM, Koirala J, Dahal RK, Mishra SK, Khadga PK, Tuladhar NR. Multidrug-resistant and extended-spectrum beta-lactamase (ESBL)-producing Salmonella enterica (serotypes Typhi and Paratyphi A) from blood isolates in Nepal: surveillance of resistance and a search for newer alternatives. Int J Infect Dis. 2006; 10(6):434–8. Epub 2006/09/19. https://doi.org/10.1016/j.ijid.2006.07.001 PMID: 16978898.

45. Holt KE, Thomson NR, Wain J, Phan MD, Nair S, Hasan R, et al. Multidrug-Resistant Salmonella enterica Serovar Paratyphi A Harbors IncHI1 Plasmids Similar to Those Found in Serovar Typhi. 2007; 189(11):4257–64. https://doi.org/10.1128/JB.00232-07 PMID: 17384186

46. Molloy A, Nair S, Cooke FJ, Wain J, Farrington M, Lehner PJ, et al. First report of Salmonella enterica serotype paratyphi A azithromycin resistance leading to treatment failure. Journal of clinical microbiology. 2010; 48(12):4655–7. Epub 2010/10/15. https://doi.org/10.1128/JCM.00648-10 PMID: 20943875; PubMed Central PMCID: PMC3008474.

47. Patel SR, Bharti S, Pratap CB, Nath G. Drug Resistance Pattern in the Recent Isolates of Salmonella Typhi with Special Reference to Cephalosporins and Azithromycin in the Gangetic Plain. Journal of clinical and diagnostic research: JCDR. 2017; 11(6):Dm01–dm3. Epub 2017/08/03. https://doi.org/10.7860/JCDR/2017/23330.9973 PMID: 28764168; PubMed Central PMCID: PMC5535361.

48. 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(6858):848–52. https://doi.org/10.1038/35101607 PMID: 11677608

49. Kim M, Kim S, Ryu S. Complete genome sequence of bacteriophage SSU5 specific for Salmonella enterica serovar Typhimurium rough strains. J Virol. 2012; 86(19):10894–. https://doi.org/10.1128/JVI.01796-12 PMID: 22966187.

50. Lima NCB, Tanmoy AM, Westeel E, de Almeida LGP, Rajoharison A, Islam M, et al. Analysis of isolates from Bangladesh highlights multiple ways to carry resistance genes in Salmonella Typhi. BMC Genomics. 2019; 20(1):530. https://doi.org/10.1186/s12864-019-5916-6 PMID: 31253105

51. Kidgell C, Pickard D, Wain J, James K, Nga LT, Diep T, et al. Characterisation and distribution of a cryptic Salmonella typhi plasmid pHCM2. Plasmid. 2002; 47:159–71. https://doi.org/10.1016/s0147-619x(02)00013-6 PMID: 12151231

MICROBIAL GENOMICS

RESEARCH ARTICLE

Rahman *et al.*, *Microbial Genomics* 2021;7:000617

DOI 10.1099/mgen.0.000617

Genome-wide analysis provides a deeper understanding of the population structure of the *Salmonella enterica* serotype Paratyphi B complex in Bangladesh

Sadia Isfat Ara Rahman[1]†, Alyce Taylor-Brown[2]†, Farhana Khanam[1], Ashraful Islam Khan[1], Gal Horesh[2], Zoe A. Dyson[2,3,4,5], Yasmin Ara Begum[1], Emran Kabir Chowdhury[6], Firdausi Qadri[1]‡, Gordon Dougan[2,3]‡ and Nicholas R. Thomson[2,4,*,‡]

Abstract

The *Salmonella enterica* serotype Paratyphi B complex causes a wide range of diseases, from gastroenteritis to paratyphoid fever, depending on the biotypes Java and *sensu stricto*. The burden of Paratyphi B biotypes in Bangladesh is still unknown, as these are indistinguishable by *Salmonella* serotyping. Here, we conducted the first whole-genome sequencing (WGS) study on 79 *Salmonella* isolates serotyped as Paratyphi B that were collected from 10 nationwide enteric disease surveillance sites in Bangladesh. Placing these in a global genetic context revealed that these are biotype Java, and the addition of these genomes expanded the previously described PG4 clade containing Bangladeshi and UK isolates. Importantly, antimicrobial resistance (AMR) genes were scarce amongst Bangladeshi *S.* Java isolates, somewhat surprisingly given the widespread availability of antibiotics without prescription. This genomic information provides important insights into the significance of *S.* Paratyphi B biotypes in enteric disease and their implications for public health.

DATA SUMMARY

The genome sequence data generated in this study have been deposited at the European Nucleotide Archive (ENA) under accession numbers ERR4339057–ERR4619485 (accession numbers and metadata available in Tables S1 and S2, available in the online version of this article). Custom R and Python scripts used for comparative pan-genome analysis are available at https://github.com/ghoresh11/Salmonella_ParaB.

INTRODUCTION

The genus *Salmonella*, which belongs to the family *Enterobacteriaceae*, is commonly associated with bacterial foodborne illness in developed countries. The species *Salmonella enterica* consists of several subspecies, the first of which, *S. enterica* subspecies enterica, is commonly split into typhoidal and non-typhoidal *Salmonella* (NTS), based on the disease syndrome [1]. The *Salmonella enterica* serotype Paratyphi B complex (*S.* Paratyphi B complex) causes both potentially life-threatening invasive paratyphoid fever and non-invasive gastroenteritis; both typhoidal and non-typhoidal types share the same somatic O antigen formula (1,4,[5],12 with the b:1,2-type of flagellar H antigen) [2], resulting in a point of confusion for microbiologists. This serotype has been further subdivided into two biotypes based on the ability to ferment dextrorotatory tartrate (*d*Ta) and to form a slime

Received 24 November 2020; Accepted 23 May 2021; Published 22 September 2021

Author affiliations: [1]Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka, Bangladesh; [2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; [3]Department of Medicine, University of Cambridge, Cambridge, UK; [4]London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK; [5]Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004, Australia; [6]Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh.

***Correspondence:** Nicholas R. Thomson, nrt@sanger.ac.uk

Keywords: *Salmonella* Paratyphi B; serotyping; surveillance; whole-genome sequencing; enteric disease.

Abbreviations: AMR, antimicrobial resistance; dTa, dextrorotatory tartrate; ETEC, Enterotoxigenic *E. coli*; icddr,b, International Centre for Diarrhoeal Disease Research, Bangladesh; IEDCR, Institute of Epidemiology, Disease Control and Research; LWS, loose watery stool; NTS, non-typhoidal *Salmonella*; RWS, rice watery stool; SNP, single nucleotide polymorphism; WGS, whole genome sequencing.

Genome sequence data generated in this study have been deposited at the European Nucleotide Archive (ENA) under accession numbers ERR4339057–ERR4619485.

†These authors contributed equally to this work

‡These authors share senior authorship.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Six supplementary tables and four supplementary figures are available with the online version of this article.

wall: *Salmonella enterica* serotype Paratyphi B variant *sensu stricto* (*S.* Paratyphi B *sensu stricto; d*Ta$^-$, slime wall-positive) and *Salmonella enterica* serotype Paratyphi B variant Java (*S.* Java; *d*Ta$^+$; slime wall-negative), which collectively comprise the *S.* Paratyphi B complex [2–4].

Whilst the D-tartrate reaction is used clinically to distinguish these biotypes, it can be unreliable and provides no phylogenetic resolution. Therefore, isolates of the serotype *S.* Paratyphi B complex have been further subtyped by phage typing [5], IS*200* profiling [6], multilocus sequence typing (MLST) [7], clustered regularly interspaced short palindromic repeats (CRISPR) typing [8] and also whole-genome sequencing (WGS). In a recent study, WGS analysis revealed that the *S.* Paratyphi B complex is represented by 10 distinct lineages (phylogroups; PGs), based on the core gene phylogeny [2]: the invasive *S.* Paratyphi B *sensu stricto* (*d*Ta$^-$) isolates grouped into a single lineage (PG1), while the remaining PGs comprised diverse lineages of biotype *S.* Java (*d*Ta$^+$). However, the pathogenic properties of *S.* Paratyphi B *sensu stricto* and *S.* Java remain poorly understood especially in low-income settings, like Bangladesh.

There has been an increase in reports of *S.* Java infections, especially from poultry sources, observed in Germany, the Netherlands and Belgium since 1990 [9] and in the UK since 2010 [10]. In addition, non-European countries such as Saudi Arabia [11] and Bangladesh [12] have also noted an increasing incidence of *S.* Java in poultry farms. In Bangladesh, very few data exist on the prevalence and incidence of *S.* Paratyphi B compared to other typhoidal *Salmonella* such as *S.* Typhi [13] and *S.* Paratyphi A. To provide a genomic snapshot of the *S.* Paratyphi B complex in Bangladesh, we sequenced 79 *Salmonella* isolates previously serotyped as Paratyphi B in nationwide hospital-based enteric disease surveillance in Bangladesh between 2014 and 2018. This is the first WGS-based study characterizing the genetic diversity of *S.* Paratyphi B isolates causing diarrhoeal disease in Bangladesh.

Impact Statement

Salmonella enterica serotype Paratyphi B complex (*S.* Paratyphi B complex) has long been a source of confusion for microbiologists, as the two biotypes in this serotype have until now been indistinguishable by *Salmonella* serotyping. Further, there is still very little molecular information available to understand the population structure of the *S.* Paratyphi B complex in many regions. In 2016, Connor *et al.* reported the utility of whole-genome sequencing (WGS) to distinguish this serotype into two biotypes, *sensu stricto* and Java, which cause, respectively, paratyphoid fever and gastroenteritis. Our study is the first to apply genomics to the *S.* Paratyphi B complex in a hospital-based surveillance study in sites across Bangladesh, where WGS analysis classified these serotyped Paratyphi B as biotype Java, associated with diarrhoeal symptoms. This study reiterates the advantage of WGS studies in addition to molecular and phenotypic methods.

METHODS

Ethics statement

Ethical approval was obtained from the Research Review Committee (RRC) and Ethical Review Committee (ERC) of the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) (reference number PR#12060). Informed written consent was taken from adult participants and the legal guardians of child participants under 18 years old.

Study settings, sample collection and bacteria isolation

This study utilized samples collected from an established nationwide enteric disease surveillance system being carried out in 10 hospitals across 8 divisions of Bangladesh in a collaboration between the Institute of Epidemiology, Disease Control and Research (IEDCR) and icddr,b (Fig. 1, Table 1). The surveillance sites were selected based on reports of acute watery diarrhoea according to the national District Health Information Software v2 Database from Directorate General of Health Services (DGHS) [14, 15]. It is a large, longitudinal, multi-pathogen surveillance study that included diarrhoeal patients infected with a variety of enteric pathogens: *Vibrio cholerae*, ETEC, *Shigella*, and typhoidal and non-typhoidal *Salmonella*. Patients were enrolled into the enteric disease surveillance study if they were over 2 months old, and attended hospital with either (a) loose or liquid stools ≥3 times; (b) loose or liquid stools causing dehydration <3 times; or (c) at least one bloody loose stool in the previous 24 h [14, 15]. Demographic and clinical information, including age, gender, date of illness onset and date of sample collection, was obtained from each participant (Tables 1 and S1).

Stool samples collected from individuals exhibiting diarrhoeal symptoms were cultured by streaking on MacConkey agar and *Salmonella–Shigella* (SS) agar. After overnight incubation at 37 °C, non-lactose-fermenting colonies were inoculated for biochemical testing and those showing typical characteristics of *Salmonella* spp. were serotyped using *Salmonella*-specific somatic O and flagellar H antiserum (Denka Sieken Tokyo, Japan) [16] for confirmation of *S.* Paratyphi B.

DNA extraction, WGS and dataset compilation

Genomic DNA was extracted from the *S.* Paratyphi B strains using the Wizard Genomic DNA kit (Promega, Madison, WI, USA) according to the manufacturer's instructions for genomic analysis. WGS was performed at the Wellcome Sanger Institute (WSI) using the Illumina Hiseq 2500 platform (Illumina, San Diego, CA, USA) to generate 150 bp paired-end reads. Sequence data quality was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

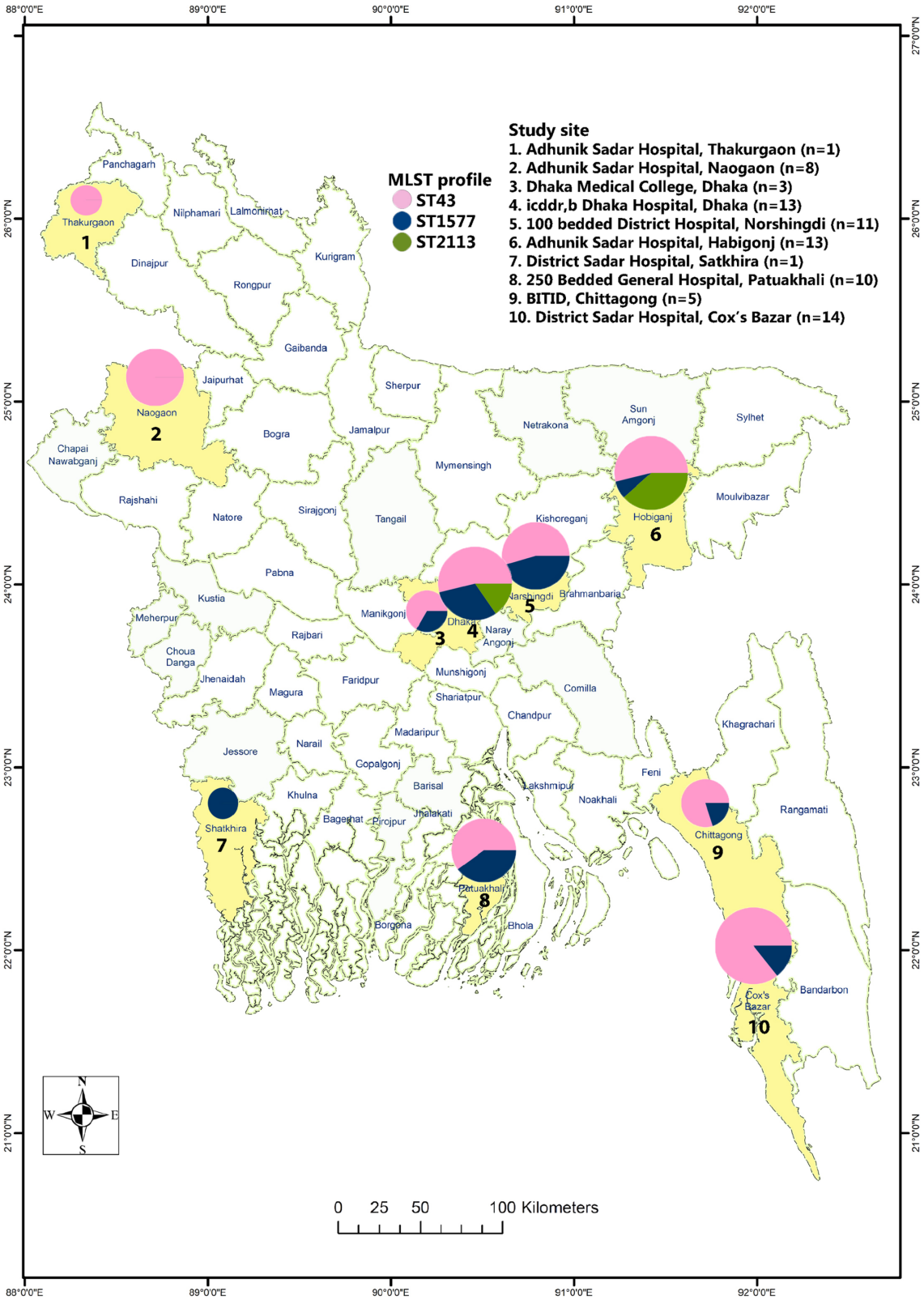

Fig. 1. Map of nationwide study surveillance sites in Bangladesh, June 2014–June 2018. Pie charts at each site depict the MLST distribution. The number of *S*. Java-positive cases (*n*) at each study site is also shown in the key.

Table 1. Prevalence of *S.* Paratyphi B (*n*=107) collected from 10 hospital-based enteric surveillance sites in Bangladesh from June 2014 to June 2018

Study site	Division	Enrolled diarrhoeal patients, *n*	*S.* Paratyphi B-positive, *n* (%)
100 bedded District Hospital, Narshingdi	Dhaka	1156	13 (1.12)
Dhaka Medical College, Dhaka	Dhaka	737	4 (0.54)
icddr,b* Hospital, Dhaka	Dhaka	14889	31 (0.21)
District Sadar Hospital, Cox's Bazar	Chittagong	2089	14 (0.67)
BITID†, Chittagong	Chittagong	1860	8 (0.43)
Adhunik Sadar Hospital, Naogaon	Rajshahi	1605	10 (0.62)
Adhunik Sadar Hospital, Thakurgaon	Rangpur	1715	1 (0.06)
Adhunik Sadar Hospital, Habigonj	Sylhet	2247	14 (0.62)
250 Bedded General Hospital, Patuakhali	Barisal	1805	10 (0.55)
District Sadar Hospital, Satkhira	Khulna	1434	2 (0.14)
Total		29537	107 (0.36%)

*icddr,b; International Centre for Diarrhoeal Disease Research, Bangladesh.
†BITID; Bangladesh Institute of Tropical and Infectious Diseases.

To provide global context for the Bangladeshi *S.* Paratyphi B genomes, 180 *S.* Paratyphi B complex genomes from Connor *et al.* [2] and 12 *S.* Paratyphi B *sensu stricto* genomes from patients presenting with enteric fever symptoms from Higginson *et al.* [4] were also included in this study (Table S2).

Read alignment and SNP calling

Illumina reads for all 271 genomes were mapped against the reference *S.* Paratyphi B strain SPB7 (accession number CP000886) using SMALT v0.7.4 [17], with PCR duplicate reads flagged using Picard v1.92 (http://broadinstitute.github.io/picard). Candidate single-nucleotide polymorphisms (SNPs) relative to the reference having a quality score >30, consensus base quality >20 and read depth >5 were identified using SAMtools [18] and were extracted using SNP-sites [19]. SNPs called in prophage regions and repetitive sequences, or in recombinant regions as detected by Gubbins (v2.3.2) [20], were excluded, resulting in a final SNP alignment of 132593 bp for the 271 *S.* Paratyphi B complex genomes.

Phylogenetic, population genetic analysis and statistical analyses

Maximum-likelihood (ML) phylogenetic trees were inferred from the SNP alignments using RAxML (v8.2.8) [21]. A generalized time-reversible model and a gamma distribution were used to model site-specific rate variation (GTR+Γ substitution model; GTRGAMMA in RAxML) with 100 bootstrap pseudoreplicates used to assess branch support for the ML phylogeny. SNP alleles from PG10 isolates reported in Connor *et al.* [2] were included as an outgroup to root the tree. The resulting phylogenies were visualized and annotated using FigTree (available at: http://tree.bio.ed.ac.uk/software/figtree), iTOL [22] and the R package ggtree [23].

We performed hierarchical Bayesian Analysis of Population Structure (BAPS) implemented in RhierBAPS [24] to redefine the *S.* Paratyphi B subpopulation structure.

To determine the statistical relationships between sequence type distribution and epidemiological factors, we conducted Fisher's exact tests implemented in STATA [25].

De novo genome assembly, annotation and comparative pan-genome analysis

Raw sequence reads were assembled *de novo* using Unicycler (v0.3.0b) [26] and annotation was performed by PROKKA (v1.5) [27]. The quality of genome assemblies were assessed using QUAST (v5.0.2) [28] and the detailed quality reports are summarized in Table S1. The pan-genome was determined with Roary [29] from the annotated assemblies, using a BLASTP percentage identity of 95% and a core definition of 95% of the included isolates. To estimate the openness of the pan-genome(s), we used the Heaps function within the Micropan R package [30], which calculates the curve fit constant according to Heaps' law [31]: $n=k*N^{-\alpha}$, where n is pan-genome size, N is the number of genomes and k, γ are curve-specific constants [32]. The curve specific constant, $\alpha=1-\gamma$ determines whether the pan-genome of a bacterial variant (e.g. species, biotype or lineage) is closed ($\gamma<0$, $\alpha>1$) or open ($0<\gamma<1$, $\alpha<1$).

Comparative pan-genome analysis, using custom R and Python scripts available at https://github.com/ghoresh11/Salmonella_ParaB, was performed to identify biotype- and clade-specific genes. The frequency of each gene in the pan-genome amongst all *S.* Paratyphi B *sensu stricto* and amongst all *S.* Java isolates was calculated. Similarly, the frequency of all genes in BAPS cluster 1.1 (PG3/PG4 clades) relative to the rest of the clades were calculated. A gene was defined

as core and specific to a biotype or clade if it was present in more than 95% of one biotype/clade and absent from more than 95% of the isolates of the other biotype/clade. To investigate the synteny in the loci containing the biotype- or clade-specific genes, a synteny graph similar to that presented elsewhere [33] was constructed in the regions of interest. A region was defined by the two flanking genes that were consistently identified upstream of and downstream to the biotype/clade-specific loci. A graph was constructed from the annotation files such that each node in the graph is a gene, and the weighted edge between two genes represents the number of times they were adjacent to each other across all genomes. The results of these analyses were visualized using Phandango [34] and Cytoscape [35]. In addition, we used the basic local alignment search tool (BLAST) [36] to identify the distribution of group-specific genes throughout the species, InterProScan (v5) [37] to predict the function of the group-specific hypothetical proteins and EffectiveDB [38] to predict secreted proteins.

Antimicrobial resistance (AMR) gene detection, plasmid detection, virulence factor detection, *in silico* serotype prediction and MLST analysis

We detected AMR genes and plasmid replicons using ARIBA [39] in conjunction with the comprehensive antibiotic resistance database (CARD) [40] and the PlasmidFinder [41] database, respectively. We used the same approach to detect virulence factors, using the Virulence Factor Database (VFDB) [42]. The Salmonella In Silico Typing Resource (SISTR) [43], implemented in PathogenWatch [44], was used for *in silico* serotype prediction of the sequenced genomes. The mapping-based allele typer SRST2 [45] was used to assign sequence types (STs) to each genome according to the *S. enterica* MLST database.

RESULTS

Demographic and clinical characteristics of *S.* Paratyphi B strains isolated from diarrhoeal patients and their genome assembly metrics

The goal of this study was to investigate the prevalence and genomic diversity of *S.* Paratyphi B complex in Bangladesh. A total of 29537 diarrhoeal patients presenting to 1 of 10 sentinel surveillance sites were enrolled into this study between June 2014 and June 2018 (see the Methods section for further details). Of these patients, 0.36% (107/29537) were confirmed as *Salmonella enterica* serotype Paratyphi B-positive by serotyping, with the antigenic formula O1,4,5,12:Hb:1,2. The percentage of patients presenting with *S.* Paratyphi B at each surveillance site was low, ranging from 0.06–1.12%, compared to other enteric infections (for example the equivalent range for *Vibrio cholerae* was 1.10–18.3% of patients [14]) (Table 1). The Narshingdi district hospital in the Dhaka division had the highest percentage of *S.* Paratyphi B-positive patients (1.12%) in this study.

Of the 107 *S.* Paratyphi B-positive samples, only 79 *S.* Paratyphi B strains were available for sequencing in this study. Sequencing of these genomes produced assemblies containing on average 36.39 contigs (≥1000 bp) and the total assembly lengths (consisting of contigs ≥1000 bp) ranged from 4619574 to 4792431 bp (an average of 4664961 bp); the expected size for *S.* Paratyphi B genomes. The mean of scaffold N50 sizes was 375447 bp (range 289235 to 399505 bp) (Table S1). For our downstream analysis we utilized all contigs over 1000 bp in length.

Patient metadata including age and sex data as well as clinical symptom data were available for most patients (Tables 2 and S1). Among these, 63 patients were adults (17–85 years of age; median age 31 years; Table 2), with 39 of these aged between 17 and 35, and 11 were young children (≤5 years old). The majority of the patients were female (*n*=50; 68%). Loose watery stool (LWS) was reported more frequently, with a longer duration of diarrhoea (*n*=43 with average duration 2.61 days), than rice watery stool (RWS) (*n*=33 with average duration 1.72 days). The most common combination of clinical symptoms, recorded in 12 patients, was RWS with vomiting, some or severe dehydration and abdominal cramping. No bloody diarrhoea was reported among the Bangladeshi *S.* Paratyphi B-positive population (Table 2). We did not observe any co-infections with any other enteric pathogens targeted in the surveillance study.

Population structure of the *S.* Paratyphi B complex in Bangladesh

To further classify the biotype of the isolates serotyped as *S.* Paratyphi B in the surveillance study, and investigate the phylogenetic relationships between the Bangladeshi and global isolates belonging to the *S.* Paratyphi B complex, we constructed a global phylogeny which included 192 contextual *S.* Paratyphi B complex genomes originating from over 20 countries [2, 4] and the 79 Bangladeshi genomes sequenced in this study (Fig. 2). Previously, Connor *et al.* reported that PG1 comprised *S.* Paratyphi B *sensu stricto*, while *S.* Java genomes were represented by PG2 to PG10 [2]. Our WGS data revealed that all 79 Bangladeshi isolates serotyped as *S.* Paratyphi B were classified as biotype Java. This is consistent with the clinical data, which showed that all isolates were taken from patients presenting with non-invasive diarrhoeal disease (Table 2). Our genomes clustered within two of the previously described *S.* Java clades; either PG3 (*n*=2) or PG4 (*n*=77) in the previously published global phylogeny [2], with up to 4709 SNPs separating genomes in the two clades (median SNP distance of 2850 bp).

Furthermore, the Bangladeshi *S.* Java isolates contributed substantially to an expansion of the known PG4 diversity, which originally mainly comprised *S.* Java isolates originating from the UK. PG3, on the other hand, only contained two Bangladeshi *S.* Java genomes, with the remainder of the clade comprising isolates from the UK, continental Europe, the USA, and South and Southeast Asia (Fig. 2). Whilst our genomes fall within the previously defined PGs, the addition of our 79 Bangladeshi *S.* Java isolates to the published

Table 2. Demographic and clinical characteristics of *S.* Paratyphi B-positive patients (*n*=79) in this study

Characteristics	*S.* Paratyphi B-positive, *n* (%)
Demographic factors	
Age (years)*	
0–5	11 (14.86)
17–85	63 (85.14)
Median age of patients (IQR)	31 (24.22)
Sex†	
Male	23 (31.51)
Female	50 (68.49)
Clinical factors	
Stool nature‡	
Loose watery	43 (55.84)
Rice watery	33 (42.85)
Bloody	0 (0.0)
Formed	1 (1.29)
Dehydration status‡	
None	14 (18.18)
Some	48 (62.33)
Severe	15 (19.48)
Abdominal cramp§	
Yes	47 (67.14)
No	23 (32.85)
Vomiting‡	
Yes	57 (74.02)
No	20 (25.97)
Duration of diarrhoea (days)‖	
No diarrhoea	6 (9.09)
1	19 (28.79)
2	23 (34.85)
3	12 (18.18)
4	3 (4.55)
5	3 (4.55)

*This information was available for 74 patients.
†This information was available for 73 patients.
‡This information was available for 77 patients.
§This information was available for 70 patients.
‖This information was available for 66 patients.

phylogeny disagreed with the original definition of phylogroups PG3 and PG4 by Bayesian hierarchical clustering, which in our updated phylogeny, merged PG3 and PG4 into a single cluster at BAPS level 1 (ascribed BAPS cluster 1.1) (Fig. 2). Potential reasons for this discrepancy are noted in the Discussion.

To analyse the Bangladeshi *S.* Java population structure in finer detail, we constructed a phylogeny of the *S.* Java isolates belonging to BAPS cluster 1.1 (*n*=123) from 20908 chromosomal SNPs across the whole genome (Fig. 3). This revealed that the population structure of *S.* Java in Bangladesh is characterized by three STs: ST43 (*n*=53, 67.1%), ST2113 (*n*=19, 24.1%) and ST1577 (*n*=7, 8.9%) (Fig. 1). ST43 is a globally distributed ST, while STs 2113 and 1577 have only been detected in the UK and Bangladesh; the former was detected for the first time in Bangladesh in this study (Fig. 3).

We then defined seven sub-clusters at BAPS level 2: PG3 isolates grouped into sub-clusters 2.2, 2.3 and 2.7 and PG4 grouped into 2.1, 2.4, 2.5 and 2.6. BAPS sub-clusters 2.6, 2.5 and 2.4 corresponded with STs 43, 2113 and 1577, respectively (Fig. 3). We did not observe any phylogeographical clustering of isolates within Bangladesh, nor was there a significant difference between the ST distributions in the Dhaka sites combined (*n*=3 sites) relative to their distribution in the rest of the regions in Bangladesh (*n*=7 sites) (*P*=0.121, Fisher's exact test) (Fig. 1). Only two hospital sites (icddr,b hospital and Habigonj district hospital) harboured all three STs. ST43 was distributed throughout all study sites in Bangladesh except Satkhira, from which only one isolate was obtained. We did not observe any significant association between age group (children, adult) or sex (male, female) and the distribution of *S.* Java STs, respectively, (*P*=0.361 and *P*=0.469, using Fisher's exact test) (Table S1).

Examining ST distribution in the context of clinical characteristics, isolates typed as ST2113 and ST1577 were commonly associated with patients with LWS (*n*=7, 100% and *n*=13, 72%, respectively). Whereas, LWS and RWS were reported at similar frequencies (*n*=22 and *n*=28, respectively) for ST43 isolates. Of note, among the 15 cases with severe dehydration, 73% (*n*=11) were associated with ST43. Furthermore, the duration of diarrhoea differed throughout the surveillance sites: for example, the average duration of diarrhoea in Habigonj was 2.83 days (*n*=12) compared to only 1.43 days (*n*=14) in Cox's Bazar (Table S1).

We explored the distribution of virulence factors (VFs) throughout the Bangladeshi *S.* Java genomes in relation to the clinical characteristics and the BAPS sub-clusters. A detailed list of the virulence genes detected in a total of 271 genomes is summarized in (Table S3). RWS was observed more frequently in BAPS sub-cluster 2.6 than the other sub-clusters (*n*=28 and 5, respectively; Table S1). Interestingly, the genomes in this sub-cluster harboured the *tcfABCD* genes, which were absent from all other lineages except for PG7 (Fig. S1). These genes encode the Typhi colonization factor and are also present in other NTS serovars [46, 47]. Interestingly, these same genomes lacked the fimbrial *stfACDEFG* genes [48], which were present in all other lineages except for PG6. We did not identify VFs that were specific to Bangladeshi isolates, and all other differences we observed between the lineages were reported by Connor *et al.* [2].

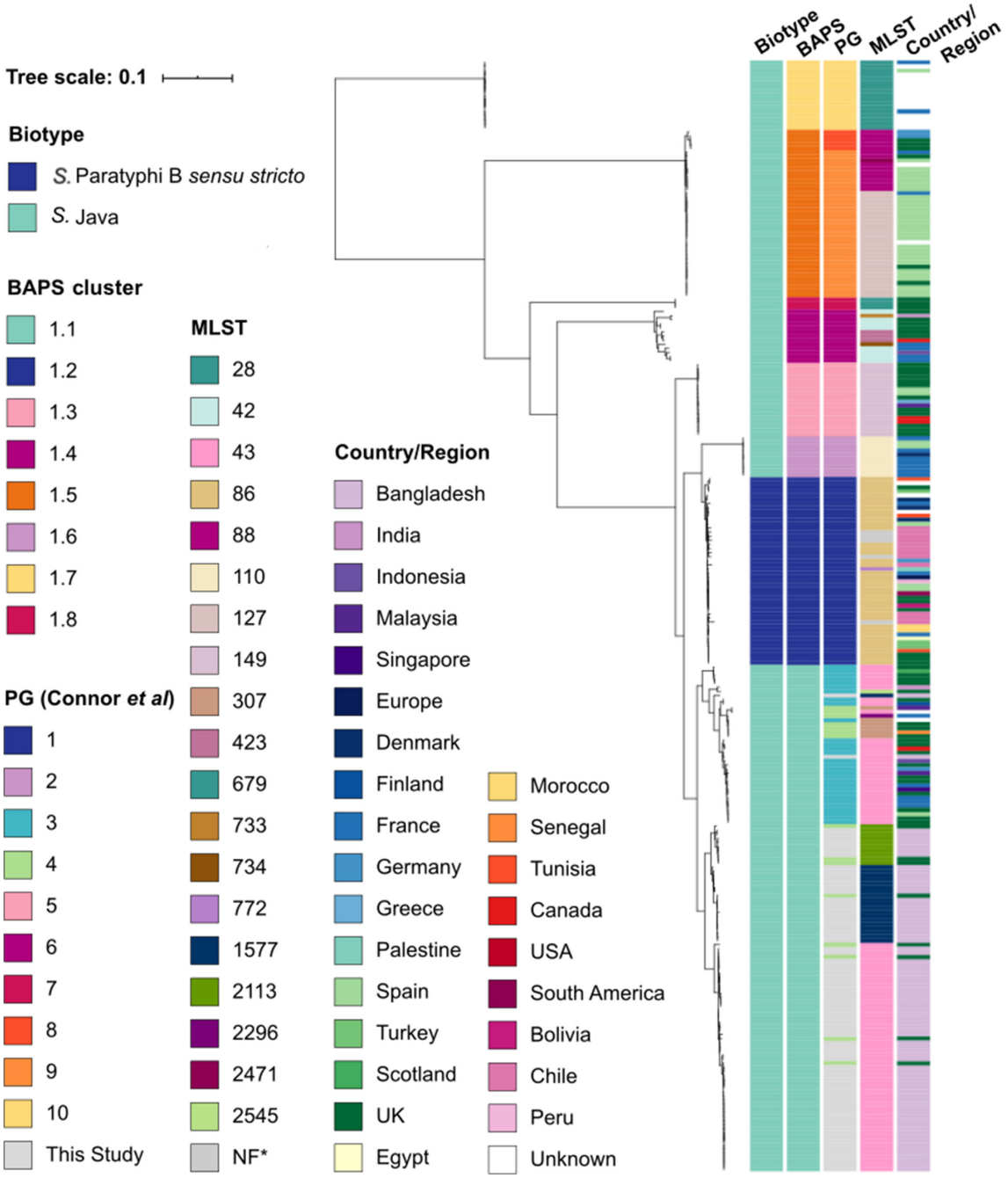

Fig. 2. Maximum-likelihood outgroup-rooted phylogenetic tree of 271 *S.* Paratyphi B strains from the global collection, including Bangladeshi *S.* Java isolates from this study. Whole-genome SNP tree with recombination regions removed and outgroup rooted with PG10/BAPS1.7. The coloured strips show the biotype, BAPS cluster (this study), PG (Connor *et al.* [2]), MLST and country or region of isolation for each isolate; see colour legend. The tree scale bar indicates the estimated mean number of nucleotide substitutions per site.

Comparative pan-genome analysis

To investigate gene distribution among the *S.* Paratyphi B complex biotypes, we conducted a core- and pan-genome analysis on all 271 genomes. This revealed that 11929 genes comprised the *S.* Paratyphi B complex pan-genome (271 genomes; PG1-10), with 3706 genes in the core genome (present in ≥95% of genomes) and 8223 in the accessory genome (present in <95% of genomes). Further, the pan-genome sizes of *S.* Paratyphi B *sensu stricto* (46 genomes; PG1) and *S.* Java (225 genomes; PG2-10) were 4930 and 11625 genes, respectively. Within these, we determined 4141 core and 789 accessory genes for *S.* Paratyphi B *sensu stricto* and 3787 core and 7838 accessory genes for *S.* Java (Fig. 4a).

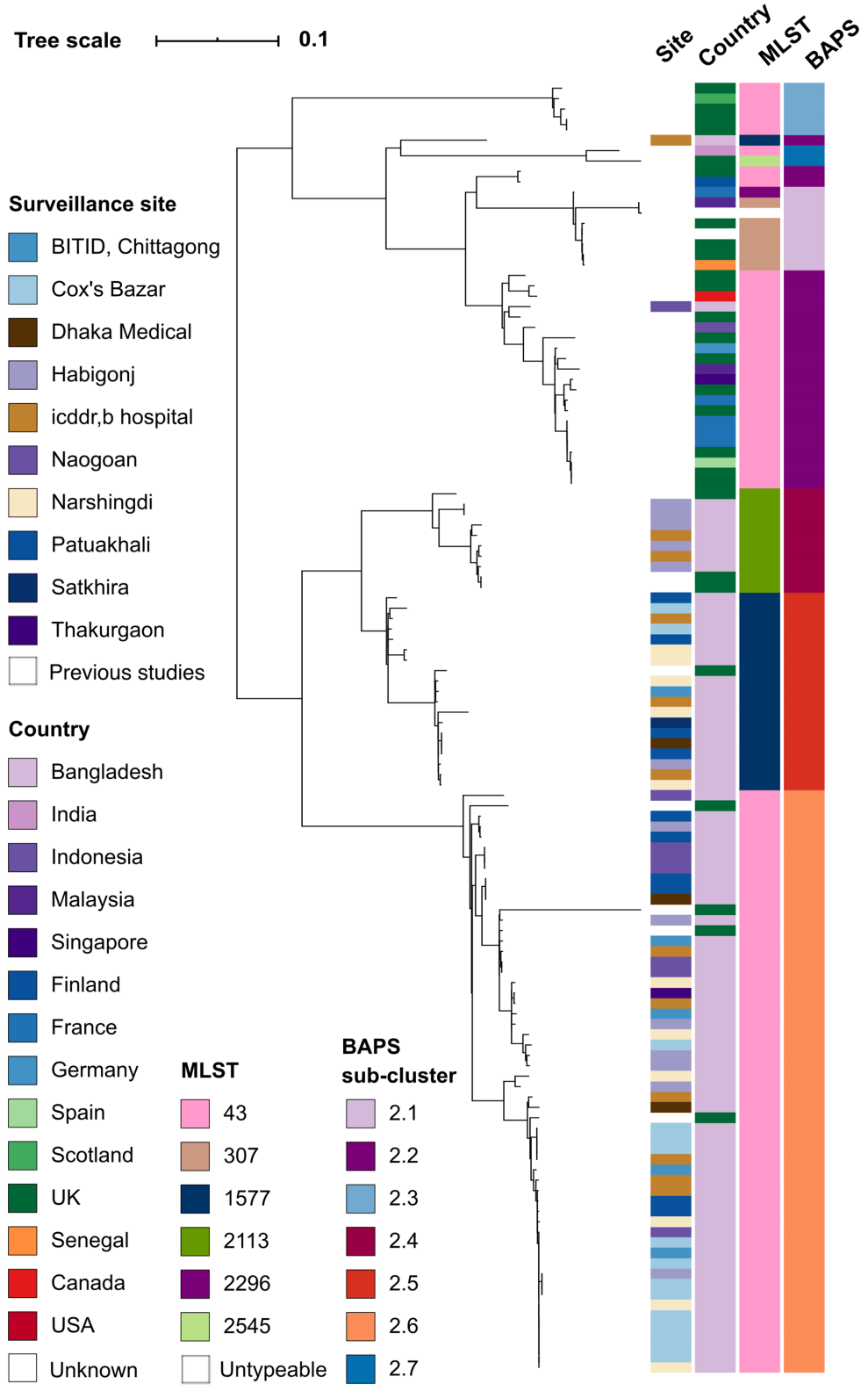

Fig. 3. Mid-point-rooted maximum-likelihood phylogeny of *S.* Java cluster 1.1 (PG3-4). Whole-genome SNP tree with recombination regions removed and mid-point rooted. The coloured strips alongside the tree show the surveillance site, country or region of isolation, MLST and BAPS sub-cluster for each isolate; see colour legend. The tree scale bar indicates the estimated mean number of nucleotide substitutions per site.

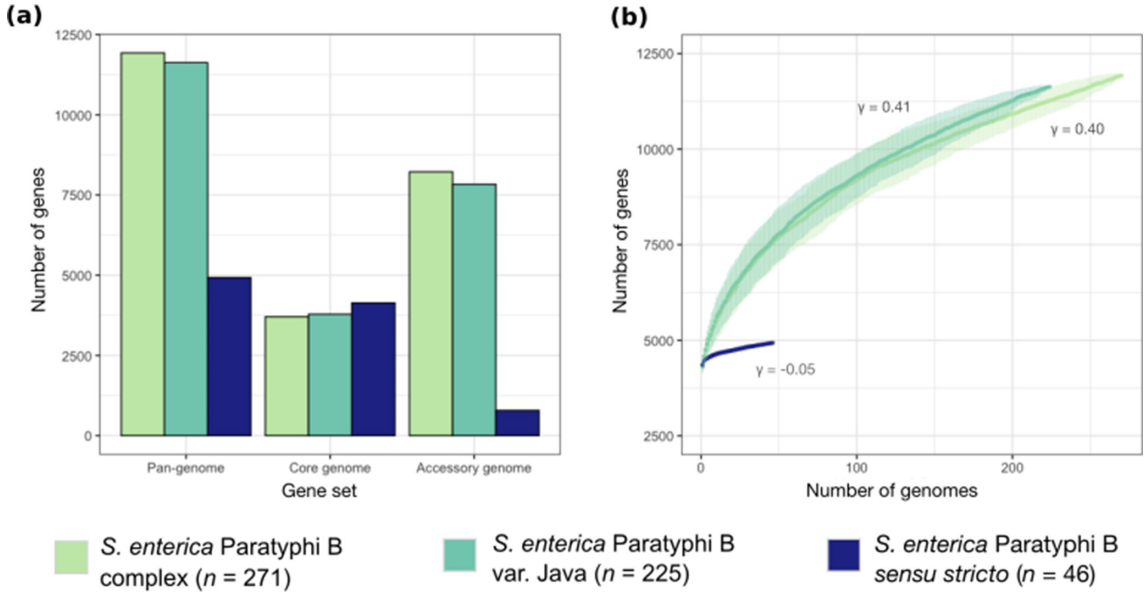

Fig. 4. Pan-genome dynamics of the *S*. Paratyphi B complex. (a) The pan, core and accessory genomes and (b) gene accumulation curves are depicted for the *S*. Paratyphi B complex and each biotype; see key for colour. Here, core genes are defined as genes present in ≥95% of strains and accessory genes are present in <95% of strains. Error bars above and below the median are depicted by shading above and below the curve in (b).

The gene accumulation curve for the *S*. Paratyphi B complex, carried out in accordance with Heaps' law [31, 32] (see Methods section) is driven by the diversity within *S*. Java. This is demonstrated by similar curve fitting parameter values for the complex and *S*. Java of γ=0.40 and γ=0.41, respectively, while that of *S*. Paratyphi B *sensu stricto* is much lower (γ=−0.05) (Fig. 4b). This suggests that within the *S*. Paratyphi B complex, the pan-genome of *S*. Java remains open and, as more strains are sequenced, new genes will be identified among these organisms, while the gene accumulation curve of *S*. Paratyphi B *sensu stricto* converged rapidly and is approaching closed.

To investigate gene flux within the *S*. Paratyphi B complex, we identified biotype-specific genes and loci (Table S4). We defined a core and specific gene as one present in ≥95% of genomes of one group and absent in >95% of the other group. Based on these criteria, we identified 20 core and specific genes for *S*. Paratyphi B *sensu stricto* (Fig. S2a) and 30 core and specific genes for *S*. Java (Fig. S2b). We confirmed that the *S*. Paratyphi B *sensu stricto*-specific genes form a single gene block, at the same locus. This gene block was assembled on the same contig in ≥93% of genomes and split over different contigs in the remaining ~7% of genomes. On the other hand, the *S*. Java-specific genes are clustered at three genetic loci (in ≥99% of genomes).

The three *S*. Java-specific loci are located within a 181 CDS-long chromosomal region and differ in length (Fig. S3). The first of these loci contains only a single 1134 bp gene (hypothetical protein; group_270; *S*. enterica serotype Paratyphi B str SPB7 v1 locus tag 01289), predicted to encode a hypothetical protein with a carboxymuconolactone decarboxylase (CMD) domain that may have peroxidase activity. The gene is positioned between *rnb* and *fabI* (Figs S2a and S4a). Given its proximity to the SPI-2 effector *steC*, we ran this gene through T3SS effector prediction software, which predicted that the gene product may be secreted.

The second *S*. Java-specific locus carried genes predicted to encode the ABC transporters *ArtM*, *YecS* and *FliY* (also referred to as *TcyJ*), involved in amino acid transport, as well as the transcriptional repressor *FrmR* and the SPI-2 type III secretion system effector protein *SseJ* (Figs S2a and S4b). Homologues of some of these genes, excluding *sseJ*, are found at other distinct, conserved, loci in *S*. Paratyphi B *sensu stricto*, suggesting that these genes are not essential for pathogenicity and have been lost by *S*. Paratyphi B *sensu stricto*. Interestingly, synteny analysis showed that in BAPS 1.7 (PG10), a primarily animal-associated clade, one of the hypothetical proteins (group_1829) has been replaced by group_2748 (Fig. S4b).

The final locus encodes numerous hydrolase and oxidoreductase enzymes involved in metabolism of amino acids, carbohydrates and nitric oxide (*norV, glsA, gabD, sad, gutB, yjjL, cbh, hoxK*) and the *hyaABC* genes and their chaperones (Figs S2b and S4c), which may facilitate the ability to utilize locally generated hydrogen. Hence, this locus encodes several genes involved in tolerance to stressors often associated with the gut niche and also encoded on this locus are outer-membrane protein (*ompC*) and tetracycline resistance gene (*tetA*). Although the genes in this locus were absent from the *S*. Paratyphi B *sensu*

stricto genomes (Figs S2b and S3), BLAST analysis revealed that some genes in this group have homologues in other *Salmonella* serovars, including *S.* Enteritidis, *S.* Typhimurium and *S.* Kentucky. This suggests they have potentially been lost by *S.* Paratyphi B *sensu stricto,* rather than gained by *S.* Java. Moreover, this locus is flanked at one end by the non-coding RNA STnc560, includes STnc170 and the Hfq binding RNA *isrF*, and has a selenocysteine insertion sequence SECIS_3 upstream. The diversity observed with respect to the gene arrangement at this locus further supports the gene loss hypothesis, with clade-specific variations noted (Fig. S4c).

The *S.* Paratyphi B *sensu stricto*-specific locus appears to be a bacteriophage/prophage, spanning approximately 44000 bp, containing approximately 59 genes (Fig. S2a, Table S4), many of which have phage-related annotations, while others are predicted to encode hypothetical proteins. The borders of this locus are difficult to define as (a) the gene order is not conserved in all genomes, (b) assemblies are fragmented in this region and (c) some genes within this region have homologues in other clades. Among numerous hypothetical proteins in this locus are bacteriophage-related genes such as the bacteriophage Mu F-like protein, proposed to be involved in viral capsid assembly. This locus also encodes SopE, which is a SPI-2 secreted effector protein also encoded by *S.* Typhi that induces nitric oxide synthetase (iNOS) in the host intestine, leading to inflammation [49]. It has been shown that this effector can be transferred between unrelated phages associated with different serovars [50]. No other SPI-2 effectors appear to be missing from the *S.* Java genomes. Further, BLAST analysis of the ~33000 bp region on the forward strand flanked by *dicA* and *sopE* revealed a high nucleotide identity to *S.* enterica serovar Typhi genomes (94–97% across 51–62% of the query region). A homologue of *sopE*, *sopE2*, which activates a different set of Rho GTPases to SopE, is encoded in the majority of Spanish isolates in BAPS cluster 1.5 (PG9), but none of the Bangladeshi isolates. This gene is also encoded by *S.* Typhimurium.

A second locus was found to be specific to *S.* Paratyphi B *sensu stricto* (BAPS cluster 1.2; PG1) and its close relative, BAPS cluster 1.6 (PG2) (Fig. S1a). This locus is predicted to encode some sugar metabolizing and transport enzymes (*gutB, yggF, cmtB, mtlA*) and provides evidence of compensatory mechanisms with respect to the sugar metabolizing enzymes in the *S.* Java-specific gene set.

Using this same approach, we identified five genes that were specific and core for BAPS cluster 1.1 (containing all the Bangladeshi isolates), relative to all other BAPS level 1 clusters. One of the predicted genes, the SPI-2 effector *sseI*, is located in a small gene cluster with a hypothetical protein and a transposase. The other two genes appear to be associated with a clade-specific phage that is also present in PG7, and/or some of PG6, mostly in isolates from the UK.

AMR and plasmid profiles of *S.* Paratyphi B complex

Next, we examined AMR and plasmid replicon gene distribution among the *S.* Paratyphi B complex (Fig. 5). All allelic variants of AMR genes including *gyrA* point mutations and plasmid replicons detected in the 271 genomes are summarized in Tables S5 and S6. The majority of genotypic antimicrobial resistance in *S.* Java was encoded by genomes in the animal-associated clade (PG10/BAPS cluster 1.7; Fig. 5). Besides that, we observed a stark lack of evidence for widespread extrinsic resistance gene and plasmid acquisition in the human-associated clades of the *S.* Paratyphi B complex, particularly in the Bangladeshi *S.* Java isolates. Three (3.9%) Bangladeshi *S.* Java genomes carried *mphA, qnrB,* $bla_{\text{DHA-7}}$ and *sul1* genes, which are predicted to confer resistance to macrolide, fluoroquinolone, beta-lactam and sulfonamide antibiotic classes, respectively (Fig. 5) [40]. Additionally, two of these isolates also carried *mphE* and *msrE,* encoding resistance to macrolide or erythromycin and streptogramin, respectively [40]. These gene sets were distinct from those seen in global *S.* Java isolates (n=10) carrying bla_{CARB}, *aadA*, *floR* and *sul1* genes (Fig. 5), which are predicted to confer resistance to beta-lactam, aminoglycoside, chloramphenicol and sulfonamide, respectively.

While some of the Bangladeshi *S.* Java isolates harbouring AMR genes also carried plasmids, these AMR gene sets could not be co-located on the same contigs as the rep genes that define the plasmid Inc groups, due to fragmentation of the assemblies. Among the plasmid Inc types we found in the Bangladeshi *S.* Java population were one IncFIB (pHCM2), two IncI and three IncFII (S), IncFIB plasmids (Fig. 5). These plasmids generally encoded putative genes related to DNA metabolism and replication rather than virulence-associated determinants and AMR genes [51, 52], which could provide another explanation as to why we did not find AMR genes co-located with rep genes.

DISCUSSION

The serovar *S.* Paratyphi B is a source of confusion as biotype *sensu stricto* is a cause of invasive paratyphoid fever, while biotype Java is associated with non-invasive gastroenteritis. In this study, we coupled an existing nationwide enteric disease surveillance study across Bangladesh with a WGS approach to investigate the genomic epidemiology of *S. enterica* serotype Paratyphi B complex. Our surveillance showed that, in Bangladesh, the prevalence of *S.* Paratyphi B isolates is low (0.36%) compared to other enteric pathogens in this surveillance study [14]. Moreover, where previous studies have only reported serotyping results of *S.* Paratyphi B strains [53], this is the first study in Bangladesh to distinguish between *S.* Java and *S.* Paratyphi B *sensu stricto* biotypes, using WGS to confirm that *S.* Java is the variant responsible for the diarrhoeal disease in Bangladesh. This is in line with previous reports in which *S.* Java, not *S.* Paratyphi B *sensu stricto*, is the aetiological agent of non-invasive gastroenteritis [2–4, 6], and fits with the clinical characteristics displayed by patients in our study.

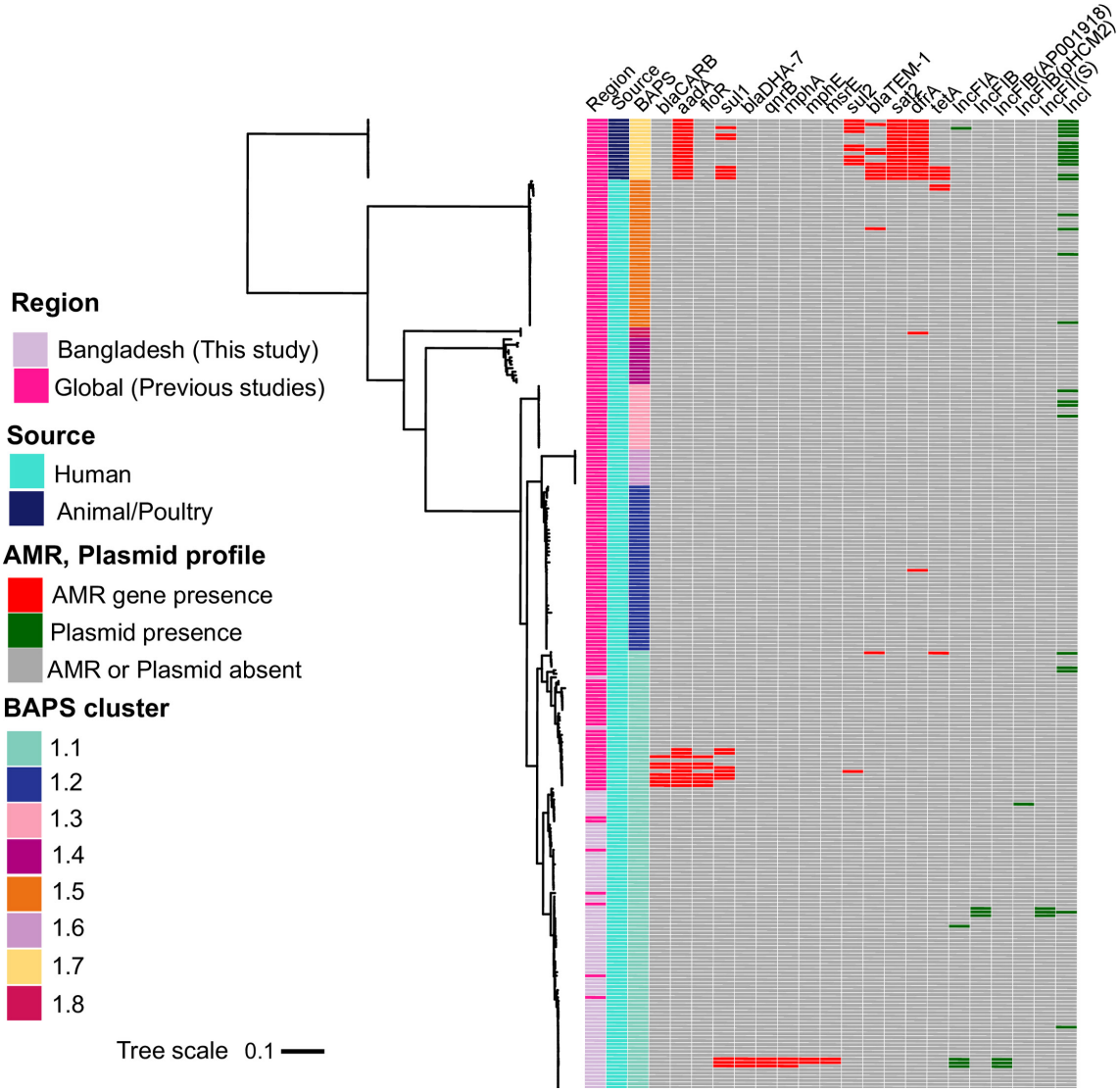

Fig. 5. Antimicrobial resistance gene and plasmid replicon distribution among the *S.* Paratyphi B complex. A maximum-likelihood outgroup-rooted tree of 271 strains from the global collection, including Bangladeshi *S.* Java isolates from this study, alongside a presence/absence matrix of AMR genes and plasmid replicons for each isolate. Only genes that were differentially detected between global and Bangladeshi human-associated *S.* Java (BAPS cluster 1.1), or global human- and animal-associated global *S.* Java (BAPS cluster 1.7), are shown, in order to observe the differences in AMR gene profiles between these subgroups. We omitted genes that were ubiquitous throughout the phylogeny, as well as genes that were only present in three or fewer genomes, unless they were Bangladeshi *S.* Java genomes. The full gene matrices can be found in Tables S4–S6. BAPS clusters, geographical region and sample source are also depicted by the colour strips (see colour legend). The tree scale bar indicates the estimated mean number of nucleotide substitutions per site.

All except two of the Bangladeshi *S.* Java isolates clustered with isolates from the UK. This may be evidence of inter-continental long-range transmission; however, this is more likely explained by the lack of genome sequencing for isolates within this complex, and provides support for continued multi-pathogen genomic surveillance efforts. Our data show that *S.* Java is a globally relevant enteric pathogen in both high- and low-income settings, with clear signs of recent population expansions in clinically relevant lineages. The two *S.* Java lineages present in Bangladesh were linked with three STs, the most prevalent of which, ST43, is a globally distributed ST; seen in Singapore, Indonesia, India, Bangladesh, Malaysia, France, Finland, Germany, Spain, the UK and Canada (Fig. 2). On the other hand, STs 1577 and 2113 were only observed in Bangladesh and the UK. In addition, STs 43 and 1577 were reported previously in both poultry- and human-associated *S.* Java isolates in Bangladesh [12]. The finding of the same STs in human and animal *S.* Java isolates in Bangladesh suggests that further sampling and WGS of a variety of sources could provide insights into reservoirs of

global STs, with WGS data providing higher discriminatory power for lineage placement than MLST alone. While our WGS study allowed us to describe the population dynamics of *S.* Java in Bangladesh, and identify in what proportions globally distributed or endemic STs are present, MLST is a lower-cost alternative to genomic surveillance and such information will facilitate long-term tracking of the population dynamics, supported by genomic surveillance where possible.

While our phylogenetic analysis closely resembled that of Connor *et al.* [2], the addition of these genomes into the existing phylogeny resulted in some minor changes to the placement of genomes within the previously described PGs 3 and 4, which upon updated BAPS analysis, were merged together into what we have called here BAPS cluster 1.1. There are three main differences between our updated phylogeny and the one in Connor *et al.* [2] that would account for these discrepancies: first, we did not use genomes from additional *Salmonella* serovars; second, our tree is constructed from an alignment of SNPs relative to the reference genome *S.* Paratyphi B SPB7 (whereas theirs is constructed from an alignment of core gene SNPs); and last, our tree contains 79 *S.* Java isolates from Bangladesh – a country that until now has been under-represented for bacterial pathogen WGS studies [2]. The differences between these phylogenetic trees reflect the differences in the aims of the studies.

Despite widespread and unregulated mis-use of antimicrobials in Bangladesh, we were surprised to observe little evidence of horizontal gene transfer of AMR genes. This contrasts with earlier studies in Scotland, England, Wales and the Channel Islands – where antimicrobial stewardship is tighter – which reported a range of AMR spectra in both in *S.* Java-infected patients and poultry since 2000 [54]. The latter study, however, did not take a genomic approach, and hence it is unknown (a) which lineages these samples belonged to and b) on which genetic element they were encoded. These factors could help explain the discrepancy between their results and ours. Importantly, our findings are supported by a recent study reporting chromosomal integron class 2-mediated vertical AMR gene inheritance in the PG10 poultry-associated *S.* Java lineage but limited acquisition of AMR genes in human *S.* Java lineages (Fig. 5) [2, 55]. There are several possible explanations for the lack of AMR genes in Bangladeshi *S.* Java. First, in general, NTS infection is a self-limiting disease, mostly manifesting as gastroenteritis, and rarely requires antimicrobial therapy, although they are sometimes taken to reduce the acuteness of symptoms [56]. Second, plasmids, which are known to carry AMR genes and act as a mode of dissemination of resistance determinants in enteric bacteria, were found rarely in Bangladeshi *S.* Java. The AMR genes we did find in our dataset are generally chromosomally encoded [2]. Worryingly, three *S.* Java isolates carried fluoroquinolone resistance genes. Fluoroquinolone-resistant *Salmonellae* are on the World Health Organization (WHO) priority list of bacteria for which new antibiotics are urgently needed. These findings highlight selective pressure towards resistance in circumstances where control and antimicrobial stewardship are challenging.

While our sample size was relatively small, we are confident that it is representative of the prevalence and dynamics of *S.* Java in Bangladesh, having covered eight divisions and a 4-year timespan. However, due to the low sample numbers, neither the ST distribution across the sites, nor the relationship between STs with symptoms or severity, reached statistical significance. A further limitation of our study is the fragmentation of assemblies: this prevented us from confidently assigning AMR genes to plasmids. Future work could include long-read sequencing to address this. Lastly, we were unlikely to detect *S.* Paratyphi B *sensu stricto* as the surveillance study underpinning the genomics was targeted to enteric pathogens. However, *S.* Paratyphi B *sensu stricto* is very rare globally [4], and symptoms can include diarrhoea; future surveillance efforts to include blood samples will increase the likelihood of our detecting *S.* Paratyphi B *sensu stricto* if it is present in Bangladesh.

This study highlights the importance of developing genomic surveillance systems in all settings in order to answer changing patterns of disease both nationally and internationally. We have used this approach to characterize the polyphyletic population structure of *S. enterica* serotype Paratyphi B and resolve the confusion associated with the spectrum of clinical symptoms. Our study provides a framework for future hospital surveillance-based genomic epidemiology studies in low-income countries. Continued molecular-based surveillance incorporating both WGS and MLST approaches will provide further information that can be used to design and implement better diagnostic tests, hence facilitating treatment options, and informing public health interventions in poor resource settings such as Bangladesh.

Funding information
This work was supported by the Wellcome Sanger Institute, Cambridge, UK and International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b). This study was supported by the grants from the Wellcome Trust (STRATAA; grant 206194) and the Bill and Melinda Gates Foundation (grant no. OPP1135223) (TyVAC). G. D. was also supported by NIHR BRC funding for AMR research.

Acknowledgements
We acknowledge the support of the dedicated field and laboratory workers at the icddr,b involved in this study. The icddr,b is grateful to the governments of Bangladesh, Canada, Sweden and the UK for providing core/unrestricted support. We would like to thank Matthew Dorman and Sally Kay as well as the members of the Pathogen Informatics and core sequencing teams at the Wellcome Sanger Institute. We also thank Kate Mellor and Grace Blackwell for helpful discussions around AMR gene and plasmid distribution and detection methods.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. *Salmonella* nomenclature. *J Clin Microbiol* 2000;38:2465–2467.
2. Connor T, Owen SV, Langridge G, Connell S, Nair S. What's in a name? Species wide whole genome sequencing resolves invasive and non-invasive *Salmonella* Paratyphi B. *mBio* 2016;7.
3. Chart H. The pathogenicity of strains of *Salmonella* paratyphi B and Salmonella java. *J Appl Microbiol* 2003;94:340–348.

4. Higginson EE, Ramachandran G, Hazen TH, Kania DA, Rasko DA, *et al*. Improving our understanding of salmonella enterica Serovar paratyphi B through the engineering and testing of a live attenuated vaccine strain. *mSphere* 2018;3:e00474-00418.
5. Barker RM, Kearney GM, Nicholson P, Blair AL, Porter R. Types of *Salmonella* paratyphi B and their phylogenetic significance. *J Med Microbiol* 1988;26:285–293.
6. Ezquerra E, Burnens A, Jones C, Stanley J. Genotypic typing and phylogenetic analysis of *Salmonella* paratyphi B and S. java with IS200. *Microbiology* 1993;139:2409–2414.
7. 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.
8. Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. *PloS one* 2012;7:e36995.
9. Authority EFS. Report of the Task Force on Zoonoses Data Collection on the Analysis of the baseline survey on the prevalence of *Salmonella* in broiler flocks of Gallus gallus, in the EU, 2005–2006-Part B: factors related to *Salmonella* flock prevalence, distribution of *Salmonella* serovars, and antimicrobial resistance patterns. *EFSA Journal* 2007;5:101r.
10. Gobin M, Launders N, Lane C, Kafatos G, Adak B. National outbreak of *Salmonella* Java phage type 3b variant 9 infection using parallel case-control and case-case study designs, United Kingdom, July to October 2010. *Euro Surveill* 2011;16:20023.
11. Al-Nakhli H, Al-Ogaily Z, Nassar T. Representative *Salmonella* serovars isolated from poultry and poultry environments in Saudi Arabia. *Revue Scientifique et Technique-Office International des Epizooties* 1999;18:700–709.
12. Barua H, Biswas PK, Talukder KA, Olsen KE, Christensen JP. Poultry as a possible source of non-typhoidal *Salmonella* enterica serovars in humans in Bangladesh. *Vet Microbiol* 2014;168:372–380.
13. Rahman SIA, Dyson ZA, Klemm EJ, Khanam F, Holt KE. 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.
14. Khan AI, Rashid MM, Islam MT, Afrad MH, Salimuzzaman M, *et al*. Epidemiology of cholera in Bangladesh: findings from Nationwide Hospital-based Surveillance, 2014–2018. *Clin Infect Dis* 2019.
15. Sack RB, Siddique AK, Longini IM, Nizam A, Yunus M. A 4-year study of the epidemiology of Vibrio cholerae in four rural areas of Bangladesh. *J Infect Dis* 2003;187:96–101.
16. Khanam F, Sheikh A, Sayeed MA, Bhuiyan MS, Choudhury FK, *et al*. Evaluation of a typhoid/paratyphoid diagnostic assay (TPTest) detecting anti-*Salmonella* IgA in secretions of peripheral blood lymphocytes in patients in Dhaka, Bangladesh. *PLoS Negl Trop Dis* 2013;7:e2316.
17. Ponstingl H, Ning Z. SMALT-a new mapper for DNA sequencing reads. *F1000 Posters* 2010;1:313.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
19. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, *et al*. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* 2016;2:e000056.
20. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.
21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
22. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 2007;23:127–128.
23. Yu G, Smith DK, Zhu H, Guan Y, Lam TT, *et al*. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2016;8:28–36.
24. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Res* 2018;3:93.
25. StataCorp L. *Stata Statistical Software: Release 12.0*. College Station, TX: StataCorp LP; 2011.
26. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
27. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
29. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
30. Snipen L, Liland KH. micropan: an R-package for microbial pan-genomics. *BMC Bioinformatics* 2015;16:79.
31. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* 2008;11:472–477.
32. Baddam R, Kumar N, Shaik S, Lankapalli AK, Ahmed N. Genome dynamics and evolution of *Salmonella* Typhi strains from the typhoid-endemic zones. *Sci Rep* 2014;4:7457.
33. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21:180.
34. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, *et al*. Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* 2017;34:292–293.
35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
36. Donkor ES, Dayie NT, Adiku TK. Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA). *J Bioinform Seq Anal* 2014;6:1–6.
37. Jones P, Binns D, Chang H-Y, Fraser M, Li W. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;30:1236–1240.
38. Eichinger V, Nussbaumer T, Platzer A, Jehl M-A, Arnold R, *et al*. EffectiveDB--updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res* 2016;44:D669:D669-74.:.
39. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, *et al*. ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3:e000131.
40. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* 2013;57:3348–3357.
41. Carattoli A, Zankari E, García-Fernández A, Larsen MV, Lund O, *et al*. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother (Bethesda)* 2014;58:3895–3903.
42. Chen L, Yang J, Yu J, Yao Z, Sun L, *et al*. VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Research* 2005;33:D325–D328.
43. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VPJ, *et al*. The *Salmonella in silico* Typing Resource (SISTR): An open web-accessible tool for rapidly typing and subtyping Draft *Salmonella* genome assemblies. *PLoS One* 2016;11:e0147101.
44. 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. *bioRxiv* 2020;2020.2007.2003.186692.

45. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, *et al*. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;6:90.
46. Azriel S, Goren A, Shomer I, Aviv G, Rahav G. The Typhi colonization factor (Tcf) is encoded by multiple non-typhoidal *Salmonella* serovars but exhibits a varying expression profile and interchanging contribution to intestinal colonization. *Virulence* 2017;8:1791–1807.
47. Robinson N. Typhi colonization factor (Tcf) genetically conserved yet functionally diverse. *Virulence* 2017;8:1511–1512.
48. Dhanani AS, Block G, Dewar K, Forgetta V, Topp E, *et al*. Genomic Comparison of Non-Typhoidal *Salmonella* enterica Serovars Typhimurium, Enteritidis, Heidelberg, Hadar and Kentucky Isolates from Broiler Chickens. *PLoS One* 2015;10:e0128773.
49. Lou L, Zhang P, Piao R, Wang Y. *Salmonella* pathogenicity Island 1 (SPI-1) and its complex regulatory network. *Front Cell Infect Microbiol* 2019;9:270.
50. Mirold S, Rabsch W, Tschäpe H, Hardt WD. Transfer of the *Salmonella* type III effector sopE between unrelated phage families. *J Mol Biol* 2001;312:7–16.
51. Kidgell C, Pickard D, Wain J, James K, Diem Nga LT. Characterisation and distribution of a cryptic *Salmonella* typhi plasmid pHCM2. *Plasmid* 2002;47:159–171.
52. Zhang D, Zhao Y, Feng J, Hu L, Jiang X, *et al*. Replicon-based typing of inci-complex plasmids, and comparative genomics analysis of inciγ/k1 plasmids. *Front Microbiol* 2019;10:48.
53. Khanam F, Rajib NH, Tonks S, Khalequzzaman M, Pollard AJ, *et al*. Case report: *Salmonella* enterica serovar paratyphi B infection in a febrile ILL child during enhanced passive surveillance in an urban slum in Mirpur, Dhaka. *Am J Trop Med Hyg* 2020;103:231–233.
54. Threlfall J, Levent B, Hopkins KL, de Pinna E, Ward LR, *et al*. Multidrug-resistant *Salmonella* Java. *Emerging Infect Dis* 2005;11:170–171.
55. Kloska F, Beyerbach M, Klein G. Infection dynamics and antimicrobial resistance profile of *Salmonella* paratyphi B d-tartrate positive (JAVA) in a persistently infected broiler barn. *Int J Environ Res Public Health* 2017;14.
56. LH S, Chiu CH. *Salmonella*: clinical importance and evolution of nomenclature. *Chang Gung Med J* 2007;30:210–219.